Lecture Notes

Microeconomic Theory

Guoqiang TIAN
Department of Economics
Texas A&M University
College Station, Texas 77843
(gtian@tamu.edu)

August, 2002/Revised: February 2018

1This lecture notes are for the purpose of my teaching and convenience of my students in class.
Contents

I Preliminary Knowledge and Methods

1 Nature of Modern Economics

1.1 Economics and Modern Economics

1.1.1 What is economics about?
1.1.2 The Four Basic Questions to Must Be Answered
1.1.3 What is Modern Economics?
1.1.4 The Difference between Economics and Natural Science

1.2 Two Categories of Economic Theories

1.2.1 Benchmark Theory and Relative Realistic Theory
1.2.2 Three Roles of Economic Theory
1.2.3 What is Microeconomic Theory?

1.3 Modern Economics and Modern Market System Governance

1.3.1 Market and Market Mechanism
1.3.2 Three Functions of Price Mechanism
1.3.3 The Superiority of Market System

1.4 Governance Boundary of Government, Market and Society

1.5 Three Institutional Arrangements for Comprehensive Governance

1.6 Modern Economics and Ancient Chinese Economic Thought

1.7 The Cornerstone Assumption in Modern Economics

1.8 Key Points in Modern Economics

1.8.1 Scarcity of Resources
1.8.2 Information Asymmetry and Decentralized Decision-making
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.8.3</td>
<td>Economic Freedom and Voluntary Exchange</td>
<td>30</td>
</tr>
<tr>
<td>1.8.4</td>
<td>Doing Things under Constraints</td>
<td>31</td>
</tr>
<tr>
<td>1.8.5</td>
<td>Incentive Compatibility of Self-interest and Mutual Benefits</td>
<td>32</td>
</tr>
<tr>
<td>1.8.6</td>
<td>Property Rights Incentive</td>
<td>33</td>
</tr>
<tr>
<td>1.8.7</td>
<td>Outcome Fairness and Equity in Opportunity</td>
<td>34</td>
</tr>
<tr>
<td>1.8.8</td>
<td>Allocative Efficiency of Resources</td>
<td>35</td>
</tr>
<tr>
<td>1.9</td>
<td>A Proper Understanding of Modern Economics</td>
<td>36</td>
</tr>
<tr>
<td>1.9.1</td>
<td>On the Scientific Nature of Modern Economics</td>
<td>36</td>
</tr>
<tr>
<td>1.9.2</td>
<td>On the Mathematical Nature of Modern Economics</td>
<td>38</td>
</tr>
<tr>
<td>1.9.3</td>
<td>How to Regard the Economic Theory Correctly?</td>
<td>38</td>
</tr>
<tr>
<td>1.9.4</td>
<td>How to Regard Experiments in Economics?</td>
<td>40</td>
</tr>
<tr>
<td>1.10</td>
<td>Basic Analytical Framework of Modern Economics</td>
<td>43</td>
</tr>
<tr>
<td>1.10.1</td>
<td>Specifying Economic Environments</td>
<td>45</td>
</tr>
<tr>
<td>1.10.2</td>
<td>Making Behavioral Assumptions</td>
<td>47</td>
</tr>
<tr>
<td>1.10.3</td>
<td>Setting Economic Institutional Arrangements</td>
<td>49</td>
</tr>
<tr>
<td>1.10.4</td>
<td>Determining Equilibrium</td>
<td>51</td>
</tr>
<tr>
<td>1.10.5</td>
<td>Having Evaluations</td>
<td>52</td>
</tr>
<tr>
<td>1.11</td>
<td>Basic Research Methodologies in Modern Economics</td>
<td>53</td>
</tr>
<tr>
<td>1.11.1</td>
<td>Establishing Benchmarks</td>
<td>54</td>
</tr>
<tr>
<td>1.11.2</td>
<td>Setting Reference System</td>
<td>55</td>
</tr>
<tr>
<td>1.11.3</td>
<td>Developing Studying Platform</td>
<td>56</td>
</tr>
<tr>
<td>1.11.4</td>
<td>Providing Analytical Tools</td>
<td>58</td>
</tr>
<tr>
<td>1.11.5</td>
<td>Constructing Rigorous Analytical Models</td>
<td>59</td>
</tr>
<tr>
<td>1.11.6</td>
<td>Positive Analysis and Normative Analysis</td>
<td>59</td>
</tr>
<tr>
<td>1.12</td>
<td>Practical Role of the Analytic Framework and Methodologies</td>
<td>60</td>
</tr>
<tr>
<td>1.13</td>
<td>Basic Requirements for Emulating Good Modern Economics</td>
<td>60</td>
</tr>
<tr>
<td>1.14</td>
<td>Distinguishing Sufficient and Necessary Conditions</td>
<td>61</td>
</tr>
<tr>
<td>1.15</td>
<td>The Role of Mathematics in Modern Economics</td>
<td>62</td>
</tr>
<tr>
<td>1.16</td>
<td>Conversion between Economic and Mathematical Language</td>
<td>66</td>
</tr>
<tr>
<td>1.17</td>
<td>Reference</td>
<td>67</td>
</tr>
</tbody>
</table>
2 Knowledge and Methods of Mathematics

2.1 Basic Set Theory ........................................... 71
2.2 Basic Linear Algebra ...................................... 71
2.3 Basic Topology .............................................. 72
2.4 Single-Valued Function and Its properties ............... 72
   2.4.1 Separating Hyperplane Theorems ....................... 74
   2.4.2 Concave and Convex Function ......................... 74
2.5 Multi-Valued Functions and Its properties ............... 76
   2.5.1 Point-to-Set Mappings ............................... 76
2.6 Static Optimization ......................................... 81
   2.6.1 Unconstrained Optimization ......................... 82
   2.6.2 Optimization with Equality Constraints ............. 82
   2.6.3 Optimization with Inequality Constraints .......... 83
   2.6.4 The Envelope Theorem .............................. 85
   2.6.5 Maximum Theorems .................................. 86
   2.6.6 Continuous Selection Theorems ...................... 87
   2.6.7 Fixed Point Theorems ............................... 87
   2.6.8 Variation Inequality ................................. 89
   2.6.9 FKKM Theorems .................................. 89
2.7 Dynamic Optimization ...................................... 90
2.8 Differential Equation ...................................... 90
2.9 Difference Equation ....................................... 90
2.10 Basic Probability ........................................ 90
   2.10.1 Probability and Conditional Probability ............ 90
   2.10.2 Mathematical Expectation and Variance ........... 90
   2.10.3 Continuous Distributions ........................... 90
2.11 Stochastic Dominance and Affiliation .................. 91
   2.11.1 Hazard Rates .................................. 91
   2.11.2 Stochastic Dominance ............................ 92
   2.11.3 Hazard Rate Dominance ........................... 96
### II Individual Decision Making

#### 3 Consumer Theory

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Introduction</td>
<td>108</td>
</tr>
<tr>
<td>3.2 Consumption Set and Budget Constraint</td>
<td>109</td>
</tr>
<tr>
<td>3.2.1 Consumption Set</td>
<td>109</td>
</tr>
<tr>
<td>3.2.2 Budget Constraint</td>
<td>110</td>
</tr>
<tr>
<td>3.3 Preferences and Utility</td>
<td>111</td>
</tr>
<tr>
<td>3.3.1 Preferences</td>
<td>111</td>
</tr>
<tr>
<td>3.3.2 The Utility Function</td>
<td>116</td>
</tr>
<tr>
<td>3.4 Utility Maximization and Optimal Choice</td>
<td>122</td>
</tr>
<tr>
<td>3.4.1 Utility Maximization</td>
<td>122</td>
</tr>
<tr>
<td>3.4.2 Consumer’s Optimal Choice</td>
<td>122</td>
</tr>
<tr>
<td>3.4.3 First Order-Conditions for Utility Maximization</td>
<td>124</td>
</tr>
<tr>
<td>3.4.4 Sufficiency for Utility Maximation</td>
<td>127</td>
</tr>
<tr>
<td>3.4.5 Non-Ordering Preference and Its Maximization</td>
<td>131</td>
</tr>
<tr>
<td>3.5 Indirect Utility, and Expenditure, and Money Metric Utility Functions</td>
<td>132</td>
</tr>
<tr>
<td>3.5.1 The Indirect Utility Function</td>
<td>132</td>
</tr>
<tr>
<td>3.5.2 The Expenditure Function and Hicksian Demand</td>
<td>134</td>
</tr>
<tr>
<td>3.5.3 The Money Metric Utility Functions</td>
<td>138</td>
</tr>
<tr>
<td>3.5.4 Some Important Identities</td>
<td>140</td>
</tr>
<tr>
<td>3.5.5 Homothetic Utility Functions</td>
<td>144</td>
</tr>
<tr>
<td>3.6 Duality Between Direct and Indirect Utility</td>
<td>145</td>
</tr>
</tbody>
</table>
3.7 Properties of Consumer Demand

3.7.1 Income Changes and Consumption Choice

3.7.2 Price Changes and Consumption Choice

3.7.3 Income-Substitution Effect, Slutsky Equation and Hicksian Equation

3.7.4 Continuity and Differentiability of Demand Functions

3.7.5 Inverse Demand Functions

3.8 Change in Consumer Surplus and its Measurement

3.9 Integrability

3.9.1 Obtaining Utility Function from Demand Function

3.9.2 Revealed Preference

3.9.3 Recoverability

3.10 Topics in Demand Behavior

3.10.1 Income-Leisure Choice Model

3.10.2 Aggregation of Commodities

3.10.3 Aggregation of Consumers’ Demand

3.11 Reference

4 Production Theory

4.1 Introduction

4.2 Production Technology

4.2.1 Specification of Technology

4.2.2 Common Properties of Production Sets

4.2.3 The Marginal Rate of Technical Substitution

4.2.4 The Elasticity of Substitution

4.3 Profit Maximization

4.3.1 Producer Behavior

4.3.2 Producer’s Optimal Choice

4.3.3 First-Order Conditions for Profit Maximization

4.3.4 Sufficiency for Profit Maximization
4.3.5 Properties of Net Supply Functions .................................. 199
4.3.6 Weak Axiom of Profit Maximization ................................. 200
4.3.7 Profit Function ....................................................... 204
4.4 Cost Minimization ...................................................... 207
4.4.1 First-Order Conditions of Cost Minimization ..................... 207
4.4.2 Sufficiency of Cost Minimization ................................. 209
4.4.3 Cost Functions and Their Properties ............................... 213
4.4.4 Produce Surplus ................................................... 221
4.5 Duality in Production ................................................. 221
4.5.1 Recovering a Production Set from a Cost Function ............... 222
4.5.2 Characterization of Cost Functions ............................... 226
4.5.3 The Integrability for Cost Functions ............................... 227
4.6 Reference .......................................................... 229

5 Choice Under Uncertainty ................................................. 231
5.1 Introduction .......................................................... 231
5.2 Three Basic Theoretical Models ...................................... 232
5.2.1 Basic Description of Choice Under Uncertainty .................. 232
5.2.2 Von Neumann-Morgenstern Expected Utility Model ............. 232
5.2.3 Savage Subjective Probability Model ............................. 232
5.2.4 Anscombe-Aumann State Dependent Utility Model ............. 232
5.3 Von Neumann-Morgenstern Expected Utility Theory ................. 232
5.3.1 Model and Basic Axioms ........................................ 232
5.3.2 Expected Utility Presentation Theorem .......................... 232
5.4 Risk Aversion ......................................................... 238
5.4.1 Absolute Risk Aversion .......................................... 238
5.4.2 Global Risk Aversion ............................................ 242
5.4.3 Relative Risk Aversion .......................................... 245
5.4.4 Anscombe-Aumann State Dependent Utility Theory ............. 248
5.4.5 Model and Basic Axioms ........................................ 248
III Game Theory and Market Theory

6 Game Theory

6.1 Introduction ........................................... 261
6.2 Description of a game ................................. 262
6.2.1 Strategic Form .................................. 262
6.3 Solution Concepts .................................... 266
6.3.1 Mixed Strategies and Pure Strategies ............. 266
6.3.2 Nash equilibrium ................................. 268
6.3.3 Dominant strategies .............................. 271
6.4 Repeated games ........................................ 272
6.5 Refinements of Nash equilibrium ....................... 275
6.5.1 Elimination of dominated strategies ............... 275
6.5.2 Sequential Games and Subgame Perfect Equilibrium ........... 276
6.5.3 Repeated games and subgame perfection ............. 282
6.6 Games with incomplete information .................... 283
6.6.1 Bayes-Nash Equilibrium ........................ 283
6.6.2 Discussion of Bayesian-Nash equilibrium .......... 286
6.7 Reference .............................................. 287
9.5.8 Vertical Product Differentiation Model ................................ 324
9.5.9 Market Entry Deterrence ............................................ 324
9.5.10 Price Competition with Asymmetric Information ............. 324
9.5.11 Limit Pricing with Asymmetric Information: Dynamic Mar-
ket Structure .......................................................... 324
9.6 Reference ................................................................. 324

IV General Equilibrium Theory and Social Welfare 327

10 Positive Theory of Equilibrium: Existence, Uniqueness, and Sta-
bility ........................................................................ 329
10.1 Introduction ............................................................. 329
10.2 The Structure of General Equilibrium Model ..................... 332
  10.2.1 Economic Environments ........................................ 332
  10.2.2 Institutional Arrangement: Private Market Mechanism .... 335
  10.2.3 Individual Behavior Assumptions: .......................... 336
  10.2.4 Competitive Equilibrium ....................................... 337
10.3 Some Examples of GE Models: Graphical Treatment ............ 340
  10.3.1 Pure Exchange Economies ...................................... 341
  10.3.2 The Economy With One Consumer And One Producer/Robinson
        Crusoe Economy .................................................. 348
10.4 Existence of Competitive Equilibrium ............................. 351
  10.4.1 The Existence of CE for Aggregate Excess Demand Function 352
  10.4.2 The existence of CE under individual characteristics ..... 360
  10.4.3 Examples of Computing CE .................................... 369
  10.4.4 The existence of CE for discontinuous demand function ... 373
  10.4.5 Existence of CE for Aggregate Excess Demand Correspon-
        dences .............................................................. 380
  10.4.6 Existence of CE for General Production Economies ....... 382
  10.4.7 The Existence of CE Under Non-Ordered Preference ...... 385
10.5 Uniqueness of Competitive Equilibria . . . . . . . . . . . . . . . . 386
10.6 Stability of Competitive Equilibrium . . . . . . . . . . . . . . . . 391
10.7 Abstract Economy . . . . . . . . . . . . . . . . . . . . . . . . . . 399
  10.7.1 Equilibrium in Abstract Economy . . . . . . . . . . . . . . 399
  10.7.2 Existence of Equilibrium for General Preferences . . . . . . 401
10.8 Reference . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 404

  11.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . 409
  11.2 Pareto Efficiency of Allocation . . . . . . . . . . . . . . . . . . 410
  11.3 The First Fundamental Theorem of Welfare Economics . . . . . 417
  11.4 Calculations of Pareto Optimum by First-Order Conditions . . . 424
    11.4.1 Exchange Economies . . . . . . . . . . . . . . . . . . . . . 425
    11.4.2 Production Economies . . . . . . . . . . . . . . . . . . . . 426
  11.5 The Second Fundamental Theorem of Welfare Economics . . . . 427
  11.6 Non-Convex Production Technologies and Marginal Cost Pricing . 435
  11.7 Pareto Optimality and Social Welfare Maximization . . . . . . . 440
    11.7.1 Social Welfare Maximization for Exchange Economies . . . 441
    11.7.2 Welfare Maximization in Production Economy . . . . . . . 443
  11.8 Political Overtones . . . . . . . . . . . . . . . . . . . . . . . . 444
  11.9 Reference . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 446

12 Economic Core, Fair Allocations, and Social Choice Theory 448
  12.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . 448
  12.2 The Core of Exchange Economies . . . . . . . . . . . . . . . . . 449
  12.3 Fairness of Allocation . . . . . . . . . . . . . . . . . . . . . . . 455
  12.4 Social Choice Theory . . . . . . . . . . . . . . . . . . . . . . . 461
    12.4.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . 461
    12.4.2 Basic Settings . . . . . . . . . . . . . . . . . . . . . . . . . 462
    12.4.3 Arrow’s Impossibility Theorem . . . . . . . . . . . . . . . 464
    12.4.4 Some Positive Result: Restricted Domain . . . . . . . . . 467
12.4.5 Gibbard-Satterthwaite Impossibility Theorem 470
12.5 Reference 475

13 General Equilibrium Under Uncertainty 477
13.1 Introduction 477
13.2 A Market Economy with Contingent Commodities 478
13.3 Arrow-Debreu Equilibrium 483
13.4 Sequential Trade 485
13.5 Incomplete Markets 492
13.6 Reference 496

V Externalities and Public Goods 499

14 Externalities 502
14.1 Introduction 502
14.2 Consumption Externalities 503
14.3 Production Externality 511
14.4 Solutions to Externalities 513
14.4.1 Pigovian Tax 514
14.4.2 Coase' Voluntary Negotiation Property Rights Approach 514
14.4.3 Missing Market 526
14.4.4 The Compensation Mechanism 527
14.5 Reference 530

15 Public Goods 534
15.1 Introduction 534
15.2 Notations and Basic Settings 535
15.3 Discrete Public Goods 538
15.3.1 Efficient Provision of Public Goods 538
15.3.2 Free-Rider Problem 539
15.3.3 Voting for a Discrete Public Good 540
VI Information, Incentives, and Mechanism Design 561

16 Optimal Mechanism Design: Contracts with One-Agent and Hidden Information 567

16.1 Introduction ................................. 567

16.2 Basic Settings of Principal-Agent Model with Adverse Selection 569

16.2.1 Economic Environment (Technology, Preferences, and Information) ................................. 569

16.2.2 Contracting Variables: Outcomes ................................. 569

16.2.3 Timing ........................................ 570

16.3 Complete Information Optimal Contract (Benchmark) ................................. 570

16.3.1 First-Best Production Levels ........................................ 570

16.3.2 Implementation of the First-Best ........................................ 571

16.3.3 A Graphical Representation of the Complete Information Optimal Contract ................................. 572

16.4 Incentive Feasible Contracts ........................................ 574

16.4.1 Incentive Compatibility and Participation ........................................ 574

16.4.2 Special Cases ........................................ 575

16.4.3 Monotonicity Constraints ........................................ 575

16.5 Information Rents ........................................ 576

16.6 The Optimization Program of the Principal ........................................ 577
17 Optimal Mechanism Design: Contracts with One-Agent and Hidden Action

17.1 Introduction ................................................. 636
17.2 Basic Settings of Principal-Agent Model with Moral Hazard .... 638
  17.2.1 Effort and Production ................................ 638
  17.2.2 Incentive Feasible Contracts .......................... 638
  17.2.3 The Complete Information Optimal Contract .......... 640
17.3 Risk Neutrality and First-Best Implementation ................. 641
17.4 The Trade-Off Between Limited Liability Rent Extraction and Efficiency .......................................................... 643
17.5 The Trade-Off Between Insurance and Efficiency ............... 645
  17.5.1 Optimal Transfers ........................................ 646
  17.5.2 The Optimal Second-Best Effort ...................... 647
17.6 More than Two Levels of Performance ........................ 648
  17.6.1 Limited Liability ....................................... 649
  17.6.2 Risk Aversion ........................................... 650
17.7 Contract Theory at Work .................................... 652
  17.7.1 Efficiency Wage ......................................... 652
  17.7.2 Sharecropping ......................................... 652
  17.7.3 Wholesale Contracts .................................... 655
  17.7.4 Financial Contracts .................................... 656
17.8 A Continuum of Performances ................................ 658
17.9 A Mixed Model of Moral Hazard and Adverse Selection ....... 659
  17.9.1 Optimal Wage Contract with Unobservable Efforts Only 660
  17.9.2 Optimal Wage Contract with Unobservable Efforts and Risk Aversion .......................................................... 663
  17.9.3 Optimal Wage Contract with Unobservable Efforts and Cost 668
17.10 Further Extension .......................................... 672
18 General Mechanism Design: Contracts with Multi-Agents 676
18.1 Introduction 676
18.2 Basic Settings 678
  18.2.1 Economic Environments 678
  18.2.2 Social Goal 680
  18.2.3 Economic Mechanism 680
  18.2.4 Solution Concept of Self-Interested Behavior 682
  18.2.5 Implementation and Incentive Compatibility 683
18.3 Examples 685
18.4 Dominant Strategy and Truthful Revelation Mechanisms 686
18.5 Gibbard-Satterthwaite Impossibility Theorem 692
18.6 Hurwicz Impossibility Theorem 693
18.7 Vickrey-Clark-Groves Mechanisms 696
  18.7.1 Vickrey-Clark-Groves Mechanisms for Discrete Public Good 697
  18.7.2 Vickrey-Clark-Groves Mechanisms with Continuous Public Goods 701
  18.7.3 Uniqueness of VCG for Efficient Decision 706
  18.7.4 Balanced VCG Mechanisms 707
18.8 Nash Implementation 710
  18.8.1 Nash Equilibrium and General Mechanism Design 710
  18.8.2 Characterization of Nash Implementation 712
18.9 Better Mechanism Design 718
  18.9.1 Groves-Ledyard Mechanism 719
  18.9.2 Walker’s Mechanism 720
  18.9.3 Tian’s Mechanism 722
  18.9.4 Informational Efficiency and Uniqueness of Competitive Market Mechanism 726
18.10 Reference 732
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>Incomplete Information and Bayesian-Nash Implementation</td>
<td>744</td>
</tr>
<tr>
<td>19.1</td>
<td>Introduction</td>
<td>744</td>
</tr>
<tr>
<td>19.2</td>
<td>Basic Analytical Framework</td>
<td>745</td>
</tr>
<tr>
<td>19.2.1</td>
<td>Model</td>
<td>745</td>
</tr>
<tr>
<td>19.2.2</td>
<td>Bayesian Incentive-Compatibility and Bayesian Implementation</td>
<td>746</td>
</tr>
<tr>
<td>19.3</td>
<td>Truthful Implementation of Pareto Efficient Outcomes</td>
<td>750</td>
</tr>
<tr>
<td>19.4</td>
<td>Characterization of DIC and BIC under Linear Mode</td>
<td>752</td>
</tr>
<tr>
<td>19.5</td>
<td>Impossibility of BIC Pareto Efficient Optimal Contract</td>
<td>756</td>
</tr>
<tr>
<td>19.5.1</td>
<td>Participation Constraints</td>
<td>756</td>
</tr>
<tr>
<td>19.5.2</td>
<td>Myerson-Satterthwaite Impossibility Theorem</td>
<td>758</td>
</tr>
<tr>
<td>19.6</td>
<td>The Revenue Equivalence Theorem in Auctions</td>
<td>762</td>
</tr>
<tr>
<td>19.7</td>
<td>Cremer-McLean Full Surplus Extraction Theorem</td>
<td>766</td>
</tr>
<tr>
<td>19.8</td>
<td>Characterization of Bayesian Implementability</td>
<td>770</td>
</tr>
<tr>
<td>19.9</td>
<td>Reference</td>
<td>778</td>
</tr>
<tr>
<td>20</td>
<td>Dynamic Mechanism Design</td>
<td>783</td>
</tr>
<tr>
<td>20.1</td>
<td>Introduction</td>
<td>783</td>
</tr>
<tr>
<td>20.2</td>
<td>Dynamic Contracts under Full Commitment</td>
<td>784</td>
</tr>
<tr>
<td>20.2.1</td>
<td>Constant Types</td>
<td>784</td>
</tr>
<tr>
<td>20.2.2</td>
<td>Independent Dynamic Types</td>
<td>786</td>
</tr>
<tr>
<td>20.2.3</td>
<td>Correlated Types</td>
<td>789</td>
</tr>
<tr>
<td>20.3</td>
<td>Dynamic Contracts under Different Commitment Power</td>
<td>794</td>
</tr>
<tr>
<td>20.3.1</td>
<td>Contracting with Full Commitment</td>
<td>796</td>
</tr>
<tr>
<td>20.3.2</td>
<td>Dynamic Contracting with No Commitment</td>
<td>797</td>
</tr>
<tr>
<td>20.3.3</td>
<td>Dynamic Contracts with Partial Commitment</td>
<td>803</td>
</tr>
<tr>
<td>20.4</td>
<td>Sequential Screening</td>
<td>805</td>
</tr>
<tr>
<td>20.4.1</td>
<td>An Example of Sequential Screening</td>
<td>806</td>
</tr>
<tr>
<td>20.4.2</td>
<td>Sequential Screening under Incomplete Information</td>
<td>808</td>
</tr>
<tr>
<td>20.5</td>
<td>Efficient Budget-Balanced Dynamic Mechanism</td>
<td>813</td>
</tr>
</tbody>
</table>
20.5.1 Efficient Budget-Balanced Static Mechanism ................. 816
20.5.2 Incentive Problem in Dynamic Environments ............... 818
20.5.3 Efficient Budget-Balanced Dynamic Mechanism ............ 819
20.6 References .................................................................. 821

VII Market Design ................................................................ 824

21 Auction Theory ................................................................ 827

21.1 Introduction ............................................................... 827
21.2 Some Useful Mathematics ............................................ 832
  21.2.1 Continuous Distributions ........................................ 832
  21.2.2 Hazard Rates ...................................................... 833
  21.2.3 Stochastic Dominance ............................................ 834
  21.2.4 Hazard Rate Dominance ....................................... 838
  21.2.5 Reverse Hazard Rate Dominance ......................... 838
  21.2.6 Order Statistics .................................................. 839
  21.2.7 Affiliation ......................................................... 840
21.3 Private Value Auctions for Single Object ....................... 843
  21.3.1 Basic Analytical Framework ................................. 843
  21.3.2 First-Price Sealed-Bid Auction ............................. 845
  21.3.3 Second-Price Sealed-Bid Auction ........................... 847
  21.3.4 Revenue Comparison of First and Second Auctions .... 849
  21.3.5 Reserve Price in First and Second Price Auctions ..... 851
  21.3.6 Efficient Allocation and Revenue Equivalence Principle ... 854
  21.3.7 Applications of Revenue Equivalence Principle ......... 857
  21.3.8 Optimal Auction Mechanism Design ..................... 863
  21.3.9 Factors Affecting Revenue Equivalence and Efficiency .... 867
21.4 Interdependent Value Auctions for Single Object ............ 871
  21.4.1 Basic Analytical Framework ................................. 873
  21.4.2 Second-Price Auction ......................................... 879
Part I

Preliminary Knowledge and Methods
In order to enable readers to grasp the contents in the notes more effectively, learn the modern economics, understand its profound economic thoughts and the theoretical models as well as their proofs of economic problems rigorously, this part introduces the preliminary knowledge and methods of economics and mathematics.

Chapter 1 briefly introduces the essence, category, thoughts and methods of modern economics discipline, as well as the compatibility of economic thought with China’s extensive and profound Chinese intellectual wisdom. We will discuss the ideas and methods of modern economics, especially those that are involved in this book. Many existing books only focus on learning, in a great extent, ignoring the profound economic thoughts behind economics. Combining both, we hope the notes to be thoughtful and academic. Economics is not only a science but also art. Learning economics, not only know well the contents of economics, but also master its profound thoughts, become a wisdom person.

Chapter 2 introduces the basic knowledge of mathematics required for advanced economics. They are used for rigorous analysis of various economic problems, especially to provide the necessary mathematical knowledge and tools for economic models, axiomatic and scientific, to derive and prove many theoretical results in microeconomics.
Chapter 1

Nature of Modern Economics

In this chapter, we first set out some basic preliminaries and methodologies imposed in modern economics in general and in the lecture notes in particular. We briefly discusses the nature and methodology of modern economics, especially the scope, knowledge, ideas and methods involved in this notes. We will discuss the basic terminologies used in modern economics, its research object of market system and its common ground with ancient Chinese economic thoughts, core assumptions, standard analytical framework, methodologies and techniques, as well as some key points one should pay attention to. The methodologies for studying modern economics and key points include: providing benchmark, establishing reference system, setting up studying platforms, developing analytical tools, making positive and normative analysis, noting the basic requirements, role, generality and limitation of an economic theory, distinguishing necessary and sufficient conditions for a statement, understanding the role of mathematics, and grasping the conversion between economic and mathematical language.
1.1 Economics and Modern Economics

1.1.1 What is economics about?

*Economics* is a social science that studies how to make decisions in face of scarce resources. Specifically, it studies individuals’ economic behavior, economic phenomena, as well as how individual agents, such as consumers, families, firms, organizations and government agencies, make trade-off choices that allocate limited resources among competing uses.

It is actually this fundamental inconsistence/conflict between individuals’ unlimited desires (i.e., wants) and scarce/limited resources that leads to economics. The core idea throughout economics is that under the basic constraint of limited resources (limited information, limited capital, limited time, limited capacity and limited freedom) versus unlimited desires of people, individuals must make best trade-off choices in resource allocation to make the best of limited resources for maximizing the satisfaction of individuals’ needs.

1.1.2 The Four Basic Questions to Must Be Answered

Any economic system, whether it is the planned economy the government play a decisive role, the free economy the market play a decisive role, or the semi-market and semi-state mixed economy in which the state-owned economy play a leading role, in all cannot do without the following four basic issues in the allocation of resources:

1. What goods and services should be produced and in what quantity?
2. How should the product be produced?
3. For whom should it be produced and how should it be distributed?
4. Who makes the decision?
Whether an institution can well answer these questions depends on the fundamental factors of whether it can properly solve the information and incentive problems.

There are two basic economic institutions that have been so far used in the real world:

(1) Planned economic institution: All the four questions are answered by the government that determines most economic activities with decision-making monopoly, including decisions on: product catalog, infrastructure investment allocation, individual work assignment, product price and employment wage, etc. The risk is borne by the government.

(2) Market economic institution: Most economic activities are organized through free exchange system. The decisions on producing what product, how to produce and for whom to produce are mainly made by decentralized firms and consumers, and the risk is borne by individuals.

In the real world, almost every economic institution is somewhere in between these two institutions. The fundamental flaw of the planned economic institution is that it cannot solve the problems of information and incentive, while market economic institution can be a good solution. This is why countries that adopted planned economic institution inevitably failed and why China carries out market-oriented reform.

1.1.3 What is Modern Economics?

*Modern economics*, mainly developed since the 1940s, systematically studies individuals’ economic behavior and social economic phenomena by a scientific studying method – observation → theory → observation – and through the use of various mathematical analytical tools. It is a branch of science that represents scientific analytical framework and research methods. Such systematic study both
involves the form of theory and provides analytical tools for examining economic data.

It will be very important to correctly understand and grasp the general knowledge of modern economics and the contents of the lecture notes in particular for theoretical creation and practical applications of modern economics. It is not only useful for studying and analyzing economic problems, explaining economic phenomena and individuals’ economic behavior; but more importantly, it can draw conclusions and make relatively accurate predictions through analysis of inherent logic in accordance with the causes.

Modern economics is referred to as the “crown” of social sciences because its basic ideas, analytical framework and research methodologies are powerful enough to study economic problems and phenomena under individuals’ behavior patterns of different countries, regions, customs and cultures, and can be applied to almost all social sciences and daily life. It can even be helpful for good leadership, management and work so that it is jokingly called “economics imperialism” or “omnipotent” discipline.

1.1.4 The Difference between Economics and Natural Science

There are three major differences between social science, especially modern economics, and natural science:

(1) Economics often requires the study of human behavior, need to conduct hypothesis, and natural science generally do not involve the behavior of human being (of course, this division is not absolute, such as biology and medicine sometimes involve human behavior. However, these disciplines are not from the perspective of utilitarianism, while economics considers human behavior mainly from the perspective of utility). Once the individuals are involved, the information is extremely asymmetric and easy to disguise because their behavior are unpredictable, and it will become very difficult and complicated to deal with.
(2) In the discussion and study of economic problems, descriptive positive analysis and normative analysis of value judgment are both needed. As individuals have different values and their own interests, controversies often emerge, while natural science generally makes descriptive empirical analysis only and the conclusions can be verified through practice.

(3) Society cannot be taken for experiment or test of most conclusions in economics because policies have broad impact and large externalities, while this is not a problem for almost all branches of natural science.

These three differences make the study of economics more complex and difficult, and a more detailed discussion will be carried out later.

1.2 Two Categories of Economic Theories

Modern economic theory is an axiomatic way to study economic problems. Similar to mathematics, it is a logical deduction from the presupposed assumptions. It consists of assumptions/conditions, analytical frameworks and models, and some conclusions that provide interpretation and/or prediction. These conclusions are strictly derived from the assumptions and analytical frameworks and models, so it is an analytical method with inherent logic. This logic analysis method is very helpful to clearly explain the problem, and can avoid many unnecessary complexity.

Modern economics is based on economic theory to explain, evaluate and speculate on the observed economic phenomena.

1.2.1 Benchmark Theory and Relative Realistic Theory

The modern economic theory can be divided into two categories according to the function. One is to provide various benchmark/baseline economic theories that establish reference points or reference\textsuperscript{1}, relatively far away from the reality, and

\textsuperscript{1}We will come back to discuss the role of the benchmark and the reference system in more detail.
dealing with the ideal situation. The first half of the book mainly discusses such benchmark theories. The other category is to provide realistic economic theories for solving the practical issues so that hypotheses are more close to reality, which are usually modifications to the benchmark theories. The second half of the book will devote to discuss relatively practical and realistic microeconomic theories. As such, both of the two types of theories are very important. All they are used to obtain logical conclusion and to do prediction, and they are a progressive and complementary relationship of development and extension. The second category of theory is relatively realistic through revising for the first category of benchmark theory, which makes the modern economics more complete and realistic.

To be completed

1.2.2 Three Roles of Economic Theory

Economic theory has at least three roles.

The first role is to provide a benchmark and a reference system to catch up or create a goal so as to point a direction for improving. Through a theory guides the reform or innovation to impel a realistic economy to close to the ideal state.

The second role is that it can be used to learn and understand the real economic world, and to explain economic phenomena and economic behavior in the real world, which is the major content of modern economics.

The third role is that it can be used to make inherent logical inferences and predictions. Through logical analysis of economic theory, we can make scientific inferences and predictions on the possible outcomes under given real economic environments, behavior of economic agents and economic institutional arrangements. This will guide us to better solve economic problems in reality. As long as the assumptions in the theoretic model are roughly met, we can obtain scientific logical conclusions, and make correct predictions accordingly, so that we may know the results without experiments. For instance, the theory of unfeasible planned economy by Hayek has this kind of insight.

A good theory can deduce the inherent logic result without experimenting.
This can solve the problem that economics cannot be tested through experiment in society to a great extent. What we need to do is to check whether the assumptions made on economic environments and behavior are reasonable (experimental economics that is popular in recent years is mainly engaged in fundamental theoretical research such as testing individual behavioral assumptions). For example, we are not allowed to issue currency recklessly for the sake of studying the relationship between inflation and unemployment. Like astronomers and biologists, most of the time economists can only utilize the data incidentally provided by the real world to develop and test theory.

### 1.2.3 What is Microeconomic Theory?

An *economic theory*, which can be considered an axiomatic approach to study economic issues, consists of a set of assumptions and conditions, an analytical framework and models, and conclusions (explanations and/or predications) that are derived from the assumptions, the analytical framework and models. It is an analytical approach of inherent logic.

Economics is concerned with the explanation of observed phenomena and also makes economic assessments and predictions based on economic theories.

*Microeconomic theory* aims to model and analyze economic activities as the interaction of individual economic agents pursuing their private interests. A notable feature of microeconomic theory is to set up theoretical hypothesis or modeling for economic activities of self-interested individuals, especially in market economy, and conduct rigorous analysis, and examine how the market works.

The whole microeconomics runs through a main theme – price or pricing: which factors affect pricing? whether enterprises have pricing power? how to have pricing advantages? and how to make optimal pricing? Therefore, we need to study the demand, supply, characteristics and functions of market, and how to make profits in all kinds of markets and various economic environments, so as to maximize profits or utilities. As a result, microeconomics is also called price theory.
Microeconomics is the core of economics and the theoretical foundation of all branches of modern economics, including macroeconomics, financial economics, applied econometrics, etc.. It can make us use simplified assumptions for in-depth analysis of complex problems, so as to find clues from the complex, making the complex problems become relatively simple. It can help us from things related to extract the most useful information, think of various issues using the method of economics, so as to make explanations and prediction about the real world.

1.3 Modern Economics and Modern Market System Governance

A main purpose of modern economics is to study the objective laws of market and individuals (such as consumers and firms)’ behavior in the market. Specifically, it studies how to realize harmony by individuals pursuing their own interests in the market, how the market allocates resources, and how to achieve economic stability and sustainable growth, etc.

1.3.1 Market and Market Mechanism

Market: Market is a trade mode where buyer and seller conduct voluntary exchanges. It refers not only to the place where buyer and seller conduct exchanges, but also to any form of trading activities.

Market mechanism: Market mechanism is a mechanism where individuals make decentralized decisions guided by price. As a form of economic organization featuring information decentralized decision-making, voluntary cooperation and voluntary exchange of products and services, it is one of the greatest inventions in human history and by far the most successful means for human beings to solve their own economic problems. The establishment of market mechanism is not a conscious, purposeful human design, but a natural process of evolution. The emergence, development and further progress of modern economics are mainly
based on the study of market system.

In the market system, decisions on resource allocation are independently made by producers and consumers pursuing their own interests under the guidance of market price. No command or order is imposed. Market system can unknowingly solve the four inevitable basic questions facing any economic system: what to produce, how to produce, for whom to produce, and who makes the decision.

In a competitive market system, firms and individuals make the decision on voluntary exchange and cooperation. Consumers seek the maximal satisfaction of consumptions of goods while firms pursue profits. In order to maximize profits, firms must have meticulous plans for the most effective use of resources. That is to say, for resources with similar utility or effect, they will use the ones at the maximal satisfaction or lowest possible cost. The best uses from the standpoint of firms and society respectively are originally irrelevant, but price links them up. The price level reflects the supply and demand of resources in society, and further the scarcity of resources. In case of inadequate timber supply and ample steel supply in society, timber will be expensive while steel will be cheap. To reduce expenses and make more profits, firms will try to use more steel and less timber. In doing so, firms does not take the interests of society into consideration, but the outcome is totally in line with social interests, which is precisely the underlying role of resource price. Resource price coordinates the interests of firms and the whole society, and solves the problem of how to produce. Price system also guides firms to make production decisions in the public interest of the society. Concerning what to produce, there is only one concern for firms: to produce the product with high price. Yet in the market system, the price level exactly reflects the social needs. For instance, poor harvests and the corresponding rising grain price will encourage farmers to produce more grain. As such, profit-pursuing producers are guided to the right track of “remedy”, and the problem of what to produce is solved. Moreover, the market system also addresses the problem of how to distribute products among consumers. If a consumer really needs a shirt, he or she will offer a higher price than others. Profit-pursuing producers will
definitely sell the shirt to the consumer who needs it most. Thus, the problem of for whom to produce is addressed. All these decisions are made by producers and consumers in a decentralized manner – the problem of who makes the decision is also solved.

As such, market mechanism easily coordinates the seemingly incompatible individual interest and social interest. As early as two hundred years ago, Adam Smith, Father of Modern Economics, saw the harmony of market mechanism (Adam Smith, 1776). He compared the competitive market mechanism as an “invisible hand”. Under the guidance of the invisible hand, individuals pursuing their own interests unintentionally head for a common goal and thus achieve the maximization of social welfare:

“Every individual necessarily labours to render the annual revenue of the society as great as he can. He generally, indeed, neither intends to promote the public interest, nor knows how much he is promoting it … He intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention. Nor is it always the worse for the society that it was no part of it. By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it.”

The first welfare theorem of economics in the general equilibrium theory as we will discuss in the lecture notes rigorously states the above argument of Adam Smith that under certain conditions, a fully competitive market leads to efficient allocation of resources.

1.3.2 Three Functions of Price Mechanism

The market system fulfills its role via the price mechanism. As analyzed by the Nobel laureate in economics Milton Friedman, price fulfills three correlated functions when organizing economic activities:

(1) Convey information: convey production and consumption information in the most efficient way;
(2) Provide incentive: motivate people to involve in consumption and production in an optimal way;

(3) Determine income distribution: resource endowment, price and the efficiency of economic activities determine the income distribution.

**Function of Price: Convey Information**

Price guides the decision-making of the participants, and conveys information of change of supply and demand. With demand (not quantity demanded) increasing, price will go up, so the suppliers will use more production factors to produce this commodity, making the message of increasing demand for this commodity available to related parties. The price system conveys information in a highly efficient way. It only conveys information to those who need it. Meanwhile, not only can price convey information, it can also give incentive to guarantee smooth transmission so that information will not be held up by those who do not need it. Those who pass on information are internally motivated to look for people who are in need of information while those who need information are internally motivated to acquire information.

**Function of Price: Provide Incentive**

Price can provide incentive as well, which will make people react to the change of demand and supply. When the demand of a commodity increases, price will be increasing, and then it should provide incentive to the suppliers to increase production. One of the advantages of the price system is that price, while conveying information, also gives people incentive, making people willing to react to information for the sake of their own interest. Thus consumers are motivated to consume in an optimal way while producers are motivated to conduct production in the most efficient way. This function is closely related to the third function of price: determining the income distribution. If the gains increased due to increased production exceed the increase of cost, producers will continue to increase production until the two are made even. In this way their profit is maximized.

**Function of Price: Determining Income Distribution**


In a market economy, a person’s income depends upon the resources he/she owns (such as assets, labor), and the results of the economic activities he/she engages in. When it comes to income distribution, people always want to separate the function of income distribution from the functions of conveying information and providing incentive, with a purpose of more equal income. However, the three functions are closely related and all are indispensible. Once price’s effect on income disappears, its functions of conveying information and providing incentive will no longer exist. If one’s income does not depend upon the price of labor or commodities he offers to others, why does he have to take the efforts to acquire the information of price and market demand and supply, and react to such information? If one gets the same income no matter how he works, then who would like to work hard and well? If no benefits are given for creation and inventions, then who are willing to make efforts in this regard? In a word, if price has no impact on income distribution, it will also lose the other two functions.

1.3.3 The Superiority of Market System

In progress

1.4 Governance Boundary of Government, Market and Society

Of course, there is no complete laissez-faire market economy that is totally independent of the government in the world. Good market operation needs the effective integration of the three-dimensional structure of government, market and society. The complete laissez-faire market that is totally independent of the government is not almighty. As we discuss in Part IV, under many circumstances, such as monopoly, unfair income distribution, the polarization of rich and poor, externality, unemployment and inadequate supply of public goods etc., it will become ineffective, leading to inefficient allocation of resources and various social
problems.

Market economy can be classified into “good market economy” and “bad market economy”. Whether it is good or bad depends on whether the governance boundary of government, market and society is clearly and reasonably defined. In a good market economy, the government can make the market fully play its role; however, in case of market failure, government can compensate to play the role. Good and effective modern economy is a contract economy and rule-of-law economy, which constrained by commodity exchange contract, market operation rules and reputation. The effect of price mechanism and the pursuit of self-interest of individuals require that it is a must to obey its rules, making the market lead to effective allocation of resources and the maximization of social welfare. Thus, modern market economy is established based on the rule of law that has two roles. First, it binds arbitrary government intervention in market economic activities, which is the fundamental role. Second, it further supports and promotes the market, including the definition and protection of property rights, the enforcement of contracts and laws, the maintenance of fair market competition, etc., so as to give full play to the fundamental role of market in resource allocation, as well as the three basic roles of price in conveying information, providing incentives and determining income distribution.

On the other hand, in a bad market economy, for lack of adequate ruling and governance capacity in the economic and social transformation, government fails to provide necessary and sufficient public goods and services to make up for market failure, thus leading to numerous rent-seeking and corruption phenomena. As a result, social and economic fairness and justice is greatly impaired. This breeds the so-called “State Capture”, which refers to the phenomenon that economic agent interferes in the decision on laws, rules and regulations through providing personal interests for government officials. As a consequence, the economic agent can make personal preferences into the basis of game rules of the whole market economy without going through free competition under the market environment, which gives rise to a large number of policy arrangements that produce high
monopoly profits for specific individuals. The price behind this is enormous social costs and the decrease of government credibility. As a result, the inefficient balance in public choice can continue for long.

Therefore, in the three-dimensional framework of government, market and society, government as an institutional arrangement with strong positive and negative externalities plays a vital role. It can make the market efficient, become the impetus for economic development, help construct a harmonious society, and achieve balanced and sustainable development. On the other hand, it may also make the market inefficient, lead to various social contradictions and become huge resistance. Almost all countries in the world adopt market economy, yet a majority of market economy countries did not achieve sound and rapid development. Among many reasons, the most fundamental one is that there lacks a reasonable and clear definition of the governance boundary between government, market and society, but exists the over-playing, under-playing and mis-playing of government role. Therefore, it is possible to define a reasonable governance boundary of government, market and society only when the government loosens its omnipresent “visible hand”, and the functions and governance boundary of the government are scientifically and reasonably defined.

How to reasonably define the governance boundary of government, market and society? The answer is to let the market do whatever it can do while the government does not participate in the economic activities directly (yet it is necessary for the government to maintain market order and guarantee the strict implement of contracts and rules); as for those things that the market cannot do, or considering other factors such as national security that the market is not suitable for, government can directly participate in the economic activities. That is to say, while considering the construction of a harmonious society and the harmonious development of economy, or while the government transforms its functions and innovates the management mode, we should consider according to the definition of the boundary of government and market. For instance, the government should exit from the competitive sectors, although the sector may
exist for long even the government does not exit. Only in case of market failure should the government play its role in solving the problems in the market by itself or along with the market.

Under modern market economy, the basic functions and role of government can be generalized as “maintenance” and “service”, that is, making the fundamental rules, ensuring national security and social order and stability, and providing public goods and services. As Hayek pointed out, government has two basic functions: on one hand, government has to undertake the functions of enforcing the laws and resisting foreign aggression; on the other hand, government has to provide the service that market is unable to provide or sufficiently provide. Meanwhile, he also stated that “it is indeed most important that we keep clearly apart these altogether different tasks of government and do not confer upon it in its service functions the authority which we concede to it in the enforcement of the law and defence against enemies.” This requires that in addition to undertaking the necessary functions, government should separate its powers to market and society. The great President Abraham Lincoln in American history gave a clear and incisive definition of the functions of the government:

“The legitimate object of government is to do for a community of people, whatever they need to have done, but cannot do, at all, or cannot, so well do, for themselves—in their separate, and individual capacities. In all that the people can individually do as well for themselves, government ought not to interfere.”

Meanwhile, a good and efficient modern market economy and the national governance mode need an independent civil society with strong ability of interest coordination as an auxiliary non-institutional arrangement. Otherwise, a severe consequence may be caused that the explicit and implicit transaction cost of market economy activities increase, thus it will be hard to establish the most basic trust relationship in society.

In summary, a reasonable and clear definition of the governance boundary between government, market and society is a prerequisite for establishing a good and efficient market economy system and achieving efficiency, fairness and sustainable
development. Of course, the transition to an efficient modern market system is often a long process. Due to various constraints, the governance boundary of government, market and society cannot be clearly defined instantly, which calls for a series of transitional institutional arrangements. However, for transitional economies, with the deepening of transition, some transitional institutional arrangements may decline in efficiency, and may even become invalid institutional arrangements or negative institutional arrangements. If the governance boundary of government, market and society cannot be timely and appropriately clarified, while some temporary, transitional institutional arrangements (such as government-led economic development) are fixed as permanent and ultimate institutional arrangements, it is then impossible to achieve an efficient market and a harmonious society. Modern economics’ analytical framework and research methods can play an irreplaceable role in studying how to reasonably and clearly defining the governance boundary of government, market and society and conduct comprehensive governance.

1.5 Three Institutional Arrangements for Comprehensive Governance

As discussed above, in order to achieve a well-functioning market and establish an efficient modern market system, it is necessary to coordinate and integrate the relationship of the three basic coordination mechanisms of government, market and society to regulate and guide individuals’ behavior and conduct comprehensive governance. Government, market and society correspond exactly to the three basic elements of governance, incentive and social norms in an economy. The overlap, comprehensive governance and long-term accumulation of these formal institutional arrangements of mandatory public governance and incentive market mechanism will help guide and shape the normative informal institutional arrangements, enhance the predictability and certainty of socio-economic activities, and significantly save transaction costs.
Regulatory governance, as the basic institutional arrangement and management rule, is mandatory. The basic criterion for whether to formulate such rules and regulations is whether they are easy and clear to define (information transparency and symmetry or not), and whether the costs of information acquisition, regulation enforcement and supervision are excessively high. If a regulation is too costly to supervise, it is not feasible for implementation. Protection of property rights, contract implementation and proper supervision all call for formulating rules and regulations, which thus requires a third party to oversee the enforcement of rules and regulations. The third party is then the government. In order to maintain market order, the role of government is inevitable. As government is also an economic agent, being both the referee and the player, it has large externalities. This requires clear procedures and rules on government behavior that should be clearly and accurately formulated. Regulations on other economic agents and market should be quite the opposite. Due to information asymmetry, regulations in this regard should not be detailed, so as to give people more economic freedom and policy space.

Incentive mechanism such as market mechanism is inducing, which is most widely applied. Due to the information asymmetry and the high cost of information acquisition, the specific operation rules should arouse individuals’ incentives through inducing incentive-compatible mechanisms such as market, realizing incentive compatibility and making people work hard for their own interests as well as others’ and social interests. Reputation and integrity under the market incentive mechanism is also a kind of punishment incentive mechanism. That integrity is essential in doing business does not mean that the business owners are subjectively willing to emphasize on integrity, but have to. Otherwise they will be eliminated from the market. Besides, integrity can save the economic costs and lower transaction costs.

Social norms are a kind of (informal) institutional arrangement that is not compulsory and does not need incentives. Keeping solving problems in accordance with the mandatory rules and inducing incentive mechanism, a social norm, faith
and culture such as corporate culture, folk custom, religious faith, ideology and concept pursuit which need not enforcement nor incentive will gradually form. This is the way that minimizes the institutional transition cost. Especially, when the concepts come to agreement, problems will be much easier to be solved and the work efficiency will be greatly improved. When the concepts cannot come to agreement, even one problem is solved with the mandatory command or inducing incentive mechanism, there will be new problems which need to be solved with large implementation cost.

The three basic ways need to be comprehensively applied and adjusted under specific circumstances with concrete analysis of specific issues. Which way to take is determined by the importance of laws and rules, the degree of information symmetry and the transaction costs of supervision and law enforcement, etc. In a word, the three systems each has its boundary conditions, “to enlighten with reason” depends on whether the information is easy to be symmetric and whether the laws are easy to be supervised. If the laws are of high costs of supervision and implementation, or are not enforced by people, they have no meaning of existence.

1.6 Modern Economics and Ancient Chinese Economic Thought

Many basic concepts of market economy and conclusions of economics, including the idea that commodity price is determined by market and the invisible hand of Adam Smith, have been stated in a remarkably profound way thousands of years ago. China has advocated simple free market economy and believed that price was determined by the market as early as the beginning of the ancient Chinese culture. There have been a great number of particularly insightful ideas of market economy and incentive compatible, dialectical strategies of national governance in China. Almost all the important fundamental ideas, core assumptions and basic conclusions of modern economics, such as self-interest assumption, economic
freedom, governance by the invisible hand, social division of labor, the intrinsic relationship between national prosperity and individual wealth, development and stability, as well as government and market, have been stated by ancient Chinese philosophers. Some examples are given as follows.

As early as over 3000 years ago, JIANG Shang (also known as JIANG Ziya and JIANG Taigong, an ancient Chinese strategist and adviser) believed that “avoiding risks and pursuing interests” is the innate nature of human beings, that is to say, “In general, people resent death and enjoy life, welcome virtue and chase profits.” He held that “the state is not the property of one man but of all people. The person who shares interests with all the people will win the state.” This actually reveals the people-centered thought of dialectical unity between national prosperity and individual wealth and a fundamental law and strategy of national governance that the government should take the common interests, risks, welfare and livelihood of the populace as its own, so as to obtain an incentive compatible result that the populace shares the same interests and risks with the government. JIANG Shang also gave an incisive answer to the relationship and priority of national and nationals’ wealth, that is, “A kingly state makes its people rich, a ruling state makes lower-rank officials wealthy, a barely surviving state makes higher-rank officials affluent, and an unprincipled, ill-governed state only makes itself prosperous.” This advice was followed by King Wen of Zhou Dynasty then, who ordered to open the barns to help the poor and reduce taxes to enrich the people. The Western Zhou thus became a growing power.

Over 2600 years ago, GUAN Zhong (a Legalist chancellor and reformer of the State of Qi in ancient China) presented the law of demand in his book Guan Zi by stating that “The devaluation comes from the excess, while the value from the scarcity”, and drew the basic conclusion that people’s wealth leads to national stability, security, prosperity and power by saying that “Only at times of plenty will people observe the etiquette. Only when they are well-clad and fed will they have a sense of honor and shame.” He further pointed out that “All remedies of government come to delivering people out of poverty. When the people become
well-off, restoring order will be easy. If they are kept in poverty, order will be difficult to maintain.”

“Usually, an orderly state is rolling in prosperity while a disorderly one is deep in poverty. So a king versed in ruling a state must give priority to making people wealthy over governance itself.”

About 2400 years ago, although the first chapter “Laying Plans” of the book The Art of War by SUN Tzu (a military general, strategist and philosopher in ancient China) mainly discussed the military strategies and tactics, it coincides, to a large extent, with the basic analytical framework of modern economics, and can be fully adopted in other contexts. It can also be the rule of making right decisions, accomplishing goals and winning competitions in governing a country, managing an enterprise or organization. Meanwhile, he also gave the basic conclusion of information economics: It is possible to achieve the optimal outcome (“the best is first best”) only under complete information; under information asymmetry, we can at most obtain suboptimal outcome (“the best is second best”). Hence the saying, “If you know the enemy and yourself, you need not fear the result of a hundred battles. If you know yourself but not the enemy, for every victory gained you will also suffer a defeat. If you know neither the enemy nor yourself, you will succumb in every battle.”

In the same period, a more remarkable fact was that LAO Tzu (a famous philosopher of ancient China, the founder of Taoism) presented the supreme law of comprehensive governance: “Govern the state with fairness, use tactics of surprise in war, and win the world by non-intervention.” (Chapter 57, Tao Te Ching) This is the essential way of governing a state or administering an organization, which can be abstracted in common parlance as “to be righteous in deed, flexible in practice and minimal in intervention” of government. Lao Tzu considered “Tao” as the invisible inner law of nature, while “Te” (meaning the inherent character, integrity, virtue) as the concrete embodiment of “Tao”. He deemed that the governance of state and people should follow the Way of Heaven (referring to the objective laws governing nature or the manifestations of heavenly will), the Virtue of Earth and the Principle of Non-intervention by saying that “Man follows the
Earth; Earth follows the Heaven; Heaven follows the Tao; The Tao follows what it is.” (Chapter 25, *Tao Te Ching*) In addition, he also pointed out that “Difficult tasks always stem (and should be tackled) from easy parts, and great undertakings always start with small beginnings.” (Chapter 63, *Tao Te Ching*) That is to say, whatever we do, success lies in details. All these statements mentioned above demonstrate that Lao Tzu’s Non-action thought does not mean “doing nothing” as is commonly regarded, which is a negative misinterpretation of Lao Tzu’s real intention. The non-action discussed by Lao Tzu is a relative concept, which indicates non-intervention in major aspects but action and carefulness in specific aspects. In other words, we should never lose sight of the general goal, and begin by tackling small practical problems at hand.

About 2100 years ago, SIMA Qian (a Chinese historian of the Han dynasty who is considered the father of Chinese historiography) made an incredible statement in his work *Shiji: Huozhi Liezhuan* (*Records of the Grand Historian: Biographies of Merchants*), “Jostling and joyous, the whole world comes after profit; racing and rioting, after profit the whole world goes,” and demonstrated the economic thought of achieving social welfare through social division of labor based on self-interest, which is similar to that of Adam Smith. SIMA Qian investigated the development of social and economic life, and realized the importance of social division of labor. He wrote that “All the goods mentioned above are what people prefer, which are the necessities for life.” Thus, “People rely on the farmer for food, the planter for wood, the craftsman for utensil, and the businessman for circulation.” Moreover, he believed that the whole social economy composed of agriculture, forestry, industry and commerce should develop in a natural course without the constraint of administrative order.

Also in his work *Records of the Grand Historian: Biographies of Merchants*, SIMA Qian continued to write that “What need is there for orders and instructions, mobilizations of labor, or periodical assemblies? Each man makes his efforts to satisfy his own needs based on his own abilities. People seek for purchase with something cheap and sale with something expensive. With diligence and com-
mitment, each man delights in his own business, like water flowing downwards ceaselessly. People gather together on their own initiative, and produce various goods without any orders. Isn’t it the proof of compliance with the law of nature?"

In addition, SIMA Qian gave a very incisive conclusion on the importance of economic freedom as well as the order of several basic institutional arrangements in *Records of the Grand Historian: Biographies of Merchants*, “The master way is to follow the natural law and not to intervene, to guide with interests comes second, to teach with moral comes third, to rule by regulations comes next and to compete for profits comes last.”

Ancient Chinese economic thought is extremely profound. What Adam Smith discussed was also addressed by ancient Chinese philosophers a long time ago. Yet as those statements were just summaries of experience, they did not develop rigorous scientific systems, boundary conditions and scopes for conclusions, or make strict inherent logical analysis. As a result, little is known to the outside world.

### 1.7 The Cornerstone Assumption in Modern Economics

The cornerstone and key assumption in economics is that individual’s behavior is self-interested in normal situations. This is not only an assumption, but is also a biggest reality as well as the cornerstone of the modern economics. Any social science needs to make some assumptions on individual behavior, and has it as the logical starting point of the theory system. The essential difference between social science and natural science lies in that the former usually needs to study human behavior and makes assumptions on human behavior while the latter studies the natural world and things instead of human beings. Economics is a very special subject, which not only needs to study and explain economic phenomena and makes empirical analyses, but also studies human behavior so as to make better
predictions and value judgment.

The so-called self-interest assumption refers to the assumption that an individual (no matter an individual as a country, a unit, an enterprise or a person), in normal cases, tends to be self-interested and pursues its/his own interest. It means that as a rational individual, any person, unit or country will try to pursue maximal benefits in activities. This assumption stands true at any level, and it is especially true when we handle the relationship between countries, units, families and people. Therefore, this is an objective reality or constraint that must be taken into consideration when we study or solve political, social or economic issues. For instance, when considering and handling the relationship between countries, a citizen needs to speak and act for the sake of the interest of his/her own country; one is liable to sentence if he/she discloses state secrets. When handling the relationship between enterprises, an employee has to safeguard the interest of his/her own enterprise; one is liable to punishment depending on the severity of consequences if he/she reveals corporate secrets to their rival. The self-interest assumption is often challenged. Since the human is rational and self-interested, and only pursues his/her own interest, why are there families? As a matter of fact, on the family level, everybody acts for the sake of his/her family’s interest. In other words, under normal circumstances, people focus on their own family instead of others’. When studying individual issues, most people also start out from their own interest. All these show that self-interest is a relative concept, which is often used when we consider issues at the same level. However, a lot of people are mistaken about this assumption and simply understand it as an assumption for individuals no matter on which level the issues are.

Certainly we need to point out that self-interest assumption also has its boundaries. Selflessness is not contradictory to self-interest. It is a different behavioral reaction in a different environment. In case of irregular conditions such as disasters and accidents, people tend to be selfless, which is another form of rationality. For instance, lots of people are willing to risk their own life and fight for their country during foreign invasions. Many people are happy to offer a helping hand...
when others encounter a crisis or disaster. When Japan invaded China in the first half of the 20th century, the Chinese people stood up and fought against the invaders. Many even lost their lives. No country, no home. In a state of national emergency people tend to be selfless. After the Wenchuan Earthquake of China in 2008, the Chinese people rendered various help at the first place. However, in contrast, the individuals tend to focus on their own interest when conducting economic activities in the normal and peaceful environment. All these serve as solid evidence that self-interest and selflessness are both natural reactions to different circumstances and conditions. There is no conflict between the two.

As such, self-interest and altruism are both relative. In fact, such duality can also been seen in animals. For example, when the wild goats were chased to the edge of a cliff, the elder goats sacrificed themselves by making the first jump, so that the younger goats or goatlings could jump on them and have a chance of running away. Even animals are willing to make sacrifice, not to mention people. When the country is in a crisis, or others fall in danger or need help, we should stand up bravely and offer a helping hand. Adam Smith was not only the author of the fundamental book *The Wealth of Nations*, he also wrote *The Theory of Moral Sentiments*, in which he related that people should have compassion and a sense of justice. He wrote and revised the two books for several rounds until his death, and the two works are complementary to each other in the academic thought of Adam Smith.

It is worth noting that “self-interested” does not equal to “harmful to others”. For rational self-interested behavior, the observation of social norms is regarded as a necessary constraint. We cannot agree more that people should be educated so as not to violate public order in their pursuit of personal interest. We also agree that we should safeguard public interest based on the individual rationality. However, we disagree that the policies are so made that they ignore personal interest and become economic idealism, and we also disagree that rational interest of the individuals can be violated in the name of maintaining collective interest. In a word, we should differentiate the self-interested behavior under the framework
of laws, policies and regulations from the selfish behavior which are against laws, policies and regulations, and detrimental to others. While the former should be protected, the latter should be opposed.

Even with self-interested behavior, there is a difference in the extent - the less, the better. However, we should understand it is impossible that self-interest disappears. This is the logical starting point for economics. We may safely say that if all human beings are unselfish and always considerate to others, then economics involving human behavior will find no use. Industrial engineering or input-output analysis maybe suffices. For instance, China carries out reform and opening-up and makes the transition from planned economy to market economy, because it takes into account that self-interest is an objective reality, and people tend to pursue their own interest when involving in economic activities. As a matter of fact, the incredible achievements China has made through reform in the past three decades have much to do with the fact that Chinese reformers (such as DENG Xiaoping) recognize self-interest and adopt market system.

1.8 Key Points in Modern Economics

Economists usually make some of the following key assumptions and conditions when they study economic issues:

(1) Scarcity of resources: individuals confront scarce resources;

(2) Information asymmetry and decentralized decision-making: decentralized decision-making is preferred;

(3) Economic freedom: voluntary cooperation and voluntary exchange;

(4) Decision-making under constraints;

(5) Incentive compatibility among parties: the system or economic mechanism should solve the problem of interest conflicts among individuals or economic units;

(6) Well-defined property rights;
(7) Equity in opportunity;

(8) Allocative efficiency of resources.

Relaxing any of these assumptions may result in different conclusions. The consideration and application of these assumptions is also useful for dealing with daily affairs. Although they seem to be simple, it is not easy to thoroughly understand and skillfully use them in reality. In the following, we will discuss these key assumptions, conditions, and principles respectively.

1.8.1 Scarcity of Resources

Economics stems from the fact that resources are limited in the world (at least the mass of earth is finite). As long as a person is self-interested and his material desires are infinite (the more, the better), it is impossible to achieve distribution according to wants, and the problem of how to use limited resources to meet the wants has to be addressed. Hence we need economics.

1.8.2 Information Asymmetry and Decentralized Decision-making

In addition to the self-interested nature of individuals, another fundamental objective reality is that in most cases, information is asymmetric among economic individuals. A person’s words do not equal to what he really intends. People are unpredictable and most difficult to deal with. Coupled with the self-interested nature, the result is often the conflict of interests between individuals, which makes social science, especially economics, much more complex and difficult to study than natural science. This is why in modern society, due to dishonesty and fraud, many people are reluctant to communicate with other people but spend more time with animals, thinking that animals will not lie to them. As a result, centralized decision-making is often ineffective, while decentralized decision-making is required, such as the use of market mechanism to solve economic problems.
Almost only by complete information acquisition can the outcome be the first best. As stated in information economics, it is possible to achieve the optimal outcome (“the best is first best”) in general only under complete information. However, information symmetry is often difficult to obtain, so incentive mechanism is needed to induce truthful information. Information acquisition incurs costs; as such, we can only obtain suboptimal outcome (“the best is second best”). This is the basic result of the principal-agent theory, optimal contract theory, and optimal mechanism design theory that will be discussed in Part V of the lecture notes. In most cases, information is asymmetric, so there will be market failure and the principal-agent problem. Yet whatever the approach, the outcome is suboptimal due to information asymmetry. Without reasonable institutional arrangements, there will be incentive distortion where inducing information incurs costs and prices; therefore, the outcome cannot be optimal (first best). Information symmetry is particularly important, without which misunderstandings may arise. By communicating with others and letting others understand you (signaling), you get to know others (screening) so as to obtain information symmetry, clear up misunderstandings and reach consensus, which is the fundamental premise of obtaining a good outcome.

The excessive intervention of government in economic activities and the overplaying of government role have led to low efficiency. This, in root, is caused by information asymmetry. There are many problems in information acquisition and discrimination. If the decision-makers are able to have all related information, centralized decision-making featuring direct control would not be problematic. That would be a simple question of optimal decision-making. However, it is impossible that the decision-makers have all related information at hand. That is why people prefer decentralized decision-making. That is also why economists stress that the incentive mechanism, a decentralized decision-making method featuring indirect control, should be used to motivate (inspire) people to do as the decision-makers desire, or to achieve the goals the decision-makers aspire. We will focus on the issue of information and incentive in Part V.
It is worth mentioning that centralized decision-making also has its advantages in some aspects, especially in the decision-making regarding major changes. For instance, centralized decision-making is more efficient when a country, unit or an enterprise is making a big decision on the future plan, orientation or strategy. Such major changes might bring about huge success or severe mistakes. For example, the decision of adopting reform and opening-up policy has led to rapid development of the Chinese economy and unprecedented achievements. In contrast, the decision of “Cultural Revolution” almost pushed the Chinese economy to the verge of collapse. One solution to this problem is to give public opinion full respect and select outstanding leaders.

1.8.3 Economic Freedom and Voluntary Exchange

As economic agents pursue their own interests under asymmetric information, institutional arrangements of the mandatory “stick” style are often not effective, and we need to give people more freedom of economic choice. Thus, we should mobilize economic agents with inducing incentive mechanism such as market through free economic options based on voluntary cooperation and exchange. Therefore, the freedom of economic choice (i.e., “relaxation”) plays a vital role in market mechanism with decentralized decision-making (i.e., “decentralization”). It is a prerequisite for the normal operation of market mechanism, and also a fundamental precondition for ensuring the optimal allocation of resources by competitive market mechanism.

In fact, the Economic Core Theorem to be discussed in Chapter 9 reveals that once full economic freedom is given and free competition, voluntary cooperation and exchange are allowed, even without the consideration of any institutional arrangement in advance, the outcome of resource allocation driven by the self-interested behavior of individuals is consistent with the equilibrium result of a perfectly competitive market. The essence of the Economic Core Theorem can be summarized as follows: under the rationality assumption, as long as freedom and competition is given while institutional arrangement is not considered, the
economic core thus obtained is market competition equilibrium.

China’s reform and opening up over the past 30 years has proved this theorem in practice. An analysis of the reason for China’s remarkable economic achievements shows that despite any other crucial issue, the utmost critical issue is to give people more freedom of economic choice. Reform practice from rural to urban areas indicates that wherever there are looser policies and a greater degree of economic freedom to producers and consumers, there will be higher economic efficiency. China’s miraculous economic growth actually stems from the government’s decentralization to the market, while in reality an imperfect market usually results from excessive government intervention and inadequate or inappropriate government regulation and institutional supply.

In addition to the freedom of economic choice, the right of economic freedom also embraces the property right and the basic survival right.

1.8.4 Doing Things under Constraints

Doing things under constraints is one of the most fundamental principles in economics, as the saying that people should bow under the eaves. Everything has its objective constraints, i.e., individuals make trade-off choice under constraints. As one of the basic principles in economics, people’s choice is determined by the objective constraints and subjective preferences. In economics, one embodiment of the basic idea of the constraint condition is the budget constraint line of consumer theory as we discussed in Part I. The development of a person or a country is faced with various restrictions and constraints, including political, social, cultural, environmental, and resource constraints. If we do not make the constraint condition clear, it is hard to make things done.

While introducing a reform measure or institutional arrangement, it is a must to consider the feasibility and meet the objective constraints; meanwhile, the implementation risk is hopefully to be reduced as much as possible, without causing social, political and economic turmoil. Feasibility means that it is necessary to consider the various constraint conditions in doing things, otherwise it is not
practical. Feasibility, therefore, is a necessary condition that is used to judge whether a reform measure or institutional arrangement is beneficial to economic development and the smooth transition of economic system. In a country’s economic transformation, an institutional arrangement is feasible because it conforms to the institutional environment of the particular stage of development of the country. Again, use China as an example. In China, the reform must adapt to China’s national conditions and fully consider the various constraint conditions, including the limited ideological level of people and participation constraint, etc.

Participation constraint is very important when considering incentive mechanism design, which means that an economic agent can benefit from economic activities, or at least not get hurt in terms of his interest, otherwise he will follow, or oppose the rules or policies. Individuals who pursue the maximization of self-interest will not automatically accept an institutional arrangement, but will make a choice between acceptance and refusal. Only under an institutional arrangement where the individual’s satisfaction is not less than its retained level (not accept the arrangement) will the individual be willing to work, product, trade, distribute, and consume. If a reform measure or institutional arrangement does not meet the participation constraint, individuals may give up. If everyone is reluctant to accept the reform measure or institutional arrangement, it cannot be successfully implemented. Mandatory reform may arouse opposition and cause social instability; thus development cannot be achieved. Therefore, participation constraint that is closely related to social stability is a basic judgment of stability or not in development.

1.8.5 Incentive Compatibility of Self-interest and Mutual Benefits

Each individual has his own self-interest; therefore, the individual interests are often conflicting with social interests. The reason is that in a given institutional arrangement or under the rules of game, individuals will make optimal choice
according to their own interests. However, the choice will not automatically satisfy the interests or the targets of others and the society. The incompleteness of information makes social optimum hard to achieve through instruction. Therefore, to implement an individual or social target, proper rules of game need to be given, making people reach the target while pursuing their own interests. This is the so-called incentive compatibility, which is unifying self-interest and mutual benefits between people, making individuals achieve the goals of others and the society in pursuit of their self-interest.

Everyone benefits from and pays the price of what he does. Therefore, by comparing benefits and costs, he will make a reasonable incentive reaction to the rules of game. A good institutional arrangement or rule can guide the self-interested individuals to act subjectively for themselves and objectively for others, making individuals’ social economic behaviors beneficial to the country, the people, the public and themselves. This is the core content in mechanism design theory to be discussed in Part V. The problem of incentive arises in every social economic unit. Everything that a person does involves interests and costs (benefits and costs). As long as the benefits and costs are not equal, different incentive reactions will appear. Since the interests of the individuals, society and economic organizations cannot be completely the same, how to combine the self-interest, mutual benefits and social interests organically? It then requires incentive compatibility, with which the reform measures and institutional arrangements adopted can greatly mobilize people’s production and work enthusiasm. We will focus on the issue of how to achieve incentive compatibility in Part V of the notes.

1.8.6 Property Rights Incentive

Property rights are an important component of market economy. Property rights include the ownership, the right of use and the decision-making of property. A clear definition of the property rights will help clearly define the profits attribution. Thus, there are incentives for property owners to consume and produce in the most effective way, to provide quality products and good service, to build
reputation and credibility, and to maintain their own commodity, house and equipment. If the property rights are not clearly defined, the enterprises’ enthusiasm will be harmed, giving rise to incentive distortion and moral risk. In the market mechanism, incentives are given to people mainly through the ownership of property and profits. The Coase Theorem to be discussed in Chapter 12 of the notes is a benchmark theorem in property rights theory, which claims that when there is no transaction cost nor income effect, as long as property rights are clearly defined, an efficient allocation of resources can be achieved through voluntary coordination and cooperation.

1.8.7 Outcome Fairness and Equity in Opportunity

“Outcome equity” is a goal that an ideal society wants to achieve as fairness. However, this “outcome equity” brings low efficiency to the human society with self-interested behavior. In what sense can fairness be consistent with economic efficiency? The answer is, if people use “equity in opportunity” as the value judgment standard, equality and efficiency can be consistent. “Equity in opportunity” means that there should not be any barrier to hinder individuals pursuing their goals in their capacity, and there should be an equal competition starting point as much as possible for every individual. The Outcome Fairness Theorem to be introduced in Chapter 12 of the notes tells us that as long as everyone has equal value of initial endowments, through the operation of competitive market, even if individuals pursue self-interest, allocation of resources of both efficiency and fairness can be achieved. A concept similar to “equity in opportunity” is “individual equality” (also known as, “All men are equal before God”), which means that, although people are born with different value, gender, physical condition, cultural background, capacity, way of life, “individual equality” requires due respect for the individual differences.

As everyone has his own preferences, the seemingly equal distribution of milk and bread may not satisfy everyone. Therefore, in addition to defining fairness by the absolute equalitarianism concept of equal allocation, the concept of fair-
ness in other sense is also used in the discussion of economic issues. The equal allocation to be introduced in Chapter 12 of the notes considers both subjective and objective factors, which means that everyone is satisfied with his own gains.

1.8.8 Allocative Efficiency of Resources

Whether social resources are efficiently allocated is a basic criterion to evaluate an economic system. In economics, the efficient allocation of resources usually refers to the Pareto efficiency or optimality, which means that there does not exist other resource allocation scheme where at least someone better off without hurting others. As such, it requires not only efficient consumption and production, but also making products best meet the needs of the consumers.

It may be remarked that, when it comes to economic efficiency, we should distinguish between three types of efficiency: individual firm’s production efficiency, industrial production efficiency and social resource allocation efficiency. The efficiency of individual firm production means to maximize the output with a given input, and conversely, to minimize the input with a given output. Industry is the sum of all firms that produce a category of commodity, of which the efficiency can be similarly defined. Note that the efficiency of an individual firm does not equal to the efficiency of the whole industry. The reason is that if the means of production of a firm with outdated technology are used into firms with advanced technology, there will be more output for the whole industry. At the same time, even if the whole industry production is efficient, there may not be (Pareto) efficiency in the allocation of social resources.

The concept of Pareto efficient allocation of resources is applicable to any economic institution. It provides a basic criterion of value judgment for an economic institution from the viewpoint of social benefit and makes evaluation on the economic effect from the angle of feasibility. This can be applied to planned economy, market economy and mixed economy. The first welfare theorem to be introduced in Chapter 8 of the notes proves that when individuals pursue self-interest, a fully competitive market will lead to efficient allocation of resources.
1.9 A Proper Understanding of Modern Economics

Correctly and profoundly understanding of modern economics can help people correctly use the basic principles and analytical methods of economics to study various kinds of economic problems under different economic environments, behavioral assumptions and institutional arrangements. The different schools and theories themselves show the universality and generality of the analytical framework and methodologies. Different economic environments require adopting different assumptions and specific models. Only in this way can the theory developed explain different economic phenomena and individuals’ economic behavior, and more importantly, make logical analyses, draw conclusions of inherent logic, or make scientific predictions and reasoning under various economic environments that are close to the theoretical assumptions.

1.9.1 On the Scientific Nature of Modern Economics

Criticisms are often heard that there exist too many different economic theories coming from different economic schools and it is difficult to figure out which is right and which is wrong. Some people, while disavowing modern economics and its scientific characteristics, scorch economists by saying that 100 economists will have 101 opinions. The reality is that they do not truly understand modern economics. Different economic, social and political environments require different economic theories, models and institutional arrangements. The fact that different opinions will arise for the same problem just shows the precision and perfection of modern economics because when the premise and environment changes, the conclusion should also change. There does not exist a universal conclusion that can be used in all situations with great generality, otherwise we would not need to analyze the concrete case in the exact situation.

Different economic, political and social environments will give rise to different economic theories and models, but not different “Economics”. For instance,
in China, it is often heard that China’s specific national characteristics call for the development of “Chinese economics”. Well, there are numerous buildings in the world, and even the buildings designed by the same person are quite different. Does that mean we need different “architecture sciences”? The answer is definitely no, because the basic principles and methods for construction designs are of no essential difference. The same is true of the study on economic issues, where the same analytical framework and research methods are adopted, be it Chinese or foreign economic issue. There only exist “Chinese issues”, “Chinese path” and “Chinese characteristics”, but not the so-called “Chinese Economics” and “Western Economics”.

The basic analytical framework and methodologies of modern economics have no bound of areas or countries, and there are no framework and methodologies independent of other countries. Some of the basic principles, methodologies and analytical framework can be used to study a variety of economic issues under any economic environment and institutional arrangement as well as the economic behaviors and phenomena in a specific area and period. The analytical framework and methodologies to be introduced in the following can be used to study and conduct comparative analysis on almost every economic phenomenon and issue. Thus, the economic problems, say in China’s actual economic environment, can also be analyzed by the framework of modern economics. In fact, this is exactly where the power and glamour of the analytical framework of modern economics lies: Its essence is that the economic, political and social environment conditions at a specific time and place must be taken into consideration and clearly defined in study. Modern economics can be used to study the economic issues and phenomena under human behavior as manifested in different countries and regions, customs and cultures. Its basic analytical framework and methodologies can also be used to study other social phenomena and human decision making. It is proven that because of the universality and normality of the analytical framework and methodologies of modern economics, in the past few decades, many analytical methods and theories have been used in political science, sociology,
and humanities, etc.

1.9.2 On the Mathematical Nature of Modern Economics

1.9.3 How to Regard the Economic Theory Correctly?

As seen from our introduction to the basic framework of modern economics, each theory or model in economics is composed of a set of assumptions on economic environments, behavior patterns and institutional arrangements, as well as the conclusions based on these assumptions. Considering the complexity of economic environments and diversity of individual preferences in reality, the more general the assumptions of a theory are, the more helpful and powerful the theory is. If the assumptions of a theory are too restrictive, the theory will lack generality and is thus less useful in reality. As economics is used to serve the society and the government, a necessary condition for a good theory is its generality: the more general, the more applicable and useful. General equilibrium theory is such a theory. It proves the existence of competitive equilibrium that leads to optimal resource allocation under very general preferences and production technologies.

Similarly, theories of social sciences, especially of modern economics, like all theorems in mathematics, have their boundary conditions. We need to pay attention to the preconditions and application scope of a theory when discussing issues or applying an economic theory. No conclusions of an economic theory are absolute and they only hold when the preconditions are satisfied. A basic judgment to see if an economist is well-trained is to see whether he is aware of this point when discussing economic issues. Due to their close relation with daily life, even ordinary people can talk a lot about some economic issues such as inflation, good or bad economic situation, balance or imbalance of supply and demand, unemployment and stock market.

For this reason, many people do not regard economics as a science. Indeed, “economics” would not a science if it did not take into account any constraints, or did not makes inherent logical analysis based on accurate data and rigorous
theories. People in doing so are not real economists. A well-trained economist always discusses issues based on a certain economic theories and is fully aware of the boundary conditions for the relationship between economic variables and the inherent logic of the conclusions. Otherwise, he cannot distinguish the differences between the theory and the reality, and is liable to go to two extremes: either simply applying the theory to reality regardless of the constraints in reality, or denying the value of modern economic theory.

The first extreme view is overestimating the role of theory and abusing a theory. No matter how general a theory and the behavioral assumptions are, they have application boundaries and limitations and thus should not be used indiscriminately. Especially for those theories developed based on an ideal state for establishing reference system, benchmark and goal, we should not apply them directly; otherwise, we might get wrong conclusions. If a person lacks the sense of social responsibility or good training in economics, overestimates the role of theory, blindly applies or misuses economic theory in real economy without taking their assumptions into account, there might be severe consequences, affecting social and economic development and leading to huge negative externalities. For instance, the conclusion of first welfare theorem that competitive market leads to efficient allocation of resources is subject to a series of preconditions and its abuse might give rise to severe policy errors and serious damage to the real economy.

Another extreme view is underestimating the role of theory and denying the significance of modern economics. Some people often negate modern economics by arguing that some assumptions and principles in modern economics are inapplicable to the real world. In fact, no discipline in the world can claim that all of its assumptions and principles coincide with reality perfectly (like the concept of free fall without air resistance in physics we have mentioned above). We should not deny the usefulness of a discipline just based on this, and this is so for modern economics as well. We learn modern economics not only for its basic principles and usefulness, but also for understanding its methods of how to think, find and solve problems.
The value of some economic theories lies not in directly explaining the reality, but in providing studying platform and reference system for further developing new theories to explain the real world. From the methods, we can learn how to solve the problems in reality. In addition, as discussed in the previous section, a theory applicable to one country or region may not be applicable to another country or region due to different environments. Thus, we need to modify and innovate the original theory to develop new theories according to the national/regional economic environment and individual behavior pattern instead of imitating and applying the theory indiscriminately.

It is often heard some people claim that they have toppled an existing theory or conclusion. As some conditions of the theory are not in line with the reality, they consider that the theory is incorrect and has been toppled. Generally speaking, this logic is not scientific or even wrong. No assumption can coincide fully with reality or cover every possible case. A theory may be applicable to a local economic environment, but inapplicable to the economic environment of another nation or region. As long as there is no inherent logical error, we cannot say the theory is wrong and need to be overturned. We may only state that it is not applicable in this case. People can criticize a theory for being too limited or unrealistic, yet what economists should do is to relax or modify the assumptions and adjust the model to improve or develop the original theory. We cannot claim the new theory topples the original one, but should say that the new theory improves or extends the old theory to more general or different economic environments.

A more common mistake is trying to get a general theory just based on some specific examples. This is a wrong methodology.

1.9.4 How to Regard Experiments in Economics?

Economic theory relies mainly upon analysis of internal logic to draw conclusions and inferences, and they cannot be tested through experiment in society. In the field of experimental economics that is popular in recent years, we mainly test human behavior and whether the human behavior assumption is rational through
experiments. Seldom do we prove theories by making experiments, as it is difficult to test the conclusions of economic theory in society. Because of tremendous externalities, in case of failure, there might be policy mistakes bringing about huge risks to economy and society.

This is the greatest difference from natural sciences, as in natural sciences, natural phenomena and objects can be studied through experiments, and by so doing theories can be tested and further developed. Astronomy might be the only exception, but it involves no individual behavior. Once that is involved, things become complicated. Moreover, one can be highly accurate when it comes to the application of theories in natural sciences. For instance, when one constructs a building or a bridge, make a missile or nuclear weapons, we can be as accurate as we wish. All the parameters are controllable, and the interrelationship between variables can be experimented. However, in economics, many factors affecting economic phenomena are uncontrollable, and economists are often criticized for inaccurate economic forecasts.

We can explain this from two perspectives. From the subjective perspective, some economists lack the capability of figuring out the main causes of problems and making correct logical analyses and inferences when they discuss and try to solve economic problems, as they did not receive systematic and rigorous good training on modern economic theories. Thus they make mistakes in proposing solutions. From the objective perspective, some economic factors that influence the results may suddenly and uncontrollably change, thus making the predictions inaccurate, even though they are made by those well-trained economists with good intuition and insight.

An economic issue involves not only human behavior which might complicate the matter, but also many other uncontrollable factors. Although an economist might be wise and insightful, his predictions are liable to deviation due to these uncontrollable factors that are subject to changes. For instance, a nation’s leader, respected as he is, may manage the affairs within his nation quite well but fail to do so in other countries. Likewise, even for a good economist with sound judg-
ment, his economic forecasts might become quite inaccurate once sudden changes take place in the economic, political or social environment. Since experiment is in general not applicable to economics, on what should we rely to predict economic situation or make accurate forecasts? The answer is the internal logical analysis of economics.

In this method, we first gain a full understanding and characterization of the related circumstances (economic environments, situations and current status) for a problem we want to solve, figure out the causes, apply specific and proper economic theories, draw logical conclusions, and based on that make scientific and accurate predictions and correct inferences. As long as the current status is in line with the cause (economic environments and behavioral assumptions) assumed in the economic model, we can obtain the inherent logical result by the application of economic theory, and thus provide solutions (certain institutional arrangement) to different circumstances (which might be varied with time, place, people, and affairs). By this way and based on economic theory we can make scientific and logical references on possible results in a given real economic and social environment, with given behavior pattern of economic agents and under given economic institutional arrangement, thus offering guidance for solving economic problems in reality. In other words, once we make clear the problem and its causes, apply specific and proper economic theory (like the prescription), and conduct concrete and comprehensive analysis, we will obtain logical conclusions and make accurate predictions and correct inferences. Otherwise severe consequences might arise.

To sum up, society is not to be under experiment in economics, nor do data have the only say. Practice is the sole criterion for testing truth; however, it is not the criterion for predicting truth. In modern economics, we rely upon analysis of inherent logic. Like a doctor prescribing for his patient or a mechanic repairing a car, the hardest part is diagnosis of disease or cause of failure. The criterion for a good doctor lies in whether he can accurately find the true cause of disease. Once this is done, it is relatively easier to prescribe for the patient, unless he is indeed incompetent. For economic problems, the prescription is economic
theories. Once we truly understand the characteristics of economic environment, thoroughly study the circumstances, and accurately assume human behavior, we will get twice the result with half the effort.

1.10 Basic Analytical Framework of Modern Economics

There are basic laws for everything that we do. The methods modern economics utilizes to study and solve problems are similar to those that people use to deal with personal, family, economic, political and social affairs. As is known to all, in order to do something well or deal with people, the first step is to learn about the national conditions and customs, i.e., to have some information about the actual situation and the conduct and personality of the person to deal with. Then, based on the information one decides what the best strategy is. This involves weighing the trade-offs in choosing one action over the other to obtain the optimal outcome. Finally, we evaluate the strategy and the outcome it generates. The basic analytical framework and research methods of modern economics completely use this mode to study economic phenomena, human behavior and how people weigh trade-offs and make decisions. Of course, there is a major difference that modern economics strictly defines the inherent logical relationship between assumptions and conclusions by formal models and rigorous argument. Such analytical framework has great normality and consistency.

Of course, a standard academic article needs to first spell out the problems to be studied and solved or the economic phenomena to be explained. That is, economists need to first identify the research objectives and their significance, provide readers with information on the overview and progress of the issues under study through literature review, and illustrate the innovation on technical analysis, or theory of the paper. Then, they discuss how to address the issues raised and draw conclusions.

The economic issues under study may be quite different, but the basic frame-
work used can be the same. The basic framework for an economic theory in modern economics consists of the following five parts or steps: (1) Specifying economic environments; (2) Making behavioral assumptions; (3) Setting institutional arrangements; (4) Determining equilibrium; and (5) Having evaluations. Any economics paper written with clarity and logical consistency is basically made up of these five parts, especially the former four parts, no matter what the conclusion is and whether the author realizes it or not. So to speak, writing an economics paper is quite similar to filling in blanks of these five parts with logical structure. Once you understand these components, you will grasp the basic writing pattern of modern economics papers that can help you study modern economics more easily. These five steps are also helpful for understanding economic theory and its proof, selecting research topics and writing standard economics papers.

Before discussing the five components one by one, we should first define the term “institution”. It is usually defined as a set of rules related with social, political and economic activities that dominate and restrict the behavior of all agents (Schultz, 1968; Ruttan, 1978; North, 1990). When people consider an issue, they always treat some factors as given exogenous variables or parameters, while others as endogenous variables or dependent variables. These endogenous variables depend on the exogenous variables, and thus are functions of those exogenous variables. In line with the classification method of Davis-North (1971, pp 6-7) and the issue to be studied, we can divide institution into two categories: institutional environments and institutional arrangements. Institutional environment is a set of a series of basic economic, political, social and legal rules that form the basis for formulating production, exchange and distribution rules. Of these rules, the basic rules and policies that govern economic activities, property and contract rights constitute the economic institutional environment. Institutional arrangement is the set of rules that dominate the potential cooperation and competition between economic participants. It can be interpreted as the generally known rules of the game, with different rules leading to different reactions. In the
long run, institutional environment and institutional arrangement will affect each other and change. Yet in most cases, as Davis-North clearly pointed out, people usually regard economic institutional environments as given exogenous variable, while consider economic institutional arrangement (such as market system arrangement) as exogenous or endogenous depending on the issues to be studied and discussed.

1.10.1 Specifying Economic Environments

The primary important component of the analytical framework of modern economics is to specify the economic environments where the issue or object to be studied lies. An economic environment is usually composed of economic agents, their characteristics, the institutional environment of economic society, the informational structure and so on, which are given as exogenous variables and parameters. They cannot change in the short term, but may evolve in the long run. The basic idea on constraints is fully reflected here.

How to specify an economic environment? It can be divided into two levels: (1) objective and realistic description of economic environments and (2) concise and acute characterization of the essential features. The former is science and the latter is art, between which there should be combination and balance. That is, the specification of economic environments requires combining the objective description and the characterization of main features according to purpose. The more clear and accurate the description of economic environments is, the greater the likelihood of obtaining correct theoretical conclusions. The more refined and acute the characterization of economic environments is, the simpler and easier the arguments and conclusions to understand. Only by combining these two levels together can we capture the essence of issues under study, as specifically discussed below:

**Description of economic environment**: The first step in every economic theory of modern economics is to objectively describe economic environments where the issue or object to be studied lies. A reasonable, useful economic theory
should exactly and properly describe the specific economic environment. Though
different countries and areas have different economic environments, which usually
lead to different conclusions, the basic analytical framework and methodologies
utilized are the same. A basic common point of studying economic issues is to
describe the economic environment. The more clear and accurate the descrip-
tion of economic environments is, the greater the likelihood of obtaining correct theoretical conclusions.

Characterization of economic environment: When describing the eco-
nomic environment, an important question is how to clearly and accurately de-
scribe the economic environment, while at the same time concisely and acutely characterizing economic environments in order to capture the essence of problem. There is no need for completely objective description of the economic environ-
ment, as people may be confused by trivial matters. If we exactly depict all aspects, it can be said that we describe economic situation or environment with great accuracy. However, this kind of simple listing cannot capture the key points and essence of problem, but presents numerous confusing facts. In order to avoid trivial aspects and focus on the most critical and central issues, we need to char-
acterize economic environments specifically according to the demands of the issue to be studied. For example, when discussing consumer behavior in Chapter 2, we simply describe the consumers as composed of preference relation, consumption space and initial endowment, regardless of the gender, age or wealth of consumers. When discussing the theory of the firm in Chapter 3, a firm’s characteristics can be described as the production possibilities set. When studying the transition-
al economic issues, such as those of China, we cannot simply mimic and apply the conclusions derived under mature market economic environment, but need to characterize the basic features of transitional economies. Yet we can still use the basic analytical framework and methodologies of modern economics.

We often hear people criticize modern economics as useless because it uses a few simple assumptions to plainly summarize complex situations. In fact, this is also the basic research methodology of physics. In the study of the relationship
between two physical variables, both theoretical research and experimental operation will fix the rest of variables that influence the object of study. In many cases, it is unnecessary to clarify every aspect (even unrelated aspect), and it may even lead to the loss of focus. This is just like drawing maps for different purposes. People need a tourist map for traveling, a traffic map for driving, a military map for war. Although these maps all describe some characteristics of a region, they are not the whole picture of the real world. Why people need tourist map, traffic map and military map? The reason is that they are for different purposes. If people depict the entire real world into a map, although it completely describes the objective reality, what is the use of that map?

Economics uses the concise and profound characterization of economic environments to describe the causes of problems, conduct analysis of inherent logic, and thus obtain logical conclusions and inferences. A good economist is sophisticated in accurately grasping the most essential characteristics of the current economic situation in his/her study. Only when we truly make clear the reason and current situation can we solve the concrete problems with specific proper remedy (economic theory adopted). Of course, to do this, one needs to have the basic training in economic theory, which is one of main purposes of the notes.

1.10.2 Making Behavioral Assumptions

The second basic component of the analytical framework of modern economics is to make assumptions on individuals’ behavior. This is the key difference between economics and natural science. Whether an economic theory is convincing, feasible and whether an institutional arrangement or economic policy is conducive to sustainable and rapid economic development mainly depends on whether the assumed individual behavior truly reflects the behavior of most people.

In general, under a given environment and game rules, individuals will make trade-offs according to their behavioral disposition. Thus, when deciding the rules of game, policies, regulations or institutional arrangements, we need to take into account the behavioral pattern of participants and make correct judgments.
Like dealing with different people in daily life, we need to know whether they are selfless, honest or not. Different rules of game should be instituted when faced with different participants. When facing an honest person who tends to tell the truth, the way to deal with him or the rules imposed on him may often be comparatively simple. One does not need to invest much energy (in designing the rules of game) to deal with him and the rules may seem not so important. On the contrary, when facing a cunning and dishonest person, the way to deal with him will be quite different and careful so that the rules will be much more complicated. We need to pay special attention. As such, making right judgments on individuals’ behavior is a very important step for the study of how people react to incentives and make trade-off choice.

As mentioned above, under normal circumstances, a reasonable and realistic assumption about individuals’ behavior used by most economists is self-interest assumption, or the stronger rationality assumption, i.e., economic agents pursue the maximization of benefits. Bounded rationality assumption is to make the optimal choice according to the knowledge and information an agent has, which belongs to the category of rationality assumption in any case. In the consumer theory to be discussed later, we assume that consumer pursues the maximization of utility/satisfaction; in the producer theory, we assume that producer pursues the maximization of profits; in game theory, there are various equilibrium solution concepts describing the behavior of economic agents. These concepts are given based on different behavioral assumptions. Any individual, in his contact with others, implicitly assumes others’ behavior.

The assumption of (bounded) rationality is largely reasonable in some sense. From a practical point of view, as mentioned before, there are three basic kinds of institutional arrangements: mandatory regulation institutional arrangements (for situations with small operation cost and relatively easy information symmetry), incentive mechanism (for information asymmetry), and social norms (composed of ideology, ideal, morals, customs, etc. that give people self-discipline norms regulating their behavior). If all the people are of very high ideological level and
are selfless, there is no need for the rigid “stick-style” mandatory regulation or the flexible inducing incentive (such as market) system.

However, the ideal state is not equal to the reality. Assuming that individual is rational and self-interested not only conforms to the reality, but more importantly, the risk to the society is minimal. Even if self-interest assumption is wrong, it will not lead to serious consequences. On the contrary, once altruistic behavior assumption is inappropriate, when facing individuals with inconsistent words and deeds, the consequences will be much more serious than those of incorrect self-interest assumption. Especially, if some ill-minded and tricky people have their way, it will cause serious, or even disastrous loss to the state, firms and individuals.

The essential reason for China’s implementation of market economy is that under normal circumstances, individuals are self-interested, and market economy is in conformity with the self-interest assumption. This is also the foundation for the institutional arrangements we will discuss later.

1.10.3 Setting Economic Institutional Arrangements

The third basic component of the analytical framework of modern economics is to set up the economic institutional arrangements, which are usually referred to as the rules of game. Different strategies or game rules should be taken for different situations, different environments, and individuals with different behavioral manners. When the situation or environment changes, the strategies or game rules will also change accordingly in most cases. Determining rules of game is very important. Different rules of game will lead to different reactions, trade-off decisions and results. This also applies to the study of economics. When an economic environment is given, agents need to decide the economic rules of game, which is called economic institutional arrangement in economics. Modern economics studies and gives various economic institutional arrangements (economic mechanisms) according to different economic environments and behavioral assumptions. Depending on the issue under consideration, an economic institutional arrangement could be exogenously given or endogenously determined.
Any theory of modern economics involves economic institutional arrangements. Standard modern economics mainly focuses on the market institutional arrangements. It studies how individuals make trade-off decisions in a market system (such as the consumer theory, firm theory and general equilibrium theory) and under what economic environments will market equilibrium exist. It also makes value judgments on the allocation of resources under different market structures (the criterion is based on whether the allocation results are optimal and fair). In these studies, market institutional arrangements are normally assumed to be exogenously given. By so doing we can simplify the issues so as to focus on the study of individuals’ economic behavior and how people make trade-offs.

Of course, the exogeneity assumption of institutional arrangements is not entirely reasonable in many cases. Different economic institutional arrangements should be given depending on different economic environments and individuals’ behavioral patterns. As to be discussed in Parts 4 and 5 of the notes, there will be market failure (i.e., inefficient allocation of resources and non-existence of market equilibrium) in many situations, which makes people try to find an alternative mechanism or better mechanism. In that case we need to treat institutional arrangement as an endogenous variable which is determined by economic environments and individual behavior. Thus, economists should consider and give a variety of alternative economic institutional arrangements for different choices.

When studying the economic behavior and choice issues of a specific economic organization, economic institutional arrangements should especially be endogenously determined. New institutional economics, transition economics, modern theory of the firm, especially the economic mechanism design theory, information economics, optimal contract theory and auction theory that have developed in the last few decades, study and give various economic institutional arrangements according to different economic environments and behavioral assumptions for a wide range from the state to the family.
1.10.4 Determining Equilibrium

The fourth basic component of the analytical framework of modern economics is to make trade-off choices and determine the “best” outcome agent thinks. Given an economic environment, institutional arrangement (rules of game) and other constraints, individuals will react to incentives based on their own behavior, weigh and choose an outcome from many available or feasible outcomes. Such an outcome is called equilibrium. In fact, the concept of equilibrium is not hard to understand. It means that from the various feasible and available choices, people need to choose one, and the one being finally chosen is the equilibrium. Those who are self-interested will choose the best one for themselves; those who are altruistic may choose an outcome that is favorable to others. Thus, the so-called equilibrium, which refers to a state without incentive for deviation to all economic agents, is a static concept.

The equilibrium defined above may be the most general definition in economics. It embraces the equilibria reached by independent decision under the drive of self-interested motivation and all kinds of technology or budget constraints in textbooks. For instance, under market system, a profit maximization production plan under the constraint of technology of a firm is called equilibrium production plan; a utility maximization consumption bundle under the budget constraint of a consumer is called equilibrium consumption. When producers, consumers and their interaction reach a state where there is no incentive for deviation, market competitive equilibrium for every commodity is obtained.

It should be noted that equilibrium is a relative concept. The equilibrium outcome depends on economic environments, individuals’ behavior (whether in terms of rationality assumption, bounded rationality assumption, or other behavioral assumptions), and the rules of game by which individuals react, which is the “best” choice relative to these factors. Note that due to bounded rationality, it may not the optimal choice in objective reality, but the “best” one regarded by individuals themselves.
1.10.5 Having Evaluations

The fifth basic component of the analytical framework of modern economics is to make evaluations and value judgments on the chosen equilibrium outcome under some economic institutional arrangement and trade-off choice. After making their choices, individuals usually hope to evaluate the chosen equilibrium and compare it with the ideal “best” possible outcome (for instance, efficient resource allocation, fair resource allocation, incentive compatibility, informational efficiency, etc.), and then make further assessments and value judgments on the economic institutional arrangement—whether the economic institutional arrangement adopted leads to some “optimal” outcome.

Besides, they need to test whether the theoretical results are consistent with the empirical reality and whether it can provide correct prediction or practical significance; finally, they make conclusions about whether the economic institution and rules are good or bad in order to find out if there is room for improvement. In short, to achieve better results for doing something, after finishing it, we should evaluate the effects, whether it is worth continuing, and whether there is possibility for improvement. Thus, we need to make evaluations and value judgments on equilibrium outcome under some economic institutional arrangement and trade-off choice in order to find out the institutions that are best suited to national development.

When making evaluation on an economic mechanism or institutional arrangement, one of the most important criteria adopted in modern economics is whether the institutional arrangement is in line with the principle of efficiency. Yet Pareto optimality is not the sole criterion, another value judgment is equality or fairness. Market system achieves efficient allocation of resources, but it also faces many problems, such as social injustice caused by huge wealth gap. There are a variety of definitions of equality and fairness. The fair allocation to be introduced in Chapter 9 of the notes takes into account both the objective equality and the subjective factors, and more importantly, it can solve the problems of fairness and efficiency at the same time. This is the basic conclusion of the Outcome
Fairness Theorem to be introduced in Chapter 9. Another important criterion for evaluating an economic institutional arrangement is incentive compatibility.

In summary, the five components discussed above constitute the analytical framework underlying almost all standard economic theories, no matter how much mathematics is used, or whether the institutional arrangement is exogenously given or endogenously determined. In the study of economic issues, we should first define the economic environment, and then examine how the self-interested behavior of individuals affects each other under exogenously given or endogenously determined mechanism. Economists usually take “equilibrium”, “efficiency”, “information” and “incentive compatibility” as the key aspects of consideration, observe the effects of different mechanisms on individual behavior and economic organizations, explain how individual behaviors achieve equilibrium, and evaluate the equilibrium. Using such a basic analytical framework in economic issues is not only compatible in methodology, but may also lead to surprising (but logically consistent) conclusions.

1.11 Basic Research Methodologies in Modern Economics

We have discussed the five components of the basic analytical framework of modern economics: specifying economic environments, imposing behavioral assumptions, providing institutional arrangements, choosing the equilibrium, and making evaluations. Broadly speaking, any economic theory consists of these five aspects. Then the question turns out to be how to combine them appropriately according to scientific research methodologies, to gradually and thoroughly study various economic phenomena and to develop new economic theories. This is what we will discuss in this section: the basic research methodologies and key points, which include establishing benchmarks and reference system, providing studying platform,
developing analytical tools, and conducting positive and normative analyses, well grasping the basic requirements, functions and key points of modern economic theories, distinguishing sufficient and necessary conditions, and understanding the relationship between modern economics and mathematics, etc.

In the following we discuss some basic research methodologies of modern economics. First, we provide basic studying platforms for all levels and aspects, and then establish reference system so as to present the criterion to evaluate an equilibrium outcome and institutional arrangements. Providing studying platform and establishing reference system are of great importance to the construction and development of any subject, and economics is no exception.

1.11.1 Establishing Benchmarks

In order to study and compare various economic issues in reality, we need to first establish benchmarks. Benchmark refers to a relatively ideal or simple economic environment as the first step to study more realistic economic issues. To study more realistic issues and develop new theories, we usually need to first figure out what the results or theories are in a friction-free ideal economic environment. Then we discuss the results in a non-ideal economic environment with frictions, which is closer to reality, develop a more general theory and compare it with the previous theories. In this sense, benchmark is a relative term in relation to irregular economic environments and new theories that are to be developed and closer to reality.

For instance, complete information is the benchmark for the study of incomplete information. When studying economic issues under information asymmetry, first and foremost we need to fully understand the situation of complete information (highly unrealistic as it is). Only when we are thoroughly clear of the situation of complete information can we study well how things are going in a situation of incomplete information. So is the case with paper writing. We start from the ideal state or simple scenarios before considering more realistic or general scenarios; we learn from others’ study results before innovating on the
existing theories. Similar to the natural science, when it comes to vital economic theories, we start from the friction-free ideal state or simple scenarios before considering more realistic, non-ideal state with frictions or general scenarios. New theories are always based on the prior research findings and results. For instance, we first have Newton’s mechanics, then Einstein’s theory of relativity, and later on the non-conservation of parity put forward by Chen-Ning Franklin Yang and Tsung-Dao Lee.

1.11.2 Setting Reference System

Reference system refers to the standard economic model in an ideal situation, which leads to the ideal result, such as efficient resource allocation. Setting reference system is of great importance to the construction and development of any subject, including economics. Although for the economic theories as the reference system, there are many discrepancies between the assumptions and the reality, at least they help in: (1) simplifying the issue and capturing its characteristics directly; (2) establishing the measurement criterion for evaluating theoretical model, understanding reality, and making improvement; (3) theoretical innovation and further analysis.

Although the economic theories as the reference system might have many unrealistic assumptions, they are very useful and can serve as the reference for further analysis. This is similar to our practice of setting role models in life. The importance of the reference does not lie in the fact that it accurately describes the reality or not, but consists in establishing the measurement for better understanding the reality. Like a mirror, it helps reveal the gap between the theoretical outcomes or realistic economic institutions and the ideal state. It is essentially important in the sense that it points out the direction as well as the extent of efforts and adjustments. Suppose that if a person has no target, and is unaware of where to put efforts and where the future lies, how can he make improvements or be motivated to do anything? Not to mention to achieve his goal.

General equilibrium theory we will study in the notes is such a reference sys-
tem. We know that a perfectly competitive market will lead to efficient allocation of resources. Although there is no such market in reality, if we make efforts toward that direction, efficiency will be enhanced. That is why we have anti-trust laws as the institutional arrangement protecting market competition. By virtue of the reference system with perfectly competitive market as a benchmark, we can study what results can be obtained from economic institutional arrangements that are closer to the reality (such as some kind of monopolistic or transitional economic institution arrangements) with invalid assumptions in the general equilibrium theory (incomplete information, imperfect competition, externalities), and compare them with the results obtained from the general equilibrium theory in the ideal state. By comparing with the ideal institutional arrangement of perfectly competitive market, we will know whether an economic institutional arrangement (be it theoretical or realistic) is efficient in allocating resources and utilizing information, and how far away the economic institutional arrangement adopted in reality is from the ideal situation. Based on this, we can make policy suggestions accordingly. In this sense, general equilibrium theory also serves as a criterion for evaluating institutional arrangements and the corresponding economic policies in reality.

1.11.3 Developing Studying Platform

A studying platform in modern economics consists of some basic economic theories and methods, and also provides the basis for deeper analysis. The methodology of modern economics is very similar to that of physics, i.e., simplifying the issue first to capture the core essence. In case that many factors breed an economic phenomenon, we need to make clear the impact of every factor. This can be done by studying the effect of one factor at a time while assuming other factors remain unchanged. The theoretical foundation of modern economics is modern microeconomics, and the most fundamental theory in microeconomics is individual choice theory-consumer theory and firm theory, which are the basic studying platform and cornerstone of modern economics. That is the reason why
almost all the textbooks in modern economics start from consumer theory and firm theory. They provide the fundamental theories explaining how the consumer and firm make choices and establish the studying platform for further study of individual choice.

Generally speaking, the equilibrium choice of individual depends not only on one’s own choice, but also on others’ choices. In order to study individual choice, we need to clearly understand how an individual make his choice in the absence of influence by other agents. The consumer theory and firm theory to be discussed in the notes is obtained through this approach. In these theoretical models, economic agents are assumed to be in the perfectly competitive market institutional arrangement. Therefore, every agent will take price as a given parameter, and individual choice will not be affected by others’ choice. The optimal choice is determined by subjective factors (such as the pursuit of maximum utility or profit) and objective factors (such as budget line or production constraints).

Many people find this research method bewildering. They think the assumptions in the theories seem too unpractical and thus doubt whether modern economic theory is of any use. Actually, this sort of criticism indicates that they haven’t truly understood this scientific research methodology yet. This methodology that involves simplification and idealization forms a platform for extending the research. It is like the approach in physics: in order to study a problem, we go to the essence first, start with the simplest situations, and then consider the more general and complicated cases. In microeconomics, the theory of market structures like monopoly, oligopoly and monopolistic competition is attained in the more general cases, where producers can interact with each other. To study the choice issue in the more general situations, economists have developed a very powerful analytical tool – game theory.

General equilibrium theory is a more sophisticated studying platform based on consumer theory and firm theory. The consumer theory and firm theory provide a fundamental platform for studying individual choice problems, while general equilibrium theory provides a fundamental platform for analyzing how to reach
market equilibrium through the interaction of all the goods in the market. The mechanism design theory provides a higher-level platform for studying, designing and evaluating various institutional arrangements and economic mechanisms (whether it is public ownership, private ownership, or mixed ownership). It can be used to study and prove the optimality and uniqueness of perfectly competitive market mechanism in resource allocation and information utilization, and more importantly, in case of market failure, it offers how to design alternative mechanisms. Under some regularity conditions, perfect competition market institutional arrangement free of externalities leads to efficient allocation of resources. Moreover, from the perspective of utilizing information (institutional implementation cost and transaction cost), it uses the least information, which demonstrates that it is the most efficient way of information utilization. In other circumstances of market failure, we need to design a variety of alternative mechanisms under different economic environments. Furthermore, the studying platform also creates conditions for providing reference systems for evaluating various kinds of economic institutional arrangements. In other words, it provides a criterion for measuring the gap between reality and the ideal state.

1.11.4 Providing Analytical Tools

For the research of economic phenomena and economic behavior, we also need various analytical tools besides the analytical framework, benchmark, measurement criteria and studying platform. Analytical tools are usually given by mathematical or graphic models. The advantages of these tools lie in that they help us to analyze the intricate economic phenomena and economic behavior through a simple and clear diagram or mathematical model. Examples include: the demand-supply curve, game theory, principal-agent theory for studying asymmetric information, Paul Samuelson’s overlapping generations model, dynamic optimality theory, etc. Of course, there are exceptions, which are not expressed with analytical tools. For instance, the original Coase theorem is established and demonstrated through words and basic logical deduction only.
1.11.5 Constructing Rigorous Analytical Models

Logical and rigorous theoretical analysis is needed when we give explanation to economic phenomena or economic behavior and make conclusions or economic inferences. As is mentioned above, each theory holds true under certain conditions. Modern economics demands not only qualitative analysis but also quantitative analysis. We need to define the boundaries within which each theory holds in order to avoid the abuse of theory. This is similar to medication and pharmacology where we have to be clear of the application and functions of the medicines. Therefore, we need to establish a rigorous analytical model clearly defining the conditions under which a theory holds true. Lack of related mathematical knowledge will make it difficult for us to have an accurate understanding of the connotation of a definition, or to make discussions on related issues, not to mention defining the boundary conditions or constraints for the research work. No wonder mathematics and mathematical statistics are used as basic analytical tools, and they are also one of the most important research methods in modern economics.

1.11.6 Positive Analysis and Normative Analysis

As per the research method, economic analysis can be divided into two types: positive or descriptive analysis; normative or value analysis. A major difference between economics and natural science is that the latter only makes positive analysis while the former involves both positive analysis and normative analysis in its discussions.

Positive analysis only explains what economics is. It only offers the facts and gives explanations (thus verifiable), but does not make value assessment to economic phenomena. For example, an important task of modern economics is to describe, compare and analyze such phenomena as production, consumption, unemployment and price, and to predict possible outcomes of different policies. Consumer theory, theory of the firm and game theory are typical examples of
positive analysis.

Normative analysis makes judgments on economic phenomena. It not only explains what economics is, but also provides opinions or value judgments. Therefore, it always involves value judgments and preferences of the economists and is thus not verifiable through facts. For instance, some economists lay more emphasis on economic benefits while others focus on equality or social justice. With the differences between the two methods in mind, we can avoid many disputes while discussing economic issues. Economic mechanism design theory is a typical example of normative analysis.

Positive analysis is the foundation for normative analysis, while normative analysis is the extension of positive analysis. In this sense, the foremost task of economics is to make positive analysis and then normative analysis. General equilibrium theory to be discussed in the notes includes both the former (the existence, stability and uniqueness of competitive equilibrium) and the latter (the first and second welfare theorems).

1.12 Practical Role of the Analytic Framework and Methodologies

1.13 Basic Requirements for Emulating Good Modern Economics

There are three basic requirements for understanding modern economic theory:

1. You have to learn well the basic concepts and definitions, which is a reflection of a clear mind. This is the prerequisite not only for the discussion and analysis, but also for a good command of economics. Otherwise, different definitions of the term may cause great confusion and thus lead to unnecessary disputes.

2. The statement of all theorems or propositions should be clear, and the basic
conclusions and their conditions should be explicit. Otherwise, a small error in applying economic theories to the analysis of issues will lead to serious mistakes. There is a scope of application for medicines. Likewise, any theory or institution has its application boundary. If we go beyond that boundary, problems are liable to arise, bringing about tremendous negative externalities. Many of the social or economic problems occur because economists misuse some theories without a good understanding of their boundaries and applicable conditions. In this sense, an eligible economist is like a qualified doctor who prescribes for his patients and needs to know well the properties and efficacy of different kinds of medicines.

3. You have to grasp how the basic theorems or propositions (ideas and processes) are proved. An excellent economist, like a good doctor, should know the medicine properties and pathology as well. Only when he is aware of both the effect and the cause can he gain a deeper understanding and a better command of the theories he has learned.

If you meet the above-mentioned requirements, it would be pretty easy to refresh your memory even if you forget the proof of some conclusions of propositions. Economics relies mainly upon analysis of internal logic rather than experiments, which is also the power of modern economic theory. That is why it is important to grasp well the theories and their application scope.

1.14 Distinguishing Sufficient and Necessary Conditions

When discussing economic issues, it is very important to distinguish between sufficient conditions and necessary conditions. It can help people think clearly and avoid unnecessary debates. A necessary condition is a condition that is indispensable in order for a proposition to be true. A sufficient condition is a condition that guarantees the proposition to be true. For instance, some people often negate market economy based on the fact that many countries adopt market economy but remains poor. They argue that China should not embark on the
path of market economy. These people do not realize the difference between sufficient and necessary conditions: the adoption of market economy is a necessary condition rather than a sufficient condition for a country to become prosperous and strong. In other words, if a country wants to be prosperous, it must adopt market economy. This is because one cannot find any wealthy country in the world that is not a market economy.

However, market mechanism is just a necessary condition for prosperity. We must also admit that market mechanism does not necessarily lead to national prosperity unless it clearly defines the governance boundary among government, market and society. As mentioned above, there is the distinction between good market economy and bad market economy. The reason is that although (based on observation of reality) market mechanism is indispensable for national prosperity, there are many other factors such as the degree of government intervention, political system, legal system, religion, culture and social structure that can affect the result. Hence, market mechanism may be labeled as good or bad.

1.15 The Role of Mathematics in Modern Economics

Almost every field in modern economics uses a lot of mathematics, statistics and econometrics. The width and depth of the use of mathematics even exceed that in physical sciences. The reasons for this include: the fact that modern economics is increasingly becoming a science, the use of mathematical analytical tools, as well as the more complex and influential social systems. Thus, when considering and studying economic issues, we need to make inherent logical analysis by rigorous theoretical analytical models and conduct empirical testing by quantitative analysis methods, and figure out the specific conditions for obtaining a conclusion. Hence, it is not surprising that mathematics and mathematical statistics are used as the basic analytical tools, and also become the most important analytical tools in modern economics. People who study modern economics and conduct research
must grasp the necessary knowledge of mathematics and mathematical statistics.

Modern Economics mainly adopts mathematical language to make assumptions on economic environments and individual behavior patterns, uses mathematical expressions to illustrate the logical relations between each economic variable and economic rules, builds mathematical models to study economic issues, and finally follows the logic of mathematical language to deduce conclusions. Without the related mathematical knowledge, it is hard to grasp the essence of concepts and discuss the related issues, let alone conduct research and figure out the necessary boundary or constraints before giving conclusions. Therefore, it is of great necessity to master sufficient mathematical knowledge if you want to learn modern economics well, conduct economic research and become a good economist.

Many people have little knowledge of mathematics and are unable to master the basic theories and analytical tools of modern economics, or understand advanced economic textbooks or papers. Thus, they deny the function of mathematics in economic studies with excuses such as it is important to produce economic thoughts, or mathematics is far away from the practical economic issues. No one could deny the importance of economic thoughts. It is the output of research. However, without taking mathematics as a tool, how could we figure out the boundary conditions and applicable scope of economic thought or conclusions? How could we defend against the abuse or misuse of economic thoughts or conclusions? How many people could develop such profound economic thoughts without using mathematical models as those offered by Adam Smith and Ronald H. Coase? Even if such people do exist, economists have never stopped seeking restraints to the conclusions. Besides, the time in which we are living is different, and modern economics has become a quite rigorous discipline in social sciences. Without strict argument, the thoughts or conclusions could not be widely recognized. As mentioned above, economic thoughts presented by philosophers in ancient China like JIANG Shang, LAO Tzu, SUN Tzu, GUAN Zhong and SIMA Qian are extremely profound, and some thoughts of Adam Smith had already
been mentioned by these philosophers long ago. However, these thoughts have never been recognized by the outside world because being conclusions of experience, they did not form a scientific system nor being logically analyzed by strict scientific methods.

There is another misunderstanding that it is beyond reality to apply mathematics into economic studies. Most of the mathematical knowledge is developed from solving real world problems. People who have basic knowledge of physics, physical history or mathematical history will know that either primary or advanced mathematics is originated from the demands of scientific development and reality. As such, why not use mathematics to study practical economic issues? The foundation of mathematics and modern economics is very important for people who want to be a good economist. If you know mathematics well and master the fundamental analytical framework and research methodology of modern economics, you could learn modern economics more easily and greatly enhance study efficiency.

The functions of mathematics in theoretical analysis of modern economics are as follows: (1) It makes the language more precise and concise and the statement of assumptions more clear, which can reduce many unnecessary debates resulting from inaccurate definitions. (2) It makes the analytical logic more rigorous and clearly states the boundary, application scope and conditions for a conclusion to hold. Otherwise, the abuse of a theory may occur. (3) Mathematics can help obtain the results that cannot be easily attained through intuition. For example, from intuition, according to the law of supply and demand, competitive markets will achieve market equilibrium by the “invisible hand” through adjusting market prices as long as the supply and demand are not equal. Yet this conclusion is not always true. Scarf (1960), as we will discuss in Chapter 7, gave a counterexample to prove that this result may not be true in some cases. (4) It helps improve and extend the existing economic theory. The examples in this respect are plenty in the study of economic theory. For instance, economic mechanism design theory is an improvement and extension of general equilibrium theory.
Qualitative theoretical analysis and quantitative empirical analysis are both needed for economic issues. Economic statistics and econometrics play an important role in these analyses. Economic statistics focuses more on data collection, description, sorting and providing statistical methods, whereas econometrics focuses more on testing economic theory, evaluating economic policy, making economic forecasts and identifying the causal relationship between economic variables. In order to better evaluate economic models and make more accurate predictions, theoretical econometricians have been developing increasingly powerful econometric tools.

It is, however, noteworthy that economics is not mathematics. Mathematics in economics is only used as a tool to consider or study economic behavior or phenomenon. Economists just employ mathematics to express their opinions and theories more rigorously and concisely and to analyze the interdependent relationship among each economic variable by mathematical models. With the metrization of economics and the precision of various assumptions, economics has become a social science with a rigorous system.

However, a good master of mathematics cannot guarantee to be a good economist. It also requires fully understanding the analytical framework and research methodologies of modern economics, and having a good intuition and insight of real economic environments and economic issues. The study of economics not only calls for the understanding of some terms, concepts and results from the perspective of mathematics (including geometry), but more importantly, even when those are given by mathematical language or geometric figure, we need to get to their economic meaning and the underlying profound economic thought. Thus we should avoid being confused by the mathematical formulas or symbols in the study of economics. To become a good economist, you need to be of original, creative research and academic way of thinking.
1.16 Conversion between Economic and Mathematical Language

The product of economic research is economic judgment and conclusion. A standard economic paper usually consists of three parts: (1) it raises questions, states the significance and identifies the research objective; (2) it establishes economic models, rigorously expresses and proves the conclusions; (3) it uses non-technical language to explain the conclusions and if relevant, provides policy suggestions.

That is to say, an economic conclusion is usually obtained through the following three stages: non-mathematical language stage - mathematical language stage - non-mathematical language stage. The first stage proposes economic ideas, concepts or conjectures, which may stem from economic intuition or historical and geographical observations. As they have not been proved in this stage, they can be regarded as the primary products of general production. This stage is very important because it is the origin of theoretical research and innovation.

The second stage verifies whether the proposed economic ideas or conjectures hold or not. The verification requires economists to give formal and rigorous proofs through economic models and analytical tools, and if possible, to test them with real empirical data. The results obtained are usually expressed in mathematical language or technical terms, which may not be understandable to non-experts. Therefore, they may not be adopted by the public, government officials or policy makers. Thus, these conclusions expressed by technical language can be regarded as the intermediate products of general production.

Economic studies serves for the real economic society. Therefore, the third phase is to express the conclusions and inferences by common language rather than technical language, making them more understandable to noneconomists. The final products of economics are the policy implications, profound meanings of conclusions and the insightful inferences conveyed through non-technical language. It is notable that the economic ideas and conclusions are presented by common, non-technical and non-mathematical language both in the first and
third phase, but the third phase is a kind of enhancement of the first phase. As a matter of fact, the three-stage model of common language - technical language - common language is a normal research method widely adopted by many disciplines.

1.17 Reference

Books and Monographs:


弗里德利希·冯·哈耶克(2000). 法律、立法与自由（第二、三卷）. 中译本，邓正来等译. 北京：中国大百科全书出版社.


Friedman, M. and R. Friedman (1980). Free to Choose, HBJ.


Smith, Adam (1759). *The Theory of Moral Sentiments*. (中译本：亚当·斯密. 道德情操论. 谢宗林译. 北京：中央编译出版社.)


Papers:

乔尔·S·赫尔曼. 转型经济中对抗政府俘获和行政腐败的战略. 见：叶谦，宾建成编译. 经济社会体制比较，2009年第2期.

热若尔·罗兰. 理解制度变迁：迅捷变革的制度与缓慢演进的制度. 南大商学院评论. 2005年第5辑.

钱颖一. 理解现代经济学. 经济社会体制比较，2002年第2期.

钱颖一. 市场与法治. 经济社会体制比较，2000年第3期.

田国强. 内生产权所有制理论与经济体制的平稳转型. 经济研究，1996年第11期.

田国强. 现代经济学的基本分析框架与研究方法. 经济研究，2005年第2期.

田国强. 和谐社会的构建与现代市场体系的完善——效率、公平与法治. 经济研究，2007年第3期.

田国强. 从拨乱反正、市场经济到和谐社会构建——效率、公平与和解发展的关键是合理界定政府与市场的边界. 《文汇报》、《解放日报》及上海管理科学研究院“中国改革开放与发展30年”征文优秀论文稿，2008年7月.

田国强. 经济学的思想与方法. 论文稿，2009年10月.

田国强. 中国经济发展中的深层次问题. 学术月刊，2011年第3期.

田国强. 中国下一步的改革与政府职能转变. 人民论坛·学术前沿. 2012年第3期.

田国强. 中国经济转型的内涵特征与现实瓶颈解读. 人民论坛，2012年第35期.

田国强. 世界变局下的中国改革与政府职能转变. 学术月刊，2012年第6期.

田国强. 中国改革的未来之路及其突破口. 比较，2013年第1期.

田国强. “中等收入陷阱”与国家公共治理模式重构. 人民论坛，2013年第8期.
田国强. 法治：现代治理体系的重要基石. 人民论坛·学术前沿，2013年第23期.

田国强. 现代国家治理视野下的中国政治体制改革. 学术月刊，2014年第3期.

田国强. 近现代中国的四次社会经济大变革——国企改革的镜鉴与反思. 探索与争鸣，2014年第6期.

田国强. 当前中国经济增速的合理区间探讨——发展和治理两大逻辑如何统筹兼顾. 人民论坛·学术前沿，2015年第6期.

田国强. 重构新时期政商关系的抓手. 人民论坛·学术前沿，2015年第5期.

田国强，夏纪军，陈旭东. 破除中国模式迷思 坚持市场导向改革. 比较，2010年第50辑.

田国强，杨立岩. 对“幸福—收入之谜”的一种解释：理论与实证.经济研究，2006年第11期.

王一江. 国家与经济. 比较，2005年第18辑。


Chapter 2

Knowledge and Methods of Mathematics

This chapter discusses some basic mathematics preliminaries and results that will be used in the lecture notes. It reviews some basic mathematics results such as: continuity and concavity of functions, Separating Hyperplane Theorem, optimization, correspondences (point to set mappings), fixed point theorems, KKM lemma, maximum theorem, etc., which will be used to prove some results in the lecture notes. For good references about the materials discussed in this section, see appendixes in Hildenbrand and Kirman (1988), Mas-Colell (1995), and Varian (1992).

2.1 Basic Set Theory

In progress

2.2 Basic Linear Algebra

In progress
2.3 Basic Topology

In progress

2.4 Single-Valued Function and Its properties

Let $X$ and $Y$ be two subsets of Euclidian spaces. In this text, vector inequalities, $\geq$, $\geq$, and $>$, are defined as follows: Let $a, b \in \mathbb{R}^n$. Then $a \geq b$ means $a_s \geq b_s$ for all $s = 1, \ldots, n$; $a \geq b$ means $a \geq b$ but $a \neq b$; $a > b$ means $a_s > b_s$ for all $s = 1, \ldots, n$.

Definition 2.4.1 A function $f : X \to \mathbb{R}$ is said to be continuous if at point $x_0 \in X$,

$$\lim_{x \to x_0} f(x) = f(x_0),$$

or equivalently, for any $\epsilon > 0$, there is a $\delta > 0$ such that for any $x \in X$ satisfying $|x - x_0| < \delta$, we have

$$|f(x) - f(x_0)| < \epsilon$$

A function $f : X \to \mathbb{R}$ is said to be continuous on $X$ if $f$ is continuous at every point $x \in X$.

The idea of continuity is pretty straightforward: There is no disconnected point if we draw a function as a curve. A function is continuous if “small” changes in $x$ produces “small” changes in $f(x)$.

The so-called upper semi-continuity and lower semi-continuity continuities are weaker than continuity. Even weak conditions on continuity are transfer continuity which characterize many optimization problems and can be found in Tian (1992, 1993, 1994) and Tian and Zhou (1995), and Zhou and Tian (1992).

Definition 2.4.2 A function $f : X \to \mathbb{R}$ is said to be upper semi-continuous if at point $x_0 \in X$, we have

$$\limsup_{x \to x_0} f(x) \leq f(x_0),$$
or equivalently, for any \( \epsilon > 0 \), there is a \( \delta > 0 \) such that for any \( x \in X \) satisfying \( |x - x_0| < \delta \), we have

\[
f(x) < f(x_0) + \epsilon.
\]

Although all the three definitions on the upper semi-continuity at \( x_0 \) are equivalent, the second one is easier to be versified.

A function \( f : X \to \mathbb{R} \) is said to be upper semi-continuous on \( X \) if \( f \) is upper semi-continuous at every point \( x \in X \).

**Definition 2.4.3** A function \( f : X \to \mathbb{R} \) is said to be lower semi-continuous on \( X \) if \( -f \) is upper semi-continuous.

It is clear that a function \( f : X \to \mathbb{R} \) is continuous on \( X \) if and only if it is both upper and lower semi-continuous, or equivalently, for all \( x \in X \), the upper contour set \( U(x) \equiv \{ x' \in X : f(x') \geq f(x) \} \) and the lower contour set \( L(x) \equiv \{ x' \in X : f(x') \leq f(x) \} \) are closed subsets of \( X \).

Let \( f \) be a function on \( \mathbb{R}^k \) with continuous partial derivatives. We define the gradient of \( f \) to be the vector

\[
Df(x) = \left[ \frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_k} \right].
\]

Suppose \( f \) has continuous second order partial derivatives. We define the Hessian of \( f \) at \( x \) to be the \( n \times n \) matrix denoted by \( D^2 f(x) \) as

\[
D^2 f(x) = \left[ \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right],
\]

which is symmetric since

\[
\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}.
\]

**Definition 2.4.4** A function \( f : X \to \mathbb{R} \) is said to be homogeneous of degree \( k \) if for every \( t \), \( f(tx) = t^k f(x) \).

An important result concerning homogeneous function is the following:

**Theorem 2.4.1 (Euler’s Theorem)** If a function \( f : \mathbb{R}^n \to \mathbb{R} \) is homogeneous of degree \( k \) if and only if

\[
kf(x) = \sum_{i=1}^{n} \frac{\partial f(x)}{\partial x_i} x_i.
\]
2.4.1 Separating Hyperplane Theorems

A set $X \subset \mathbb{R}^n$ is said to be **compact** if it is bounded and closed. A set $X$ is said to be **convex** if for any two points $x, x' \in X$, the point $tx + (1 - t)x' \in X$ for all $0 \leq t \leq 1$. Geometrically the convex set means every point on the line segment joining any two points in the set is also in the set.

**Theorem 2.4.2 (Separating Hyperplane Theorem)** Suppose that $A, B \subset \mathbb{R}^m$ are convex and $A \cap B = \emptyset$. Then, there is a vector $p \in \mathbb{R}^m$ with $p \neq 0$, and a value $c \in \mathbb{R}$ such that

$$px \leq c \leq py \quad \forall x \in A \& y \in B.$$  

Furthermore, suppose that $B \subset \mathbb{R}^m$ is convex and closed, $A \subset \mathbb{R}^m$ is convex and compact, and $A \cap B = \emptyset$. Then, there is a vector $p \in \mathbb{R}^m$ with $p \neq 0$, and a value $c \in \mathbb{R}$ such that

$$px < c < py \quad \forall x \in A \& y \in B.$$

2.4.2 Concave and Convex Function

Concave, convex, and quasi-concave functions arise frequently in microeconomics and have strong economic meanings. They also have a special role in optimization problems.

**Definition 2.4.5** Let $X$ be a convex set. A function $f : X \to \mathbb{R}$ is said to be **concave on** $X$ if for any $x, x' \in X$ and any $t$ with $0 \leq t \leq 1$, we have

$$f(tx + (1 - t)x') \geq tf(x) + (1 - t)f(x')$$

The function $f$ is said to be **strictly concave on** $X$ if

$$f(tx + (1 - t)x') > tf(x) + (1 - t)f(x')$$

for all $x \neq x' \in X$ an $0 < t < 1$.

A function $f : X \to \mathbb{R}$ is said to be (strictly) convex on $X$ if $-f$ is (strictly) concave on $X$. 

74
Remark 2.4.1 A linear function is both concave and convex. The sum of two concave (convex) functions is a concave (convex) function.

Remark 2.4.2 When a function $f$ defined on a convex set $X$ has continuous second partial derivatives, it is concave (convex) if and only if the Hessian matrix $D^2f(x)$ is negative (positive) semi-definite on $X$. It is strictly concave (strictly convex) if the Hessian matrix $D^2f(x)$ is negative (positive) definite on $X$.

Remark 2.4.3 The strict concavity of $f(x)$ can be checked by verifying if the leading principal minors of the Hessian must alternate in sign, i.e.,

$$
\begin{vmatrix}
  f_{11} & f_{12} \\
  f_{21} & f_{22}
\end{vmatrix} > 0,
$$

$$
\begin{vmatrix}
  f_{11} & f_{12} & f_{13} \\
  f_{21} & f_{22} & f_{23} \\
  f_{31} & f_{32} & f_{33}
\end{vmatrix} < 0,
$$

and so on, where $f_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$. This algebraic condition is useful for checking second-order conditions.

In economic theory quasi-concave functions are used frequently, especially for the representation of utility functions. Quasi-concave is somewhat weaker than concavity.

Definition 2.4.6 Let $X$ be a convex set. A function $f : X \to \mathbb{R}$ is said to be quasi-concave on $X$ if the set

$$\{x \in X : f(x) \geq c\}$$

is convex for all real numbers $c$. It is strictly quasi-concave on $X$ if

$$\{x \in X : f(x) > c\}$$

is convex for all real numbers $c$.

A function $f : X \to \mathbb{R}$ is said to be (strictly) quasi-convex on $X$ if $-f$ is (strictly) quasi-concave on $X$. 

75
**Remark 2.4.4** The sum of two quasi-concave functions in general is not a quasi-concave function. Any monotonic function defined on a subset of the one dimensional real space is both quasi-concave and quasi-convex.

**Remark 2.4.5** When a function $f$ defined on a convex set $X$ has continuous second partial derivatives, it is strictly quasi-concave (convex) if the naturally ordered principal minors of the bordered Hessian matrix $\bar{H}(x)$ alternate in sign, i.e.,

$$
\begin{vmatrix}
0 & f_1 & f_2 \\
f_1 & f_{11} & f_{12} \\
f_2 & f_{21} & f_{22}
\end{vmatrix} > 0,
$$

$$
\begin{vmatrix}
0 & f_1 & f_2 & f_3 \\
f_1 & f_{11} & f_{12} & f_{13} \\
f_2 & f_{21} & f_{22} & f_{23} \\
f_3 & f_{31} & f_{32} & f_{33}
\end{vmatrix} < 0,
$$

and so on.

### 2.5 Multi-Valued Functions and Its properties

#### 2.5.1 Point-to-Set Mappings

When a mapping is not a single-valued function, but is a point-to-set mapping, it is called a correspondence, or multi-valued functions. That is, a correspondence $F$ maps point $x$ in the domain $X \subseteq \mathbb{R}^n$ into sets in the range $Y \subseteq \mathbb{R}^m$, and it is denoted by $F : X \rightarrow 2^Y$. We also use $F : X \rightarrow Y$ to denote the mapping $F : X \rightarrow 2^Y$ in this lecture notes.

**Definition 2.5.1** A correspondence $F : X \rightarrow 2^Y$ is: (1) *non-empty valued* if the set $F(x)$ is non-empty for all $x \in X$; (2) *convex valued* if the set $F(x)$ is a convex set for all $x \in X$; (3) *closed valued* if the set $F(x)$ is a closed set for all $x \in X$; (4) *compact valued* if the set $F(x)$ is a compact set for all $x \in X$. 

76
Intuitively, a correspondence is continuous if small changes in $x$ produce small changes in the set $F(x)$. Unfortunately, giving a formal definition of continuity for correspondences is not so simple. Figure 2.1 shows a continuous correspondence.

![Figure 2.1: A continuous correspondence.](image)

The notions of hemi-continuity are usually defined in terms of sequences (see Debreu (1959) and Mask-Collell et al. (1995)), but, although they are relatively easy to verify, they are not intuitive and depend on the assumption that a correspondence is compacted-valued. The following definitions are more formal (see Border, 1988).

**Definition 2.5.2** A correspondence $F : X \to 2^Y$ is upper hemi-continuous at $x$ if for each open set $U$ containing $F(x)$, there is an open set $N(x)$ containing $x$ such that if $x' \in N(x)$, then $F(x') \subset U$. A correspondence $F : X \to 2^Y$ is upper hemi-continuous if it is upper hemi-continuous at every $x \in X$, or equivalently, if the set $\{x \in X : F(x) \subset V\}$ is open in $X$ for every open subset $V$ of $Y$.

**Remark 2.5.1** Upper hemi-continuity captures the idea that $F(x)$ will not “suddenly contain new points” just as we move past some point $x$, in other words, $F(x)$ does not suddenly becomes much larger if one changes the argument $x$ slightly. That is, if one starts at a point $x$ and moves a little way to $x'$, upper
hemi-continuity at $x$ implies that there will be no point in $F(x')$ that is not close to some point in $F(x)$.

**Definition 2.5.3** A correspondence $F : X \to 2^Y$ is said to be *lower hemi-continuous at $x$* if for every open set $V$ with $F(x) \cap V \neq \emptyset$, there exists a neighborhood $N(x)$ of $x$ such that $F(x') \cap V \neq \emptyset$ for all $x' \in N(x)$. A correspondence $F : X \to 2^Y$ is *lower hemi-continuous* if it is lower hemi-continuous at every $x \in X$, or equivalently, the set $\{x \in X : F(x) \cap V \neq \emptyset\}$ is open in $X$ for every open set $V$ of $Y$.

**Remark 2.5.2** Lower hemi-continuity captures the idea that any element in $F(x)$ can be “approached” from all directions, in other words, $F(x)$ does not suddenly becomes much smaller if one changes the argument $x$ slightly. That is, if one starts at some point $x$ and some point $y \in F(x)$, lower hemi-continuity at $x$ implies that if one moves a little way from $x$ to $x'$, there will be some $y' \in F(x')$ that is close to $y$.

**Remark 2.5.3** Based on the following two facts, both notions of hemi-continuity can be characterized by sequences.

(a) If a correspondence $F : X \to 2^Y$ is compacted-valued, then it is upper hemi-continuous if and only if for any $\{x_k\}$ with $x_k \to x$ and $\{y_k\}$ with $y_n \in F(x_k)$, there exists a converging subsequence $\{y_{km}\}$ of $\{y_k\}$, $y_{km} \to y$, such that $y \in F(x)$.

(b) A correspondence $F : X \to 2^Y$ is said to be *lower hemi-continuous at $x$* if and only if for any $\{x_k\}$ with $x_k \to x$ and $y \in F(x)$, there is a sequence $\{y_k\}$ with $y_k \to y$ and $y_n \in F(x_k)$.

**Definition 2.5.4** A correspondence $F : X \to 2^Y$ is said to be *closed at $x$* if for any $\{x_k\}$ with $x_k \to x$ and $\{y_k\}$ with $y_k \to y$, $y_n \in F(x_k)$ implies $y \in F(x)$. $F$ is said to be closed if $F$ is closed for all $x \in X$ or equivalently

$$Gr(F) = \{(x,y) \in X \times Y : y \in F(x)\}$$

is closed.
Remark 2.5.4 Regarding the relationship between upper hemi-continuity and closed graph, the following facts can be proved.

(i) If $Y$ is compact and $F : X \to 2^Y$ is closed-valued, then $F$ has closed graph implies it is upper hemi-continuous.

(ii) If $X$ and $Y$ are closed and $F : X \to 2^Y$ is closed-valued, then $F$ is upper hemi-continuous implies that it has closed graph.

Because of fact (i), a correspondence with closed graph is sometimes called upper hemi-continuity in the literature. But one should keep in mind that they are not the same in general. For example, let $F : \mathbb{R}_+ \to 2^\mathbb{R}$ be defined by

$$F(x) = \begin{cases} \{\frac{1}{x}\} & \text{if } x > 0 \\ \{0\} & \text{if } x = 0. \end{cases}$$

The correspondence is closed but not upper hemi-continuous.

Also, define $F : \mathbb{R}_+ \to 2^\mathbb{R}$ by $F(x) = (0, 1)$. Then $F$ is upper hemi-continuous but not closed.

Figure 2.2 shows the correspondence is upper hemi-continuous, but not lower hemi-continuous. To see why it is upper hemi-continuous, imagine an open interval $U$ that encompasses $F(x)$. Now consider moving a little to the left of $x$ to a point $x'$. Clearly $F(x') = \{\hat{y}\}$ is in the interval. Similarly, if we move to a point $x'$ a little to the right of $x$, then $F(x)$ will inside the interval so long as $x'$ is sufficiently close to $x$. So it is upper hemi-continuous. On the other hand, the correspondence it not lower hemi-continuous. To see this, consider the point $y \in F(x)$, and let $U$ be a very small interval around $y$ that does not include $\hat{y}$. If we take any open set $N(x)$ containing $x$, then it will contain some point $x'$ to the left of $x$. But then $F(x') = \{\hat{y}\}$ will contain no points near $y$, i.e., it will not intersect $U$. Thus, the correspondence is not lower hemi-continuous.

Figure 2.3 shows the correspondence is lower hemi-continuous, but not upper hemi-continuous. To see why it is lower hemi-continuous: For any $0 \leq x' \leq x$, note that $F(x') = \{\hat{y}\}$.

Then $x_n > 0$ for
sufficiently large $n$, $x_n \to x'$, $y_n \to \hat{y}$, and $y_n \in F(x_n) = \{\hat{y}\}$. So it is lower 
hemi-continuous. It is clearly lower hemi-continuous for $x_i > x$. Thus, it is lower 
hemi-continuous on $X$. On the other hand, the correspondence is not upper 
hemi-continuous. If we start at $x$ by noting that $F(x) = \{\hat{y}\}$, and make a small 
move to the right to a point $x'$, then $F(x')$ suddenly contains many points that 
are not close to $\hat{y}$. So this correspondence fails to be upper hemi-continuous.
Combining upper and lower hemi-continuity, we can define the continuity of a correspondence.

**Definition 2.5.5** A correspondence $F : X \to 2^Y$ at $x \in X$ is said to be *continuous* if it is both upper hemi-continuous and lower hemi-continuous at $x \in X$. A correspondence $F : X \to 2^Y$ is said to be *continuous* if it is both upper hemi-continuous and lower hemi-continuous.

**Remark 2.5.5** As it turns out, the notions of upper and hemi-continuous correspondence both reduce to the standard notion of continuity for a function if $F(\cdot)$ is a single-valued correspondence, i.e., a function. That is, $F(\cdot)$ is a single-valued upper (or lower) hemi-continuous correspondence if and only if it is a continuous function.

**Definition 2.5.6** A correspondence $F : X \to 2^Y$ is said to be *open* if its graph

$$Gr(F) = \{(x, y) \in X \times Y : y \in F(x)\}$$

is open.

**Definition 2.5.7** A correspondence $F : X \to 2^Y$ is said to have *upper open sections* if $F(x)$ is open for all $x \in X$.

A correspondence $F : X \to 2^Y$ is said to have *lower open sections* if its inverse set $F^{-1}(y) = \{x \in X : y \in F(x)\}$ is open.

**Remark 2.5.6** If a correspondence $F : X \to 2^Y$ has an open graph, then it has upper and lower open sections. If a correspondence $F : X \to 2^Y$ has lower open sections, then it must be lower hemi-continuous.

### 2.6 Static Optimization

Optimization is a fundamental tool for the development of modern economic analysis. Most economic models are described by the optimization models, and the analyses are based on the solution of optimization problems. Results of this section are used throughout the text.
2.6.1 Unconstrained Optimization

The basic optimization problem is that of maximizing or minimizing a function on some set. The basic and central result is the existence theorem of Weierstrass.

**Theorem 2.6.1 (Weierstrass Theorem)** Any upper (lower) semi continuous function reaches its maximum (minimum) on a compact set, and the set of maximum is compact.

2.6.2 Optimization with Equality Constraints

**EQUALITY CONSTRAINED OPTIMIZATION**

An optimization problem with equality constraints has the form

\[
\max f(x) \\
\text{such that} \quad h_1(x) = d_1 \\
h_2(x) = d_2 \\
\vdots \\
h_k(x) = d_k,
\]

where \(f, h_1, \ldots, h_k\) are differentiable functions defined on \(\mathbb{R}^n\) and \(k < n\) and \(d_1, \ldots, d_k\) are constants.

The most important result for constrained optimization problems is the Lagrange multiplier theorem, giving necessary conditions for a point to be a solution.

Define the Lagrange function:

\[
L(x, \lambda) = f(x) + \sum_{i=1}^{k} \lambda_i [d_i - h_i(x)],
\]

where \(\lambda_1, \ldots, \lambda_k\) are called the Lagrange multipliers.

The necessary conditions for \(x\) to solve the maximization problem is that there are \(\lambda_1, \ldots, \lambda_k\) such that the first-order conditions (FOC) are held:

\[
\frac{L(x, \lambda)}{\partial x_i} = \frac{\partial f(x)}{\partial x_i} - \sum_{l=1}^{k} \lambda_l \frac{\partial h_l(x)}{\partial x_i} = 0 \quad i = 1, 2, \ldots, n.
\]
2.6.3 Optimization with Inequality Constraints

INEQUALITY CONSTRAINED OPTIMIZATION

Consider an optimization problem with inequality constraints:

$$\max f(x)$$

such that $$g_i(x) \leq d_i \quad i = 1, 2, \ldots, k.$$ 

A point $$x$$ making all constraints held with equality (i.e., $$g_i(x) = d_i$$ for all $$i$$) is said to satisfy the constrained qualification condition if the gradient vectors, $$Dg_1(x), Dg_2(x), \ldots, Dg_k(x)$$ are linearly independent.

Theorem 2.6.2 (Kuhn-Tucker Theorem) Suppose $$x$$ solves the inequality constrained optimization problem and satisfies the constrained qualification condition. Then, there are a set of Kuhn-Tucker multipliers $$\lambda_i \geq 0, i = 1, \ldots, k$$, such that

$$Df(x) = \sum_{i=1}^{k} \lambda_i Dg_i(x).$$

Furthermore, we have the complementary slackness conditions:

$$\lambda_i \geq 0 \quad \text{for all } i = 1, 2, \ldots, k$$

$$\lambda_i = 0 \quad \text{if } g_i(x) < d_i.$$

Comparing the Kuhn-Tucker theorem to the Lagrange multipliers in the equality constrained optimization problem, we see that the major difference is that the signs of the Kuhn-Tucker multipliers are nonnegative while the signs of the Lagrange multipliers can be anything. This additional information can occasionally be very useful.

The Kuhn-Tucker theorem only provides a necessary condition for a maximum. The following theorem states conditions that guarantee the above first-order conditions are sufficient.

Theorem 2.6.3 (Kuhn-Tucker Sufficiency) Suppose $$f$$ is concave and each $$g_i$$ is convex. If $$x$$ satisfies the Kuhn-Tucker first-order conditions specified in the above theorem, then $$x$$ is a global solution to the constrained optimization problem.
We can weaken the conditions in the above theorem when there is only one constraint. Let \( C = \{ x \in \mathbb{R}^n : g(x) \leq d \} \).

**Proposition 2.6.1** Suppose \( f \) is quasi-concave and the set \( C \) is convex (this is true if \( g \) is quasi-convex). If \( x \) satisfies the Kuhn-Tucker first-order conditions, then \( x \) is a global solution to the constrained optimization problem.

Sometimes we require \( x \) to be nonnegative. Suppose we had optimization problem:

\[
\begin{align*}
\text{max } & f(x) \\
n\text{such that } & g_i(x) \leq d_i \quad i = 1, 2, \ldots, k \\
& x \geq 0.
\end{align*}
\]

Then the Lagrange function in this case is given by

\[
L(x, \lambda) = f(x) + \sum_{i=1}^{k} \lambda_i [d_i - h_i(x)] + \sum_{j=1}^{n} \mu_j x_j,
\]

where \( \mu_1, \ldots, \mu_k \) are the multipliers associated with constraints \( x_j \geq 0 \). The first-order conditions are

\[
\frac{L(x, \lambda)}{\partial x_i} = \frac{\partial f(x)}{\partial x_i} - \sum_{i=1}^{k} \lambda_i \frac{\partial g_i(x)}{\partial x_i} + \mu_i = 0 \quad i = 1, 2, \ldots, n
\]

\[
\lambda_i \geq 0 \quad l = 1, 2, \ldots, k
\]

\[
\lambda_i = 0 \quad \text{if } g_i(x) < d_i
\]

\[
\mu_i \geq 0 \quad i = 1, 2, \ldots, n
\]

\[
\mu_i = 0 \quad \text{if } x_i > 0.
\]

Eliminating \( \mu_i \), we can equivalently write the above first-order conditions with nonnegative choice variables as

\[
\frac{L(x, \lambda)}{\partial x_i} = \frac{\partial f(x)}{\partial x_i} - \sum_{i=1}^{k} \lambda_i \frac{\partial g_i(x)}{\partial x_i} \leq 0 \quad \text{with equality if } x_i > 0 \quad i = 1, 2, \ldots, n,
\]
or in matrix notation,
\[
Df - \lambda Dg \leq 0
\]
\[
x[Df - \lambda Dg] = 0
\]
where we have written the product of two vector \(x\) and \(y\) as the inner production, i.e., \(xy = \sum_{i=1}^{n} x_i y_i\). Thus, if we are at an interior optimum (i.e., \(x_i > 0\) for all \(i\)), we have
\[
Df(x) = \lambda Dg.
\]

### 2.6.4 The Envelope Theorem

Consider an arbitrary maximization problem where the objective function depends on some parameter \(a\):

\[
M(a) = \max_x f(x, a).
\]

The function \(M(a)\) gives the maximized value of the objective function as a function of the parameter \(a\).

Let \(x(a)\) be the value of \(x\) that solves the maximization problem. Then we can also write \(M(a) = f(x(a), a)\). It is often of interest to know how \(M(a)\) changes as \(a\) changes. The envelope theorem tells us the answer:

\[
\frac{dM(a)}{da} = \frac{\partial f(x, a)}{\partial a} \bigg|_{x=x(a)}.
\]

This expression says that the derivative of \(M\) with respect to \(a\) is given by the partial derivative of \(f\) with respect to \(a\), holding \(x\) fixed at the optimal choice. This is the meaning of the vertical bar to the right of the derivative. The proof of the envelope theorem is a relatively straightforward calculation.

Now consider a more general parameterized constrained maximization problem of the form

\[
M(a) = \max_{x_1, x_2, a} g(x_1, x_2, a)
\]

such that \(h(x_1, x_2, a) = 0\).
The Lagrangian for this problem is

\[ \mathcal{L} = g(x_1, x_2, a) - \lambda h(x_1, x_2, a), \]

and the first-order conditions are

\[
\begin{align*}
\frac{\partial g}{\partial x_1} - \lambda \frac{\partial h}{\partial x_1} &= 0 \\
\frac{\partial g}{\partial x_2} - \lambda \frac{\partial h}{\partial x_2} &= 0 \\
h(x_1, x_2, a) &= 0.
\end{align*}
\]

These conditions determine the optimal choice functions \((x_1(a), x_2(a), a)\), which in turn determine the maximum value function

\[ M(a) \equiv g(x_1(a), x_2(a)). \] (2.2)

The envelope theorem gives us a formula for the derivative of the value function with respect to a parameter in the maximization problem. Specifically, the formula is

\[
\frac{dM(a)}{da} = \left. \frac{\partial \mathcal{L}(x, a)}{\partial a} \right|_{x=x(a)} = \left. \frac{\partial g(x_1, x_2, a)}{\partial a} \right|_{x=x_i(a)} - \lambda \left. \frac{\partial h(x_1, x_2, a)}{\partial a} \right|_{x=x_i(a)}.
\]

As before, the interpretation of the partial derivatives needs special care: they are the derivatives of \(g\) and \(h\) with respect to \(a\) holding \(x_1\) and \(x_2\) fixed at their optimal values.

### 2.6.5 Maximum Theorems

In many optimization problems, we need to check if an optimal solution is continuous in parameters, say, to check the continuity of the demand function. We can apply the so-called Maximum Theorem.
Theorem 2.6.4 (Berg’s Maximum Theorem) Suppose \( f(x,a) \) is a continuous function mapping from \( A \times X \rightarrow \mathbb{R} \), and the constraint set \( F : A \rightarrow X \) is a continuous correspondence with non-empty compact values. Then, the optimal valued function (also called marginal function):

\[
M(a) = \max_{x \in F(a)} f(x,a)
\]

is a continuous function, and the optimal solution:

\[
\phi(a) = \arg \max_{x \in F(a)} f(x,a)
\]

is a upper hemi-continuous correspondence.

2.6.6 Continuous Selection Theorems

The continuous selection theorem is a powerful tool to prove the existence of equilibrium, and it is closely related to the fixed point theorem to be introduced below. The basic conclusion of a continuous selection theorem is that if a correspondence is lower hemi-continuous with non-empty convex values, there is a continuous function, so that for all points at the region, the function value is a corresponding subset.

Theorem 2.6.5 (Michael, 1956) Let \( X \subset \mathbb{R}^n \) be convex. Suppose \( F : X \rightarrow 2^{\mathbb{R}^m} \) is a lower hemi-continuous correspondence with non-empty and convex values. Then there exists a continuous function \( f : X \rightarrow 2^{\mathbb{R}^m} \) such that \( f(x) \in F(x) \) for all \( x \in X \).

2.6.7 Fixed Point Theorems

To show the existence of a competitive equilibrium for the continuous aggregate excess demand function, we will use the following fixed-point theorem. The generalization of Brouwer’s fixed theorem can be found in Tian (1991) that gives necessary and sufficient conditions for a function to have a fixed point.
Theorem 2.6.6 (Brouwer’s Fixed Theorem) Let \( X \) be a non-empty, compact, and convex subset of \( \mathbb{R}^m \). If a function \( f : X \to X \) is continuous on \( X \), then \( f \) has a fixed point, i.e., there is a point \( x^* \in X \) such that \( f(x^*) = x^* \).

![Figure 2.4: Fixed points are given by the intersections of the 45° line and the curve of the function. There are three fixed points in the case depicted.](image)

**Example 2.6.1** \( f : [0, 1] \to [0, 1] \) is continuous, then \( f \) has a fixed point \( (x) \). To see this, let 

\[
g(x) = f(x) - x.
\]

Then, we have 

\[
g(0) = f(0) \geq 0
\]
\[
g(1) = f(1) - 1 \leq 0.
\]

From the mean-value theorem, there is a point \( x^* \in [0, 1] \) such that 

\[
g(x^*) = f(x^*) - x^* = 0.
\]

When a mapping is a correspondence, we have the following version of fixed point theorem.

**Theorem 2.6.7 (Kakutani’s Fixed Point Theorem)** Let \( X \) be a non-empty, compact, and convex subset of \( \mathbb{R}^m \). If a correspondence \( F : X \to 2^X \) is an upper hemi-continuous correspondence with non-empty compact and convex values on \( X \), then \( F \) has a fixed point, i.e., there is a point \( x^* \in X \) such that \( x^* \in F(x^*) \).
2.6.8 Variation Inequality

2.6.9 FKKM Theorems

The Knaster-Kuratowski-Mazurkiewicz (KKM) lemma is quite basic and in some ways more useful than Brouwer’s fixed point theorem. The following is a generalized version of KKM lemma due to Ky Fan (1984).

**Theorem 2.6.8 (FKKM Theorem)** Let $Y$ be a convex set and $\emptyset \not\subset X \subset Y$. Suppose $F : X \to 2^Y$ is a correspondence such that

1. $F(x)$ is closed for all $x \in X$;
2. $F(x_0)$ is compact for some $x_0 \in X$;
3. $F$ is FS-convex, i.e., for any $x_1, \ldots, x_m \in X$ and its convex combination $x_\lambda = \sum_{i=1}^{m} \lambda_i x_i$, we have $x_\lambda \in \cup_{i=1}^{m} F(x_i)$.

Then $\cap_{x \in X} F(x) \neq \emptyset$.

Here, The term FS is for Fan (1984) and Sonnenschein (1971), who introduced the notion of FS-convexity.

2.7 Dynamic Optimization

2.8 Differential Equation

2.9 Difference Equation

2.10 Basic Probability

2.10.1 Probability and Conditional Probability

2.10.2 Mathematical Expectation and Variance

2.10.3 Continuous Distributions

Given a random variable $X$, which takes on values in $[0, \omega]$, we define its cumulative distribution function $F : [0, \omega] \rightarrow [0, 1]$ by

$$F(x) = \text{Prob}[X \leq x]$$

the probability that $X$ takes on a value not exceeding $x$. By definition, the function $F$ is nondecreasing and satisfies $F(0) = 0$ and $F(\omega) = 1$ (if $\omega = \infty$, then $\lim_{x \to \infty} F(x) = 1$). In this course, we always suppose that $F$ is increasing and continuously differentiable.

The derivative of $F$ is called the associated probability density function and is usually denoted by the corresponding lowercase letter $f \equiv F'$. By assumption, $f$ is continuous and we will suppose that for all $x \in (0, \omega)$, $f(x)$ is positive. The interval $[0, \omega]$ is called the support of the distribution.

If $X$ is distributed according to $F$, then the expectation of $X$ is

$$E(X) = \int_{0}^{\omega} xf(x)dx$$

and if $\gamma : [0, \omega] \rightarrow \mathbb{R}$ is some arbitrary function, then the expectation of $\gamma(X)$ is analogously defined as

$$E[\gamma(X)] = \int_{0}^{\omega} \gamma(x)f(x)dx.$$
Sometimes the expectation of $\gamma(X)$ is also written as

$$E[\gamma(X)] = \int_0^\omega \gamma(x) dF(x).$$

The conditional expectation of $X$ given that $X < x$ is

$$E[X|X < x] = \frac{1}{F(x)} \int_0^x tf(t)dt$$

and so

$$F(x)E[X|X < x] = \int_0^x tf(t)dt = xF(x) - \int_0^x F(t)dt$$

which is obtained by integrating the right-hand side of the first equality by parts.

### 2.11 Stochastic Dominance and Affiliation

#### 2.11.1 Hazard Rates

Let $F$ be a distribution function with support $[0, \omega]$. The hazard rate of $F$ is the function $\lambda : [0, \omega) \to \mathbb{R}_+$ defined by

$$\lambda(x) \equiv \frac{f(x)}{1 - F(x)}.$$

If $F$ represents the probability that some event will happen before time $x$, then the hazard rate at $x$ represents the instantaneous probability that the event will happen at $x$, given that it has not happened until time $x$. The event may be the failure of some component—a light bulb, for instance—and hence it is sometimes also known as the “failure rate”.

Solving for $F$, we have

$$F(x) = 1 - \exp \left( - \int_0^x \lambda(t) dt \right).$$

This shows that any arbitrary function $\lambda : [0, \omega) \to \mathbb{R}_+$ such that for all $x < \omega$,

$$\int_0^x \lambda(t) dt < \infty, \quad \lim_{x \to \omega} \int_0^x \lambda(t) dt = \infty,$$

is the hazard rate of some distribution.
Closely related to the hazard rate is the function $\sigma : (0, \omega] \to \mathcal{R}_+$ defined by

$$\sigma(x) \equiv \frac{f(x)}{F(x)},$$

sometimes known as the reverse hazard rate or is referred to as the inverse of the Mills’ ratio. Similarly, Solving for $F$, we have

$$F(x) = \exp \left( - \int_x^\omega \sigma(t)dt \right),$$

This shows that any arbitrary function $\sigma : (0, \omega] \to \mathcal{R}_+$ such that for all $x > 0$,

$$\int_x^\omega \sigma(t)dt < \infty \text{ 同时 } \lim_{x \to 0} \int_x^\omega \sigma(t)dt = \infty.$$

is the “reverse hazard rate” of some distribution.

### 2.11.2 Stochastic Dominance

#### First-Order Stochastic Dominance

**Definition 2.11.1 (First-Order Stochastic Dominance)** Given two distribution functions $F$ and $G$, we say that $F$ first-order stochastically dominates $G$ if for all $z \in [0, \omega]$, $F(z) \leq G(z)$.

The first-Order stochastic dominance means that for any outcome $x$, the probability of obtaining at least $x$ under $F(\cdot)$ is at least as high as that under $G(\cdot)$, i.e., $F(\cdot) \leq G(\cdot)$ implies that the probability in lower part under $F(\cdot)$ is smaller than under $G(\cdot)$, or that the probability in higher part under $F(\cdot)$ is larger than under $G(\cdot)$. This is analogous to the monotonicity concept under certainty.

There is another test criterion for $F$ to first-order stochastically dominate $G$: Does every expected utility maximizer with an increasing utility function prefer $F(\cdot)$ over $G(\cdot)$?

The following theorem shows that these two criterions are equivalent.

**Theorem 2.11.1** $F(\cdot)$ first-order stochastically dominates $G(\cdot)$ if and only if for any function $u : [0, \omega] \to \mathcal{R}$ that is (weakly) increasing and differentiable, we have

$$\int u(z)dF(z) \geq \int u(z)dG(z).$$
Proof. Define $H(z) = F(z) - G(z)$. We need to prove that $H(z) \leq 0$ if and only if $\int u(z)dH(z) \geq 0$ for any increasing and differentiable function $u(\cdot)$.

Sufficiency: We prove this by way of contradiction. Suppose there is a $\hat{z}$ such that $H(\hat{z}) > 0$. We choose a weakly increasing and differentiable function $u(z)$ as

$$u(z) = \begin{cases} 
0, & z \leq \hat{z}, \\
1, & z > \hat{z},
\end{cases}$$

then immediately $\int u(z)dH(z) = -H(\hat{z}) < 0$, a contradiction.

Necessity:

$$\int u(z)dH(z) = [u(z)H(z)]_{\infty}^{\infty} - \int u'(z)H(z)dz = 0 - \int u'(z)H(z)dz \geq 0,$$

in which the first equality is based on the formula of integration by parts, the second equality is based on

$$F(-\infty) = G(-\infty) = 0, \quad F(\infty) = G(\infty) = 1,$$

while the inequality is based on the assumptions that $u(\cdot)$ is weakly increasing ($u'(\cdot) \geq 0$) and $H(z) \leq 0$. ■

Since a monotone function can be arbitrarily approximated by a sequence of monotone and differentiable functions, the differentiability requirement imposed on $u$ is not necessary. For any two gambles $F$ and $G$, as long as an agent’s utility is (weakly) increasing in outcomes, he prefers the one that first-order stochastically dominates the other one.

Second-Order Stochastic Dominance

Definition 2.11.2 (Second-Order Stochastic Dominance) Given two distribution functions $F$ and $G$ with the same expectation, we say that $F(\cdot)$ second-order stochastically dominates $G(\cdot)$ if

$$\int_{-\infty}^{z} F(r)dr \leq \int_{-\infty}^{z} G(r)dr$$

for all $z$. 

93
It is clear that first-order stochastic dominance implies second-order stochastic dominance. The second-order stochastic dominance implies not only monotonicity but also lower risk. To do so, we introduce the notion of “Mean-Preserving Spreads”.

Suppose $X$ is a random variable with distribution function $F$. Let $Z$ be a random variable whose distribution conditional on $X = x$, $H(\cdot | X = x)$, is such that for all $x$, $E[Z|X = x] = 0$. Suppose $Y = X + Z$ is the random variable obtained from first drawing $X$ from $F$ and then for each realization $X = x$, drawing a $Z$ from the conditional distribution $H(\cdot | X = x)$ and adding it to $X$. Let $G$ be the distribution of $Y$ so defined. We will then say that $G$ is a mean-preserving spread of $F$.

While random variables $X$ and $Y$ have the same mean, namely $E[X] = E[Y]$, variable $Y$ is “more spread-out” than $X$ since it is obtained by adding a “noise” variable $Z$ to $X$. Now suppose $u : [0, \omega] \to \mathcal{R}$ is a concave function. Using Jensen’s inequality

$$E(u(X)) \leq E(u(Y)),$$

we obtain

$$E_Y[u(Y)] = E_X[E_Z[u(X + Z)]|X = x]$$

$$\leq E_X[u(E_Z[X + Z|X = x])]$$

$$= E_X[u(X)].$$

As such, similar to Theorem 21.2.1, we have the following conclusion for the second-order stochastic dominance.

**Theorem 2.11.2** If distributions $F(\cdot)$ and $G(\cdot)$ have the same mean, then the following statements are equivalent.

1. $F(\cdot)$ second-order stochastically dominates $G(\cdot)$;
2. for any nondecreasing concave function $u : \mathcal{R} \to \mathcal{R}$, we have $\int u(z)dF(z) \geq \int u(z)dG(z)$;
3. $G(\cdot)$ is a mean-preserving spread of $F(\cdot)$. 

94
Proof. (3)⇒(2): It is obtained by using
\[
\int u(z)dF(z) = \int u \left( \int (x+z)dH_z(x) \right) dF(z) \\
\geq \int \left( \int u(x+z)dH_z(x) \right) dF(z) \\
= \int u(z)dG(z),
\]
in which the inequality follows from the concavity of \(u(\cdot)\).

(1)⇒(2): For expositional convenience, we set \(\omega = 1\). We have
\[
\int u(z)dF(z) - \int u(z)dG(z) \\
= -u'(1) \int_0^1 (F(z) - G(z))dz + \int \left( \int_0^z (F(x) - G(x))dx \right) u''(z)dz \\
= \int \left( \int_0^z (F(x) - G(x))dx \right) u''(z)dz \\
\geq 0,
\]
in which the inequality follows from the definition of second-order stochastic dominance, namely
\[
\int \int F(r)dr \leq \int \int G(r)dr
\]
, and also \(u''(\cdot) \leq 0\) for any \(z\). We thus have
\[
\int u(z)dF(z) - \int u(z)dG(z) \geq 0.
\]

(1)⇒(3): We just show the case with discrete distributions.
Define
\[
S(z) = G(z) - F(z), \\
T(x) = \int_0^x S(z)dz.
\]
By the definition of second-order stochastic dominance, we have \(T(x) \geq 0\) and \(T(1) \geq 0\), which imply that there exists some \(\hat{z}\) such that \(S(z) \geq 0\) for \(z \leq \hat{z}\) and \(S(z) \leq 0\) for \(z \geq \hat{z}\).
Since the random variable follows a discrete distribution, \( S(z) \) must be a step function. Let \( I_1 = (a_1, a_2) \) be the first interval over which \( S(z) \) is positive, and \( I_2 = (a_3, a_4) \) be the first interval over which \( S(z) \) is negative. If no such \( I_1 = (a_1, a_2) \) exists, then \( S(z) \equiv 0 \) and hence statement (3) is immediate. If \( I_1 = (a_1, a_2) \) does exist, then \( I_2 = (a_3, a_4) \) must exist as well.

So, \( S(z) \equiv \gamma_1 > 0 \) for \( z \in I_1 \), and \( S(z) \equiv -\gamma_2 < 0 \) for \( z \in I_2 \). By \( T(x) \geq 0 \), we must have \( a_2 < a_3 \). If \( \gamma_1 (a_2 - a_1) \geq \gamma_2 (a_4 - a_3) \), then there exist \( a_1 < \hat{a}_2 \leq a_2 \) and \( \hat{a}_4 = a_4 \) such that \( \gamma_1 (\hat{a}_2 - a_1) = \gamma_2 (\hat{a}_4 - a_3) \). If \( \gamma_1 (a_2 - a_1) < \gamma_2 (a_4 - a_3) \), then there exists \( a_3 < \hat{a}_4 \leq a_4 \) such that \( \gamma_1 (\hat{a}_2 - a_1) = \gamma_2 (\hat{a}_4 - a_3) \).

Letting

\[
S_1(z) = \begin{cases} 
\gamma_1, & \text{if } a_1 < z < \hat{a}_2, \\
-\gamma_2, & \text{if } a_3 < z < \hat{a}_4, \\
0, & \text{otherwise.}
\end{cases}
\]

If \( F_1 = F + S_1 \), then \( F_1 \) is a mean-preserving spread of \( F \). Letting \( S^1 = G - F_1 \), then we can similarly construct \( S_2(z) \) and \( F_2 \). Since \( S(z) \) is a step function, then there exists an \( n \) such that \( F_0 = F, F_n = G, \) and \( F_{i+1} \) is a mean-preserving spread of \( F_i \). Also, a finite summation of mean-preserving spreads is still a mean-preserving spread. ■

Though a continuous function can be arbitrarily approximated by step functions, the formal proof is complicated, and Rothschild and Stiglitz (1971) provide a complete proof for the case with continuous distributions.

2.11.3 Hazard Rate Dominance

**Definition 2.11.3 (Hazard Rate Dominance)** For any two distributions \( F \) and \( G \) with hazard rates \( \lambda_F \) and \( \lambda_G \), respectively. We say that \( F \) dominates \( G \) in terms of the hazard rate if \( \lambda_F(x) \leq \lambda_G(x) \) for all \( x \). This order is also referred in short as hazard rate dominance.
If $F$ dominates $G$ in terms of the hazard rate, then
\[ F(x) = 1 - \exp \left( -\int_0^x \lambda_F(t) dt \right) \leq 1 - \exp \left( -\int_0^x \lambda_G(t) dt \right) = G(x), \]
and hence $F$ stochastically dominates $G$. Thus, hazard rate dominance implies first-order stochastic dominance.

### 2.11.4 Reverse Hazard Rate Dominance

**Definition 2.11.4 (Reverse Hazard Rate Dominance)** For any two distributions $F$ and $G$ with reverse hazard rates $\sigma_F$ and $\sigma_G$, respectively. We say that $F$ dominates $G$ in terms of the reverse hazard rate if $\sigma_F(x) \geq \sigma_G(x)$ for all $x$. This order is also referred in short as reverse hazard rate dominance.

If $F$ dominates $G$ in terms of the reverse hazard rate, then
\[ F(x) = \exp \left( -\int_x^\infty \sigma_F(t) dt \right) \leq \exp \left( -\int_x^\infty \sigma_G(t) dt \right) = G(x), \]
and hence, again, $F$ stochastically dominates $G$. Thus, reverse hazard rate dominance also implies first-order stochastic dominance.

### 2.11.5 Order Statistics

Let $X_1, X_2, \ldots, X_n$ be $n$ independently draws from a distribution $F$ with associated density $f$. Let $Y_1^{(n)}, Y_2^{(n)}, \ldots, Y_n^{(n)}$ be a rearrangement of these so that
\[ Y_1^{(n)} \geq Y_2^{(n)} \geq \cdots \geq Y_n^{(n)}. \]
These random variables $Y_k^{(n)}$, $k = 1, 2, \cdots, n$ are referred to as order statistics.

Let $F_k^{(n)}$ denote the distribution of $Y_k^{(n)}$, with corresponding probability density function $f_k^{(n)}$. When the “sample size” $n$ is fixed and there is no ambiguity, we simply write $Y_k$ instead of $Y_k^{(n)}$, $F_k$ instead of $F_k^{(n)}$ and $f_k$ instead of $f_k^{(n)}$. In auction theory, we will typically be interested in properties of the highest and second highest order statistics, namely $Y_1$ and $Y_2$. 

97
2.11.6 Order Statistic

Highest Order Statistic

The distribution of the highest order statistic $Y_1$ is easy to derive. The event that $Y_1 \leq y$ is the same as the event: for all $k$, $X_k \leq y$. Since each $X_k$ is an independent draw from the same distribution $F$, we have that

$$F_1(y) = F(y)^n.$$  

The associated probability density function is

$$f_1(y) = nF(y)^{n-1}f(y).$$

Observe that if $F$ stochastically dominates $G$, and $F_1$ and $G_1$ are the distributions of the highest order statistics of $n$ draws from $F$ and $G$, respectively, then $F_1$ stochastically dominates $G_1$.

Second-Highest Order Statistic

The distribution of the second-highest order statistic $Y_2$ can also be easily derived. The event that $Y_2 \leq y$ is the union of the following disjoint events: (1) all $X_k$’s are less than or equal to $y$; and (2) $n - 1$ of the $X_k$’s are less than or equal to $y$ and one is greater than $y$. There are $n$ different ways in which (2) can occur, so we have that

$$F_2(y) = \underbrace{F(y)^n}_{(i)} + \underbrace{nF(y)^{n-1}(1 - F(y))}_{(ii)} = nF(y)^{n-1} - (n - 1)F(y)^n.$$  

The associated probability density function is

$$f_2(y) = n(n - 1)(1 - F(y))F(y)^{n-2}f(y).$$

Again, it can be verified that if $F$ stochastically dominates $G$ and also $F_2$ and $G_2$ are the distributions of the second-highest order statistics of $n$ draws from $F$ and $G$, respectively, then $F_2$ stochastically dominates $G_2$.
2.11.7 Affiliation

Affiliation is a basic assumption used to study auction with interdependent values which are non-negative correlated.

**Definition 2.11.5** Suppose the random variables $X_1, X_2, \cdots, X_n$ are distributed on some product of intervals $\mathcal{X} \subseteq \mathbb{R}^n$ according to the joint density function $f$. The variables $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ are said to be affiliated if for all $\mathbf{x}', \mathbf{x}'' \in \mathcal{X}$,

$$f(\mathbf{x}' \vee \mathbf{x}'')f(\mathbf{x}' \wedge \mathbf{x}'') \geq f(\mathbf{x}')f(\mathbf{x}''),$$

(2.3)
in which

$$\mathbf{x}' \vee \mathbf{x}'' = (\max(x'_1, x''_1), \cdots, \max(x'_n, x''_n))$$
denotes the component-wise maximum of $\mathbf{x}'$ and $\mathbf{x}''$, and

$$\mathbf{x}' \wedge \mathbf{x}'' = (\min(x'_1, x''_1), \cdots, \min(x'_n, x''_n))$$
denotes the component-wise minimum of $\mathbf{x}'$ and $\mathbf{x}''$. If (2.1) is satisfied, then we also say that $f$ is affiliated.

Suppose that the density function $f : \mathcal{X} \to \mathbb{R}_+$ is strictly positive in the interior of $\mathcal{X}$ and twice continuously differentiable. It is “easy” to verify that $f$ is affiliated if and only if, for all $i \neq j$,

$$\frac{\partial^2}{\partial x_i \partial x_j} \ln f \geq 0.$$ 

In other words, the off-diagonal elements of the Hessian of $\ln f$ are nonnegative.

**Proposition 2.11.1** Let $X_1, X_2, \cdots, X_n$ be random variables, and $Y_1, Y_2, \cdots, Y_{n-1}$ be the largest, second largest, ..., smallest order statistics from among $X_2, X_3, \cdots, X_n$. If $X_1, X_2, \cdots, X_n$ are symmetrically distributed and affiliated, then we have

1. variables in any subset of $X_1, X_2, \cdots, X_n$ are also affiliated;
2. $X_1, Y_1, Y_2, \cdots, Y_{n-1}$ are affiliated.
Monotone Likelihood Ratio Property

Suppose the two random variables $X$ and $Y$ have a joint density $f : [0, \omega]^2 \to \mathcal{R}$. If $X$ and $Y$ are affiliated, then for all $x' \geq x$ and $y' \geq y$, we have

$$f(x', y)f(x, y') \leq f(x, y)f(x', y') \iff \frac{f(x, y')}{f(x, y)} \leq \frac{f(x', y')}{f(x', y)}$$ \hspace{1cm} (2.4)

and

$$\frac{f(y|x)}{f(y|x')} \leq \frac{f(y'|x)}{f(y'|x')}$$

so the likelihood ratio

$$\frac{f(\cdot|x')}{f(\cdot|x)}$$

is increasing and this is referred to as the monotone likelihood ratio property.

Likelihood Ratio Dominance

**Definition 2.11.6 (Likelihood Ratio Dominance)** The distribution function $F$ dominates $G$ in terms of the likelihood ratio if for all $x < y$,

$$\frac{f(x)}{f(y)} \leq \frac{g(x)}{g(y)}.$$

We thus have the following conclusion.

**Proposition 2.11.2** If $X$ and $Y$ are affiliated, the following properties hold:

1. For all $x' \geq x$, $F(\cdot|x')$ dominates $F(\cdot|x)$ in terms of hazard rate; that is,

$$\lambda(y|x') \equiv \frac{f(y|x')}{1 - F(y|x')} \leq \frac{f(y|x)}{1 - F(y|x)} \equiv \lambda(y|x).$$

Or equivalently, for all $y$, $\lambda(y|\cdot)$ is nonincreasing.

2. For all $x' \geq x$, $F(\cdot|x')$ dominates $F(\cdot|x)$ in terms of the reverse hazard rate; that is,

$$\sigma(y|x') \equiv \frac{f(y|x')}{F(y|x')} \leq \frac{f(y|x)}{F(y|x)} \equiv \sigma(y|x),$$

or equivalently, for all $y$, $\sigma(y|\cdot)$ is nondecreasing.
(3) For all \( x' \geq x \), \( F(\cdot|x') \) stochastically dominate \( F(\cdot|x) \); that is,

\[
F(y|x') \leq F(y|x),
\]

or equivalently, for all \( y \), \( F(y|\cdot) \) is nonincreasing.

All of these results extend in a straightforward manner to the case where the number of conditioning variables is more than one. Suppose \( Y, X_1, X_2, \ldots, X_n \) are affiliated and let \( F_Y(\cdot|x) \) denote the distribution of \( Y \) conditional on \( X = x \). Then, using the same arguments as above, it can be deduced that for all \( x' \geq x \), \( F_Y(\cdot|x') \) dominates \( F_Y(\cdot|x) \) in terms of the likelihood ratio. The other dominance relationships then follow as usual.

2.12 Reference

Books and Monographs:


**Papers:**


Part II

Individual Decision Making
Part II is devoted to the theories of individual decision making and consists of three chapters: consumer theory, producer theory, and choice under uncertainty. It studies how a consumer or producer selects an appropriate action or making an appropriate decision. Microeconomic theory is founded on the premise that these individuals behave rationally, making choices that are optimal for themselves. Throughout this part, we restrict ourselves to an ideal situation (benchmark case) where the behavior of the others are summarized in non-individualized parameters – the prices of commodities, each individual makes decision independently by taking prices as given and individuals’ behavior are indirectly interacted through prices.

We will treat consumer theory first, and at some length – both because of its intrinsic importance, and because its methods and results are paradigms for many other topic areas. Producer theory is next, and we draw attention to the many formal similarities between these two important building blocks of modern microeconomics. Finally, we conclude our treatment of the individual consumer by looking at the problem of choice under uncertainty. It is greatly acknowledged that the materials are largely drawn from the references in the end of each charter, especially from Varian (1992).
Chapter 3

Consumer Theory

3.1 Introduction

In this chapter, we will explore the essential features of modern consumer theory—a bedrock foundation on which so many theoretical structures in economics are build, and it is also central to the economists’ way of thinking.

A consumer can be characterized by many factors and aspects such as sex, age, lifestyle, wealth, parentage, ability, intelligence, etc. But which are most important ones for us to study consumer’s behavior in making choices? To grasp the most important features in studying consumer behavior and choices in modern consumer theory, it is assumed that the key characteristic of a consumer consists of three essential components: the consumption set, initial endowments, and the preference relation. Consumer’s characteristic together with the behavior assumption are building blocks in any model of consumer theory. The consumption set represents the set of all individually feasible alternatives or consumption plans and sometimes also called the choice set. An initial endowment represents the amount of various goods the consumer initially has and can consume or trade with other individuals. The preference relation specifies the consumer’s tastes or satisfactions for the different objects of choice. The behavior assumption expresses the guiding principle the consumer uses to make final choices and identifies the
ultimate objects in choice. It is generally assumed that the consumer seeks to identify and select an available alternative that is most preferred in the light of his/her personal tastes/interests.

3.2 Consumption Set and Budget Constraint

3.2.1 Consumption Set

We consider a consumer faced with possible consumption bundles in consumption set $X$. We usually assume that $X$ is the nonnegative orthant in $R^L$ as shown in the right figure in Figure 3.1, but more specific consumption sets may be used. For example, it may allow consumptions of some good in a suitable interval as such leisure as shown in the left figure in Figure 3.1, or we might only include bundles that would give the consumer at least a subsistence existence or that consists of only integer units of consumptions as shown in Figure 3.2.

We assume that $X$ is a closed and convex set unless otherwise stated. The convexity of a consumption set means that every good is divisible and can be consumed by fraction units.
Figure 3.2: The left figure: A consumption set that reflects survival needs. The right figure: A consumption set where good 2 must be consumed in integer amounts.

### 3.2.2 Budget Constraint

In the basic problem of consumer’s choice, not all consumptions bundles are affordable in a limited resource economy, and a consumer is constrained by his/her wealth. In a market institution, the wealth may be determined by the value of his/her initial endowment and/or income from stock-holdings of firms. It is assumed that the income or wealth of the consumer is fixed and the prices of goods cannot be affected by the consumer’s consumption when discussing a consumer’s choice. Let $m$ be the fixed amount of money available to a consumer, and let $p = (p_1,\ldots,p_L)$ be the vector of prices of goods, $1,\ldots,L$. The set of affordable alternatives is thus just the set of all bundles that satisfy the consumer’s budget constraint. The set of affordable bundles, the budget set of the consumer, is given by

$$B(p,m) = \{x \in X : px \leq m, \}$$

where $px$ is the inner product of the price vector and consumption bundle, i.e., $px = \sum_{l=1}^{L} p_l x_l$ which is the sum of expenditures of commodities at prices $p$. The ratio, $p_i/p_k$, may be called the economic rate of substitution between goods $i$ and $j$. Note that multiplying all prices and income by some positive number
does not change the budget set.

Thus, the budget set reflects consumer’s objective ability of purchasing commodities and the scarcity of resources. It significantly restricts the consumer choices. To determine the optimal consumption bundles, one needs to combine consumer objective ability of purchasing various commodities with subjective taste on various consumption bundles which are characterized by the notion of preference or utility.

3.3 Preferences and Utility

3.3.1 Preferences

The consumer is assumed to have preferences on the consumption bundles in $X$ so that he can compare and rank various goods available in the economy. When we write $\mathbf{x} \succeq \mathbf{y}$, we mean “the consumer thinks that the bundle $\mathbf{x}$ is at least as good as the bundle $\mathbf{y}$.” We want the preferences to order the set of bundles. Therefore, we need to assume that they satisfy the following standard properties.

COMPLETE. For all $\mathbf{x}$ and $\mathbf{y}$ in $X$, either $\mathbf{x} \succeq \mathbf{y}$ or $\mathbf{y} \succeq \mathbf{x}$ or both.

REFLEXIVE. For all $\mathbf{x}$ in $X$, $\mathbf{x} \succeq \mathbf{x}$. 
TRANSITIVE. For all \( x, y \) and \( z \) in \( X \), if \( x \succeq y \) and \( y \succeq z \), then \( x \succeq z \).

The first assumption is just says that any two bundles can be compared, the second is trivial and says that every consumption bundle is as good as itself, and the third requires the consumer’s choice be consistent. A preference relation that satisfies these three properties is called a preference ordering.

Given an ordering \( \succeq \) describing “weak preference,” we can define the strict preference \( \succ \) by \( x \succ y \) to mean not \( y \succeq x \). We read \( x \succ y \) as “\( x \) is strictly preferred to \( y \).” Similarly, we define a notion of indifference by \( x \sim y \) if and only if \( x \succeq y \) and \( y \succeq x \).

Given a preference ordering, we often display it graphically, as shown in Figure 3.4. The set of all consumption bundles that are indifferent to each other is called an indifference curve. For a two-good case, the slope of an indifference curve at a point measures marginal rate of substitution between goods \( x_1 \) and \( x_2 \). For a \( L \)-dimensional case, the marginal rate of substitution between two goods is the slope of an indifference surface, measured in a particular direction.

![Figure 3.4: Preferences in two dimensions.](image)

For a given consumption bundle \( y \), let \( P(y) = \{x \in X : x \succeq y\} \) be the set of all bundles on or above the indifference curve through \( y \) and it is called the upper contour set at \( y \), \( P_s(y) = \{x \in X : x \succ y\} \) be the set of all bundles above the indifference curve through \( y \) and it is called the strictly upper contour set at
$y$, $L(y) = \{x \in X : x \preceq y\}$ be the set of all bundles on or below the indifference curve through $y$ and it is called the **lower contour set** at $y$, and $L_s(y) = \{x \in X : x \preceq y\}$ be the set of all bundles on or below the indifference curve through $y$ and it is called the **strictly lower contour set** at $y$.

We often wish to make other assumptions on consumers’ preferences; for example.

**CONTINUITY.** For all $y$ in $X$, the upper and lower contour sets $P(y)$ and $L(y)$, are closed. It follows that the strictly upper and lower contour sets, $P_s(y)$ and $L_s(y)$, are open sets.

This assumption is necessary to rule out certain discontinuous behavior; it says that if $(x^i)$ is a sequence of consumption bundles that are all at least as good as a bundle $y$, and if this sequence converges to some bundle $x^*$, then $x^*$ is at least as good as $y$. The most important consequence of continuity is this: if $y$ is strictly preferred to $z$ and if $x$ is a bundle that is close enough to $y$, then $x$ must be strictly preferred to $z$.

**Example 3.3.1 (Lexicographic Ordering)** An interesting preference ordering is the so-called lexicographic ordering defined on $\mathbb{R}^L$, based on the way one orders words alphabetically. It is defined follows: $x \succeq y$ if and only if there is a $l$, $1 \leq l \leq L$, such that $x_i = y_i$ for $i < l$ and $x_l > y_l$ or if $x_i = y_i$ for all $i = 1, \ldots, L$.

Essentially the lexicographic ordering compares the components on at a time, beginning with the first, and determines the ordering based on the first time a different component is found, the vector with the greater component is ranked highest. However, the lexicographic ordering is not continuous or even not upper semi-continuous, i.e., the upper contour set is not closed. This is easily seen for the two-dimensional case by considering the upper contour correspondence to $y = (1, 1)$, that is, the set $P(1, 1) = \{x \in X : x \succeq (1, 1)\}$ as shown in Figure 3.5. It is clearly not closed because the boundary of the set below $(1, 1)$ is not contained in the set.

There are two more assumptions, namely, monotonicity and convexity, that
WEAK MONOTONICITY. If $x \succeq y$ then $x \succeq y$.

MONOTONICITY. If $x > y$, then $x \succ y$.

STRONG MONOTONICITY. If $x \succeq y$ and $x \neq y$, then $x \succ y$.

Weak monotonicity says that “at least as much of everything is at least as good,” which ensures a commodity is a “good”, but not a “bad”. Monotonicity says that strictly more of every good is strictly better. Strong monotonicity says that at least as much of every good, and strictly more of some good, is strictly better.

Another assumption that is weaker than either kind of monotonicity or strong monotonicity is the following:

LOCAL NON-SATIATION. Given any $x$ in $X$ and any $\epsilon > 0$, then there is some bundle $y$ in $X$ with $|x - y| < \epsilon$ such that $y \succ x$.

NON-SATIATION. Given any $x$ in $X$, then there is some bundle $y$ in $X$ such that $y \succ x$.

Remark 3.3.1 Monotonicity of preferences can be interpreted as individuals’ desires for goods: the more, the better. Local non-satiation says that one can
always do a little bit better, even if one is restricted to only small changes in the consumption bundle. Thus, local non-satiation means individuals’ desires are unlimited. You should verify that (strong) monotonicity implies local non-satiation and local non-satiation implies non-satiation, but not vice versa.

We now give various types of convexity properties used in the consumer theory.

**STRICT CONVEXITY.** Given \(x, x'\) in \(X\) such that \(x' \succeq x\), then it follows that \(tx + (1 - t)x' > x\) for all \(0 < t < 1\).

**CONVEXITY.** Given \(x, x'\) in \(X\) such that \(x' \succ x\), then it follows that \(tx + (1 - t)x' > x\) for all \(0 \leq t < 1\).

**WEAK CONVEXITY.** Given \(x, x'\) in \(X\) such that \(x' \succeq x\), then it follows that \(tx + (1 - t)x' \succeq x\) for all \(0 \leq t \leq 1\).

**Remark 3.3.2** The convexity of preferences implies that people want to diversify their consumptions (the consumer prefers averages to extremes), and thus, convexity can be viewed as the formal expression of basic measure of economic markets for diversification. Note that convex preferences may have indifference curves that exhibit “flat spots,” while strictly convex preferences have indifference curves that are strictly rotund. The strict convexity of \(\succ\) implies the neoclassical assumption of “diminishing marginal rates of substitution” between any two goods as shown in Figure 3.9.
3.3.2 The Utility Function

Sometimes it is easier to work directly with the preference relation and its associated sets. But other times, especially when one wants to use calculus methods, it is easier to work with preferences that can be represented by a utility function; that is, a function $u: X \to R$ such that $x \succeq y$ if and only if $u(x) \geq u(y)$. In the following, we give some examples of utility functions.

**Example 3.3.2 (Cobb-Douglas Utility Function)** A utility function that is used frequently for illustrative and empirical purposes is the Cobb-Douglas utility function,

$$u(x_1, x_2, \ldots, x_L) = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_L^{\alpha_L}$$

with $\alpha_l > 0$, $l = 1, \ldots, L$. This utility function represents a preference ordering that is continuous, strictly monotonic, and strictly convex in $\mathbb{R}_{++}^L$. 
Figure 3.8: “Thick” indifference curves are weakly convex, but not convex.

Figure 3.9: The marginal rate of substitution is diminishing when we the consumption of good 1 increases.

**Example 3.3.3 (Linear Utility Function)** A utility function that describes perfect substitution between goods is the linear utility function,

\[ u(x_1, x_2, \ldots, x_L) = a_1 x_1 + a_2 x_2 + \ldots + a_L x_L \]

with \( a_l \geq 0 \) for all \( l = 1, \ldots, L \) and \( a_l > 0 \) for at least \( l \). This utility function represents a preference ordering that is continuous, monotonic, and convex in \( \mathbb{R}_+^L \).

**Example 3.3.4 (Leontief Utility Function)** A utility function that describes perfect complement between goods is the Leontief utility function,

\[ u(x_1, x_2, \ldots, x_L) = \min \{a_1 x_1, a_2 x_2, \ldots, a_L x_L\} \]
with \( a_l \geq 0 \) for all \( l = 1, \ldots, L \) and \( a_l > 0 \) for at least \( l \). This represents a preference that all commodities should be used together in order to increase consumer utility. This utility function represents a preference ordering that is also continuous, monotonic, and convex in \( \mathbb{R}_+^L \).

Not all preference orderings can be represented by utility functions, but it can be shown that any (upper semi-)continuous preference ordering can be represented by a (upper semi-)continuous utility function.\(^1\) We now state a weaker version of this assertion. The following theorem shows the existence of a utility function when a preference ordering is continuous.

**Theorem 3.3.1 (Existence of Continuous Utility Function)** Suppose preferences are complete, reflexive, transitive, and continuous. Then there exists a continuous utility function \( u: \mathbb{R}_+^k \rightarrow R \) which represents those preferences.

**Proof.** For simplicity and intuition, here we only give the proof for strong monotonicity of preference. A complete and elementary proof can be seen in Jaffray (1975).

Let \( e \) be the vector in \( \mathbb{R}_+^k \) consisting of all ones. Then given any vector \( x \) let \( u(x) \) be that number such that \( x \sim u(x)e \). We have to show that such a number exists and is unique.

Let \( B = \{ t \in \mathbb{R}: te \succeq x \} \) and \( W = \{ t \in \mathbb{R}: x \succeq te \} \). Then strong monotonicity implies \( B \) is nonempty; \( W \) is certainly nonempty since it contains 0. Continuity implies both sets are closed. Since the real line is connected, there is some \( t_x \) such that \( t_xe \sim x \). We have to show that this utility function actually represents the underlying preferences. Let

\[
\begin{align*}
u(x) &= t_x \quad \text{where} \quad t_xe \sim x \\
u(y) &= t_y \quad \text{where} \quad t_ye \sim y.
\end{align*}
\]

\(^1\)the first proof for existence of continuous utility function was given by Debreu (1964). A unified proof for the existence of upper semi-continuous and continuous utility function was given by Bosi and Mehta (2002).
Then if \( t_x < t_y \), strong monotonicity shows that \( t_x e < t_y e \), and transitivity shows that

\[
{x \sim t_x e < t_y e \sim y.}
\]

Similarly, if \( x \succ y \), then \( t_x e \succ t_y e \) so that \( t_x \) must be greater than \( t_y \).

Finally, we show that the function \( u \) defined above is continuous. Suppose \( \{x_k\} \) is a sequence with \( x_k \to x \). We want to show that \( u(x_k) \to u(x) \). Suppose not. Then we can find \( \epsilon > 0 \) and an infinite number of \( k' \)’s such that \( u(x_k) > u(x) + \epsilon \) or an infinite set of \( k' \)’s such that \( u(x_k) < u(x) - \epsilon \). Without loss of generality, let us assume the first of these. This means that \( x_k \sim u(x_k)e \succ (u(x) + \epsilon)e \sim x + \epsilon e \).

So by transitivity, \( x_k \succ x + \epsilon e \). But for a large \( k \) in our infinite set, \( x + \epsilon e > x_k \), so \( x + \epsilon e \succ x_k \), contradiction. Thus \( u \) must be continuous.

The following is an example of the non-existence of utility function when preference ordering is not continuous.

**Example 3.3.5 (Non-Representation of Lexicographic Ordering by a Function)**

Given a upper semi-continuous utility function \( u \), the upper contour set \( \{x \in X: u(x) \geq \bar{u}\} \) must be closed for each value of \( \bar{u} \). It follows that the lexicographic ordering defined on \( \mathbb{R}^L \) discussed earlier cannot be represented by a upper semi-continuous utility function because its upper contour sets are not closed.

The role of the utility function is to efficiently record the underlying preference ordering. The actual numerical values of \( u \) have essentially no meaning: only the sign of the difference in the value of \( u \) between two points is significant. Thus, a utility function is often a very convenient way to describe preferences, but it should not be given any psychological interpretation. The only relevant feature of a utility function is its ordinal character. Specifically, we can show that a utility function is unique only to within an arbitrary, strictly increasing transformation.

**Theorem 3.3.2 (Invariance of Utility Function to Monotonic Transforms)**

If \( u(x) \) represents some preferences \( \succeq \) and \( f: \mathbb{R} \to \mathbb{R} \) is strongly monotonic increasing, then \( f(u(x)) \) will represent exactly the same preferences.
Proof. This is because \( f(u(x)) \geq f(u(y)) \) if and only if \( u(x) \geq u(y) \).

This invariance theorem is useful in many aspects. For instance, as it will be shown, we may use it to simplify the computation of deriving a demand function from utility maximization.

We can also use utility function to find the marginal rate of substitution between goods. Let \( u(x_1, ..., x_k) \) be a utility function. Suppose that we increase the amount of good \( i \); how does the consumer have to change his consumption of good \( j \) in order to keep utility constant?

Let \( dx_i \) and \( dx_j \) be the differentials of \( x_i \) and \( x_j \). By assumption, the change in utility must be zero, so

\[
\frac{\partial u(x)}{\partial x_i} dx_i + \frac{\partial u(x)}{\partial x_j} dx_j = 0.
\]

Hence

\[
\frac{dx_j}{dx_i} = -\frac{\frac{\partial u(x)}{\partial x_i}}{\frac{\partial u(x)}{\partial x_j}} = -\frac{MU_{x_i}}{MU_{x_j}}.
\]

which gives the marginal rate of substitution between goods \( i \) and \( j \) and is defined as the ratio of the marginal utility of \( x_i \) and the marginal utility of \( x_j \).

**Remark 3.3.3** The marginal rate of substitution does not depend on the utility function chosen to represent the underlying preferences. To prove this, let \( v(u) \) be a monotonic transformation of utility. The marginal rate of substitution for this utility function is

\[
\frac{dx_j}{dx_i} = -\frac{\frac{v'(u)\frac{\partial u(x)}{\partial x_i}}{v'(u)\frac{\partial u(x)}{\partial x_j}}}{\frac{\partial(x)}{\partial x_i}} = -\frac{\frac{\partial(x)}{\partial x_i}}{\frac{\partial(x)}{\partial x_j}}.
\]

The important properties of a preference ordering can be easily verified by examining utility function. The properties are summarized in the following proposition.

**Proposition 3.3.1** Let \( \succeq \) be represented by a utility function \( u : X \rightarrow \mathbb{R} \). Then:

1. An ordering is strongly monotonic if and only if \( u \) is strictly monotonic.
(2) An ordering is continuous if and only if \( u \) is continuous.

(3) An ordering is weakly convex if and only if \( u \) is quasi-concave.

(4) An ordering is strictly convex if and only if \( u \) is strictly quasi-concave.

Note that a function \( u \) is quasi-concavity if for any \( c \), \( u(x) \geq c \) and \( u(y) \geq c \) implies that \( u(tx + (1-t)y) \geq c \) for all \( t \) with \( 0 < t < 1 \). A function \( u \) is strictly quasi-concavity if for any \( c \) \( u(x) \geq c \) and \( u(y) \geq c \) implies that \( u(tx + (1-t)y) > c \) for all \( t \) with \( 0 < t < 1 \).

**Remark 3.3.4** The strict quasi-concave of \( u(x) \) can be checked by verifying if the naturally ordered principal minors of the bordered Hessian alternate in sign, i.e.,

\[
\begin{vmatrix}
0 & u_1 & u_2 \\
u_1 & u_{11} & u_{12} \\
u_2 & u_{21} & u_{22}
\end{vmatrix} > 0, \\
\begin{vmatrix}
0 & u_1 & u_2 & u_3 \\
u_1 & u_{11} & u_{12} & u_{13} \\
u_2 & u_{21} & u_{22} & u_{23} \\
u_3 & u_{31} & u_{32} & u_{33}
\end{vmatrix} < 0,
\]

and so on, where \( u_i = \frac{\partial u}{\partial x_i} \) and \( u_{ij} = \frac{\partial^2 u}{\partial x_i \partial x_j} \).

**Example 3.3.6** Suppose the preference ordering is represented by the Cobb-Douglas utility function: \( u(x_1, x_2) = x_1^\alpha x_2^\beta \) with \( \alpha > 0 \) and \( \beta > 0 \). Then, we
have

\[ u_x = \alpha x^{\alpha - 1} y^\beta \]
\[ u_y = \beta x^\alpha y^{\beta - 1} \]
\[ u_{xx} = \alpha (\alpha - 1) x^{\alpha - 2} y^\beta \]
\[ u_{xy} = \alpha \beta x^{\alpha - 1} y^{\beta - 1} \]
\[ u_{yy} = \beta (\beta - 1) x^\alpha y^{\beta - 2} \]

and thus

\[
\begin{vmatrix}
0 & u_x & u_y \\
 u_x & u_{xx} & u_{xy} \\
 u_y & u_{xy} & u_{yy}
\end{vmatrix}
= \begin{vmatrix}
0 & \alpha x^{\alpha - 1} y^\beta & \beta x^\alpha y^{\beta - 1} \\
\alpha x^{\alpha - 1} y^\beta & \alpha (\alpha - 1) x^{\alpha - 2} y^\beta & \alpha \beta x^{\alpha - 1} y^{\beta - 1} \\
\beta x^\alpha y^{\beta - 1} & \alpha \beta x^{\alpha - 1} y^{\beta - 1} & \beta (\beta - 1) x^\alpha y^{\beta - 2}
\end{vmatrix}
= x^{3\alpha - 2} y^{3\beta - 2} [\alpha \beta (\alpha + \beta)] > 0 \text{ for all } (x, y) > 0,
\]

which means \( u \) is strictly quasi-concave in \( \mathbb{R}_+^2 \).

### 3.4 Utility Maximization and Optimal Choice

#### 3.4.1 Utility Maximization

A foundational hypothesis on individual behavior in modern economics in general and the consumer theory in particular is that a rational agent will always choose a most preferred bundle from the set of affordable alternatives. We will derive demand functions by considering a model of utility-maximizing behavior coupled with a description of underlying economic constraints.

#### 3.4.2 Consumer’s Optimal Choice

In the basic problem of preference maximization, the set of affordable alternatives is just the set of all bundles that satisfy the consumer’s budget constraint we
discussed before. That is, the problem of preference maximization can be written as:

\[ \max u(x) \]
\[ \text{such that} \quad px \leq m \]
\[ x \text{ is in } X. \]

There will exist a solution to this problem if the utility function is continuous and that the constraint set is closed and bounded. The constraint set is certainly closed. If \( p_i > 0 \) for \( i = 1, \ldots, k \) and \( m > 0 \), it is not difficult to show that the constraint set will be bounded. If some price is zero, the consumer might want an infinite amount of the corresponding good.

**Proposition 3.4.1** Under the local nonsatiation assumption, a utility-maximizing bundle \( x^* \) must meet the budget constraint with equality.

Proof. Suppose we get an \( x^* \) where \( px^* < m \). Since \( x^* \) costs strictly less than \( m \), every bundle in \( X \) close enough to \( x^* \) also costs less than \( m \) and is therefore feasible. But, according to the local nonsatiation hypothesis, there must be some bundle \( x \) which is close to \( x^* \) and which is preferred to \( x^* \). But this means that \( x^* \) could not maximize preferences on the budget set \( B(p, m) \).

This proposition allows us to restate the consumer’s problem as

\[ \max u(x) \]
\[ \text{such that} \quad px = m. \]

The value of \( x \) that solves this problem is the consumer’s **demanded bundle**: it expresses how much of each good the consumer desires at a given level of prices and income. In general, the optimal consumption is not unique. Denote by \( x(p, m) \) the set of all utility maximizing consumption bundles and it is called the consumer’s **demand correspondence**. When there is a unique demanded bundle for each \( (p, m) \), \( x(p, m) \) becomes a function and thus is called the consumer’s **demand function**. We will see from the following proposition that strict convexity of preferences will ensure the uniqueness of optimal bundle.
Proposition 3.4.2 (Uniqueness of Demanded Bundle) If preferences are strictly convex, then for each \( p > 0 \) there is a unique bundle \( x \) that maximizes \( u \) on the consumer’s budget set, \( B(p, m) \).

**Proof.** Suppose \( x' \) and \( x'' \) both maximize \( u \) on \( B(p, m) \). Then \( \frac{1}{2}x' + \frac{1}{2}x'' \) is also in \( B(p, m) \) and is strictly preferred to \( x' \) and \( x'' \), which is a contradiction.

Since multiplying all prices and income by some positive number does not change the budget set at all and thus cannot change the answer to the utility maximization problem.

Proposition 3.4.3 (Homogeneity of Demand Function) The consumer’s demand function \( x(p, m) \) is homogeneous of degree 0 in \( (p, m) > 0 \), i.e., \( x(tp, tm) = x(p, m) \).

Note that a function \( f(x) \) is homogeneous of degree \( k \) if \( f(tx) = t^k f(x) \) for all \( t > 0 \).

### 3.4.3 First Order-Conditions for Utility Maximization

We can characterize optimizing behavior by calculus, as long as the utility function is differentiable. We will analyze this constrained maximization problem using the method of Lagrange multipliers. The Lagrangian for the utility maximization problem can be written as

\[
\mathcal{L} = u(x) - \lambda(px - m),
\]

where \( \lambda \) is the Lagrange multiplier. Suppose preference is locally non-satiated. Differentiating the Lagrangian with respect to \( x_i \), gives us the first-order conditions for the interior solution

\[
\frac{\partial u(x)}{\partial x_i} - \lambda p_i = 0 \quad \text{for } i = 1, \ldots, L \quad (3.1)
\]

\[
px = m \quad (3.2)
\]
Using vector notation, we can also write equation (3.1) as

$$Du(x) = \lambda p.$$ 

Here

$$Du(x) = \left( \frac{\partial u(x)}{\partial x_1}, \ldots, \frac{\partial u(x)}{\partial x_L} \right)$$

is the gradient of $u$: the vector of partial derivatives of $u$ with respect to each of its arguments.

In order to interpret these conditions we can divide the $i^{th}$ first-order condition by the $j^{th}$ first-order condition to eliminate the Lagrange multiplier. This gives us

$$\frac{\partial u(x^\ast)}{\partial x_i} \frac{\partial u(x^\ast)}{\partial x_j} = \frac{p_i}{p_j} \quad \text{for } i, j, = 1, \ldots, L. \quad (3.3)$$

The fraction on the left is the marginal rate of substitution between good $i$ and $j$, and the fraction on the right is the economic rate of substitution between goods $i$ and $j$. Maximization implies that these two rates of substitution should be equal. Suppose they were not; for example, suppose

$$\frac{\partial u(x^\ast)}{\partial x_i} \frac{\partial u(x^\ast)}{\partial x_j} = 1 \neq \frac{2}{1} = \frac{p_i}{p_j}. \quad (3.4)$$

Then, if the consumer gives up one unit of good $i$ and purchases one unit of good $j$, he or she will remain on the same indifference curve and have an extra dollar to spend. Hence, total utility can be increased, contradicting maximization.

Figure 3.10 illustrates the argument geometrically. The budget line of the consumer is given by $\{x: p_1x_1 + p_2x_2 = m\}$. This can also be written as the graph of an implicit function: $x_2 = m/p_2 - (p_1/p_2)x_1$. Hence, the budget line has slope $-p_1/p_2$ and vertical intercept $m/p_2$. The consumer wants to find the point on this budget line that achieves highest utility. This must clearly satisfy the tangency condition that the slope of the indifference curve equals the slope of the budget line so that the marginal rate of substitution of $x_1$ for $x_2$ equals the economic rate of substitution of $x_1$ for $x_2$.  

125
Figure 3.10: Preference maximization. The optimal consumption bundle will be at a point where an indifference curve is tangent to the budget constraint.

Remark 3.4.1 The calculus conditions derived above make sense only when the choice variables can be varied in an open neighborhood of the optimal choice and the budget constraint is binding. In many economic problems the variables are naturally nonnegative. If some variables have a value of zero at the optimal choice, the calculus conditions described above may be inappropriate. The necessary modifications of the conditions to handle boundary solutions are not difficult to state. The relevant first-order conditions are given by means of the so-called Kuhn-Tucker conditions:

\[
\frac{\partial u(x)}{\partial x_i} - \lambda p_i \leq 0 \quad \text{with equality if } x_i > 0 \quad i = 1, \ldots, L. \\
px \leq m \quad \text{with equality if } \lambda > 0
\] (3.5)  

Thus the marginal utility from increasing \( x_i \) must be less than or equal to \( \lambda p_i \), otherwise the consumer would increase \( x_i \). If \( x_i = 0 \), the marginal utility from increasing \( x_i \) may be less than \( \lambda p_i \) — which is to say, the consumer would like to decrease \( x_i \). But since \( x_i \) is already zero, this is impossible. Finally, if \( x_i > 0 \) so that the nonnegativity constraint is not binding, we will have the usual conditions for an interior solution.
3.4.4 Sufficiency for Utility Maximization

The above first-order conditions are merely necessary conditions for a local optimum. However, for the particular problem at hand, these necessary first-order conditions are in fact sufficient for a global optimum when a utility function is quasi-concave. We then have the following proposition.

**Proposition 3.4.4** Suppose that \( u(x) \) is differentiable and quasi-concave on \( \mathbb{R}^L_+ \) and \( (p,m) > 0 \). If \((x,\lambda)\) satisfies the first-order conditions given in (3.5) and (3.6), then \( x \) solves the consumer’s utility maximization problem at prices \( p \) and income \( m \).

**Proof.** Since the budget set \( B(p,m) \) is convex and \( u(x) \) is differentiable and quasi-concave on \( \mathbb{R}^L_+ \), by Proposition 2.6.1, we know \( x \) solves the consumer’s utility maximization problem at prices \( p \) and income \( m \).

With the sufficient conditions in hand, it is enough to find a solution \((x, \lambda)\) that satisfies the first-order conditions (3.5) and (3.6). The conditions can typically be used to solve the demand functions \( x_i(p,m) \) as we show in the following examples.

**Example 3.4.1** Suppose the preference ordering is represented by the Cobb-Douglas utility function: \( u(x_1, x_2) = x'^a_1x'^{1-a}_2 \), which is strictly quasi-concave on \( \mathbb{R}^2_+ \). Since any monotonic transform of this function represents the same preferences, we can also write \( u(x_1, x_2) = a \ln x_1 + (1 - a) \ln x_2 \).

The demand functions can be derived by solving the following problem:

\[
\max a \ln x_1 + (1 - a) \ln x_2 \\
\text{such that } p_1 x_1 + p_2 x_2 = m.
\]

The first–order conditions are

\[
\frac{a}{x_1} - \lambda p_1 = 0
\]
or

\[
\frac{a}{p_1 x_1} = \frac{1 - a}{p_2 x_2}.
\]

127
Cross multiply and use the budget constraint to get

\[ a p_2 x_2 = p_1 x_1 - a p_1 x_1 \]
\[ a m = p_1 x_1 \]
\[ x_1(p_1, p_2, m) = \frac{a m}{p_1}. \]

Substitute into the budget constraint to get the demand function for the second commodity:

\[ x_2(p_1, p_2, m) = \frac{(1 - a) m}{p_2}. \]

For Cobb-Douglas utility function with \( L \) commodities,

\[ u(x_1, x_2, \ldots, x_L) = \prod_{l=1}^{L} x_l^{a_l}, \quad a_l > 0, \quad \sum_{l=1}^{L} a_l = 1, \]

It can be verified that the demand function are given by

\[ x_l(p, m) = \frac{a_l m}{p_l}. \]

More generally, when Cobb-Douglas utility function with \( L \) commodities is given by:

\[ u(x_1, x_2, \ldots, x_L) = \prod_{l=1}^{L} x_l^{a_l}, \quad a_l > 0 \]

By monotonic transformation, the utility function can be rewritten as:

\[ u(x_1, x_2, \ldots, x_L) = \prod_{l=1}^{L} \frac{x_l^{a_l}}{\sum_{l=1}^{L} x_l^{a_l}}, \]

we then have

\[ x_l(p, m) = \frac{a_l m}{\sum_{l=1}^{L} p_l}. \]

**Example 3.4.2** Suppose the preference ordering is represented by the Leontief utility function: \( u(x_1, x_2) = \min\{ax_1, bx_2\} \). Since the Leontief utility function is not differentiable, so the maximum must be found by a direct argument. Assume \( p > 0 \).
The optimal solution must be at the kink point of the indifference curve. That is,

\[ ax_1 = bx_2. \]

Substituting \( x_1 = \frac{b}{a} x_2 \) into the budget constraint \( px = m \), we have

\[ \frac{b}{a} p_1 x_2 + p_2 x_2 = m \]

and thus the demand functions are given by

\[ x_2(p_1, p_2, m) = \frac{am}{bp_1 + ap_2} \]

and

\[ x_1(p_1, p_2, m) = \frac{bm}{bp_1 + ap_2} \]

For Leontief utility function with \( L \) commodities,

\[ u(x_1, x_2, \ldots, x_L) = \min\{a_1 x_1, a_2 x_2, \ldots, a_L x_L\}, \quad a_i > 0 \]

It can be verified that the demand function are given by

\[ x_i(p, m) = \frac{m}{\sum_{i=1}^{L} \frac{p_i}{a_i}}. \]

**Example 3.4.3** Now suppose the preference ordering is represented by the linear utility function:

\[ u(x, y) = ax + by. \]

Since the marginal rate of substitution is \( a/b \) and the economic rate of substitution is \( p_x/p_y \) are both constant, they cannot be in general equal. So the first-order condition cannot hold with equality as long as \( a/b \neq p_x/p_y \). In this case the answer to the utility-maximization problem typically involves a boundary solution: only one of the two goods will be consumed. It is worthwhile presenting a more formal solution since it serves as a nice example of the Kuhn-Tucker theorem in action. The Kuhn-Tucker theorem is the appropriate tool to use here, since we will almost never have an interior solution.
The Lagrange function is

\[ L(x, y, \lambda) = ax + by + \lambda(m - p_x x - p_y y) \]

and thus

\[
\begin{align*}
\frac{\partial L}{\partial x} &= a - \lambda p_x \quad (3.7) \\
\frac{\partial L}{\partial y} &= b - \lambda p_y \quad (3.8) \\
\frac{\partial L}{\partial \lambda} &= m - p_x x - p_y y \quad (3.9) \\
\end{align*}
\]

There are four cases to be considered:

**Case 1.** \( x > 0 \) and \( y > 0 \). Then we have \( \frac{\partial L}{\partial x} = 0 \) and \( \frac{\partial L}{\partial y} = 0 \). Thus, \( \frac{a}{b} = \frac{p_x}{p_y} \).

Since \( \lambda = \frac{a}{p_x} > 0 \), we have \( p_x x + p_y y = m \) and thus all \( x \) and \( y \) that satisfy \( p_x x + p_y y = m \) are the optimal consumptions.

**Case 2.** \( x > 0 \) and \( y = 0 \). Then we have \( \frac{\partial L}{\partial x} = 0 \) and \( \frac{\partial L}{\partial y} \leq 0 \). Thus, \( \frac{a}{b} \geq \frac{p_x}{p_y} \).

Since \( \lambda = \frac{a}{p_x} > 0 \), we have \( p_x x + p_y y = m \) and thus \( x = \frac{m}{p_x} \) is the optimal consumption.

**Case 3.** \( x = 0 \) and \( y > 0 \). Then we have \( \frac{\partial L}{\partial x} \leq 0 \) and \( \frac{\partial L}{\partial y} = 0 \). Thus, \( \frac{a}{b} \leq \frac{p_x}{p_y} \).

Since \( \lambda = \frac{b}{p_y} > 0 \), we have \( p_x x + p_y y = m \) and thus \( y = \frac{m}{p_y} \) is the optimal consumption.

**Case 4.** \( x = 0 \) and \( y = 0 \). Then we have \( \frac{\partial L}{\partial x} \leq 0 \) and \( \frac{\partial L}{\partial y} \leq 0 \). Since \( \lambda \geq \frac{b}{p_y} > 0 \), we have \( p_x x + p_y y = m \) and thus \( m = 0 \) because \( x = 0 \) and \( y = 0 \).

In summary, the demand functions are given by

\[
(x(p_x, p_y, m), y(p_x, p_y, m)) = \begin{cases} 
(m/p_x, 0) & \text{if } a/b > p_x/p_y \\
(0, m/p_y) & \text{if } a/b < p_x/p_y \\
(x, m/p_x - p_y/p_x x) & \text{if } a/b = p_x/p_y 
\end{cases}
\]

for all \( x \in [0, m/p_x] \).

**Remark 3.4.2** In fact, it is easily found out the optimal solutions by comparing relatives steepness of the indifference curves and the budget line. For instance,
as shown in Figure 3.11 below, when $a/b > p_x/p_y$, the indifference curves become steeper, and thus the optimal solution is the one the consumer spends his all income on good $x$. When $a/b < p_x/p_y$, the indifference curves become flatter, and thus the optimal solution is the one the consumer spends his all income on good $y$. When $a/b = p_x/p_y$, the indifference curves and the budget line are parallel and coincide at the optimal solutions, and thus the optimal solutions are given by all the points on the budget line.

### 3.4.5 Non-Ordering Preference and Its Maximization

To be completed
3.5 Indirect Utility, and Expenditure, and Money Metric Utility Functions

3.5.1 The Indirect Utility Function

The ordinary utility function, \( u(x) \), is defined over the consumption set \( X \) and therefore to as the direct utility function. Given prices \( p \) and income \( m \), the consumer chooses a utility-maximizing bundle \( x(p, m) \). The level of utility achieved when \( x(p, m) \) is chosen thus will be the highest level permitted by the consumer’s budget constraint facing \( p \) and \( m \), and can be denoted by

\[
v(p, m) = \max u(x) \text{ such that } px = m.
\]

The function \( v(p, m) \) that gives us the maximum utility achievable at given prices and income is called the indirect utility function and thus it is a compose function of \( u(\cdot) \) and \( x(p, m) \), i.e.,

\[
v(p, m) = u(x(p, m)). \tag{3.11}
\]

The properties of the indirect utility function are summarized in the following proposition.

**Proposition 3.5.1 (Properties of the indirect utility function)** If \( u(x) \) is continuous and monotonic on \( \mathbb{R}^L_+ \) and \( (p, m) > 0 \), the indirect utility function has the following properties:

1. \( v(p, m) \) is nonincreasing in \( p \); that is, if \( p' \geq p \), \( v(p', m) \leq v(p, m) \). Similarly, \( v(p, m) \) is nondecreasing in \( m \).

2. \( v(p, m) \) is homogeneous of degree 0 in \( (p, m) \).

3. \( v(p, m) \) is quasiconvex in \( p \); that is, \( \{p: v(p, m) \leq k\} \) is a convex set for all \( k \).
(4) \( v(p, m) \) is continuous at all \( p \gg 0, m > 0 \).

**Proof.**

(1) Let \( B = \{x: px \leq m\} \) and \( B' = \{x: p'x \leq m\} \) for \( p' \geq p \). Then \( B' \) is contained in \( B \). Hence, the maximum of \( u(x) \) over \( B \) is at least as big as the maximum of \( u(x) \) over \( B' \). The argument for \( m \) is similar.

(2) If prices and income are both multiplied by a positive number, the budget set doesn’t change at all. Thus, \( v(tp, tm) = v(p, m) \) for \( t > 0 \).

(3) Suppose \( p \) and \( p' \) are such that \( v(p, m) \leq k, v(p', m) \leq k \). Let \( p'' = tp + (1 - t)p' \). We want to show that \( v(p'', m) \leq k \). Define the budget sets:

\[
\begin{align*}
B &= \{x: px \leq m\} \\
B' &= \{x: p'x \leq m\} \\
B'' &= \{x: p''x \leq m\}
\end{align*}
\]

We will show that any \( x \) in \( B'' \) must be in either \( B \) or \( B' \); that is, that \( B \cup B' \supset B'' \). Suppose not. We then must have \( px > m \) and \( p'x > m \). Multiplying the first inequality by \( t \) and the second by \( (1 - t) \) and then summing, we find that \( tpx + (1 - t)p'x > m \) which contradicts our original assumption. Then \( B'' \) is contained in \( B \cup B' \), the maximum of \( u(x) \) over \( B'' \) is at most as big as the maximum of \( u(x) \) over \( B \cup B' \), and thus \( v(p'', m) \leq k \) by noting that \( v(p, m) \leq k \) and \( v(p', m) \leq k \).

(4) This follows from the Maximum Theorem.

**Example 3.5.1 (The General Cobb-Douglas Utility Function)** Suppose a preference ordering is represented by the Cobb-Douglas utility function is given by:

\[
u(x) = \prod_{l=1}^{L} (x_l)^{\alpha_l}, \quad \alpha_l > 0, \ l = 1, 2, \ldots, L.
\]
Since any monotonic transform of this function represents the same preferences, we can also write

\[ u(x) = \prod_{l=1}^{L} (x_l)^{\alpha_l}, \quad \alpha_l > 0, \ l = 1, 2, \ldots, L. \]

where \( \alpha = \sum_{l=1}^{L} \alpha_l \). Let \( a_l = \alpha_l / \alpha \). Then it reduces to the Cobb-Douglas utility we examined before and thus the demand functions are given by

\[ x_l(p, m) = \frac{a_l m}{p_l} = \frac{\alpha_l m}{\alpha p_l}, \ l = 1, 2, \ldots, L. \]

Substitute into the objective function and eliminate constants to get the indirect utility function:

\[ v(p, m) = \prod_{l=1}^{L} \left( \frac{\alpha_l m}{\alpha p_l} \right)^{\alpha_l} \]

The above example also shows that monotonic transformation sometimes is very useful to simplify the computation of finding solutions.

### 3.5.2 The Expenditure Function and Hicksian Demand

We note that if preferences satisfy the local nonsatiation assumption, then \( e(p, m) \) will be strictly increasing in \( m \). We then can invert the function and solve for \( m \) as a function of the level of utility; that is, given any level of utility, \( u \), we can find the minimal amount of income necessary to achieve utility \( u \) at prices \( p \). The function that relates income and utility in this way— the inverse of the indirect utility function – is known as the expenditure function and is denoted by \( e(p, u) \). Formally, the expenditure function is given by the following problem:

\[
\begin{align*}
 e(p, u) &= \min \ p x \\
 \text{such that } u(x) &\geq u.
\end{align*}
\]

The expenditure function gives the minimum cost of achieving a fixed level of utility. The solution which is the function of \( (p, u) \) is denoted by \( h(p, u) \) and called the Hicksian demand function. The Hicksian demand function tells us
what consumption bundle achieves a target level of utility and minimizes total expenditure.

A Hicksian demand function is sometimes called a compensated demand function. This terminology comes from viewing the demand function as being constructed by varying prices and income so as to keep the consumer at a fixed level of utility. Thus, the income changes are arranged to “compensate” for the price changes. Hicksian demand functions are not directly observable since they depend on utility, which is not directly observable. Demand functions expressed as a function of prices and income are observable; when we want to emphasize the difference between the Hicksian demand function and the usual demand function, we will refer to the latter as the Marshallian demand function, \( x(p, m) \). The Marshallian demand function is just the ordinary market demand function we have been discussing all along.

**Proposition 3.5.2** [Properties of the Expenditure Function.] If \( u(x) \) is continuous and locally non-satiated on \( \mathbb{R}^L_+ \) and \( (p, m) > 0 \), the expenditure function has the following properties:

1. \( e(p, u) \) is nondecreasing in \( p \).
2. \( e(p, u) \) is homogeneous of degree 1 in \( p \).
3. \( e(p, u) \) is concave in \( p \).
4. \( e(p, u) \) is continuous in \( p \), for \( p \gg 0 \).
5. For all \( p > 0 \), \( e(p, u) \) is strictly increasing in \( u \).
6. Shephard’s lemma: If \( h(p, u) \) is the expenditure-minimizing bundle necessary to achieve utility level \( u \) at prices \( p \), then \( h_i(p, u) = \frac{\partial e(p, u)}{\partial p_i} \) for \( i = 1, ..., k \) assuming the derivative exists and that \( p_i > 0 \).

**Proof.** Since the expenditure function is the inverse function of the indirect utility function, Properties (1), (4)-(5) are true by Properties (1) and (4) of the
indirect utility given in Proposition 3.5.1. We only need to show the Properties (2), (3) and (6).

(2) We show that if \( x \) is the expenditure-minimizing bundle at prices \( p \), then \( x \) also minimizes the expenditure at prices \( tp \). Suppose not, and let \( x' \) be an expenditure minimizing bundle at \( tp \) so that \( tpx' < tpx \). But this inequality implies \( px' < px \), which contradicts the definition of \( x \). Hence, multiplying prices by a positive scalar \( t \) does not change the composition of an expenditure minimizing bundle, and, thus, expenditures must rise by exactly a factor of \( t \): \( e(p, u) = tpx = te(p, u) \).

(3) Let \((p, x)\) and \((p', x')\) be two expenditure-minimizing price-consumption combinations and let \( p'' = t(p + (1 - t)p') \) for any \( 0 \leq t \leq 1 \). Now,

\[
e(p'', u) = p''x'' = tpx'' + (1 - t)p'x''.
\]

Since \( x'' \) is not necessarily the minimal expenditure to reach \( u \) at prices \( p' \) or \( p \), we have \( px'' \geq e(p, u) \) and \( p' \cdot x'' \geq e(p', u) \). Thus,

\[
e(p'', u) \geq te(p, u) + (1 - t)e(p', u).
\]

(6) Let \( x^* \) be an expenditure-minimizing bundle to achieve utility level \( u \) at prices \( p^* \). Then define the function

\[
g(p) = e(p, u) - px^*.
\]

Since \( e(p, u) \) is the cheapest way to achieve \( u \), this function is always non-positive. At \( p = p^* \), \( g(p^*) = 0 \). Since this is a maximum value of \( g(p) \), its derivative must vanish:

\[
\frac{\partial g(p^*)}{\partial p_i} = \frac{\partial e(p^*, u)}{\partial p_i} - x^*_i = 0 \quad i = 1, \ldots, L.
\]

Hence, the expenditure-minimizing bundles are just given by the vector of derivatives of the expenditure function with respect to the prices.
Remark 3.5.1 We can also prove property (6) by applying the Envelop Theorem for the constrained version. In this problem the parameter $a$ can be chosen to be one of the prices, $p_i$. Define the Lagrange function $L(x, \lambda) = px - \lambda(u - u(x))$. The optimal value function is the expenditure function $e(p, u)$. The envelope theorem asserts that

$$\frac{\partial e(p, u)}{\partial p_i} = \frac{\partial L}{\partial p_i} = x_i \bigg|_{x_i = h_i(p, u)} = h_i(p, u),$$

which is simply Shephard’s lemma.

We now give some basic properties of Hicksian demand functions:

**Proposition 3.5.3 (Negative Semi-Definite Substitution Matrix)** The matrix of substitution terms $(\partial h_j(p, u)/\partial p_i)$ is negative semi-definite.

Proof. This follows

$$\frac{\partial h_j(p, u)}{\partial p_i} = \frac{\partial^2 e(p, u)}{\partial p_i \partial p_j},$$

which is negative semi-definite because the expenditure function is concave.

Since the substitution matrix is negative semi-definite, thus it is symmetric and has non-positive diagonal terms. Then we have

**Proposition 3.5.4 (Symmetric Substitution Terms)** The matrix of substitution terms is symmetric, i.e.,

$$\frac{\partial h_j(p, u)}{\partial p_i} = \frac{\partial^2 e(p, u)}{\partial p_j \partial p_i} = \frac{\partial^2 e(p, u)}{\partial p_i \partial p_j} = \frac{\partial h_i(p, u)}{\partial p_j}.$$

**Proposition 3.5.5 (Negative Own-Substitution Terms)** The compensated own-price effect is non-positive; that is, the Hicksian demand curves slope downward:

$$\frac{\partial h_i(p, u)}{\partial p_i} = \frac{\partial^2 e(p, u)}{\partial p_i^2} \leq 0.$$
3.5.3 The Money Metric Utility Functions

There is a nice construction involving the expenditure function that comes up in a variety of places in welfare economics. Consider some prices $p$ and some given bundle of goods $x$. We can ask the following question: how much money would a given consumer need at the prices $p$ to be as well off as he could be by consuming the bundle of goods $x$? If we know the consumer’s preferences, we can simply solve the following problem:

$$m(p, x) \equiv \min_z pz$$

such that $u(z) \geq u(x)$.

That is,

$$m(p, x) \equiv e(p, u(x)).$$

This type of function is called money metric utility function. It is also known as the “minimum income function,” the “direct compensation function,” and by a variety of other names. Since, for fixed $p$, $m(p, x)$ is simply a monotonic transform of the utility function and is itself a utility function.

There is a similar construct for indirect utility known as the money metric indirect utility function, which is given by

$$\mu(p; q, m) \equiv e(p, \nu(q, m)).$$

That is, $\mu(p; q, m)$ measures how much money one would need at prices $p$ to be as well off as one would be facing prices $q$ and having income $m$. Just as in the direct case, $\mu(p; q, m)$ is simply a monotonic transformation of an indirect utility function.

Example 3.5.2 (The CES Utility Function) The CES utility function is given by $u(x_1, x_2) = (x_1^\rho + x_2^\rho)^{1/\rho}$, where $0 \neq \rho < 1$. It can be easily verified this utility function is strongly monotonic increasing and strictly concave. Since preferences are invariant with respect to monotonic transforms of utility, we could just as well choose $u(x_1, x_2) = \frac{1}{\rho} \ln(x_1^\rho + x_2^\rho)$. 138
The first-order conditions are

\[
\frac{x_1^{\rho - 1}}{x_1^\rho + x_2^\rho} - \lambda p_1 = 0 \\
\frac{x_2^{\rho - 1}}{x_1^\rho + x_2^\rho} - \lambda p_2 = 0 \\
p_1 x_1 + p_2 x_2 = m
\]

Dividing the first equation by the second equation and then solving for \(x_2\), we have

\[
x_2 = x_1 \left(\frac{p_2}{p_1}\right)^{\frac{1}{\rho - 1}}.
\]

Substituting the above equation in the budget line and solving for \(x_1\), we obtain

\[
x_1(p, m) = \frac{1}{p_1^{\frac{1}{\rho - 1}}} \frac{m}{x_1^\rho + x_2^\rho}
\]

and thus

\[
x_2(p, m) = \frac{1}{p_2^{\frac{1}{\rho - 1}}} \frac{m}{x_1^\rho + x_2^\rho}
\]

Substituting the demand functions into the utility function, we get the indirect CES utility function:

\[
v(p, m) = (p_1^{\rho/(\rho - 1)} + p_2^{\rho/(\rho - 1)})^{1 - \rho/\rho} m
\]

or

\[
v(p, m) = (p_1^r + p_2^r)^{-1/r} m
\]

where \(r = \rho/\rho - 1\). Inverting the above equation, we obtain the expenditure function for the CES utility function which has the form

\[
e(p, u) = (p_1^r + p_2^r)^{1/r} u.
\]

Consequently, the money metric direct and indirect utility functions are given by

\[
m(p, x) = (p_1^r + p_2^r)^{1/r} (x_1^\rho + x_2^\rho)^{1/\rho}
\]

and

\[
\mu(p; q, m) = (p_1^r + p_2^r)^{1/r} (q_1^r + q_2^r)^{-1/r} m.
\]
Remark 3.5.2 The CES utility function contains several other well-known utility functions as special cases, depending on the value of the parameter $\rho$.

(1) The linear utility function ($\rho = 1$). Simple substitution yields
\[ y = x_1 + x_2. \]

(2) The Cobb-Douglas utility function ($\rho = 0$). When $\rho = 0$ the CES utility function is not defined, due to division by zero. However, we will show that as $\rho$ approaches zero, the indifference curves of the CES utility function look very much like the indifference curves of the Cobb-Douglas utility function.

This is easiest to see using the marginal rate of substitution. By direct calculation,
\[ \frac{\partial u}{\partial x_2} = \frac{x_1}{x_2}. \]

(3) The Leontief utility function ($\rho = -\infty$). We have just seen that the $MRS$ of the CES utility function is given by equation (3.12). As $\rho$ approaches $-\infty$, this expression approaches
\[ \frac{x_2}{x_1}. \]

which is simply the $MRS$ for the Cobb-Douglas utility function.

3.5.4 Some Important Identities

There are some important identities that tie together the expenditure function, the indirect utility function, the Marshallian demand function, and the Hicksian demand function.
Let us consider the utility maximization problem

\[ v(p, m^*) = \max u(x) \quad \text{such that } px \leq m^*. \]  

(3.13)

Let \( x^* \) be the solution to this problem and let \( u^* = u(x^*) \). Consider the expenditure minimization problem

\[ e(p, u^*) = \min px \quad \text{such that } u(x) \geq u^*. \]  

(3.14)

An inspection of Figure 3.12 should convince you that the answers to these two problems should be the same \( x^* \). Formally, we have the following proposition.

![Figure 3.12: Maximize utility and minimize expenditure are normally equivalent.](image)

**Proposition 3.5.6 (Equivalence of Utility Max and Expenditure Min)**

*Suppose the utility function \( u \) is continuous and locally non-satiated, and suppose that \( m > 0 \). If the solutions both problems exist, then the above two problems have the same solution \( x^* \). That is,*

1. **Utility maximization implies expenditure minimization:** Let \( x^* \) be a solution to (3.13), and let \( u = u(x^*) \). Then \( x^* \) solves (3.14).
Expenditure minimization implies utility maximization. Suppose that the above assumptions are satisfied and that $x^*$ solves (3.14). Let $m = px^*$. Then $x^*$ solves (3.13).

Proof.

1 Suppose not, and let $x'$ solve (3.14). Hence, $px' < px^*$ and $u(x') \geq u(x^*)$. By local nonsatiation there is a bundle $x''$ close enough to $x'$ so that $px'' < px^* = m$ and $u(x'') > u(x^*)$. But then $x^*$ cannot be a solution to (3.13).

2 Suppose not, and let $x'$ solve (3.13) so that $u(x') > u(x^*)$ and $px' = px^* = m$. Since $px^* > 0$ and utility is continuous, we can find $0 < t < 1$ such that $ptx' < px^* = m$ and $u(tx') > u(x^*)$. Hence, $x$ cannot solve (3.14).

This proposition leads to four important identities that is summarized in the following proposition.

Proposition 3.5.7 Suppose the utility function $u$ is continuous and locally nonsatiated, and suppose that $m > 0$. Then we have

1. $e(p, v(p, m)) \equiv m$. The minimum expenditure necessary to reach utility $v(p, m)$ is $m$.

2. $v(p, e(p, u)) \equiv u$. The maximum utility from income $e(p, u)$ is $u$.

3. $x_i(p, m) \equiv h_i(p, v(p, m))$. The Marshallian demand at income $m$ is the same as the Hicksian demand at utility $v(p, m)$.

4. $h_i(p, u) \equiv x_i(p, e(p, u))$. The Hicksian demand at utility $u$ is the same as the Marshallian demand at income $e(p, u)$.

This last identity is perhaps the most important since it ties together the “observable” Marshallian demand function with the “unobservable” Hicksian demand function. Thus, any demanded bundle can be expressed either as the solution to the utility maximization problem or the expenditure minimization problem.
A nice application of one of these identities is given in the next proposition:

Roy’s identity. If \( x(p, m) \) is the Marshallian demand function, then

\[
x_i(p, m) = -\frac{\partial v(p, m)}{\partial p_i} \frac{\partial p_i}{\partial m}
\]

for \( i = 1, \ldots, k \)

provided that the right-hand side is well defined and that \( p_i > 0 \) and \( m > 0 \).

**Proof.** Suppose that \( x^* \) yields a maximal utility of \( u^* \) at \((p^*, m^*)\). We know from our identities that

\[
x(p^*, m^*) = h(p^*, u^*). \tag{3.15}
\]

From another one of the fundamental identities, we also know that

\[
u^* = v(p, e(p, u^*)).
\]

Since this is an identity we can differentiate it with respect to \( p_i \) to get

\[
0 = \frac{\partial v(p^*, m^*)}{\partial p_i} + \frac{\partial v(p^*, m^*)}{\partial m} \frac{\partial e(p^*, u^*)}{\partial p_i}.
\]

Rearranging, and combining this with identity (3.15), we have

\[
x_i(p^*, m^*) = h_i(p^*, u^*) \equiv \frac{\partial e(p^*, u^*)}{\partial p_i} \equiv -\frac{\partial v(p^*, m^*)}{\partial p_i} \frac{\partial v(p^*, m^*)}{\partial m}.
\]

Since this identity is satisfied for all \((p^*, m^*)\) and since \( x^* = x(p^*, m^*) \), the result is proved.

**Example 3.5.3 (The General Cobb-Douglas Utility Function)** Consider the indirect Cobb-Douglas utility function:

\[
v(p, m) = \prod_{l=1}^{L} \left( \frac{\alpha_l m}{\alpha_p} \right)^{\alpha_l}
\]

where \( \sum_{l=1}^{L} \alpha_l \). Then we have

\[
v_p \equiv \frac{\partial v(p, m)}{\partial p_l} = -\frac{\alpha_l}{p_j} v(p, m)
\]

\[
v_m \equiv \frac{\partial v(p, m)}{\partial m} = \frac{\alpha}{m} v(p, m)
\]

Thus, Roy’s identity gives the demand functions as

\[
x_l(p, m) = \frac{\alpha_l m}{\alpha_p}, \quad l = 1, 2, \ldots, L.
\]
Example 3.5.4 (The General Leontief Utility Function) Suppose a preference ordering is represented by the Leontief utility function is given by:

\[ u(x) = \min\{\frac{x_1}{a_1}, \frac{x_2}{a_2}, \ldots, \frac{x_L}{a_L}\} \]

which is not itself differentiable, but the indirect utility function:

\[ v(p, m) = \frac{m}{ap} \] (3.16)

where \( a = (a_1, a_2, \ldots, a_L) \) and \( ap \) is the inner product, which is differentiable. Applying Roy’s identity, we have

\[ x_l(p, m) = -\frac{v_p}{v_m} = \frac{a_l m}{ap} \]

Hence, Roy’s identity often works well even if the differentiability properties of the statement do not hold.

Example 3.5.5 (The CES Utility Function) The CES utility function is given by \( u(x_1, x_2) = (x_1^\rho + x_2^\rho)^{1/\rho} \). We have derived earlier that the indirect utility function is given by:

\[ v(p, m) = (p_1^r + p_2^r)^{-1/r}m. \]

The demand functions can be found by Roy’s law:

\[ x_l(p, m) = -\frac{\partial v(p, m)}{\partial p_l}/\frac{\partial v(p, m)}{\partial m} = \frac{1}{r}(p_1^r + p_2^r)^{-1/r}m \frac{1}{r} \frac{(p_1^r + p_2^r)^{-1/r}m}{p_l^r} = \frac{p_l^r m}{(p_1^r + p_2^r)}, \ l = 1, 2. \]

3.5.5 Homothetic Utility Functions

It can easily shown that if the utility function is homogeneous of degree 1, then the expenditure function can be written as \( e(p, u) = e(p)u \). This in turn implies that the indirect utility function can be written as \( v(p, m) = v(p)m \). Roy’s identity then implies that the demand functions take the form \( x_i(p, m) = x_i(p)m \) — i.e., they are linear functions of income. The fact that the “income effects” take this special form is often useful in demand analysis, as we will see below.
3.6 Duality Between Direct and Indirect Utility

We have seen how one can recover an indirect utility function from observed demand functions by solving the integrability equations. Here we see how to solve for the direct utility function.

The answer exhibits quite nicely the duality between direct and indirect utility functions. It is most convenient to describe the calculations in terms of the normalized indirect utility function, where we have prices divided by income so that expenditure is identically one. Thus the normalized indirect utility function is given by

\[ v(p) = \max_x u(x) \]

such that \( px = 1 \)

We then have the following proposition

**Proposition 3.6.1** Given the indirect utility function \( v(p) \), the direct utility function can be obtained by solving the following problem:

\[ u(x) = \min_p v(p) \]

such that \( px = 1 \)

Proof. Let \( x \) be the demanded bundle at the prices \( p \). Then by definition \( v(p) = u(x) \). Let \( p' \) be any other price vector that satisfies the budget constraint so that \( p'x = 1 \). Then since \( x \) is always a feasible choice at the prices \( p' \), due to the form of the budget set, the utility-maximizing choice must yield utility at least as great as the utility yielded by \( x \); that is, \( v(p') \geq u(x) = v(p) \). Hence, the minimum of the indirect utility function over all \( p \)'s that satisfy the budget constraint gives us the utility of \( x \).

The argument is depicted in Figure 3.13. Any price vector \( p \) that satisfies the budget constraint \( px = 1 \) must yield a higher utility than \( u(x) \), which is simply to say that \( u(x) \) solves the minimization problem posed above.
Example 3.6.1 (Solving for the Direct Utility Function) Suppose that we have an indirect utility function given by $v(p_1, p_2) = -a \ln p_1 - b \ln p_2$. What is its associated direct utility function? We set up the minimization problem:

$$\min_{p_1, p_2} -a \ln p_1 - b \ln p_2$$

such that $p_1 x_1 + p_2 x_2 = 1$.

The first-order conditions are

$$-a/p_1 = \lambda x_1$$
$$-b/p_2 = \lambda x_2,$$

or,

$$-a = \lambda p_1 x_1$$
$$-b = \lambda p_2 x_2.$$

Adding together and using the budget constraint yields

$$\lambda = -a - b.$$

Substitute back into the first-order conditions to find

$$p_1 = \frac{a}{(a + b)x_1}$$
$$p_2 = \frac{b}{(a + b)x_2}.$$
These are the choices of \((p_1, p_2)\) that minimize indirect utility. Now substitute these choices into the indirect utility function:

\[
\begin{align*}
u(x_1 x_2) &= -a \ln \frac{a}{(a + b)x_1} - b \ln \frac{b}{(a + b)x_2} \\
&= a \ln x_1 + b \ln x_2 + \text{constant}.
\end{align*}
\]

This is the familiar Cobb-Douglas utility function.

The duality between seemingly different ways of representing economic behavior is useful in the study of consumer theory, welfare economics, and many other areas in economics. Many relationships that are difficult to understand when looked at directly become simple, or even trivial, when looked at using the tools of duality.

### 3.7 Properties of Consumer Demand

In this section we will examine the comparative statics of consumer demand behavior: how the consumer’s demand changes as prices and income change.

#### 3.7.1 Income Changes and Consumption Choice

It is of interest to look at how the consumer’s demand changes as we hold prices fixed and allow income to vary; the resulting locus of utility-maximizing bundles is known as the income expansion path. From the income expansion path, we can derive a function that relates income to the demand for each commodity (at constant prices). These functions are called Engel curves. There are two possibilities: (1) As income increases, the optimal consumption of a good increases. Such a good is called a normal good. (2) As income increases, the optimal consumption of a good decreases. Such a good is called interior good.

For the two-good consumer maximization problem, when the income expansion path (and thus each Engel curve) is upper-ward slopping, both goods are normal goods. When the income expansion path could bend backwards, there is
one and only one good that is inferior when the utility function is locally non-satiated; increase in income means the consumer actually wants to consume less of the good. (See Figure 3.14)

![Figure 3.14: Income expansion paths with an interior good.](image)

3.7.2 Price Changes and Consumption Choice

We can also hold income fixed and allow prices to vary. If we let \( p_1 \) vary and hold \( p_2 \) and \( m \) fixed, the locus of tangencies will sweep out a curve known as the **price offer curve**. In the first case in Figure 3.15 we have the ordinary case where a lower price for good 1 leads to greater demand for the good so that the Law of Demand is satisfied; in the second case we have a situation where a decrease in the price of good 1 brings about a **decreased** demand for good 1. Such a good is called a **Giffen good**. Although it is very hard to find a real world example for giffen good, it exists either from experiment (cf. Battallio, Kagel, and Kogut (1991) or from theory. Indeed, Haagsma (2012) found a convenient utility function that is strictly increasing, strictly quasi-concave and continuous and leads to Giffen Behaviour.
Figure 3.15: Offer curves. In panel A the demand for good 1 increases as the price decreases so it is an ordinary good. In panel B the demand for good 1 decreases as its price decreases, so it is a Giffen good.

3.7.3 Income-Substitution Effect, Slutsky Equation and Hicksian Equation

In the above we see that a fall in the price of a good may have two sorts of effects: substitution effect—one commodity will become less expensive than another, and income effect — total “purchasing power” increases. A fundamental result of the theory of the consumer, the Slutsky equation, relates these two effects.

Even though the compensated demand function is not directly observable, we shall see that its derivative can be easily calculated from observable things, namely, the derivative of the Marshallian demand with respect to price and income. This relationship is known as the Slutsky equation.

Slutsky equation.

\[
\frac{\partial x_j(p, m)}{\partial p_i} = \frac{\partial h_j(p, v(p, m))}{\partial p_i} - \frac{\partial x_j(p, m)}{\partial m} x_i(p, m)
\]

Proof. Let \( x^* \) maximize utility at \((p^*, m)\) and let \( u^* = u(x^*) \). It is identically true that

\[
h_j(p^*, u^*) \equiv x_j(p, e(p, u^*)).
\]
We can differentiate this with respect to $p_i$ and evaluate the derivative at $p^*$ to get

$$\frac{\partial h_j(p^*, u^*)}{\partial p_i} = \frac{\partial x_j(p^*, m^*)}{\partial p_i} + \frac{\partial x_j(p^*, m^*)}{\partial m} \frac{\partial e(p^*, u^*)}{\partial p_i}.$$  

Note carefully the meaning of this expression. The left-hand side is how the compensated demand changes when $p_i$ changes. The right-hand side says that this change is equal to the change in demand holding expenditure fixed at $m^*$ plus the change in demand when income changes times how much income has to change to keep utility constant. But this last term, $\frac{\partial e(p^*, u^*)}{\partial p_i}$, is just $x_i^*$; rearranging gives us

$$\frac{\partial x_j(p^*, m^*)}{\partial p_i} = \frac{\partial h_j(p^*, u^*)}{\partial p_i} - \frac{\partial x_j(p^*, m^*)}{\partial m} x_i^* \Delta p_i$$

which is the Slutsky equation.

There are other ways to derive Slutsky’s equations that can be found in Varian (1992).

The Slutsky equation decomposes the demand change induced by a price change $\Delta p_i$ into two separate effects: the substitution effect and the income effect:

$$\Delta x_j \approx \frac{\partial x_j(p, m)}{\partial p_i} \Delta p_i = \frac{\partial h_j(p, u)}{\partial p_i} \Delta p_i - \frac{\partial x_j(p, m)}{\partial m} x_i^* \Delta p_i$$

As we mentioned previously, the restrictions all about the Hicksian demand functions are not directly observable. However, as indicated by the Slutsky equation, we can express the derivatives of $h$ with respect to $p$ as derivatives of $x$ with respect to $p$ and $m$, and these are observable.

Also, Slutsky’s equation and the negative semi-definite matrix on Hicksian demand functions given in Proposition 3.5.3 give us the following result on the Marshallian demand functions:

**Proposition 3.7.1** The substitution matrix $\left( \frac{\partial x_i(p, m)}{\partial p_i} + \frac{\partial x_i(p, u)}{\partial m} x_i \right)$ is a symmetric, negative semi-definite matrix.

This is a rather nonintuitive result: a particular combination of price and income derivatives has to result in a negative semidefinite matrix.
Figure 3.16: The Hicks decomposition of a demand change into two effects: the substitution effect and the income effect.

Example 3.7.1 (The Cobb-Douglas Slutsky equation) Let us check the Slutsky equation in the Cobb-Douglas case. As we’ve seen, in this case we have

\[ v(p_1, p_2, m) = mp_1^{-\alpha}p_2^{\alpha-1} \]
\[ e(p_1, p_2, u) = wp_1^{\alpha}p_2^{-\alpha} \]
\[ x_1(p_1, p_2, m) = \frac{am}{p_1} \]
\[ h_1(p_1, p_2, u) = \alpha p_1^{\alpha-1}p_2^{1-\alpha}u. \]

Thus

\[ \frac{\partial x_1(p, m)}{\partial p_1} = -\frac{\alpha m}{p_1^2} \]
\[ \frac{\partial x_1(p, m)}{\partial m} = -\frac{\alpha}{p_1} \]
\[ \frac{\partial h_1(p, u)}{\partial p_1} = \alpha(\alpha - 1)p_1^{\alpha-2}p_2^{1-\alpha}u \]
\[ \frac{\partial h_1(p, v(p, m))}{\partial p_1} = \alpha(\alpha - 1)p_1^{\alpha-2}p_2^{1-\alpha}mp_1^{-\alpha}p_2^{\alpha-1} \]
\[ = \alpha(\alpha - 1)p_1^{-2}m. \]
Now plug into the Slutsky equation to find

\[
\frac{\partial h_1}{\partial p_1} - \frac{\partial x_1}{\partial m} = \frac{\alpha(\alpha - 1)m}{p_1^2} - \frac{\alpha m}{p_1} \frac{\alpha m}{p_1} = \frac{\alpha(\alpha - 1) - \alpha^2}{p_1^2} = -\frac{\alpha m}{p_1^2} = \frac{\partial x_1}{\partial p_1}.
\]

In our study of consumer behavior we have taken income to be exogenous. But in more elaborate models of consumer behavior it is necessary to consider how income is generated. The standard way to do this is to think of the consumer as having some endowment \( \omega = (\omega_1, \ldots, \omega_L) \) of various goods which can be sold at the current market prices \( p \). This gives the consumer income \( m = p_\omega \) which can be used to purchase other goods.

The utility maximization problem becomes

\[
\max_{x} u(x) \quad \text{such that } px = p_\omega.
\]

This can be solved by the standard techniques to find a demand function \( x(p, p_\omega) \). The net demand for good \( i \) is \( x_i - \omega_i \). The consumer may have positive or negative net demands depending on whether he wants more or less of something than is available in his endowment.

In this model prices influence the value of what the consumer has to sell as well as the value of what the consumer wants to sell. This shows up most clearly in Slutsky’s equation, which we now derive. First, differentiate demand with respect to price:

\[
\frac{dx_i(p, p_\omega)}{dp_j} = \left. \frac{\partial x_i(p, p_\omega)}{\partial p_j} \right|_{p_\omega = \text{constant}} + \frac{\partial x_i(p, p_\omega)}{dm} \omega_j.
\]

The first term in the right-hand side of this expression is the derivative of demand with respect to price, holding income fixed. The second term is the derivative of demand with respect to income, times the change in income. The first term can
be expanded using Slutsky’s equation. Collecting terms we have
\[
\frac{dx_i(p, p\omega)}{dp_j} = \frac{\partial h_i(p, u)}{\partial p_j} + \frac{\partial x_i(p, p\omega)}{\partial m}(\omega_j - x_j).
\]
Now the income effect depends on the net demand for good \(j\) rather than the gross demand.

### 3.7.4 Continuity and Differentiability of Demand Functions

Up until now we have assumed that the demand functions are nicely behaved; that is, that they are continuous and even differentiable functions. Are these assumptions justifiable?

**Proposition 3.7.2 (Continuity of Demand Function)** Suppose \(\succeq\) is continuous and weakly convex, and \((p, m) > 0\). Then, \(x(p, m)\) is a upper hemi-continuous convex-valued correspondence. Furthermore, if the weak convexity is replaced by the strict convexity, \(x(p, m)\) is a continuous single-valued function.

**Proof.** First note that, since \((p, m) > 0\), one can show that the budget constrained set \(B(p, m)\) is a continuous correspondence with non-empty and compact values and \(\succeq_i\) is continuous. Then, by the Maximum Theorem, we know the demand correspondence \(x(p, m)\) is upper hemi-continuous. We now show \(x(p, m)\) is convex. Suppose \(x\) and \(x'\) are two optimal consumption bundles. Let \(x_t = tx + (1 - t)x'\) for \(t \in [0, 1]\). Then, \(x_t\) also satisfied the budget constraint, and by weak convexity of \(\succeq\), we have \(x_t = tx + (1 - t)x' \succeq x\). Because \(x\) is an optimal consumption bundle, we must have \(x_t \sim x\) and thus \(x_t\) is also an optimal consumption bundle.

Now, when \(\succ\) is strictly convex, \(x(p, m)\) is then single-valued by Proposition 3.4.2, and thus it is a continuous function since a upper hemi-continuous correspondence is a continuous function when it is single-valued.

A demand correspondence may not be continuous for non-convex preference ordering, as illustrated in Figure 3.17. Note that, in the case depicted in Figure
3.17, a small change in the price brings about a large change in the demanded bundles: the demand correspondence is discontinuous.

![Discontinuous demand](image)

Figure 3.17: Discontinuous demand. Demand is discontinuous due to non-convex preferences

Sometimes, we need to consider the slopes of demand curves and hence we would like a demand function is differentiable. What conditions can guarantee the differentiability? We give the following proposition without proof.

**Proposition 3.7.3** Suppose $x > 0$ solves the consumer’s utility maximization problem at $(p,m) > 0$. If

1. $u$ is twice continuously differentiable on $R^L_{++}$,
2. $\frac{\partial u(x)}{\partial x_l} > 0$ for some $l = 1, \ldots, L$,
3. the bordered Hessian of $u$ has nonzero determinant at $x$,

then $x(p,m)$ is differentiable at $(p,m)$.

### 3.7.5 Inverse Demand Functions

In many applications it is of interest to express demand behavior by describing prices as a function of quantities. That is, given some vector of goods $x$, we
would like to find a vector of prices \( p \) and an income \( m \) at which \( x \) would be the demanded bundle.

Since demand functions are homogeneous of degree zero, we can fix income at some given level, and simply determine prices relative to this income level. The most convenient choice is to fix \( m = 1 \). In this case the first-order conditions for the utility maximization problem are simply

\[
\frac{\partial u(x)}{\partial x_i} - \lambda p_i = 0 \quad \text{for} \ i, \ldots, k
\]

\[
\sum_{i=1}^{k} p_i x_i = 1.
\]

We want to eliminate \( \lambda \) from this set of equations.

To do so, multiply each of the first set of equalities by \( x_i \) and sum them over the number of goods to get

\[
\sum_{i=1}^{k} \frac{\partial u(x)}{\partial x_i} x_i = \lambda \sum_{i=1}^{k} p_i x_i = \lambda.
\]

Substitute the value of \( \lambda \) back into the first expression to find \( p \) as function of \( x \):

\[
p_i(x) = \left. \frac{\partial u(x)}{\partial x_i} \right|_{\sum_{j=1}^{k} \frac{\partial u(x)}{\partial x_j} x_j}.
\]

Given any vector of demands \( x \), we can use this expression to find the price vector \( p(x) \) which will satisfy the necessary conditions for maximization. If the utility function is quasi-concave so that these necessary conditions are indeed sufficient for maximization, then this will give us the inverse demand relationship.

What happens if the utility function is not everywhere quasi-concave? Then there may be some bundles of goods that will not be demanded at any price; any bundle on a non-convex part of an indifference curve will be such a bundle.

There is a dual version of the above formula for inverse demands that can be obtained from the duality between direct utility function and indirect utility function we discussed earlier. The argument given there shows that the demanded
bundle \( x \) must minimize indirect utility over all prices that satisfy the budget constraint. Thus \( x \) must satisfy the first-order conditions

\[
\frac{\partial v(p)}{\partial p_l} - \mu x_l = 0 \quad \text{for } l = 1, \ldots, L
\]

\[
\sum_{i=1}^L p_i x_i = 1.
\]

Now multiply each of the first equations by \( p_l \) and sum them to find that

\[
\mu = \sum_{l=1}^L \frac{\partial v(p)}{\partial p_l} p_l.
\]

Substituting this back into the first-order conditions, we have an expression for the demanded bundle as a function of the normalized indirect utility function:

\[
x_i(p) = \frac{\partial v(p)}{\partial p_i} \sum_{j=1}^k \frac{\partial v(p)}{\partial p_j} p_j.
\] (3.18)

Note the nice duality: the expression for the direct demand function, (3.18), and the expression for the inverse demand function (3.17) have the same form. This expression can also be derived from the definition of the normalized indirect utility function and Roy’s identity.

### 3.8 Change in Consumer Surplus and its Measurement

### 3.9 Integrability

#### 3.9.1 Obtaining Utility Function from Demand Function

Given a system of demand functions \( x(p, m) \). Is there necessarily a utility function from which these demand functions can be derived? This question is known as the integrability problem. We will show how to solve this problem by solving a differential equation and integrating back, ultimately to the utility function. The Slutsky matrix plays a key role in this process.
We have seen that the utility maximization hypothesis imposes certain observable restrictions on consumer behavior. If a demand function $x(p, m)$ is well-behaved, our previous analysis has shown that $x(p, m)$ satisfies the following five conditions:

1. Nonnegativity: $x(p, m) \geq 0$.
2. Homogeneity: $x(tp, tm) = x(p, m)$.
3. Budget Balancedness: $px(p, m) = m$.
4. Symmetry: The Slutsky matrix $S \equiv \left( \frac{\partial x_i(p, m)}{\partial p_j} + \frac{\partial x_i(p, m)}{\partial m} x_j(p, m) \right)$ is symmetric.
5. Negative Semi-definite: The matrix $S$ is negative semi-definite.

The main result of the integrability problem is that these conditions, together with some technical assumptions, are in fact sufficient as well as necessary for the integrability process as shown by Hurwicz and Uzawa (1971). This result is very important from the point of view of political economy. The utility maximization approach to the study of consumer behavior sometimes are criticized because they think the notion of utility is a psychological measurement and cannot be observed, and thus, they think the demand function from utility maximization is meaningless. The integrability result, however, tells us that a utility function can be derived from observable data on demand although the utility function is not directly observable. This impressive result warrants a formal statement.

**Theorem 3.9.1** A continuous differentiable function $x : \mathbb{R}^{L+1}_{++} \rightarrow \mathbb{R}^L_+$ is the demand function generalized by some increasing, quasi-concave utility function $u$ if (and only if, when $u$ is continuous, strictly increasing and strictly quasi-concave) it satisfy homogeneity, budget balancedness, symmetry, and negative semi-definiteness.

The proof of the theorem is somehow complicated and can be found in Hurwicz and Uzawa (1971). So it is omitted here.
To actually find a utility function from a given system of demand functions, we must find an equation to integrate. As it turns out, it is somewhat easier to deal with the integrability problem in terms of the expenditure function rather than the indirect utility function.

Recall that from Shephard’s lemma given in Proposition 3.5.2,
\[
\frac{\partial e(p, u)}{\partial p_i} = x_i(p, m) = x_i(p, e(p, u)) \quad i = 1, \ldots, L. \tag{3.19}
\]
We also specify a boundary condition of the form \( e(p^*, u) = c \) where \( p^* \) and \( c \) are given. The system of equations given in (3.19) is a system of partial differential equations. It is well-known that a system of partial differential equations of the form
\[
\frac{\partial f(p)}{\partial p_i} = g_i(p) \quad i = 1, \ldots, k
\]
has a (local) solution if and only if
\[
\frac{\partial g_i(p)}{\partial p_j} = \frac{\partial g_j(p)}{\partial p_i} \quad \text{all } i \text{ and } j.
\]
Applying this condition to the above problem, we see that it reduces to requiring that the matrix
\[
\left( \frac{\partial x_i(p, m)}{\partial p_j} + \frac{\partial x_i(p, m)}{\partial m} \frac{\partial e(p, u)}{\partial p_j} \right)
\]
is symmetric. But this is just the Slutsky restriction! Thus the Slutsky restrictions imply that the demand functions can be “integrated” to find an expenditure function consistent with the observed choice behavior.

Under the assumption that all five of the properties listed at the beginning of this section hold, the solution function \( e \) will be an expenditure function. Inverting the found expenditure function, we can find the indirect utility function. Then using the duality between the direct utility function and indirect utility function we will study in the next section, we can determine the direct utility function.

**Example 3.9.1 (The General Cobb-Douglas Utility Function)** Consider the demand functions
\[
x_i(p, m) = \frac{a_i m}{p_i} = \frac{\alpha_i m}{\alpha p_i}
\]
where $\alpha = \sum_{i=1}^{L} \alpha_i$. The system (3.19) becomes

$$\frac{\partial e(p, u)}{\partial p_i} = \frac{\alpha_m}{\alpha p_i} \quad i = 1, \ldots, L.$$  (3.20)

The $i$-equation can be integrated with respect to $p_i$ to obtain.

$$\ln e(p, u) = \frac{\alpha_i}{\alpha} \ln p_i + c_i$$

where $c_i$ does not depend on $p_i$, but it may depend on $p_j$ for $j \neq i$. Thus, combining these equations we find

$$\ln e(p, u) = \sum_{i=1}^{L} \frac{\alpha_i}{\alpha} \ln p_i + c$$

where $c$ is independent of all $p_i$'s. The constant $c$ represents the freedom that we have in setting the boundary condition. For each $u$, let us take $p^* = (1, \ldots, 1)$ and use the boundary condition $e(p^*, u) = u$. Then it follows that

$$\ln e(p, u) = \sum_{i=1}^{L} \frac{\alpha_i}{\alpha} \ln p_i + \ln u.$$  

Inverting the above equation, we have

$$\ln v(p, m) = -\sum_{i=1}^{L} \frac{\alpha_i}{\alpha} \ln p_i + \ln m$$

which is a monotonic transformation of the indirect utility function for a Cobb-Douglas we found previously.

### 3.9.2 Revealed Preference

**Axioms of Revealed Preferences**

The basic preference axioms sometimes are criticized as being too strong on the grounds that individuals are unlikely to make choices through conscious use of a preference relation. One response to this criticism is to develop an alternative theory on the basis of a weaker set of hypotheses. One of the most interesting alternative theories is that of revealed preference, which is discussed in this section.
The basic principle of revealed preference theory is that preference statements should be constructed only from observable decisions, that is, from actual choice made by a consumer. An individual preference relation, even if it exists, can never be directly observed in the market. The best that we may hope for in practice is a list of the choices made under different circumstances. For example, we may have some observations on consumer behavior that take the form of a list of prices, $p_t$, and the associated chosen consumption bundles, $x_t$ for $t = 1, \ldots, T$. How can we tell whether these data could have been generated by a utility-maximizing consumer? Revealed preference theory focuses on the choices made by a consumer, not on a hidden preference relation.

We will say that a utility function rationalizes the observed behavior $(p_t'x_t')$ for $t = 1, \ldots, T$ if $u(x_t') \geq u(x)$ for all $x$ such that $p_t'x_t' \geq p_t'x$. That is, $u(x)$ rationalizes the observed behavior if it achieves its maximum value on the budget set at the chosen bundles. Suppose that the data were generated by such a maximization process. What observable restrictions must the observed choices satisfy?

Without any assumptions about $u(x)$ there is a trivial answer to this question, namely, no restrictions. For suppose that $u(x)$ were a constant function, so that the consumer was indifferent to all observed consumption bundles. Then there would be no restrictions imposed on the patterns of observed choices: anything is possible.

To make the problem interesting, we have to rule out this trivial case. The easiest way to do this is to require the underlying utility function to be locally non-satiated. Our question now becomes: what are the observable restrictions imposed by the maximization of a locally non-satiated utility function?

**Direct Revealed Preference:** If $p_t'x_t' \geq p_t'x$, then $u(x_t') \geq u(x)$. We will say that $x_t'$ is directly revealed preferred to $x$, and write $x_t' \mathrel{R^D} x$.

This condition means that if $x_t'$ was chosen when $x$ could have been chosen, the utility of $x_t'$ must be at least as large as the utility of $x$. As a consequence of this definition and the assumption that the data were generated by utility
maximization, we can conclude that “$x^t R^D x$ implies $u(x^t) \geq u(x)$.”

Strictly Direct Revealed Preference: If $p^t x^t > p^t x$, then $u(x^t) > u(x)$. We will say that $x^t$ is strictly directly revealed preferred to $x$ and write $x^t P^D x$.

It is not hard to show that local non-satiation implies this conclusion. For we know from the previous paragraph that $u(x^t) \geq u(x)$; if $u(x^t) = u(x)$, then by local non-satiation there would exist some other $x'$ close enough to $x$ so that $p^t x^t > p^t x'$ and $u(x') > u(x) = u(x^t)$. This contradicts the hypothesis of utility maximization.

Revealed Preference: $x^t$ is said to be revealed preferred to $x$ if there exists a finite number of bundles $x_1, x_2, \ldots, x_n$ such that $x^t R^D x_1, x^t R^D x_2, \ldots, x^t R^D x$. In this case, we write $x^t R x$.

The relation $R$ constructed above by considering chains of $R^D$ is sometimes called the transitive closure of the relation $R^D$. If we assume that the data were generated by utility maximization, it follows that “$x^t R x$ implies $u(x^t) \geq u(x)$.”

Consider two observations $x^t$ and $x^s$. We now have a way to determine whether $u(x^t) \geq u(x^s)$ and an observable condition to determine whether $u(x^s) > u(x^t)$. Obviously, these two conditions should not both be satisfied. This condition can be stated as the

GENERALIZED AXIOM OF REVEALED PREFERENCE (GARP): If $x^t$ is revealed preferred to $x^s$, then $x^s$ cannot be strictly directly revealed preferred to $x^t$.

Using the symbols defined above, we can also write this axiom as

GARP: $x^t R x^s$ implies not $x^s P^D x^t$. In other words, $x^t R x^s$, implies $p^t x^s \leq p^s x^t$.

As the name implies, GARP is a generalization of various other revealed preference tests. Here are two standard conditions.

WEAK AXIOM OF REVEALED PREFERENCE (WARP): If $x^t R^D x^s$ and $x^t$ is not equal to $x^s$, then it is not the case that $x^s R^D x^t$, i.e., $p_t x_t \geq p_t x_s$ implies $p_s x_t > p_s x_s$.

STRONG AXIOM OF REVEALED PREFERENCE (SARP): If $x^t R$
If the data \((p^t, x^t)\) were generated by a utility-maximizing consumer with non-satiated preferences, the data must satisfy GARP. Hence, GARP is an observable consequence of utility maximization. But does it express all the implications of that model? If some data satisfy this axiom, is it necessarily true that it must come from utility maximization, or at least be thought of in that way? Is GARP a sufficient condition for utility maximization?

It turns out that it is. If a finite set of data is consistent with GARP, then there exists a utility function that rationalizes the observed behavior — i.e., there exists a utility function that could have generated that behavior. Hence, GARP exhausts the list of restrictions imposed by the maximization model.

We state the following theorem without proof.

**Afriat’s theorem.** Let \((p^t, x^t)\) for \(t = 1, \ldots, T\) be a finite number of observations of price vectors and consumption bundles. Then the following conditions are equivalent.

1. There exists a locally nonsatiated utility function that rationalizes the data;
2. The data satisfy GARP;
3. There exist positive numbers \((u^t, \lambda^t)\) for \(t = 1, \ldots, T\) that satisfy the Afriat inequalities:
   \[
   u^s \leq u^t + \lambda^t p^t (x^s - x^t) \quad \text{for all } t, s; 
   \]
4. There exists a locally nonsatiated, continuous, concave, monotonic utility function that rationalizes the data.
Thus, Afriat’s theorem states that a finite set of observed price and quantity data satisfy GARP if and only if there exists a locally non-satiated, continuous, increasing, and concave utility function that rationalizes the data.

Condition (3) in Afriat’s theorem has a natural interpretation. Suppose that $u(x)$ is a concave, differentiable utility function that rationalizes the observed choices. The fact that $u(x)$ is differentiable implies it must satisfy the $T$ first-order conditions

$$ Du(x^t) = \lambda^t p^t $$

(3.21)

The fact that $u(x)$ is concave implies that it must satisfy the concavity conditions

$$ u(x^t) \leq u(x^s) + Du(x^s)(x^t - x^s). $$

(3.22)

Substituting from (3.21) into (3.22), we have

$$ u(x^t) \leq u(x^s) + \lambda^s p^s (x^t - x^s). $$

Hence, the Afriat numbers $u^t$ and $\lambda^t$ can be interpreted as utility levels and marginal utilities that are consistent with the observed choices.

The reason the inequality holds for a concave function is because that, from Figure 3.18, we have

$$ \frac{u(x^t) - u(x^s)}{x^t - x^s} \leq u'(x^s). $$

(3.23)

Thus, we have $u(x^t) \leq u(x^s) + u'(x^s) (x^t - x^s)$.

The most remarkable implication of Afriat’s theorem is that (1) implies (4): if there is any locally nonsatiated utility function at all that rationalizes the data, there must exist a continuous, monotonic, and concave utility function that rationalizes the data. If the underlying utility function had the “wrong” curvature at some points, we would never observe choices being made at such points because they wouldn’t satisfy the right second-order conditions. Hence market data do not allow us to reject the hypotheses of convexity and monotonicity of preferences.
3.9.3 Recoverability

Since the revealed preference conditions are a complete set of the restrictions imposed by utility-maximizing behavior, they must contain all of the information available about the underlying preferences. It is more-or-less obvious now to use the revealed preference relations to determine the preferences among the observed choices, $x_t$, for $t = 1, \ldots, T$. However, it is less obvious to use the revealed preference relations to tell you about preference relations between choices that have never been observed.

This is easiest to see using an example, Figure 3.19 depicts a single observation of choice behavior, $(p^1, x^1)$. What does this choice imply about the indifference curve through a bundle $x^0$? Note that $x^0$ has not been previously observed; in particular, we have no data about the prices at which $x^0$ would be an optimal choice.

Let’s try to use revealed preference to “bound” the indifference curve through $x^0$. First, we observe that $x^1$ is revealed preferred to $x^0$. Assume that preferences
are convex and monotonic. Then all the bundles on the line segment connecting $x^0$ and $x^1$ must be at least as good as $x^0$, and all the bundles that lie to the northeast of this bundle are at least as good as $x^0$. Call this set of bundles $RP(x^0)$, for “revealed preferred” to $x^0$. It is not difficult to show that this is the best “inner bound” to the upper contour set through the point $x^0$.

To derive the best outer bound, we must consider all possible budget lines passing through $x^0$. Let $RW$ be the set of all bundles that are revealed worse than $x^0$ for all these budget lines. The bundles in $RW$ are certain to be worse than $x^0$ no matter what budget line is used.

![Figure 3.19: Inner and outer bounds. RP is the inner bound to the indifference curve through $x^0$; the consumption of RW is the outer bound.](image)

The outer bound to the upper contour set at $x^0$ is then defined to be the complement of this set: $NRW = \text{all bundles not in } RW$. This is the best outer bound in the sense that any bundle not in this set cannot ever be revealed preferred to $x^0$ by a consistent utility-maximizing consumer. Why? Because by construction, a bundle that is not in $NRW(x^0)$ must be in $RW(x^0)$ in which case it would be revealed worse than $x^0$. 

165
In the case of a single observed choice, the bounds are not very tight. But with many choices, the bounds can become quite close together, effectively trapping the true indifference curve between them. See Figure 3.20 for an illustrative example. It is worth tracing through the construction of these bounds to make sure that you understand where they come from. Once we have constructed the inner and outer bounds for the upper contour sets, we have recovered essentially all the information about preferences that is not aimed in the observed demand behavior. Hence, the construction of $RP$ and $RW$ is analogous to solving the integrability equations.

Figure 3.20: Inner and outer bounds. When there are several observations, the inner bound and outer bound can be quite tight.

Our construction of $RP$ and $RW$ up until this point has been graphical. However, it is possible to generalize this analysis to multiple goods. It turns out that determining whether one bundle is revealed preferred or revealed worse than another involves checking to see whether a solution exists to a particular set of linear inequalities.
3.10 Topics in Demand Behavior

In this section we investigate several topics in demand behavior. Most of these have to do with special forms of the budget constraint or preferences that lead to special forms of demand behavior. There are many circumstances where such special cases are very convenient for analysis, and it is useful to understand how they work.

3.10.1 Income-Leisure Choice Model

Suppose that a consumer chooses two goods, consumption and “leisure”. Let \( \ell \) be the number of hours and \( \bar{L} \) be the maximum number of hours that the consumer can work. We then have \( L = \bar{L} - \ell \). She also has some nonlabor income \( m \). Let \( u(c, L) \) be the utility of consumption and leisure and write the utility maximization problem as

\[
\max_{c,L} u(c, L) \quad \text{such that} \quad pc + wL = w\bar{L} + m.
\]

This is essentially the same form that we have seen before. Here the consumer “sells” her endowment of labor at the price \( w \) and then buys some back as leisure.

Slutsky’s equation allows us to calculate how the demand for leisure changes as the wage rate changes. We have

\[
\frac{dL(p, w, m)}{dw} = \frac{\partial L(p, w, u)}{\partial w} + \frac{\partial L(p, w, m)}{\partial m} [\bar{L} - L].
\]

Note that the term in brackets is nonnegative by definition, and almost surely positive in practice. This means that the derivative of leisure demand is the sum of a negative number and a positive number and is inherently ambiguous in sign. In other words, an increase in the wage rate can lead to either an increase or a decrease in labor supply.
3.10.2 Aggregation of Commodities

In many circumstances it is reasonable to model consumer choice by certain “partial” maximization problems. For example, we may want to model the consumer’s choice of “meat” without distinguishing how much is beef, pork, lamb, etc. In most empirical work, some kind of aggregation of this sort is necessary.

In order to describe some useful results concerning this kind of separability of consumption decisions, we will have to introduce some new notation. Let us think of partitioning the consumption bundle into two “subbundles” so that the consumption bundle takes the form \((x, z)\). For example, \(x\) could be the vector of consumptions of different kinds of meat, and \(z\) could be the vector of consumption of all other goods.

We partition the price vector analogously into \((p, q)\). Here \(p\) is the price vector for the different kinds of meat, and \(q\) is the price vector for the other goods. With this notation the standard utility maximization problem can be written as

\[
\max_{x, z} u(x, z) \quad \text{such that} \quad px + qz = m. \tag{3.24}
\]

The problem of interest is under what conditions we can study the demand problem for the \(x\)-goods, say, as a group, without worrying about how demand is divided among the various components of the \(x\)-goods.

One way to formulate this problem mathematically is as follows. We would like to be able to construct some scalar quantity index, \(X\), and some scalar price index, \(P\), that are functions of the vector of quantities and the vector of prices:

\[
P = f(p) \tag{3.25}
\]

\[
X = g(x).
\]

In this expression \(P\) is supposed to be some kind of “price index” which gives the “average price” of the goods, while \(X\) is supposed to be a quantity index.
that gives the average “amount” of meat consumed. Our hope is that we can find a way to construct these price and quantity indices so that they behave like ordinary prices and quantities.

That is, we hope to find a new utility function $U(X, z)$, which depends only on the quantity index of $x$-consumption, that will give us the same answer as if we solved the entire maximization problem in (3.24). More formally, consider the problem

$$\max_{X, z} U(X, z)$$

such that $PX + qz = m$.

The demand function for the quantity index $X$ will be some function $X(P, q, m)$. We want to know when it will be the case that

$$X(P, q, m) \equiv X(f(p), q, m) = g(x(p, q, m)).$$

This requires that we get to the same value of $X$ via two different routes:

1) first aggregate prices using $P = f(p)$ and then maximize $U(X, z)$ subject to the budget constraint $PX + qz = m$.

2) first maximize $u(x, z)$ subject to $px + qz = m$ and then aggregate quantities to get $X = g(x)$.

There are two situations under which this kind of aggregation is possible. The first situation, which imposes constraints on the price movements, is known as **Hicksian separability.** The second, which imposes constraints on the structure of preferences, is known as **functional separability.**

**Hicksian separability**

Suppose that the price vector $p$ is always proportional to some fixed base price vector $p^0$ so that $p = tp^0$ for some scalar $t$. If the $x$-goods are various kinds of meat, this condition requires that the relative prices of the various kinds of meat remain constant — they all increase and decrease in the same proportion.
Following the general framework described above, let us define the price and quantity indices for the \( x \)-goods by

\[
\begin{align*}
    P &= t \\
    X &= p^0x.
\end{align*}
\]

We define the indirect utility function associated with these indices as

\[
V(P, q, m) = \max_{x, z} u(x, z) \\
\text{such that } Pp^0x + qz &= m.
\]

It is straightforward to check that this indirect utility function has all the usual properties: it is quasiconvex, homogeneous in price and income, etc. In particular, a straightforward application of the envelope theorem shows that we can recover the demand function for the \( x \)-good by Roy’s identity:

\[
X(P, q, m) = -\frac{\partial V(P, q, m)/\partial P}{\partial V(P, q, m)/\partial m} = p^0x(p, q, m).
\]

This calculation shows that \( X(P, q, m) \) is an appropriate quantity index for the \( x \)-goods consumption: we get the same result if we first aggregate prices and then maximize \( U(X, z) \) as we get if we maximize \( u(x, z) \) and then aggregate quantities.

We can solve for the direct utility function that is dual to \( V(P, q, m) \) by the usual calculation:

\[
U(X, z) = \min_{P, q} V(P, q, m) \\
\text{such that } PX + qz &= m.
\]

By construction this direct utility function has the property that

\[
V(P, q, m) = \max_{X, z} U(X, z) \\
\text{such that } PX + qz &= m.
\]

Hence, the price and quantity indices constructed this way behave just like ordinary prices and quantities.
The two-good model
One common application of Hicksian aggregation is when we are studying the demand for a single good. In this case, think of the $z$-goods as being a single good, $z$, and the $x$-goods as “all other goods.” The actual maximization problem is then

$$\max_{x,z} u(x, z)$$

such that $px + qz = m$.

Suppose that the relative prices of the $x$-goods remains constant, so that $p = Pp^0$. That is, the vector of prices $p$ is some base price vector $p^0$ times some price index $P$. Then Hicksian aggregation says that we can write the demand function for the $z$-good as

$$z = z(P, q, m).$$

Since this demand function is homogeneous of degree zero, with some abuse of notation, we can also write

$$z = z(q/P, m/P).$$

This says that the demand for the $z$-good depends on the relative price of the $z$-good to “all other goods” and income, divided by the price of “all other goods.” In practice, the price index for all other goods is usually taken to be some standard consumer price index. The demand for the $z$-good becomes a function of only two variables: the price of the $z$-good relative to the CPI and income relative to the CPI.

Functional separability
The second case in which we can decompose the consumer’s consumption decision is known as the case of functional separability. Let us suppose that the underlying preference ordering has the property that

$$(x, z) \succ (x', z') \text{ if and only if } (x, z') \succ (x', z').$$
for all consumption bundles \( \mathbf{x}, \mathbf{x}', \mathbf{z} \) and \( \mathbf{z}' \). This condition says that if \( \mathbf{x} \) is preferred to \( \mathbf{x}' \) for some choices of the other goods, then \( \mathbf{x} \) is preferred to \( \mathbf{x}' \) for all choices of the other goods. Or, even more succinctly, the preferences over the \( x \)-goods are independent of the \( z \)-goods.

If this “independence” property is satisfied and the preferences are locally nonsatiated, then it can be shown that the utility function for \( \mathbf{x} \) and \( \mathbf{z} \) can be written in the form \( u(\mathbf{x}, \mathbf{z}) = U(v(\mathbf{x}), \mathbf{z}) \), where \( U(v, \mathbf{z}) \) is an increasing function of \( v \). That is, the overall utility from \( \mathbf{x} \) and \( \mathbf{z} \) can be written as a function of the subutility of \( \mathbf{x}, v(\mathbf{x}) \), and the level of consumption of the \( z \)-goods.

If the utility function can be written in this form, we will say that the utility function is weakly separable. What does separability imply about the structure of the utility maximization problem? As usual, we will write the demand function for the goods as \( \mathbf{x}(\mathbf{p}, \mathbf{q}, \mathbf{m}) \) and \( \mathbf{z}(\mathbf{p}, \mathbf{q}, \mathbf{m}) \). Let \( m_x = \mathbf{p}\mathbf{x}(\mathbf{p}, \mathbf{q}, \mathbf{m}) \) be the optimal expenditure on the \( x \)-goods.

It turns out that if the overall utility function is weakly separable, the optimal choice of the \( x \)-goods can be found by solving the following subutility maximization problem:

\[
\max v(\mathbf{x})
\text{ such that } \mathbf{p}\mathbf{x} = m_x. \tag{3.26}
\]

This means that if we know the expenditure on the \( x \)-goods, \( m_x = \mathbf{p}\mathbf{x}(\mathbf{p}, \mathbf{q}, \mathbf{m}) \), we can solve the subutility maximization problem to determine the optimal choice of the \( x \)-goods. In other words, the demand for the \( x \)-goods is only a function of the prices of the \( x \)-goods and the expenditure on the \( x \)-goods \( m_x \). The prices of the other goods are only relevant insofar as they determine the expenditure on the \( x \)-goods.

The proof of this is straightforward. Assume that \( \mathbf{x}(\mathbf{p}, \mathbf{q}, \mathbf{m}) \) does not solve the above problem. Instead, let \( \mathbf{x}' \) be another value of \( \mathbf{x} \) that satisfies the budget constraint and yields strictly greater subutility. Then the bundle \( (\mathbf{x}', \mathbf{z}) \) would give higher overall utility than \( (\mathbf{x}(\mathbf{p}, \mathbf{q}, \mathbf{m}), \mathbf{z}(\mathbf{p}, \mathbf{q}, \mathbf{m})) \), which contradicts the
definition of the demand function.

The demand functions \( x(p, m_x) \) are sometimes known as conditional demand functions since they give demand for the \( x \)-goods conditional on the level of expenditure on these goods. Thus, for example, we may consider the demand for beef as a function of the prices of beef, pork, and lamb and the total expenditure on meat. Let \( e(p, v) \) be the expenditure function for the subutility maximization problem given in (3.26). This tells us how much expenditure on the \( x \)-goods is necessary at prices \( p \) to achieve the subutility \( v \). It is not hard to see that we can write the overall maximization problem of the consumer as

\[
\max_{v, z} U(v, z) \\
\text{such that } e(p, v) + qz = m
\]

This is almost in the form we want: \( v \) is a suitable quantity index for the \( x \)-goods, but the price index for the \( x \)-goods isn’t quite right. We want \( P \) times \( X \), but we have some nonlinear function of \( p \) and \( X = v \).

In order to have a budget constraint that is linear in quantity index, we need to assume that subutility function has a special structure. For example. Suppose that the subutility function is homothetic. Then we can write \( e(p, v) \) as \( e(p)v \). Hence, we can choose our quantity index to be \( X = v(x) \), our price index to be \( P = e(p) \), and our utility function to be \( U(X, z) \). We get the same \( X \) if we solve

\[
\max_{X, z} U(X, z) \\
\text{such that } PX + qz = m
\]

as if we solve

\[
\max_{x, z} u(v(x), z) \\
\text{such that } px + qz = m,
\]

and then aggregate using \( X = v(x) \).

In this formulation we can think of the consumption decision as taking place in two stages: first the consumer considers how much of the composite commodity
(e.g., meat) to consume as a function of a price index of meat by solving the overall maximization problem; then the consumer considers how much beef to consume given the prices of the various sorts of meat and the total expenditure on meat, which is the solution to the subutility maximization problem. Such a two-stage budgeting process is very convenient in applied demand analysis.

3.10.3 Aggregation of Consumers’ Demand

We have studied the properties of a consumer’s demand function, $x(p, m)$. Now let us consider some collection of $i = 1, \ldots, n$ consumers, each of whom has a demand function for some $L$ commodities, so that consumer $i$’s demand function is a vector $x_i(p, m_i) = (x_{1i}(p, m_i), \ldots, x_{Li}(p, m_i))$ for $i = 1, \ldots, n$. Note that we have changed our notation slightly: goods are now indicated by superscripts while consumers are indicated by subscripts. The aggregate demand function is defined by $X(p, m_1, \ldots, m_n) = \sum_{i=1}^{n} x_i(p, m)$. The aggregate demand for good $l$ is denoted by $X^l(p, m)$ where $m$ denotes the vector of incomes $(m_1, \ldots, m_n)$.

The aggregate demand function inherits certain properties of the individual demand functions. For example, if the individual demand functions are continuous, the aggregate demand function will certainly be continuous. Continuity of the individual demand functions is a sufficient but not necessary condition for continuity of the aggregate demand functions.

What other properties does the aggregate demand function inherit from the individual demands? Is there an aggregate version of Slutsky’s equation or of the Strong Axiom of Revealed Preference? Unfortunately, the answer to these questions is no. In fact the aggregate demand function will in general possess no interesting properties other than homogeneity and continuity. Hence, the theory of the consumer places no restrictions on aggregate behavior in general. However, in certain cases it may happen that the aggregate behavior may look as though it were generated by a single “representative” consumer. Below, we consider a circumstance where this may happen.

Suppose that all individual consumers’ indirect utility functions take the Gor-
man form:

\[ v_i(p, m_i) = a_i(p) + b(p)m_i. \]

Note that the \( a_i(p) \) term can differ from consumer to consumer, but the \( b(p) \) term is assumed to be identical for all consumers. By Roy’s identity the demand function for good \( j \) of consumer \( i \) will then take the form

\[ x^j_i(p, m_i) = a^j_i(p) + \beta^j_i(p)m_i. \] (3.27)

where,

\[
\begin{align*}
a^j_i(p) & = - \frac{\partial a_i(p)}{b(p)} \\
\beta^j_i(p) & = - \frac{\partial b(p)}{b(p)}.
\end{align*}
\]

Note that the marginal propensity to consume good \( j \), \( \partial x^j_i(p, m_i) / \partial m_i \), is independent of the level of income of any consumer and also constant across consumers since \( b(p) \) is constant across consumers. The aggregate demand for good \( j \) will then take the form

\[ X^j(p, m^1, \ldots, m^n) = - \left[ \sum_{i=1}^{n} \frac{\partial a_i}{b(p)} \frac{\partial b(p)}{b(p)} \sum_{i=1}^{n} m_i \right]. \]

This demand function can in fact be generated by a representative consumer. His representative indirect utility function is given by

\[ V(p, M) = \sum_{i=1}^{n} a_i(p) + b(p)M = A(p) + B(p)M, \]

where \( M = \sum_{i=1}^{n} m_i \).

The proof is simply to apply Roy’s identity to this indirect utility function and to note that it yields the demand function given in equation (3.27). In fact it can be shown that the Gorman form is the most general form of the indirect utility
function that will allow for aggregation in the sense of the representative consumer model. Hence, the Gorman form is not only sufficient for the representative consumer model to hold, but it is also necessary.

Although a complete proof of this fact is rather detailed, the following argument is reasonably convincing. Suppose, for the sake of simplicity, that there are only two consumers. Then by hypothesis the aggregate demand for good $j$ can be written as

$$X^j(p, m_1 + m_2) = x^j_1(p, m_1) + x^j_2(p, m_2).$$

If we first differentiate with respect to $m_1$ and then with respect to $m_2$, we find the following identities

$$\frac{\partial^2 X^j(p, M)}{\partial M^2} = \frac{\partial x^j_1(p, m_1)}{\partial m_1} = \frac{\partial x^j_2(p, m_2)}{\partial m_2}.$$

Hence, the marginal propensity to consume good $j$ must be the same for all consumers. If we differentiate this expression once more with respect to $m_1$, we find that

$$\frac{\partial^2 X^j(p, M)}{\partial M^2} = \frac{\partial^2 x^j_1(p, m_1)}{\partial m_1^2} \equiv 0.$$

Thus, consumer 1’s demand for good $j$ – and, therefore, consumer 2’s demand – is affine in income. Hence, the demand functions for good $j$ take the form $x^j_i(p, m_i) = a^j_i(p) + \beta^j_i(p)m_i$. If this is true for all goods, the indirect utility function for each consumer must have the Gorman form.

One special case of a utility function having the Gorman form is a utility function that is homothetic. In this case the indirect utility function has the form $v(p, m) = v(p)m$, which is clearly of the Gorman form. Another special case is that of a quasi-linear utility function. A utility function $U$ is said to be quasi-linear if it has the following functional form:

$$U(x_0, x_1, \ldots, x_L) = x_0 + u(x_1, \ldots, x_L)$$

In this case $v(p, m) = v(p) + m$, which obviously has the Gorman form. Many of the properties possessed by homothetic and/or quasi-linear utility functions are also possessed by the Gorman form.
The class of quasi-linear utility function is a very important class of utility functions which plays an important role in many economics fields such as those of information economics, mechanism design theory, property rights theory due to its important property of no income effect when income changes.

### 3.11 Reference

**Books and Monographs:**


**Papers:**


Chapter 4

Production Theory

4.1 Introduction

Economic activity not only involves consumption but also production and trade. Production should be interpreted very broadly, however, to include production of both physical goods – such as rice or automobiles – and services – such as medical care or financial services.

A firm can be characterized by many factors and aspects such as sectors, production scale, ownerships, organization structures, etc. But which are most important features for us to study producer’s behavior in making choices? To grasp the most important features in studying producer behavior and choices in modern producer theory, it is assumed that the key characteristic of a firm is production set. Producer’s characteristic together with the behavior assumption are building blocks in any model of producer theory. The production set represents the set of all technologically feasible production plans. The behavior assumption expresses the guiding principle the producer uses to make choices. It is generally assumed that the producer seeks to identify and select a production that is most profitable.

We will first present a general framework of production technology. By itself, the framework does not describe how production choices are made. It only
specifies basic characteristic of a firm which defines what choices can be made; it does not specify what choices should be made. We then will discuss what choices should be made based on the behavior assumptions on firms. A basic behavior assumption on producers is profit maximization. After that, we will describe production possibilities in physical terms, which is recast into economic terms – using cost functions.

4.2 Production Technology

Production is the process of transforming inputs to outputs. Typically, inputs consist of labor, capital equipment, raw materials, and intermediate goods purchased from other firms. Outputs consists of finished products or service, or intermediate goods to be sold to other firms. Often alternative methods are available for producing the same output, using different combinations of inputs. A firm produces outputs from various combinations of inputs. In order to study firm choices we need a convenient way to summarize the production possibilities of the firm, i.e., which combinations of inputs and outputs are technologically feasible.

4.2.1 Specification of Technology

It is usually most satisfactory to think of the inputs and outputs as being measured in terms of flows: a certain amount of inputs per the period are used to produce a certain amount of outputs per unit the period at some location. It is a good idea to explicitly include the time and location dimensions in a specification of inputs and outputs. The level of detail that we will use in specifying inputs and outputs will depend on the problem at hand, but we should remain aware of the fact that a particular input or output good can be specified in arbitrarily fine detail. However, when discussing technological choices in the abstract, as we do in this chapter, it is common to omit the time and location dimensions.
The fundamental reality firms must contend with in this process is technological feasibility. The state of technology determines and restricts what is possible in combing inputs to produce outputs, and there are several ways we can represent this constraint. The most general way is to think of the firm as having a production possibility set.

Suppose the firm has $L$ possible goods to serve as inputs and/or outputs. If a firm uses $y_j^i$ units of a good $j$ as an input and produces $y_j^o$ of the good as an output, then the net output of good $j$ is given by $y_j = y_j^o - y_j^i$.

A production plan is simply a list of net outputs of various goods. We can represent a production plan by a vector $y$ in $\mathbb{R}^L$ where $y_j$ is negative if the $j^{th}$ good serves as a net input and positive if the $j^{th}$ good serves as a net output. The set of all technologically feasible production plans is called the firm’s production possibilities set and will be denoted by $Y$, a subset of $\mathbb{R}^L$. The set $Y$ is supposed to describe all patterns of inputs and outputs that are technologically feasible. It gives us a complete description of the technological possibilities facing the firm.

When we study the behavior of a firm in certain economic environments, we may want to distinguish between production plans that are “immediately feasible” and those that are “eventually” feasible. We will generally assume that such restrictions can be described by some vector $z$ in $\mathbb{R}^L$. The restricted or short-run production possibilities set will be denoted by $Y(z)$; this consists of all feasible net output bundles consistent with the constraint level $z$. The following are some examples of such restrictions.

**EXAMPLE: Input requirement set**

Suppose a firm produces only one output. In this case we write the net output bundle as $(y, -x)$ where $x$ is a vector of inputs that can produce $y$ units of output. We can then define a special case of a restricted production possibilities set, the input requirement set:

$$V(y) = \{x \in \mathbb{R}^L_+: (y, -x) \text{ is in } Y\}$$

The input requirement set is the set of all input bundles that produce at least $y$
units of output.

Note that the input requirement set, as defined here, measures inputs as positive numbers rather than negative numbers as used in the production possibilities set.

EXAMPLE: ISOQUANT

In the case above we can also define an isoquant:

\[ Q(y) = \{ x \in \mathbb{R}^n : x \text{ is in } V(y) \text{ and } x \text{ is not in } V(y') \text{ for } y' > y \} \]

The isoquant gives all input bundles that produce exactly \( y \) units of output.

EXAMPLE: SHORT-RUN PRODUCTION POSSIBILITIES SET

Suppose a firm produces some output from labor and some kind of machine which we will refer to as “capital.” Production plans then look like \((y, -l, -k)\) where \( y \) is the level of output, \( l \) the amount of labor input, and \( k \) the amount of capital input. We imagine that labor can be varied immediately but that capital is fixed at the level \( k \) in the short run. Then

\[ Y(k) = \{(y, -l, -k) \in Y : k = \bar{k}\} \]

is a short-run production possibilities set.
EXAMPLE: Production function
If the firm has only one output, we can define the production function: $f(x) = \{y \in \mathbb{R}: y$ is the maximum output associated with $-x$ in $Y\}$.

EXAMPLE: Transformation function
A production plan $y$ in $Y$ is (technologically) efficient if there is no $y'$ in $Y$ such that $y' \geq y$ and $y' \neq y$; that is, a production plan is efficient if there is no way to produce more output with the same inputs or to produce the same output with less inputs. (Note carefully how the sign convention on inputs works here.) We often assume that we can describe the set of technologically efficient production plans by a transformation function $T: R \rightarrow R$ where $T(y) = 0$ if and only if $y$ is efficient. Just as a production function picks out the maximum scalar output as a function of the inputs, the transformation function picks out the maximal vectors of net outputs.

EXAMPLE: Cobb-Douglas technology
Let $a$ be a parameter such that $0 < a < 1$. Then the Cobb-Douglas technology is defined in the following manner.
\[ Y = \{(y, -x_1, -x_2) \in \mathbb{R}^3 : y \leq x_1^\alpha x_2^{1-\alpha}\} \]

\[ V(y) = \{(x_1, x_2) \in \mathbb{R}_+^2 : y \leq x_1^\alpha x_2^{1-\alpha}\} \]

\[ Q(y) = \{(x_1, x_2) \in \mathbb{R}_+^2 : y = x_1^\alpha x_2^{1-\alpha}\} \]

\[ Y(z) = \{(y, -x_1, -x_2) \in \mathbb{R}^3 : y \leq x_1^\alpha x_2^{1-\alpha}, x_2 = z\} \]

\[ T(y, x_1 x_2) = y - x_1^\alpha x_2^{1-\alpha} \]

\[ f(x_1, x_2) = x_1^\alpha x_2^{1-\alpha}. \]

**EXAMPLE: Leontief Technology**

Let \( a > 0 \) and \( b > 0 \) be parameters. Then the **Leontief technology** is defined in the following manner.

\[ Y = \{(y, -x_1, -x_2) \in \mathbb{R}^3 : y \leq \min(ax_1, bx_2)\} \]

\[ V(y) = \{(x_1, x_2) \in \mathbb{R}_+^2 : y \leq \min(ax_1, bx_2)\} \]

\[ Q(y) = \{(x_1, x_2) \in \mathbb{R}_+^2 : y = \min(ax_1, bx_2)\} \]

\[ T(y, x_1 x_2) = y - \min(ax_1, bx_2) \]

\[ f(x_1, x_2) = \min(ax_1, bx_2). \]

### 4.2.2 Common Properties of Production Sets

Although the production possibility sets of different processes can differ widely in structure, many technologies share certain general properties. If it can be assumed that these properties are satisfied, special theoretical results can be derived. Some important properties are defined below:

**Possibility of Inaction:** \( 0 \in Y \).

Possibility of inaction means that no action on production is a possible production plan.

**Closeness:** \( Y \) is closed.
The possibility set $Y$ is closed means that, whenever a sequence of production plans $y_i, i = 1, 2, \ldots$, are in $Y$ and $y_i \to y$, then the limit production plan $y$ is also in $Y$. It guarantees that points on the boundary of $Y$ are feasible. Note that $Y$ is closed implies that the input requirement set $V(y)$ is a closed set for all $y \geq 0$.

**FREE DISPOSAL OR MONOTONICITY**: If $y \in Y$ implies that $y' \in Y$ for all $y' \leq y$, then the set $Y$ is said to satisfy the free disposal or monotonicity property.

Free disposal implies that commodities (either inputs or outputs) can be thrown away. This property means that if $y \in Y$, then $Y$ includes all vectors in the negative orthant translated to $y$, i.e. there are only inputs, but no outputs.

A weaker requirement is that we only assume that the input requirement is monotonic: If $x$ is in $V(y)$ and $x' \geq x$, then $x'$ is in $V(y)$. Monotonicity of $V(y)$ means that, if $x$ is a feasible way to produce $y$ units of output and $x'$ is an input vector with at least as much of each input, then $x'$ should be a feasible way to produce $y$.

**IRREVERSIBILITY**: $Y \cap \{-Y\} = \{0\}$.

Irreversibility means a production plan is not reversible unless it is a non-action plan.

**CONVEXITY**: $Y$ is convex if whenever $y$ and $y'$ are in $Y$, the weighted average $ty + (1 - t)y$ is also in $Y$ for any $t$ with $0 \leq t \leq 1$.

Convexity of $Y$ means that, if all goods are divisible, it is often reasonable to assume that two production plans $y$ and $y'$ can be scaled downward and combined. However, it should be noted that the convexity of the production set is a strong hypothesis. For example, convexity of the production set rules out “start up costs” and other sorts of returns to scale. This will be discussed in greater detail shortly.

**STRICT CONVEXITY**: $y$ is strictly convex if $y \in Y$ and $y' \in Y$, then
ty + (1 − t)y′ ∈ intY for all 0 < t < 1, where intY denotes the interior points of Y.

As we will show, the strict convexity of Y can guarantee the profit maximizing production plan is unique provided it exists.

A weak and more reasonable requirement is to assume that V(y) is a convex set for all outputs y_o:

**CONVEXITY OF INPUT REQUIREMENT SET:** If x and x′ are in V(y), then tx + (1 − t)x′ is in V(y) for all 0 ≤ t ≤ 1. That is, V(y) is a convex set.

Convexity of V(y) means that, if x and x′ both can produce y units of output, then any weighted average tx + (1 − t)x′ can also produce y units of output. We describe a few of the relationships between the convexity of V(y), the curvature of the production function, and the convexity of Y. We first have

**Proposition 4.2.1 (Convex Production Set Implies Convex Input Requirement Set)**

*If the production set Y is a convex set, then the associated input requirement set, V(y), is a convex set.*

**Proof.** If Y is a convex set then it follows that for any x and x′ such that (y, −x) and (y, −x′) are in Y for 0 ≤ t ≤ 1, we must have (ty + (1 − t)y, tx − (1 − t)x′) in Y. This is simply requiring that (y, (tx + (1 − t)x′)) is in Y. It follows that if x and x′ are in V(y), tx + (1 − t)x′ is in V(y) which shows that V(y) is convex.

**Proposition 4.2.2** V(y) is a convex set if and only if the production function f(x) is a quasi-concave function.

**Proof** V(y) = {x: f(x) ≥ y}, which is just the upper contour set of f(x). But a function is quasi-concave if and only if it has a convex upper contour set. Suppose that we are using some vector of inputs x to produce some output y and we decide to scale all inputs up or down by some amount t ≥ 0. What will happen to the level of output? The notions of returns to scale can be used to answer this
question. Returns to scale refer to how output responds when all inputs are varied in the same proportion so that they consider long run production processes. There are three possibilities: technology exhibits (1) constant returns to scale; (2) decreasing returns to scale, and (3) increasing returns to scale. Formally, we have

(Global) Returns to Scale. A production function \( f(x) \) is said to exhibit:

1. **Constant Returns to Scale** if \( f(tx) = tf(x) \) for all \( t \geq 0 \);
2. **Decreasing Returns to Scale** if \( f(tx) < tf(x) \) for all \( t > 1 \);
3. **Increasing Returns to Scale** if \( f(tx) > tf(x) \) for all \( t > 1 \);

Constant returns to scale (CRS) means that doubling inputs exactly double outputs, which is often a reasonable assumption to make about technologies. Decreasing returns to scale means that doubling inputs are less than doubling outputs. Increasing returns to scale means that doubling inputs are more than doubling outputs.

Note that a technology has constant returns to scale if and only if its production function is homogeneous of degree 1. Constant returns to scale is also equivalent to: the statement \( y \) in \( Y \) implies \( ty \) is in \( Y \) for all \( t \geq 0 \); or equivalent to the statement \( x \) in \( V(y) \) implies \( tx \) is in \( V(ty) \) for all \( t > 1 \).

It may be remarked that the various kinds of returns to scale defined above are global in nature. It may well happen that a technology exhibits increasing returns to scale for some values of \( x \) and decreasing returns to scale for other values. Thus in many circumstances a local measure of returns to scale is useful.

To define locally returns to scale, we first define elasticity of scale.

The **elasticity of scale** measures the percent increase in output due to a one percent increase in all inputs— that is, due to an increase in the scale of operations.

Let \( y = f(x) \) be the production function. Let \( t \) be a positive scalar, and (consider the function \( y(t) = f(tx) \). If \( t = 1 \), we have the current scale of
operation; if \( t > 1 \), we are scaling all inputs up by \( t \); and if \( t < 1 \), we are scaling all inputs down by \( t \). The elasticity of scale is given by

\[
e(x) = \frac{d}{dt} \frac{dy(t)}{y(t)} \frac{1}{t},
\]
evaluated at \( t = 1 \). Rearranging this expression, we have

\[
e(x) = \left. \frac{dy(t)}{dt} \frac{t}{y(t)} \right|_{t=1} = \left. \frac{df(tx)}{dt} \frac{t}{f(tx)} \right|_{t=1}
\]

Note that we must evaluate the expression at \( t = 1 \) to calculate the elasticity of scale at the point \( x \).

Thus, we have the following the local returns to scale:

(LOCAL) RETURNS TO SCALE. A production function \( f(x) \) is said to exhibits **locally increasing, constant, or decreasing returns to scale** as \( e(x) \) is greater, equal, or less than 1.

### 4.2.3 The Marginal Rate of Technical Substitution

Suppose that technology is summarized by a smooth production function and that we are producing at a particular point \( y^* = f(x_1^*, x_2^*) \). Suppose that we want to increase a small amount of input 1 and decrease some amount of input 2 so as to maintain a constant level of output. How can we determine this marginal rate of technical substitution (MRTS) between these two factors? The way is the same as for deriving the marginal rate of substitution of an indifference curve. Differentiating production function when output keeps constant, we have

\[
0 = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2,
\]
which can be solved for

\[
\frac{dx_2}{dx_1} = -\frac{\partial f/\partial x_1}{\partial f/\partial x_2} \equiv -\frac{MP_{x_1}}{MP_{x_2}}.
\]

This gives us an explicit expression for the marginal rate technical substitution, which is the rate of marginal production of \( x_1 \) and marginal production of \( x_2 \).
Example 4.2.1 (MRTS for a Cobb-Douglas Technology) Given that \( f(x_1, x_2) = x_1^\alpha x_2^{1-\alpha} \), we can take the derivatives to find

\[
\frac{\partial f(x)}{\partial x_1} = ax_1^{\alpha-1}x_2^{1-\alpha} = a \left[ \frac{x_2}{x_1} \right]^{1-a}
\]

\[
\frac{\partial f(x)}{\partial x_2} = (1-a)x_1^\alpha x_2^{-\alpha} = (1-a) \left[ \frac{x_2}{x_1} \right]^a.
\]

It follows that

\[
\frac{\partial x_2(x_1)}{\partial x_1} = -\frac{\partial f/\partial x_1}{\partial f/\partial x_2} = -\frac{a}{1-a} \frac{x_1}{x_2}.
\]

4.2.4 The Elasticity of Substitution

The marginal rate of technical substitution measures the slope of an isoquant.

The elasticity of substitution measures the curvature of an isoquant. More specifically, the elasticity of substitution measures the percentage change in the factor ratio divided by the percentage change in the MRTS, with output being held fixed. If we let \( \Delta(x_2/x_1) \) be the change in the factor ratio and \( \Delta \text{MRTS} \) be the change in the technical rate of substitution, we can express this as

\[
\sigma = \frac{\Delta(x_2/x_1)}{\Delta \text{MRTS}}.
\]

This is a relatively natural measure of curvature: it asks how the ratio of factor inputs changes as the slope of the isoquant changes. If a small change in slope gives us a large change in the factor input ratio, the isoquant is relatively flat which means that the elasticity of substitution is large.

In practice we think of the percent change as being very small and take the limit of this expression as \( \Delta \) goes to zero. Hence, the expression for \( \sigma \) becomes

\[
\sigma = \frac{\text{MRTS} d(x_2/x_1)}{(x_2/x_1) d\text{MRTS}} = \frac{d\ln(x_2/x_1)}{d\ln |\text{MRTS}|}.
\]

(The absolute value sign in the denominator is to convert the MRTS to a positive number so that the logarithm makes sense.)
Example 4.2.2 (The Cobb-Douglas Production Function) We have seen above that

$$MRTS = -\frac{a}{1-a} \frac{x_2}{x_1},$$

or

$$\frac{x_2}{x_1} = -\frac{1-a}{a} MRTS.$$

It follows that

$$\ln \frac{x_2}{x_1} = \ln \left(\frac{1-a}{a} + \ln |MRTS|\right).$$

This in turn implies

$$\sigma = \frac{d \ln (x_2/x_1)}{d \ln |MRTS|} = 1.$$

Example 4.2.3 (The CES Production Function) The constant elasticity of substitution (CES) production function has the form

$$y = \left[ a_1 x_1^\rho + a_2 x_2^\rho \right]^{\frac{1}{\rho}}.$$

It is easy to verify that the CES function exhibits constant returns to scale. It will probably not surprise you to discover that the CES production function has a constant elasticity of substitution. To verify this, note that the marginal rate of technical substitution is given by

$$MRTS = -\left( \frac{x_1}{x_2} \right)^{\rho-1},$$

so that

$$\frac{x_2}{x_1} = |MRTS|^{\frac{1}{1-\rho}}.$$

Taking logs, we see that

$$\ln \frac{x_2}{x_1} = \frac{1}{1-\rho} \ln |MRTS|.$$

Applying the definition of $\sigma$ using the logarithmic derivative,

$$\sigma = \frac{d \ln x_2/x_1}{d \ln |MRTS|} = \frac{1}{1 - \rho}.$$
4.3 Profit Maximization

4.3.1 Producer Behavior

A basic hypothesis on individual firm behavior in the producer theory is that a firm will always choose a most profitable production plan from the production set. We will derive input demand and output supply functions by considering a model of profit-maximizing behavior coupled with a description of underlying production constraints.

Economic profit is defined to be the difference between the revenue a firm receives and the costs that it incurs. It is important to understand that all (explicit and implicit) costs must be included in the calculation of profit. Both revenues and costs of a firm depend on the actions taken by the firm. We can write revenue as a function of the level of operations of some $n$ actions, $R(a_1, ..., a_n)$, and costs as a function of these same $n$ activity levels, $C(a_1, ..., a_n)$, where actions can be in term of employment level of inputs or output level of production or prices of outputs if the firm has a market power to set up the prices.

A basic assumption of most economic analysis of firm behavior is that a firm acts so as to maximize its profits; that is, a firm chooses actions $(a_1, ..., a_n)$ so as to maximize $R(a_1, ..., a_n) - C(a_1, ..., a_n)$. The profit maximization problem facing the firm can be then written as

$$\max_{a_1, ..., a_n} R(a_1, ..., a_n) - C(a_1, ..., a_n).$$

The first order conditions for interior optimal actions, $\mathbf{a}^* = (a_1^*, ..., a_n^*)$, is characterized by the conditions

$$\frac{\partial R(\mathbf{a}^*)}{\partial a_i} = \frac{\partial C(\mathbf{a}^*)}{\partial a_i}, \quad i = 1, \ldots, n.$$

The intuition behind these conditions should be clear: if marginal revenue were greater than marginal cost, it would pay to increase the level of the activity; if marginal revenue were less than marginal cost, it would pay to decrease the level of the activity. In general, revenue is composed of two parts: how much a firm
sells of various outputs times the price of each output. Costs are also composed of two parts: how much a firm uses of each input times the price of each input.

The firm’s profit maximization problem therefore reduces to the problem of determining what prices it wishes to charge for its outputs or pay for its inputs, and what levels of outputs and inputs it wishes to use. In determining its optimal policy, the firm faces two kinds of constraints: technological constraints that are specified by production sets and market constraints that concern the effect of actions of other agents on the firm. The firms described in the remainder of this chapter are assumed to exhibit the simplest kind of market behavior, namely that of price-taking behavior. Each firm will be assumed to take prices as given. Thus, the firm will be concerned only with determining the profit-maximizing levels of outputs and inputs. Such a price-taking firm is often referred to as a competitive firm. We will consider the general case in Chapter 9 – the theory of markets.

4.3.2 Producer’s Optimal Choice

Let \( p \) be a vector of prices for inputs and outputs of the firm. The profit maximization problem of the firm can be stated

\[
\pi(p) = \max_{y} py
\]

such that \( y \) is in \( Y \).

Note that since outputs are measured as positive numbers and inputs are measured as negative numbers, the objective function for this problem is profits: revenues minus costs. The function \( \pi(p) \), which gives us the maximum profits as a function of the prices, is called the profit function of the firm.

There are several useful variants of the profit function:

Case 1. Short-run maximization problem. In this case, we might define the short-run profit function, also known as the restricted profit function:

\[
\pi(p, z) = \max_{y} py
\]

such that \( y \) is in \( Y(z) \).
Case 2. If the firm produces only one output, the profit function can be written as

$$\pi(p, w) = \max pf(x) - wx$$

where $p$ is now the (scalar) price of output, $w$ is the vector of factor prices, and the inputs are measured by the (nonnegative) vector $x = (x_1, \ldots, x_n)$.

The value of $y$ that solves the profit problem (4.1) is in general not unique. When there is such a unique production plan, the production plan is called the net output function or net supply function, the corresponding input part is called the producer’s input demand function and the corresponding output vector is called the producer’s output supply function. We will see from the following proposition that strict convexity of production set will ensure the uniqueness of optimal production plan.

**Proposition 4.3.1** Suppose $Y$ strictly convex. Then, for each given $p \in \mathbb{R}^L_+$, the profit maximizing production is unique provide it exists.

Proof: Suppose not. Let $y$ and $y'$ be two profit maximizing production plans for $p \in \mathbb{R}^L_+$. Then, we must have $py = py'$. Thus, by the strict convexity of $Y$, we have $ty + (1 - t)y' \in \text{int}Y$ for all $0 < t < 1$. Therefore, there exists some $k > 1$ such that

$$k[ty + (1 - t)y'] \in \text{int}Y.$$  \hspace{1cm} (4.2)

Then $k[tpy + (1 - t)py'] = kpy > py$ which contradicts the fact that $y$ is a profit maximizing production plan.

**4.3.3 First-Order Conditions for Profit Maximization**

Profit-maximizing behavior can be characterized by calculus when the technology can be described by a differentiable production function. For example, the first-order conditions for the single output profit maximization problem with interior solution are

$$p \frac{\partial f(x^*)}{\partial x_i} = w_i \quad i = 1, \ldots, n.$$  \hspace{1cm} (4.3)
Using vector notation, we can also write these conditions as

\[ p \mathbf{D} f(\mathbf{x}^*) = \mathbf{w}. \]

The first-order conditions state that the “marginal value of product of each factor must be equal to its price,” i.e., marginal revenue equals marginal cost at the profiting maximizing production plan. This first-order condition can also be exhibited graphically. Consider the production possibilities set depicted in Figure 4.3. In this two-dimensional case, profits are given by \( \Pi = py - wx \). The level sets of this function for fixed \( p \) and \( w \) are straight lines which can be represented as functions of the form: \( y = \Pi/p + (w/p)x \). Here the slope of the isoprofit line gives the wage measured in units of output, and the vertical intercept gives us profits measured in units of output. The point of maximal profits the production function must lie below its tangent line at \( x^* \); i.e., it must be “locally concave.”

![Figure 4.3: Profit maximization when the slope of the isoprofit line equals the slope of the production function.](image)

Similar to the arguments in the consumer theory, the calculus conditions derived above make sense only when the choice variables can be varied in an open neighborhood of the optimal choice. The relevant first-order conditions that also include boundary solutions are given by the Kuhn-Tucker conditions:

\[ p \frac{\partial f(\mathbf{x})}{\partial x_i} - w_i \leq 0 \quad \text{with equality if } x_i > 0 \quad (4.4) \]
Remark 4.3.1 There may exist no profit maximizing production plan when a production technology exhibits constant returns to scale or increasing returns to scale. For example, consider the case where the production function is \( f(x) = x \). Then for \( p > w \) no profit-maximizing plan will exist. It is clear from this example that the only nontrivial profit-maximizing position for a constant-returns-to-scale firm is the case of \( p = w \) and zero profits. In this case, all production plans are profit-maximizing production plans. If \((y, x)\) yields maximal profits of zero for some constant returns technology, then \((ty, tx)\) will also yield zero profits and will therefore also be profit-maximizing.

4.3.4 Sufficiency for Profit Maximization

The second-order condition for profit maximization is that the matrix of second derivatives of the production function must be negative semi-definite at the optimal point; that is, the second-order condition requires that the Hessian matrix

\[
D^2 f(x^*) = \begin{pmatrix}
\frac{\partial^2 f(x^*)}{\partial x_i \partial x_j}
\end{pmatrix}
\]

must satisfy the condition that \( hD^2 f(x^*)h' \leq 0 \) for all vectors \( h \). (The prime indicates the transpose operation.) Geometrically, the requirement that the Hessian matrix is negative semi-definite means that the production function must be locally concave in the neighborhood of an optimal choice. Formerly, we have the following proposition.

Proposition 4.3.2 Suppose that \( f(x) \) is differentiable and concave on \( \mathbb{R}^L_+ \) and \((p, w) > 0 \). If \( x > 0 \) satisfies the first-order conditions given in (4.4), then \( x \) is (globally) profit maximizing production plan at prices \((p, w)\).

Remark 4.3.2 The strict concavity of \( f(x) \) can be checked by verifying if the
leading principal minors of the Hessian must alternate in sign, i.e.,

\[
\begin{vmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{vmatrix} > 0, \\
\begin{vmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \end{vmatrix} < 0, \\
\begin{vmatrix} f_{31} & f_{32} & f_{33} \end{vmatrix}
\]

and so on, where \( f_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j} \). This algebraic condition is useful for checking second-order conditions.

**Example 4.3.1 (The Profit Function for Cobb-Douglas Technology)** Consider the problem of maximizing profits for the production function of the form \( f(x) = x^a \) where \( a > 0 \). The first-order condition is

\[ pax^{a-1} = w, \]

and the second-order condition reduces to

\[ pa(a - 1)x^{a-2} \leq 0. \]

The second-order condition can only be satisfied when \( a \leq 1 \), which means that the production function must have constant or decreasing returns to scale for competitive profit maximization to be meaningful.

If \( a = 1 \), the first-order condition reduces to \( p = w \). Hence, when \( w = p \) any value of \( x \) is a profit-maximizing choice. When \( a < 1 \), we use the first-order condition to solve for the factor demand function

\[ x(p, w) = \left( \frac{w}{ap} \right)^{\frac{1}{a-1}}. \]

The supply function is given by

\[ y(p, w) = f(x(p, w)) = \left( \frac{w}{ap} \right)^{\frac{a}{a-1}}, \]

and the profit function is given by

\[ \pi(p, w) = py(p, w) - wx(p, w) = w \left( \frac{1 - a}{a} \right) \left( \frac{w}{ap} \right)^{\frac{1}{a-1}}. \]
4.3.5 Properties of Net Supply Functions

In this section, we show that the net supply functions are the solutions to the profit maximization problem that in fact have imposed certain restrictions on the behavior of the demand and supply functions.

Proposition 4.3.3 Net output functions $y(p)$ are homogeneous of degree zero, i.e., $y(tp) = y(p)$ for all $t > 0$.

Proof. It is easy to see that if we multiply all of the prices by some positive number $t$, the production plan that maximizes profits will not change. Hence, we must have $y(tp) = y(p)$ for all $t > 0$.

Proposition 4.3.4 (Negative Definiteness of Substitution Matrix) Let $y = f(x)$ be a twice differentiable and strictly concave single output production function, and let $x(p,w)$ be the input demand function. Then, the substitution matrix

$$Dx(p,w) = \left[ \frac{\partial x_i(p,w)}{\partial w_j} \right]$$

is symmetric negative definite.

Proof. Without loss of generality, we normalize $p = 1$. Then the first-order conditions for profit maximization are

$$Df(x(w)) - w \equiv 0.$$

If we differentiate with respect to $w$, we get

$$D^2 f(x(w))Dx(w) - I \equiv 0.$$

Solving this equation for the substitution matrix, we find

$$Dx(w) \equiv [D^2 f(x(w))]^{-1}.$$

Recall that the second-order condition for (strict) profit maximization is that the Hessian matrix is a symmetric negative definite matrix. It is a standard result of linear algebra that the inverse of a symmetric negative definite matrix is
a symmetric negative definite matrix. Then, $D^2f(x(w))$ is a symmetric negative definite matrix, and thus the substitution matrix $Dx(w)$ is a symmetric negative definite matrix. This means that the substitution matrix itself must be a symmetric, negative definite matrix.

**Remark 4.3.3** Note that since $Dx(p,w)$ symmetric, negative definite, we particularly have:

1. $\partial x_i/\partial w_i < 0$, for $i = 1,2,\ldots,n$ since the diagonal entries of a negative definite matrix must be negative.
2. $\partial x_i/\partial w_j = \partial x_j/\partial w_i$ by the symmetry of the matrix.

### 4.3.6 Weak Axiom of Profit Maximization

In this subsection we examine the consequences of profit-maximizing behavior. Suppose that we have are given a list of observed price vectors $p^t$, and the associated net output vectors $y^t$, for $t = 1,\ldots,T$. We refer to this collection as the *data*. In terms of the net supply functions we described before, the data are just $(p^t, y(p^t))$ for some observations $t = 1,\ldots,T$. If the firm is maximizing profits, then the observed net output choice at price $p^t$ must have a level of profit at least as, great as the profit at any other net output the firm could have chosen. Thus, a necessary condition for profit maximization is

$$p^t y^t \geq p^s y^s \text{ for all } t \text{ and } s = 1,\ldots,T.$$ 

We will refer to this condition as the **Weak Axiom of Profit Maximization** (WTAPM).

In Figure 4.4A we have drawn two observations that violate WAPM, while Figure 4.4B depicts two observations that satisfy WAPM.

WAPM is a simple, but very useful, condition; let us derive some of its consequences. Fix two observations $t$ and $s$, and write WAPM for each one. We
Figure 4.4: WAPM. Panel A shows two observations that violate WAPM since \( p^1 y^2 > p^1 y^1 \). Panel B shows two observations that satisfy WAPM.

We have

\[
\begin{align*}
    p^t (y^t - y^s) & \geq 0 \\
    -p^s (y^t - y^s) & \geq 0.
\end{align*}
\]

Adding these two inequalities gives us

\[
(p^t - p^s)(y^t - y^s) \geq 0.
\]

Letting \( \Delta p = (p^t - p^s) \) and \( \Delta y = (y^t - y^s) \), we can rewrite this expression as

\[
\Delta p \Delta y \geq 0.
\]

In other words, the inner product of a vector of price changes with the associated vector of changes in net outputs must be nonnegative.

**Recoverability**

Does WAPM exhaust all of the implications of profit-maximizing behavior, or are there other useful conditions implied by profit maximization? One way to answer this question is to try to construct a technology that generates the observed behavior \((p^t, y^t)\) as profit-maximizing behavior. If we can find such a technology for any set of data that satisfy WAPM, then WAPM must indeed exhaust the
implications of profit-maximizing behavior. We refer to the operation of constructing a technology consistent with the observed choices as the operation of recoverability.

We will show that if a set of data satisfies WAPM it is always possible to find a technology for which the observed choices are profit-maximizing choices. In fact, it is always possible to find a production set $Y$ that is closed and convex. The remainder of this subsection will sketch the proof of this assertion. Formerly, we have

**Proposition 4.3.5** For any set of data satisfies WAPM, there is a convex and closed production set such that the observed choices are profit-maximizing choices.

Proof. We want to construct a convex and closed production set that will generate the observed choices $(p^t, y^t)$ as profit-maximizing choices. We can actually construct two such production sets, one that serves as an “inner bound” to the true technology and one that serves as an “outer bound.” We start with the inner bound.

Suppose that the true production set $Y$ is convex and monotonic. Since $Y$ must contain $y^t$ for $t = 1, ..., T$, it is natural to take the inner bound to be the smallest convex, monotonic set that contains $y^1, ..., y^t$. This set is called the convex, monotonic hull of the points $y^1, ..., y^T$ and is denoted by

$$Y_I = \text{convex, monotonic hull of } \{y^t : t = 1, \cdots, T\}$$

The set $Y_I$ is depicted in Figure 4.5.

It is easy to show that for the technology $Y_I$, $y^t$ is a profit-maximizing choice prices $p^t$. All we have to do is to check that for all $t$, $p^t y^t \geq p^t y$ for all $y$ in $Y_I$.

Suppose that this is not the case. Then for some observation $t, p^t y^t < p^t y$ for some $y$ in $Y_I$. But inspecting the diagram shows that there must then exist some observation $s$ such that $p^t y^t < p^t y^s$. But this inequality violates WAPM.
Thus the set $YI$ rationalizes the observed behavior in the sense that it is one possible technology that could have generated that behavior. It is not hard to see that $YI$ must be contained in any convex technology that generated the observed behavior: if $Y$ generated the observed behavior and it is convex, then it must contain the observed choices $y^t$ and the convex hull of these points is the smallest such set. In this sense, $YI$ gives us an “inner bound” to the true technology that generated the observed choices.

It is natural to ask if we can find an outer bound to this “true” technology. That is, can we find a set $YO$ that is guaranteed to contain any technology that is consistent with the observed behavior?

The trick to answering this question is to rule out all of the points that couldn’t possibly be in the true technology and then take everything that is left over. More precisely, let us define NOTY by

$$NOTY = \{ y: p'y > p'y^t \text{ for some } t \}.$$

NOTY consists of all those net output bundles that yield higher profits than some observed choice. If the firm is a profit maximizer, such bundles couldn’t be technologically feasible; otherwise they would have been chosen. Now as our outer bound to $Y$ we just take the complement of this set:

$$YO = \{ y: p'y \leq p'y^t \text{ for all } t = 1, ..., T \}.$$

Figure 4.5: The set of $YI$ and $YO$. 

203
The set $YO$ is depicted in Figure 4.5B.

In order to show that $YO$ rationalizes the observed behavior we must show that the profits at the observed choices are at least as great as the profits at any other $y$ in $YO$. Suppose not. Then there is some $y'$ such that $p'y' < p'y$ for some $y$ in $YO$. But this contradicts the definition of $YO$ given above. It is clear from the construction of $YO$ that it must contain any production set consistent with the data $(y')$. Hence, $YO$ and $YI$ form the tightest inner and outer bounds to the true production set that generated the data.

### 4.3.7 Profit Function

Given any production set $Y$, we have seen how to calculate the profit function, $\pi(p)$, which gives us the maximum profit attainable at prices $p$. The profit function possesses several important properties that follow directly from its definition. These properties are very useful for analyzing profit-maximizing behavior.

#### Properties of the Profit Function

The properties given below follow solely from the assumption of profit maximization. No assumptions about convexity, monotonicity, or other sorts of regularity are necessary.

**Proposition 4.3.6 (Properties of the Profit Function)** The following has the following properties:

1. **Nondecreasing in output prices, nonincreasing in input prices.** If $p'_i \geq p_i$ for all outputs and $p'_j \leq p_j$ for all inputs, then $\pi(p') \geq \pi(p)$.

2. **Homogeneous of degree 1 in $p$.** $\pi(tp) = t\pi(p)$ for all $t \geq 0$.

3. **Convex in $p$.** Let $p'' = tp + (1-t)p'$ for $0 \leq t \leq 1$. Then $\pi(p'') \leq t\pi(p) + (1-t)\pi(p')$.

204
Continuous in $p$. The function $\pi(p)$ is continuous, at least when $\pi(p)$ is well-defined and $p_i > 0$ for $i = 1, \ldots, n$.

**Proof.**

1. Let $y$ be a profit-maximizing net output vector at $p$, so that $\pi(p) = py$ and let $y'$ be a profit-maximizing net output vector at $p'$ so that $\pi(p') = p'y'$. Then by definition of profit maximization we have $p'y' \geq p'y$. Since $p'_i \geq p_i$ for all $i$ for which $y_i \geq 0$ and $p'_i \leq p_i$ for all $i$ for which $y_i \leq 0$, we also have $p'y \geq py$. Putting these two inequalities together, we have $\pi(p') = p'y' \geq py = \pi(p)$, as required.

2. Let $y$ be a profit-maximizing net output vector at $p$, so that $py \geq py'$ for all $y'$ in $Y$. It follows that for $t \geq 0$, $tpy \geq tpy'$ for all $y'$ in $Y$. Hence $y$ also maximizes profits at prices $tp$. Thus $\pi(tp) = tpy = t\pi(p)$.

3. Let $y$ maximize profits at $p$, $y'$ maximize profits at $p'$, and $y''$ maximize profits at $p''$. Then we have

$$\pi(p'') = p''y'' = (tp + (1 - t)p')y'' = tpy'' + (1 - t)p'y''.$$  \hspace{1cm} (3.1)

By the definition of profit maximization, we know that

$$tpy'' \leq tpy = t\pi(p)$$

and

$$(1 - t)p'y'' \leq (1 - t)p'y' = (1 - t)\pi(p').$$

Adding these two inequalities and using (3.1), we have

$$\pi(p'') \leq t\pi(p) + (1 - t)\pi(p'),$$

as required.

4. The continuity of $\pi(p)$ follows from the Theorem of the Maximum.

205
Deriving Net Supply Functions from Profit Function

If we are given the net supply function \( y(p) \), it is easy to calculate the profit function. We just substitute into the definition of profits to find \( \pi(p) = py(p) \). Suppose that instead we are given the profit function and are asked to find the net supply functions. How can that be done? It turns out that there is a very simple way to solve this problem: just differentiate the profit function. The proof that this works is the content of the next proposition.

**Proposition 4.3.7 (Hotelling’s lemma.)** Let \( y_i(p) \) be the firm’s net supply function for good \( i \). Then

\[
y_i(p) = \frac{\partial \pi(p)}{\partial p_i} \quad \text{for } i = 1, ..., n,
\]

assuming that the derivative exists and that \( p_i > 0 \).

**Proof.** Suppose \((y^*)\) is a profit-maximizing net output vector at prices \((p^*)\). Then define the function

\[
g(p) = \pi(p) - py^*.
\]

Clearly, the profit-maximizing production plan at prices \( p \) will always be at least as profitable as the production plan \( y^* \). However, the plan \( y^* \) will be a profit-maximizing plan at prices \( p^* \), so the function \( g \) reaches a minimum value of 0 at \( p^* \). The assumptions on prices imply this is an interior minimum.

The first-order conditions for a minimum then imply that

\[
\frac{\partial g(p^*)}{\partial p_i} = \frac{\partial \pi(p^*)}{\partial p_i} - y_i^* = 0 \quad \text{for } i = 1, \ldots, n.
\]

Since this is true for all choices of \( p^* \), the proof is done.

**Remark 4.3.4** Again, we can prove this derivative property of the profit function by applying Envelope Theorem:

\[
\frac{d\pi(p)}{dp_i} \bigg|_{x=x(a)} = \frac{\partial py(p)}{\partial p_i} = y_i.
\]

206
This expression says that the derivative of $\pi$ with respect to $a$ is given by the partial derivative of $f$ with respect to $p_i$, holding $x$ fixed at the optimal choice. This is the meaning of the vertical bar to the right of the derivative.

### 4.4 Cost Minimization

An important implication of the firm choosing a profit-maximizing production plan is that there is no way to produce the same amounts of outputs at a lower total input cost. Thus, cost minimization is a necessary condition for profit maximization. This observation motives us to an independent study of the firm’s cost minimization. The problem is of interest for several reasons. First, it leads us to a number of results and constructions that are technically very useful. Second, as long as the firm is a price taker in its input market, the results flowing from the cost minimization continue to be valid whether or not the output market is competitive and so whether or not the firm takes the output price as given.

Third, when the production set exhibits nondecreasing returns to scale, the cost function and optimizing vectors of the cost minimization problem, which keep the levels of outputs fixed, are better behaved than the profit function.

To be concrete, we focus our analysis on the single-output case. We assume throughout that firms are perfectly competitive on their input markets and therefore they face fixed prices. Let $w = (w_1, w_2, \ldots, w_n) \geq 0$ be a vector of prevailing market prices at which the firm can buy inputs $x = (x_1, x_2, \ldots, x_n)$.

#### 4.4.1 First-Order Conditions of Cost Minimization

Let us consider the problem of finding a cost-minimizing way to produce a given level of output:

$$\min_{x} w^T x$$

such that $f(x) \geq y$
We analyze this constrained minimization problem using the Lagrangian function:

\[ L(\lambda, x) = wx - \lambda (f(x) - y) \]

where production function \( f \) is assumed to be differentiate and \( \lambda \) is the Lagrange multiplier. The first-order conditions characterizing an interior solution \( x^* \) are

\[ w_i - \lambda \frac{\partial f(x^*)}{\partial x_i} = 0, \quad i = 1, \ldots, n \]
\[ f(x^*) = y \]

or in vector notation, the condition can be written as

\[ w = \lambda \mathbf{D} f(x^*). \]

We can interpret these first-order conditions by dividing the \( j^{th} \) condition by the \( i^{th} \) condition to get

\[ \frac{w_i}{w_j} = \frac{\frac{\partial f(x^*)}{\partial x_i}}{\frac{\partial f(x^*)}{\partial x_j}} \quad i, j = 1, \ldots, n, \]

which means the marginal rate of technical substitution of factor \( i \) for factor \( j \) equals the economic rate of substitution factor \( i \) for factor \( i \) at the cost minimizing input bundle.

This first-order condition can also be represented graphically. In Figure 4.6, the curved lines represent iso-quants and the straight lines represent constant cost curves. When \( y \) is fixed, the problem of the firm is to find a cost-minimizing point on a given iso-quant. It is clear that such a point will be characterized by the tangency condition that the slope of the constant cost curve must be equal to the slope of the iso-quant.

Again, the conditions are valid only for interior operating positions: they must be modified if a cost minimization point occurs on the boundary. The appropriate conditions turn out to be

\[ \lambda \frac{\partial f(x^*)}{\partial x_i} - w_i \leq 0 \quad \text{with equality if} \quad x_i > 0, \quad i = 1, 2, \ldots, n \]
Remark 4.4.1 It is known that a continuous function achieves a minimum and a maximum value on a closed and bounded set. The objective function $w\mathbf{x}$ is certainly a continuous function and the set $V(y)$ is a closed set by hypothesis. All that we need to establish is that we can restrict our attention to a bounded subset of $V(y)$. But this is easy. Just pick an arbitrary value of $x$, say $x'$. Clearly the minimal cost factor bundle must have a cost less than $w\mathbf{x}'$. Hence, we can restrict our attention to the subset $\{x \in V(y): w\mathbf{x} \leq w\mathbf{x}'\}$, which will certainly be a bounded subset, as long as $w > 0$. Thus the cost minimizing input bundle always exists.

4.4.2 Sufficiency of Cost Minimization

Again, like consumer’s constrained optimization problem, the above first-order conditions are merely necessary conditions for a local optimum. However, these necessary first-order conditions are in fact sufficient for a global optimum when a production function is quasi-concave, which is formerly stated in the following proposition.

Proposition 4.4.1 Suppose that $f(x) : \mathbb{R}_+^n \rightarrow \mathbb{R}$ is differentiable and quasi-concave on $\mathbb{R}_+^n$ and $w > 0$. If $(x, \lambda) > 0$ satisfies the first-order conditions given
in (4.5) and (4.6), then \( x \) solves the firm’s cost minimization problem at prices \( w \).

**Proof.** Since \( f(x) : \mathbb{R}^n \to \mathbb{R}^n \) is differentiable and quasi-concave, the input requirement set \( V(y) = \{ x : f(x) \geq y \} \) is a convex and closed set. Further the objective function \( wx \) is convex and continuous, then by the Kuhn-Tucker theorem, the first-order conditions are sufficient for the constrained minimization problem.

Similarly, the strict quasi-concavity of \( f \) can be checked by verifying if the naturally ordered principal minors of the bordered Hessian alternative in sign, i.e.,

\[
\begin{vmatrix}
0 & f_1 & f_2 \\
f_1 & f_{11} & f_{12} \\
f_2 & f_{21} & f_{22}
\end{vmatrix} > 0,
\]

\[
\begin{vmatrix}
0 & f_1 & f_2 & f_3 \\
f_1 & f_{11} & f_{12} & f_{13} \\
f_2 & f_{21} & f_{22} & f_{23} \\
f_3 & f_{31} & f_{32} & f_{33}
\end{vmatrix} < 0,
\]

and so on, where \( f_i = \frac{\partial f}{\partial x_i} \) and \( f_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j} \).

For each choice of \( w \) and \( y \) there will be some choice of \( x^* \) that minimizes the cost of producing \( y \) units of output. We will call the function that gives us this optimal choice the **conditional input demand function** and write it as \( x(w,y) \). Note that conditional factor demands depend on the level of output produced as well as on the factor prices. The **cost function** is the minimal cost at the factor prices \( w \) and output level \( y \); that is. \( c(w,y) = wx(w,y) \).

**Example 4.4.1 (Cost Function for the Cobb-Douglas Technology)** Consider the cost minimization problem

\[
c(w,y) = \min_{x_1,x_2} w_1x_1 + w_2x_2 \\
such that \ Ax_1^a x_2^b = y.
\]

210
Solving the constraint for $x_2$, we see that this problem is equivalent to

$$\min_{x_1} w_1 x_1 + w_2 A^{-\frac{1}{\lambda}} y^\frac{1}{\lambda} x_1^{-\frac{\alpha}{\lambda}}.$$

The first-order condition is

$$w_1 - \frac{a}{b} w_2 A^{-\frac{1}{\lambda}} y^\frac{1}{\lambda} x_1^{-\frac{\alpha + b}{\lambda}} = 0,$$

which gives us the conditional input demand function for factor 1:

$$x_1(w_1, w_2, y) = A^{-\frac{1}{\lambda + \alpha}} \left[ \frac{aw_1}{bw_1} \right]^\frac{b}{\lambda + \alpha} y^\frac{1}{\lambda + \alpha}.$$

The other conditional input demand function is

$$x_2(w_1, w_2, y) = A^{-\frac{1}{\lambda + \alpha}} \left[ \frac{aw_2}{bw_1} \right]^\frac{a}{\lambda + \alpha} y^\frac{1}{\lambda + \alpha}.$$

The cost function is thus

$$c(w_1, w_2, y) = w_1 x_1(w_1, w_2, y) + w_2 x_2(w_1, w_2, y)$$

$$= A^{-\frac{1}{\lambda + \alpha}} \left[ \left( \frac{a}{b} \right)^\frac{b}{\lambda + \alpha} \left( \frac{a}{b} \right)^\frac{a}{\lambda + \alpha} \right] w_1^\frac{a}{\lambda + \alpha} w_2^\frac{b}{\lambda + \alpha} y^\frac{1}{\lambda + \alpha}.$$

When $A = 1$ and $a + b = 1$ (constant returns to scale), we particularly have

$$c(w_1, w_2 y) = K w_1^a w_2^{1-a} y,$$

where $K = a^{-a}(1 - a)^{a-1}$.

**Example 4.4.2 (The Cost function for the CES technology)** Suppose that $f(x_1, x_2) = (x_1^\rho + x_2^\rho)^{\frac{1}{\rho}}$. What is the associated cost function? The cost minimization problem is

$$\min_{x_1, x_2} w_1 x_1 + w_2 x_2$$

such that $x_1^\rho + x_2^\rho = y^\rho$

The first-order conditions are

$$w_1 - \lambda \rho x_1^{\rho - 1} = 0$$

$$w_2 - \lambda \rho x_2^{\rho - 1} = 0$$

$$x_1^\rho + x_2^\rho = y.$$
Solving the first two equations for $x_1^\rho$ and $x_2^\rho$, we have

\begin{align*}
x_1^\rho &= w_1^{\frac{\rho}{\rho-1}} (\lambda \rho)^{\frac{\rho}{\rho-1}} \quad (4.9) \\
x_2^\rho &= w_2^{\frac{\rho}{\rho-1}} (\lambda \rho)^{\frac{\rho}{\rho-1}}. \quad (4.10)
\end{align*}

Substitute this into the production function to find

\[(\lambda \rho)^{\frac{\rho}{\rho-1}} \left( w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} \right) = y^\rho.\]

Solve this for $(\lambda \rho)^{\frac{\rho}{\rho-1}}$ and substitute into equations (4.9) and (4.10). This gives us the conditional input demand functions

\begin{align*}
x_1(w_1, w_2, y) &= w_1^{\frac{1}{\rho-1}} \left[ w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} \right]^{-\frac{1}{\rho}} y \\
x_2(w_1, w_2, y) &= w_2^{\frac{1}{\rho-1}} \left[ w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} \right]^{-\frac{1}{\rho}} y.
\end{align*}

Substituting these functions into the definition of the cost function yields

\begin{align*}
c(w_1, w_2, y) &= w_1 x_1(w_1, w_2, y) + w_2 x_2(w_1, w_2, y) \\
&= y \left[ w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} \right] \left[ w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} \right]^{-\frac{1}{\rho}} y \\
&= y \left[ w_1^{\frac{\rho}{\rho-1}} + w_2^{\frac{\rho}{\rho-1}} \right]^{\frac{\rho-1}{\rho}}. \\
\end{align*}

This expression looks a bit nicer if we set $r = \rho/(\rho - 1)$ and write

\[c(w_1, w_2, y) = y[w_1^r + w_2^r]^{\frac{1}{r}}.\]

Note that this cost function has the same form as the original CES production function with $r$ replacing $\rho$. In the general case where

\[f(x_1, x_2) = [(a_1 x_1)^\rho + (a_2 x_2)^\rho]^{\frac{1}{\rho}},\]

similar computations can be done to show that

\[c(w_1, w_2, y) = [(w_1/a_1)^r + (w_2/a_2)^r]^{\frac{1}{r}} y.\]
Example 4.4.3 (The Cost function for the Leontief technology) Suppose $f(x_1, x_2) = \min\{ax_1, bx_2\}$. Since we know that the firm will not waste any input with a positive price, the firm must operate at a point where $y = ax_1 = bx_2$. Hence, if the firm wants to produce $y$ units of output, it must use $y/a$ units of good 1 and $y/b$ units of good 2 no matter what the input prices are. Hence, the cost function is given by

$$c(w_1, w_2, y) = \frac{w_1 y}{a} + \frac{w_2 y}{b} = y \left( \frac{w_1}{a} + \frac{w_2}{b} \right).$$

Example 4.4.4 (The cost function for the linear technology) Suppose that $f(x_1, x_2) = ax_1 + bx_2$, so that factors 1 and 2 are perfect substitutes. What will the cost function look like? Since the two goods are perfect substitutes, the firm will use whichever is cheaper. Hence, the cost function will have the form $c(w_1, w_2, y) = \min\{w_1/a, w_2/b\}y$.

In this case the answer to the cost-minimization problem typically involves a boundary solution: one of the two factors will be used in a zero amount. It is easy to see the answer to this particular problem by comparing the relative steepness of the isocost line and isoquant curve. If $\frac{a_1}{a_2} < \frac{w_1}{w_2}$, the firm only uses $x_2$ and the cost function is given by $c(w_1, w_2, y) = w_2 x_2 = w_2 \frac{y}{a_2}$. If $\frac{a_1}{a_2} > \frac{w_1}{w_2}$, the firm only uses $x_1$ and the cost function is given by $c(w_1, w_2, y) = w_1 x_1 = w_1 \frac{y}{a_1}$.

### 4.4.3 Cost Functions and Their Properties

The cost function measures the minimum cost of producing a given level of output for some fixed factor prices. As such it summarizes information about the technological choices available to the firms. It turns out that the behavior of the cost function can tell us a lot about the nature of the firm’s technology. In the following we will first investigate the behavior of the cost function $c(w, y)$ with respect to its price and quantity arguments. We then define a few related functions, namely the average and the marginal cost functions.
Properties of Cost Functions

You may have noticed some similarities here with consumer theory. These similarities are in fact exact when one compares the cost function with the expenditure function. Indeed, consider their definitions.

(1) Expenditure Function: \( e(p, u) \equiv \min_{x \in \mathbb{R}^n_+} px \) such that \( u(x) \geq u \)

(2) Cost Function: \( c(w, y) \equiv \min_{x \in \mathbb{R}^n_+} wx \) such that \( f(x) \geq y \)

Mathematically, the two optimization problems are identical. Consequently, for every theorem we proved about expenditure functions, there is an equivalent theorem for cost functions. We shall state these results here, but we do not need to prove them. Their proofs are identical to those given for the expenditure function.

Proposition 4.4.2 [Properties of the Cost Function.] Suppose the production function \( f \) is continuous and strictly increasing. Then the cost function has the following properties:

(1) \( c(w, y) \) is nondecreasing in \( w \).

(2) \( c(w, y) \) is homogeneous of degree 1 in \( w \).

(3) \( c(w, y) \) is concave in \( w \).

(4) \( c(w, y) \) is continuous in \( w \), for \( w > 0 \).

(5) For all \( w > 0 \), \( c(w, y) \) is strictly increasing \( y \).

(6) Shephard’s lemma: If \( x(w, y) \) is the cost-minimizing bundle necessary to produce production level \( y \) at prices \( w \), then \( x_i(w, y) = \frac{\partial c(w, y)}{\partial w_i} \) for \( i = 1, ..., n \), assuming the derivative exists and that \( x_i > 0 \).
Properties of Conditional Input Demand

As solution to the firm’s cost-minimization problem, the conditional input demand functions possess certain general properties. These are analogous to the properties of Hicksian compensation demands, so once again it is not necessary to repeat the proof.

**Proposition 4.4.3 (Negative Semi-Definite Substitution Matrix)** The matrix of substitution terms \( \frac{\partial x_j(w, y)}{\partial w_i} \) is negative semi-definite.

Again since the substitution matrix is negative semi-definite, thus it is symmetric and has non-positive diagonal terms. We then particularly have

**Proposition 4.4.4 (Symmetric Substitution Terms)** The matrix of substitution terms is symmetric, i.e.,

\[
\frac{\partial x_j(w, y)}{\partial w_i} = \frac{\partial^2 c(w, y)}{\partial w_j \partial w_i} = \frac{\partial^2 c(w, y)}{\partial w_i \partial w_j} = \frac{\partial x_i(w, y)}{\partial w_j}.
\]

**Proposition 4.4.5 (Negative Own-Substitution Terms)** The compensated own-price effect is non-positive; that is, the input demand curves slope downward:

\[
\frac{\partial x_i(w, y)}{\partial w_i} = \frac{\partial^2 c(w, y)}{\partial w_i^2} \leq 0.
\]

**Remark 4.4.2** Using the cost function, we can restate the firm’s profit maximization problem as

\[
\max_{y \geq 0} py - c(w, y). \tag{4.11}
\]

The necessary first-order condition for \( y^* \) to be profit-maximizing is then

\[
p - \frac{\partial c(w, y^*)}{\partial y} \leq 0, \text{ with equality if } y^* > 0. \tag{4.12}
\]

In other words, at an interior optimum (i.e., \( y^* > 0 \)), price equals marginal cost. If \( c(w, y) \) is convex in \( y \), then the first-order condition (4.12) is also sufficient for \( y^* \) to be the firm’s optimal output level.
**Average and Marginal Costs**

Let us consider the structure of the cost function. Note that the cost function can always be expressed simply as the value of the conditional factor demands.

\[ c(w, y) = wx(w, y) \]

In the short run, some of the factors of production are fixed at predetermined levels. Let \( x_f \) be the vector of fixed factors, \( x_v \), the vector of variable factors, and break up \( w \) into \( w = (w_v, w_f) \), the vectors of prices of the variable and the fixed factors. The short-run conditional factor demand functions will generally depend on \( x_f \), so we write them as \( x_v(w, y, x_f) \). Then the short-run cost function can be written as

\[ c(w, y, x_f) = w_v x_v(w, y, x_f) + w_f x_f. \]

The term \( w_v x_v(w, y, x_f) \) is called **short-run variable cost** (SVC), and the term \( w_f x_f \) is the **fixed cost** (FC). We can define various derived cost concepts from these basic units:

- **short-run total cost** = \( STC = w_v x_v(w, y, x_f) + w_f x_f \)
- **short-run average cost** = \( SAC = \frac{c(w, y, x_f)}{y} \)
- **short-run average variable cost** = \( SAVC = \frac{w_v x_v(w, y, x_f)}{y} \)
- **short-run average fixed cost** = \( SAFC = \frac{w_f x_f}{y} \)
- **short-run marginal cost** = \( SMC = \frac{\partial c(w, y, x_f)}{\partial y} \).

When all factors are variable, the firm will optimize in the choice of \( x_f \). Hence, the long-run cost function only depends on the factor prices and the level of output as indicated earlier. We can express this long-run function in terms of the short-run cost function in the following way. Let \( x_f(w, y) \) be the optimal choice of
the fixed factors, and let \( x_v(w, y) = x_v(w, y, x_f(w, y)) \) be the long-run optimal choice of the variable factors. Then the long-run cost function can be written as

\[
c(w, y) = w_v x_v(w, y) + w_f x_f(w, y) = c(w, y, x_f(w, y)).
\]

Similarly, we can define the long-run average and marginal cost functions:

\[
\begin{align*}
\text{long-run average cost} & = LAC = \frac{c(w, y)}{y} \\
\text{long-run marginal cost} & = LMC = \frac{\partial c(w, y)}{\partial y}.
\end{align*}
\]

Notice that “long-run average cost” equals “long-run average variable cost” since all costs are variable in the long-run; “long-run fixed costs” are zero for the same reason.

**Example 4.4.5 (The short-run Cobb-Douglas cost functions)** Suppose the second factor in a Cobb-Douglas technology is restricted to operate at a level \( k \). Then the cost-minimizing problem is

\[
\min w_1 x_1 + w_2 k
\]

such that \( y = x_1^a k^{1-a} \).

Solving the constraint for \( x_1 \) as a function of \( y \) and \( k \) gives

\[
x_1 = (yk^{a-1})^{\frac{1}{a}}.
\]

Thus

\[
c(w_1, w_2, y, k) = w_1 (yk^{a-1})^{\frac{1}{a}} + w_2 k.
\]

The following variations can also be calculated:

\[
\begin{align*}
\text{short-run average cost} & = w_1 \left( \frac{y}{k} \right)^{\frac{1-a}{a}} + \frac{w_2 k}{y} \\
\text{short-run average variable cost} & = w_1 \left( \frac{y}{k} \right)^{\frac{1-a}{a}} \\
\text{short-run average fixed cost} & = \frac{w_2 k}{y} \\
\text{short-run marginal cost} & = \frac{w_1}{a} \left( \frac{y}{k} \right)^{\frac{1-a}{a}}
\end{align*}
\]
Example 4.4.6 (Constant returns to scale and the cost function) If the production function exhibits constant returns to scale, then it is intuitively clear that the cost function should exhibit costs that are linear in the level of output: if you want to produce twice as much output it will cost you twice as much. This intuition is verified in the following proposition:

Proposition 4.4.6 (Constant returns to scale) If the production function exhibits constant returns to scale, the cost function may be written as \( c(w, y) = yc(w, 1) \).

Proof. Let \( x^* \) be a cheapest way to produce one unit of output at prices \( w \) so that \( c(w, 1) = wx^* \). We want to show that \( c(w, y) = wyx^* = yc(w, 1) \). Notice first that \( yx^* \) is feasible to produce \( y \) since the technology is constant returns to scale. Suppose that it does not minimize cost; instead let \( x' \) be the cost-minimizing bundle to produce \( y \) at prices \( w \) so that \( wx' < wyx^* \). Then \( wx'/y < wx^* \) and \( x'/y \) can produce 1 since the technology is constant returns to scale. This contradicts the definition of \( x^* \).

Thus, if the technology exhibits constant returns to scale, then the average cost, the average variable cost, and the marginal cost functions are all the same.

The Geometry of Costs

Let us first examine the short-run cost curves. In this case, we will write the cost function simply as \( c(y) \), which has two components: fixed costs and variable costs. We can therefore write short-run average cost as

\[
SAC = \frac{c(w, y, x_f)}{y} = \frac{w_f x_f}{y} + \frac{w_v x_v(w, y, x_f)}{y} = SAFC + SAVC.
\]

As we increase output, average variable costs may initially decrease if there is some initial region of economies of scale. However, it seems reasonable to suppose that the variable factors required will eventually increase by the law of diminishing marginal returns, as depicted in Figure 4.7. Average fixed costs must
of course decrease with output, as indicated in Figure 4.7. Adding together the average variable cost curve and the average fixed costs gives us the U-shaped average cost curve in Figure 4.7. The initial decrease in average costs is due to the decrease in average fixed costs; the eventual increase in average costs is due to the increase in average variable costs. The level of output at which the average cost of production is minimized is sometimes known as the **minimal efficient scale**.

In the long run all costs are variable costs and the appropriate long-run average cost curve should also be U-shaped by the facts that variable costs usually exhibit increasing returns to scale at low level of production and ultimately exhibits decreasing returns to scale.

![Figure 4.7: Average variable, average fixed, and average cost curves.](image)

Let us now consider the marginal cost curve. What is its relationship to the average cost curve? Since

$$
\frac{d}{dy} \left( \frac{c(y)}{y} \right) = \frac{yc'(y) - c(y)}{y^2} = \frac{1}{y} [c'(y) - \frac{c(y)}{y}],
$$

$$
\frac{d}{dy} \left( \frac{c(y)}{y} \right) \leq 0 (\geq 0) \text{ if and only if } c'(y) - \frac{c(y)}{y} \leq 0 (\geq 0).
$$

Thus, the average variable cost curve is decreasing when the marginal cost curve lies below the average variable cost curve, and it is increasing when the marginal cost curve lies above the average variable cost curve. It follows that average cost reach its minimum at $y^*$ when the marginal cost curve passes through the average variable cost curve, i.e.,

$$
c'(y^*) = \frac{c(y^*)}{y^*}.
$$
**Remark 4.4.3** All of the analysis just discussed holds in both the long and the short run. However, if production exhibits constant returns to scale in the long run, so that the cost function is linear in the level of output, then average cost, average variable cost, and marginal cost are all equal to each other, which makes most of the relationships just described rather trivial.

**Long-Run and Short-Run Cost Curves**

Let us now consider the relationship between the long-run cost curves and the short-run cost curves. It is clear that the long-run cost curve must never lie above any short-run cost curve, since the short-run cost minimization problem is just a constrained version of the long-run cost minimization problem.

Let us write the long-run cost function as $c(y) = c(y, z(y))$. Here we have omitted the factor prices since they are assumed fixed, and we let $z(y)$ be the cost-minimizing demand for a single fixed factor. Let $y^*$ be some given level of output, and let $z^* = z(y^*)$ be the associated long-run demand for the fixed factor. The short-run cost, $c(y, z^*)$, must be at least as great as the long-run cost, $c(y^*, z(y^*))$, for all levels of output, and the short-run cost will equal the long-run cost at output $y^*$, so $c(y^*, z^*) = c(y^*, z(y^*))$. Hence, the long- and the short-run cost curves must be tangent at $y^*$.

This is just a geometric restatement of the envelope theorem. The slope of the long-run cost curve at $y^*$ is

$$\frac{dc(y^*, z(y^*))}{dy} = \frac{\partial c(y^*, z^*)}{\partial y} + \frac{\partial c(y^*, z^*)}{\partial z} \frac{\partial z(y^*)}{\partial y}.$$  

But since $z^*$ is the *optimal* choice of the fixed factors at the output level $y^*$, we must have

$$\frac{\partial c(y^*, z^*)}{\partial z} = 0.$$  

Thus, long-run marginal costs at $y^*$ equal short-run marginal costs at $(y^*, z^*)$.

Finally, we note that if the long- and short-run cost curves are tangent, the long- and short-run *average* cost curves must also be tangent. A typical configuration is illustrated in Figure 4.8.
4.4.4 Produce Surplus

To be completed

4.5 Duality in Production

In the last section we investigated the properties of the cost function. Given any technology, it is straightforward, at least in principle, to derive its cost function: we simply solve the cost minimization problem.

In this section we show that this process can be reversed. Given a cost function we can “solve for” a technology that could have generated that cost function. This means that the cost function contains essentially the same information that the production function contains. Any concept defined in terms of the properties of the production function has a “dual” definition in terms of the properties of the cost function and vice versa. This general observation is known as the principle of duality.
4.5.1 Recovering a Production Set from a Cost Function

Given data \((w^t, x^t, y^t)\), define \(VO(y)\) as an “outer bound” to the true input requirement set \(V(y)\):

\[
VO(y) = \{x: w^t x \geq w^t x^t \text{ for all } t \text{ such that } y^t \leq y\}.
\]

It is straightforward to verify that \(VO(y)\) is a closed, monotonic, and convex technology. Furthermore, it contains any technology that could have generated the data \((w^t, x^t, y^t)\) for \(t = 1, ..., T\).

If we observe choices for many different factor prices, it seems that \(VO(y)\) should “approach” the true input requirement set in some sense. To make this precise, let the factor prices vary over all possible price vectors \(w \geq 0\). Then the natural generalization of \(VO\) becomes

\[
V^*(y) = \{x: wx \geq wx(w, y) = c(w, y) \text{ for all } w \geq 0\}.
\]

What is the relationship between \(V^*(y)\) and the true input requirement set \(V(y)\)? Of course, \(V^*(y)\) clearly contain \(V(y)\). In general, \(V^*(y)\) will strictly contain \(V(y)\). For example, in Figure 4.9A we see that the shaded area cannot be ruled out of \(V^*(y)\) since the points in this area satisfy the condition that \(wx \geq c(w, y)\).

The same is true for Figure 4.9B. The cost function can only contain information about the economically relevant sections of \(V(y)\), namely, those factor bundles that could actually be the solution to a cost minimization problem, i.e., that could actually be conditional factor demands.

However, suppose that our original technology is convex and monotonic. In this case \(V^*(y)\) will equal \(V(y)\). This is because, in the convex, monotonic case, each point on the boundary of \(V(y)\) is a cost-minimizing factor demand for some price vector \(w \geq 0\). Thus, the set of points where \(wx \geq c(w, y)\) for all \(w \geq 0\) will precisely describe the input requirement set. More formally:

**Proposition 4.5.1 (Equality of \(V(y)\) and \(V^*(y)\))** Suppose \(V(y)\) is a closed, convex, monotonic technology. Then \(V^*(y) = V(y)\).
Figure 4.9: Relationship between $V(y)$ and $V^*(y)$. In general, $V^*(y)$ will strictly contain $V(y)$.

**Proof (Sketch)** We already know that $V^*(y)$ contains $V(y)$, so we only have to show that if $x$ is in $V^*(y)$ then $x$ must be in $V(y)$. Suppose that $x$ is not an element of $V(y)$. Then since $V(y)$ is a closed convex set satisfying the monotonicity hypothesis, we can apply a version of the separating hyperplane theorem to find a vector $w^* \geq 0$ such that $w^*x < w^*z$ for all $z$ in $V(y)$. Let $z^*$ be a point in $V(y)$ that minimizes cost at the prices $w^*$. Then in particular we have $w^*x < w^*z^* = c(w^*, y)$. But then $x$ cannot be in $V^*(y)$, according to the definition of $V^*(y)$.

This proposition shows that if the original technology is convex and monotonic, then the cost function associated with the technology can be used to completely reconstruct the original technology. This is a reasonably satisfactory result in the case of convex and monotonic technologies, but what about less well-behaved cases? Suppose we start with some technology $V(y)$, possibly non-convex. We find its cost function $c(w, y)$ and then generate $V^*(y)$. We know from the above results that $V^*(y)$ will not necessarily be equal to $V(y)$, unless $V(y)$ happens to have the convexity and monotonicity properties. However, suppose we define

$$c^*(w, y) = \min_{x} wx$$

such that $x$ is in $V^*(y)$.

What is the relationship between $c^*(w, y)$ and $c(w, y)$?
Proposition 4.5.2 (Equality of $c(w, y)$ and $c^*(w, y)$) It follows from the definition of the functions that $c^*(w, y) = c(w, y)$.

Proof. It is easy to see that $c^*(w, y) \leq c(w, y)$; since $V^*(y)$ always contains $V(y)$, the minimal cost bundle in $V^*(y)$ must be at least as small as the minimal cost bundle in $V(y)$. Suppose that for some prices $w'$ the cost-minimizing bundle $x'$ in $V^*(y)$ has the property that $w'x' = c^*(w', y) < c(w', y)$. But this can’t happen, since by definition of $V^*(y)$, $w'x' \geq c(w', y)$.

This proposition shows that the cost function for the technology $V(y)$ is the same as the cost function for its convexification $V^*(y)$. In this sense, the assumption of convex input requirement sets is not very restrictive from an economic point of view.

Let us summarize the discussion to date:

(1) Given a cost function we can define an input requirement set $V^*(y)$.

(2) If the original technology is convex and monotonic, the constructed technology will be identical with the original technology.

(3) If the original technology is non-convex or monotonic, the constructed input requirement will be a convexified, monotonized version of the original set, and, most importantly, the constructed technology will have the same cost function as the original technology.

In conclusion, the cost function of a firm summarizes all of the economically relevant aspects of its technology.

Example 4.5.1 (Recovering production from a cost function) Suppose we are given a specific cost function $c(w, y) = yw_1^aw_2^{1-a}$. How can we solve for its associated technology? According to the derivative property

\[ x_1(w, y) = ayw_1^{a-1}w_2^{-a} = ay \left( \frac{w_2}{w_1} \right)^{1-a} \]

\[ x_2(w, y) = (1-a)yw_1^aw_2^{-a} = (1-a)y \left( \frac{w_2}{w_1} \right)^{-a} \]
We want to eliminate \( w_2/w_1 \) from these two equations and get an equation for \( y \) in terms of \( x_1 \) and \( x_2 \). Rearranging each equation gives

\[
\frac{w_2}{w_1} = \left( \frac{x_1}{ay} \right)^{\frac{1}{a}};
\]

\[
\frac{w_2}{w_1} = \left( \frac{x_2}{(1-a)y} \right)^{-\frac{1}{a}}.
\]

Setting these equal to each other and raising both sides to the \(-a(1-a)\) power,

\[
\frac{x_1^{-a}}{a^{-a}y^{-a}} = \frac{x_2^{-a}}{(1-a)(1-a)y^{1-a}},
\]

or,

\[
[a^a(1-a)^{1-a}]y = x_1^a x_2^{1-a}.
\]

This is just the Cobb-Douglas technology.

We know that if the technology exhibited constant returns to scale, then the cost function would have the form \( c(w)y \). Here we show that the reverse implication is also true.

**Proposition 4.5.3 (Constant returns to scale.)** Let \( V(y) \) be convex and monotonic; then if \( c(w,y) \) can be written as \( yc(w) \), \( V(y) \) must exhibit constant returns to scale.

**Proof** Using convexity, monotonicity, and the assumed form of the cost function assumptions, we know that

\[
V(y) = V^*(y) = \{ x: w \cdot x \geq yc(w) \text{ for all } w \geq 0 \}.
\]

We want to show that, if \( x \) is in \( V^*(y) \), then \( tx \) is in \( V^*(ty) \). If \( x \) is in \( V^*(y) \), we know that \( w x \geq yc(w) \) for all \( w \geq 0 \). Multiplying both sides of this equation by \( t \) we get: \( w tx \geq tyc(w) \) for all \( w \geq 0 \). But this says \( tx \) is in \( V^*(ty) \).
4.5.2 Characterization of Cost Functions

We have seen in the last section that all cost functions are nondecreasing, homogeneous, concave, continuous functions of prices. The question arises: suppose that you are given a nondecreasing, homogeneous, concave, continuous function of prices is it necessarily the cost function of some technology? The answer is yes, and the following proposition shows how to construct such a technology.

Proposition 4.5.4 Let $\phi(w, y)$ be a differentiable function satisfying

1. $\phi(tw, y) = t\phi(w, y)$ for all $t \geq 0$;
2. $\phi(w, y) \geq 0$ for $w \geq 0$ and $y \geq 0$;
3. $\phi(w', y) \geq \phi(w, y)$ for $w' \geq w$;
4. $\phi(w, y)$ is concave in $w$.

Then $\phi(w, y)$ is the cost function for the technology defined by $V^*(y) = \{ x \geq 0 : wx \geq \phi(w, y), \text{ for all } w \geq 0 \}$.

Proof. Given a $w \geq 0$ we define

$$x(w, y) = \left( \frac{\partial \phi(w, y)}{\partial w_1}, \ldots, \frac{\partial \phi(w, y)}{\partial w_n} \right)$$

and note that since $\phi(w, y)$ is homogeneous of degree 1 in $w$, Euler’s law implies that $\phi(w, y)$ can be written as

$$\phi(w, y) = \sum_{i=1}^{n} w_i \frac{\partial \phi(w, y)}{\partial w_i} = wx(w, y).$$

Note that the monotonicity of $\phi(w, y)$ implies $x(w, y) \geq 0$.

What we need to show is that for any given $w' \geq 0$, $x(w', y)$ actually minimizes $w'x$ over all $x$ in $V^*(y)$:

$$\phi(w', y) = w'x(w', y) \leq w'x \text{ for all } x \text{ in } V^*(y).$$
First, we show that \( x(w', y) \) is feasible; that is, \( x(w', y) \) is in \( V^*(y) \). By the concavity of \( \phi(w, y) \) in \( w \) we have

\[
\phi(w', y) \leq \phi(w, y) + D\phi(w, y)(w' - w)
\]

for all \( w \geq 0 \).

Using Euler’s law as above, this reduces to

\[
\phi(w', y) \leq w'x(w, y) \text{ for all } w \geq 0.
\]

It follows from the definition of \( V^*(y) \), that \( x(w', y) \) is in \( V^*(y) \).

Next we show that \( x(w, y) \) actually minimizes \( wx \) over all \( x \) in \( V^*(y) \). If \( x \) is in \( V^*(y) \), then by definition it must satisfy

\[
w x \geq \phi(w, y).
\]

But by Euler’s law,

\[
\phi(w, y) = w x(w, y).
\]

The above two expressions imply

\[
w x \geq w x(w, y)
\]

for all \( x \) in \( V^*(y) \) as required.

4.5.3 The Integrability for Cost Functions

The proposition proved in the last subsection raises an interesting question. Suppose you are given a set of functions \( (g_i(w, y)) \) that satisfy the properties of conditional factor demand functions described in the previous sections, namely, that they are homogeneous of degree 0 in prices and that

\[
\left( \frac{\partial g_i(w, y)}{\partial w_j} \right)
\]

is a symmetric negative semi-definite matrix. Are these functions necessarily factor demand functions for some technology?
Let us try to apply the above proposition. First, we construct a candidate for a cost function:

$$\phi(w, y) = \sum_{i=1}^{n} w_i g_i(w, y).$$

Next, we check whether it satisfies the properties required for the proposition just proved.

1) Is $\phi(w, y)$ homogeneous of degree 1 in $w$? To check this we look at $\phi(tw, y) = \sum_i tw_i g_i(tw, y)$. Since the functions $g_i(w, y)$ are by assumption homogeneous of degree 0, $g_i(tw, y) = g_i(w, y)$ so that

$$\phi(tw, y) = t \sum_{i=1}^{n} wg_i(w, y) = t\phi(w, y).$$

2) Is $\phi(w, y) \geq 0$ for $w \geq 0$? Since $g_i(w, y) \geq 0$, the answer is clearly yes.

3) Is $\phi(w, y)$ nondecreasing in $w_i$? Using the product rule, we compute

$$\frac{\partial \phi(w, y)}{\partial w_i} = g_i(w, y) + \sum_{j=1}^{n} w_j \frac{\partial g_i(w, y)}{\partial w_i} = g_i(w, y) + \sum_{j=1}^{n} w_j \frac{\partial g_i(w, y)}{\partial w_j}.$$

Since $g_i(w, y)$ is homogeneous of degree 0, the last term vanishes and $g_i(w, y)$ is clearly greater than or equal to 0.

4) Finally is $\phi(w, y)$ concave in $w$? To check this we differentiate $\phi(w, y)$ twice to get

$$\left( \frac{\partial^2 \phi}{\partial w_i \partial w_j} \right) = \left( \frac{\partial g_i(w, y)}{\partial w_j} \right).$$

For concavity we want these matrices to be symmetric and negative semi-definite, which they are by hypothesis.

Hence, the proposition proved in the last subsection applies and there is a technology $V^*(y)$ that yields $(g_i(w, y))$ as its conditional factor demands. This means that the properties of homogeneity and negative semi-definiteness form a complete list of the restrictions on demand functions imposed by the model of cost-minimizing behavior.
4.6 Reference

Books and Monographs:


Papers:


Chapter 5

Choice Under Uncertainty

Rewrite this chapter

5.1 Introduction

Until now, we have been concerned with the behavior of a consumer under conditions of certainty. However, many choices made by consumers take place under conditions of uncertainty. In this chapter we explore how the theory of consumer choice can be used to describe such behavior.

The board outline of this chapter parallels a standard presentation of microeconomic theory for deterministic situations. It first considers the problem of an individual consumer facing an uncertain environment. It shows how preference structures can be extended to uncertain situations and describes the nature of the consumer choice problem. We then proceed to derive the expected utility theorem, a result of central importance. In the remaining sections, we discuss the concept of risk aversion, and extend the basic theory by allowing utility to depend on states of nature underlying the uncertainty as well as on the monetary payoffs. We also discuss the theory of subjective probability, which offers a way of modelling choice under uncertainty in which the probabilities of different risky alternatives are not given to the decision maker in any objective fashion.
5.2 Three Basic Theoretical Models

5.2.1 Basic Description of Choice Under Uncertainty

5.2.2 Von Neumann-Morgenstern Expected Utility Model

5.2.3 Savage Subjective Probability Model

5.2.4 Anscombe-Aumann State Dependent Utility Model

5.3 Von Neumann-Morgenstern Expected Utility Theory

5.3.1 Model and Basic Axioms

5.3.2 Expected Utility Presentation Theorem

The first task is to describe the set of choices facing the consumer. We shall imagine that the choices facing the consumer take the form of lotteries. Suppose there are \( S \) states. Associated with each state \( s \) is a probability \( p_s \) representing the probability that the state \( s \) will occur and a commodity bundle \( x_s \) representing the prize or reward that will be won if the state \( s \) occurs, where we have \( p_s \geq 0 \) and \( \sum_{s=1}^{S} p_s = 1 \). The prizes may be money, bundles of goods, or even further lotteries. A lottery is denoted by

\[
p_1 \circ x_1 \oplus p_2 \circ x_2 \oplus \ldots \oplus p_S \circ x_S.
\]

For instance, for two states, a lottery is given \( p \circ x \oplus (1 - p) \circ y \) which means: “the consumer receives prize \( x \) with probability \( p \) and prize \( y \) with probability \( (1-p) \).” Most situations involving behavior under risk can be put into this lottery framework. Lotteries are often represented graphically by a fan of possibilities as in Figure 5.1 below.

A compound lottery is shown in Figure 5.2. This lottery is between two prizes: a lottery between \( x \) and \( y \), and a bundle \( z \).
We will make several axioms about the consumers perception of the lotteries open to him.

**A1 (Certainty)**. \( 1 \circ x \oplus (1 - 1) \circ y \sim x \). Getting a prize with probability one is equivalent to that prize.

**A2 (Independence of Order)**. \( p \circ x \oplus (1 - p) \circ y \sim (1 - p) \circ y \oplus p \circ x \). The consumer doesn’t care about the order in which the lottery is described—only the prizes and the probabilities of winning those prizes matter.

**A3 (Compounding)**. \( q \circ (p \circ x \oplus (1 - p) \circ y) \oplus (1 - q) \circ y \sim (qp) \circ x \oplus (1 - qp) \circ y \). It is only the net probabilities of receiving the a reward that matters. It is a fundamental axiom used to reduce compound lotteries—by determining the overall probabilities associated with its components. This axiom sometimes called “reduction of compound lotteries.”

Under these assumptions we can define \( \mathcal{L} \), the space of lotteries available to
the consumer. The consumer is assumed to have preferences on this lottery space: given any two lotteries, he can choose between them. As usual we will assume the preferences are complete, reflexive, and transitive so it is an ordering preference.

The fact that lotteries have only two outcomes is not restrictive since we have allowed the outcomes to be further lotteries. This allows us to construct lotteries with arbitrary numbers of prizes by compounding two prize lotteries as shown in Figure 5.2. For example, suppose we want to represent a situation with three prizes \( x, y \) and \( z \) where the probability of getting each prize is one third. By the reduction of compound lotteries, this lottery is equivalent to the lottery

\[
\frac{2}{3} \circ \left[ \frac{1}{2} \circ x \oplus \frac{1}{2} \circ y \right] \oplus \frac{1}{3} \circ z.
\]

According to the compounding axiom (A3) above, the consumer only cares about the net probabilities involved, so this is indeed equivalent to the original lottery.

Under minor additional assumptions, the theorem concerning the existence of a utility function may be applied to show that there exists a continuous utility function \( u \) which describes the consumer’s preferences; that is, \( p \circ x \oplus (1 - p) \circ y \succ q \circ w \oplus (1 - q) \circ z \) if and only if

\[
\frac{1}{2} \circ x \oplus \frac{1}{2} \circ y > \frac{1}{3} \circ z.
\]

Of course, this utility function is not unique; any monotonic transform would do as well. Under some additional hypotheses, we can find a particular monotonic transformation of the utility function that has a very convenient property, the expected utility property:

\[
u(p \circ x \oplus (1 - p) \circ y) = pu(x) + (1 - p)u(y).
\]

The expected utility property says that the utility of a lottery is the expectation of the utility from its prizes and such an expected utility function is called \textbf{von Neumann-Morgenstern utility function}. To have a utility function with the above convenient property, we need the additional axioms:
A4 (Continuity). \( \{ p \in [0, 1]: p \circ x \oplus (1 - p) \circ y \succeq z \} \) and \( \{ p \in [0, 1]: z \succeq p \circ x \oplus (1 - p) \circ y \} \) are closed sets for all \( x, y \) and \( z \) in \( \mathcal{L} \). Axiom 4 states that preferences are continuous with respect to probabilities.

A5 (Strong Independence). \( x \sim y \) implies \( p \circ x \oplus (1 - p) \circ z \sim p \circ y \oplus (1 - p) \circ z \). It says that lotteries with indifferent prizes are indifferent.

In order to avoid some technical details we will make two further assumptions.

A6 (Boundedness). There is some best lottery \( b \) and some worst lottery \( w \). For any \( x \) in \( \mathcal{L} \), \( b \succeq x \succeq w \).

A7 (Monotonicity). A lottery \( p \circ b \oplus (1 - p) \circ w \) is preferred to \( q \circ b \oplus (1 - q) \circ w \) if and only if \( p > q \).

Axiom A7 can be derived from the other axioms. It just says that if one lottery between the best prize and the worst prize is preferred to another it must be because it gives higher probability of getting the best prize.

Under these assumptions we can state the main theorem.

**Theorem 5.3.1 (Expected utility theorem)** If \( (\mathcal{L}, \succeq) \) satisfy Axioms 1-7, there is a utility function \( u \) defined on \( \mathcal{L} \) that satisfies the expected utility property:

\[
u(p \circ x \oplus (1 - p) \circ y) = pu(x) + (1 - p)u(y)
\]

**Proof.** Define \( u(b) = 1 \) and \( u(w) = 0 \). To find the utility of an arbitrary lottery \( z \), set \( u(z) = p_z \) where \( p_z \) is defined by

\[
p_z \circ b \oplus (1 - p_z) \circ w \sim z. \tag{5.1}
\]

In this construction the consumer is indifferent between \( z \) and a gamble between the best and the worst outcomes that gives probability \( p_z \) of the best outcome.

To ensure that this is well defined, we have to check two things.

(1) Does \( p_z \) exist? The two sets \( \{ p \in [0, 1]: p \circ b \oplus (1 - p) \circ w \succeq z \} \) and \( \{ p \in [0, 1]: z \succeq p \circ b \oplus (1 - p) \circ w \} \) are closed and nonempty by the continuity and boundedness axioms (A4 and A6), and every point in \( [0, 1] \) is in one or the other of the two sets. Since the unit interval is connected, there must be some \( p \) in both – but this will just be the desired \( p_z \).
(2) Is $p_z$ unique? Suppose $p_z$ and $p'_z$ are two distinct numbers and that each satisfies (5.1). Then one must be larger than the other. By the monotonicity axiom A7, the lottery that gives a bigger probability of getting the best prize cannot be indifferent to one that gives a smaller probability. Hence, $p_z$ is unique and $u$ is well defined.

We next check that $u$ has the expected utility property. This follows from some simple substitutions:

$$p \circ x \circ (1 - p) \circ y$$

$$\sim_1 p \circ [p_x \circ b \circ (1 - p_x) \circ w] \oplus (1 - p) \circ [p_y \circ b \circ (1 - p_y) \circ w]$$

$$\sim_2 [p p_x + (1 - p) p_y] \circ b \oplus [1 - p p_x - (1 - p) p_y] \circ w$$

$$\sim_3 [p u(x) + (1 - p) u(y)] \circ b \oplus (1 - p u(x) - (1 - p) u(y)) \circ w.$$

Substitution 1 uses the strong independence axiom (A5) and the definition of $p_z$ and $p_y$. Substitution 2 uses the compounding axiom (A3), which says only the net probabilities of obtaining $b$ or $w$ matter. Substitution 3 uses the construction of the utility function.

It follows from the construction of the utility function that

$$u(p \circ x \circ (1 - p) \circ y) = p u(x) + (1 - p) u(y).$$

Finally, we verify that $u$ is a utility function. Suppose that $x \succ y$. Then

$$u(x) = p_z \text{ such that } x \sim p_x \circ b \oplus (1 - p_x) \circ w$$

$$u(y) = p_y \text{ such that } y \sim p_y \circ b \oplus (1 - p_y) \circ w$$

By the monotonicity axiom (A7), we must have $u(x) > u(y)$.

We have shown that there exists an expected utility function $u : \mathcal{L} \rightarrow \mathbb{R}$. Of course, any monotonic transformation of $u$ will also be a utility function that describes the consumer’s choice behavior. But will such a monotonic transform preserve the expected utility properly? Does the construction described above characterize expected utility functions in any way?
It is not hard to see that, if $u(\cdot)$ is an expected utility function describing some consumer, then so is $v(\cdot) = au(\cdot) + c$ where $a > 0$; that is, any affine transformation of an expected utility function is also an expected utility function. This is clear since

$$v(p \cdot x \oplus (1 - p) \cdot y) = au(p \cdot x \oplus (1 - p) \cdot y) + c$$
$$= a[pu(x) + (1 - p)u(y)] + c$$
$$= pv(x) + (1 - p)v(y).$$

It is not much harder to see the converse: that any monotonic transform of $u$ that has the expected utility property must be an affine transform. Stated another way:

**Theorem 5.3.2 (Uniqueness of expected utility function)** *An expected utility function is unique up to an affine transformation.*

**Proof.** According to the above remarks we only have to show that, if a monotonic transformation preserves the expected utility property, it must be an affine transformation. Let $f : R \to R$ be a monotonic transformation of $u$ that has the expected utility property. Then

$$f(u(p \cdot x \oplus (1 - p) \cdot y)) = pf(u(x)) + (1 - p)f(u(y)),$$

or

$$f(pu(x) + (1 - p)u(y)) = pf(u(x)) + (1 - p)f(u(y)).$$

But this is equivalent to the definition of an affine transformation.

We have proved the expected utility theorem for the case where there are two outcomes to the lotteries. As indicated earlier, it is straightforward to extend this proof to the case of a finite number of outcomes by using compound lotteries. If outcome $x_i$ is received with probability $p_i$ for $i = 1, \ldots, n$, the expected utility of this lottery is simply

$$\sum_{t=1}^{n} p_i u(x_i).$$  \tag{5.2}
Subject to some minor technical details, the expected utility theorem also holds for continuous probability distributions. If \( p(x) \) is probability density function defined on outcomes \( x \), then the expected utility of this gamble can be written as

\[
\int u(x)p(x)dx.
\] (5.3)

We can subsume of these cases by using the expectation operator. Let \( X \) be a random variable that takes on values denoted by \( x \). Then the utility function of \( X \) is also a random variable, \( u(X) \). The expectation of this random variable \( Eu(X) \) is simply the expected utility associated with the lottery \( X \). In the case of a discrete random variable, \( Eu(X) \) is given by (5.2), and in the case of a continuous random variable \( Eu(X) \) is given by (5.3).

5.4 Risk aversion

Rewrite this section

5.4.1 Absolute Risk Aversion

Let us consider the case where the lottery space consists solely of gambles with money prizes. We have shown that if the consumer’s choice behavior satisfies Axioms 1-7, we can use an expected utility function to represent the consumer’s preferences on lotteries. This means that we can describe the consumer’s behavior over all money gambles by expected utility function. For example, to compute the consumer’s expected utility of a gamble \( p \circ x \oplus (1 - p) \circ y \), we just look at \( pu(x) + (1 - p)u(y) \).

This construction is illustrated in Figure 5.3 for \( p = \frac{1}{2} \). Notice that in this example the consumer prefers to get the expected value of the lottery. This is, the utility of the lottery \( u(p \circ x \oplus (1 - p) \circ y) \) is less than the utility of the expected value of the lottery, \( px + (1 - p)y \). Such behavior is called risk aversion. A consumer may also be risk loving; in such a case, the consumer prefers a lottery to its expected value.
Figure 5.3: Expected utility of a gamble.

If a consumer is risk averse over some region, the chord drawn between any two points of the graph of his utility function in this region must lie below the function. This is equivalent to the mathematical definition of a concave function. Hence, concavity of the expected utility function is equivalent to risk aversion.

It is often convenient to have a measure of risk aversion. Intuitively, the more concave the expected utility function, the more risk averse the consumer. Thus, we might think we could measure risk aversion by the second derivative of the expected utility function. However, this definition is not invariant to changes in the expected utility function: if we multiply the expected utility function by 2, the consumer’s behavior doesn’t change, but our proposed measure of risk aversion does. However, if we normalize the second derivative by dividing by the first, we get a reasonable measure, known as the Arrow-Pratt measure of (absolute) risk aversion:

\[
    r(w) = -\frac{u''(w)}{u'(w)}
\]

**Example 5.4.1 (Constant risk aversion)** Suppose an individual has a constant risk aversion coefficient \( r \). Then the utility function satisfies

\[
    u''(x) = -ru'(x).
\]
One can easily check that all solutions are

\[ u(x) = -ae^{-rx} + b \]

where \( a \) and \( b \) are arbitrary. For \( u(x) \) to be increasing in \( x \), we must take \( a > 0 \).

**Example 5.4.2** A woman with current wealth \( X \) has the opportunity to bet any amount on the occurrence of an event that she knows will occur with probability \( p \). If she wagers \( w \), she will receive \( 2w \) if the event occurs and 0 if it does not. She has a constant risk aversion coefficient utility \( u(x) = -e^{-rx} \) with \( r > 0 \). How much should she wager?

Her final wealth will be either \( X + w \) or \( X - w \). Hence she solves

\[
\max_w \{ pu(X + w) + (1 - p)u(X - w) \} = \max_w \{ -pe^{-r(X+w)} - (1 - p)e^{-r(X-w)} \}
\]

Setting the derivative to zero yields

\[
(1 - p)e^{rw} = pe^{-rw}.
\]

Hence,

\[
w = \frac{1}{2r} \ln \frac{p}{1 - p}.
\]

Note that a positive wager will be made for \( p > 1/2 \). The wager decreases as the risk coefficient increases. Note also that in this case the results is independent of the initial wealth—a particular feature of this utility function.

**Example 5.4.3 (The demand for insurance)** Suppose a consumer initially has monetary wealth \( W \). There is some probability \( p \) that he will lose an amount \( L \) for example, there is some probability his house will burn down. The consumer can purchase insurance that will pay him \( q \) dollars in the event that he incurs this loss. The amount of money that he has to pay for \( q \) dollars of insurance coverage is \( \pi q \); here \( \pi \) is the premium per dollar of coverage.

How much coverage will the consumer purchase? We look at the utility maximization problem

\[
\max pu(W - L - \pi q + q) + (1 - p)u(W - \pi q).
\]
Taking the derivative with respect to $q$ and setting it equal to zero, we find

$$pu'(W - L + q^*(1 - \pi))(1 - \pi) - (1 - p)u'(W - \pi q^*)\pi = 0$$

$$\frac{u'(W - L + (1 - \pi)q^*)}{u'(W - \pi q^*)} = \frac{(1 - p)\pi}{p(1 - \pi)}$$

If the event occurs, the insurance company receives $\pi q - q$ dollars. If the event doesn’t occur, the insurance company receives $\pi q$ dollars. Hence, the expected profit of the company is

$$(1 - p)\pi q - p(1 - \pi)q.$$ 

Let us suppose that competition in the insurance industry forces these profits to zero. This means that

$$-p(1 - \pi)q + (1 - p)\pi q = 0,$$

from which it follows that $\pi = p$.

Under the zero-profit assumption the insurance firm charges an actuarially fair premium: the cost of a policy is precisely its expected value, so that $p = \pi$. Inserting this into the first-order conditions for utility maximization, we find

$$u'(W - L + (1 - \pi)q^*) = u'(W - \pi q^*).$$

If the consumer is strictly risk averse so that $u''(W) < 0$, then the above equation implies

$$W - L + (1 - \pi)q^* = W - \pi q^*$$

from which it follows that $L = q^*$. Thus, the consumer will completely insure himself against the loss $L$.

This result depends crucially on the assumption that the consumer cannot influence the probability of loss. If the consumer’s actions do affect the probability of loss, the insurance firms may only want to offer partial insurance, so that the consumer will still have an incentive to be careful.
5.4.2 Global Risk Aversion

The Arrow-Pratt measure seems to be a sensible interpretation of local risk aversion: one agent is more risk averse than another if he is willing to accept fewer small gambles. However, in many circumstances we want a global measure of risk aversion—that is, we want to say that one agent is more risk averse than another for all levels of wealth. What are natural ways to express this condition?

The first plausible way is to formalize the notion that an agent with utility function $A(w)$ is more risk averse than an agent with utility function $B(w)$ is to require that

$$\frac{-A''(w)}{A'(w)} > \frac{-B''(w)}{B'(w)}$$

for all levels of wealth $w$. This simply means that agent $A$ has a higher degree of risk aversion than agent $B$ everywhere.

Another sensible way to formalize the notion that agent $A$ is more risk averse than agent $B$ is to say that agent $A$’s utility function is “more concave” than agent $B$’s. More precisely, we say that agent $A$’s utility function is a concave transformation of agent $B$’s; that is, there exists some increasing, strictly concave function $G(\cdot)$ such that

$$A(w) = G(B(w)).$$

A third way to capture the idea that $A$ is more risk averse than $B$ is to say that $A$ would be willing to pay more to avoid a given risk than $B$ would. In order to formalize this idea, let $\bar{e}$ be a random variable with expectation of zero: $E\bar{e} = 0$. Then define $\pi_A(\bar{e})$ to be the maximum amount of wealth that person $A$ would give up in order to avoid facing the random variable $\bar{e}$. In symbols, this risk premium is

$$A(w - \pi_A(\bar{e})) = EA(w + \bar{e}).$$

The left-hand side of this expression is the utility from having wealth reduced by $\pi_A(\bar{e})$ and the right-hand side is the expected utility from facing the gamble $\bar{e}$.
It is natural to say that person $A$ is (globally) more risk averse than person $B$ if \( \pi_A(\epsilon) > \pi_B(\epsilon) \) for all \( \epsilon \) and \( w \).

It may seem difficult to choose among these three plausible sounding interpretations of what it might mean for one agent to be “globally more risk averse” than another. Luckily, it is not necessary to do so: all three definitions turn out to be equivalent! As one step in the demonstration of this fact we need the following result, which is of great use in dealing with expected utility functions.

**Lemma 5.4.1 (Jensen’s inequality)** Let \( X \) be a nondegenerate random variable and \( f(X) \) be a strictly concave function of this random variable. Then \( Ef(X) < f(EX) \).

**Proof.** This is true in general, but is easiest to prove in the case of a differentiable concave function. Such a function has the property that at any point \( \bar{x} \), \( f(x) < f(\bar{x}) + f'(\bar{x})(x - \bar{x}) \). Let \( \bar{X} \) be the expected value of \( X \) and take expectations of each side of this expression, we have

\[
Ef(X) < f(\bar{X}) + f'(\bar{X})E(X - \bar{X}) = f(\bar{X}),
\]

from which it follows that

\[
Ef(X) < f(\bar{X}) = f(EX).
\]

**Theorem 5.4.1 (Pratt’s theorem)** Let \( A(w) \) and \( B(w) \) be two differentiable, increasing and concave expected utility functions of wealth. Then the following properties are equivalent.

1. \(-A''(w)/A'(w) > -B''(w)/B'(w)\) for all \( w \).

2. \( A(w) = G(B(w)) \) for some increasing strictly concave function \( G \).

3. \( \pi_A(\bar{\epsilon}) > \pi_B(\bar{\epsilon}) \) for all random variables \( \bar{\epsilon} \) with \( E\bar{\epsilon} = 0 \).

**Proof**
(1) implies (2). Define \( G(B) \) implicitly by \( A(w) = G(B(w)) \). Note that monotonicity of the utility functions implies that \( G \) is well defined i.e., that there is a unique value of \( G(B) \) for each value of \( B \). Now differentiate this definition twice to find

\[
A'(w) = G'(B)B'(w) \\
A''(w) = G''(B)B'(w)^2 + G'(B)B''(w).
\]

Since \( A'(w) > 0 \) and \( B'(w) > 0 \), the first equation establishes \( G'(B) > 0 \). Dividing the second equation by the first gives us

\[
\frac{A''(w)}{A'(w)} = \frac{G''(B)}{G'(B)}B'(w) + \frac{B''(w)}{B'(w)}.
\]

Rearranging gives us

\[
\frac{G''(B)}{G'(B)}B'(w) = \frac{A''(w)}{A'(w)} - \frac{B''(w)}{B'(w)} < 0,
\]

where the inequality follows from (1). This shows that \( G''(B) < 0 \), as required.

(2) implies (3). This follows from the following chain of inequalities:

\[
A(w - \pi_A) = EA(w + \bar{\epsilon}) = EG(B(w + \bar{\epsilon})) < G(EB(w + \bar{\epsilon})) = G(B(w - \pi_B)) = A(w - \pi_B).
\]

All of these relationships follow from the definition of the risk premium except for the inequality, which follows from Jensen’s inequality. Comparing the first and the last terms, we see that \( \pi_A > \pi_B \).

(3) implies (1). Since (3) holds for all zero-mean random variables \( \bar{\epsilon} \), it must hold for arbitrarily small random variables. Fix an \( \bar{\epsilon} \), and consider the family of random variables defined by \( t\bar{\epsilon} \) for \( t \) in \([0, 1]\). Let \( \pi(t) \) be the risk premium as a function of \( t \). The second-order Taylor series expansion of \( \pi(t) \) around \( t = 0 \) is given by

\[
\pi(t) \approx \pi(0) + \pi(0)t + \pi + \frac{1}{2}\pi''(0)t^2. \quad (5.4)
\]
We will calculate the terms in this Taylor series in order to see how $\pi(t)$ behaves for small $t$. The definition of $\pi(t)$ is

$$A(w - \pi(t)) \equiv EA(w + \varepsilon t).$$

It follows from this definition that $\pi(0) = 0$. Differentiating the definition twice with respect to $t$ gives us

$$-A'(w - \pi(t))\pi'(t) = E[A'(w + t\varepsilon)\varepsilon]$$

and

$$A''(w - \pi(t))\pi'(t)^2 - A'(w - \pi(t))\pi''(t) = E[A''(w + t\varepsilon)\varepsilon].$$

Evaluating the first expression when $t = 0$, we see that $\pi'(0) = 0$. Evaluating the second expression when $t = 0$, we see that

$$\pi''(0) = -\frac{EA''(w)\varepsilon^2}{A'(w)} = -\frac{A''(w)}{A'(w)}\sigma^2,$$

where $\sigma^2$ is the variance of $\varepsilon$. Plugging the derivatives into equation (5.4) for $\pi(t)$, we have

$$\pi(t) \approx 0 + 0 - A''(w)\sigma^2 t^2.$$ 

This implies that for arbitrarily small values of $t$, the risk premium depends monotonically on the degree of risk aversion, which is what we wanted to show.

### 5.4.3 Relative Risk Aversion

Consider a consumer with wealth $w$ and suppose that she is offered gambles of the form: with probability $p$ she will receive $x$ percent of her current wealth; with probability $(1 - p)$ she will receive $y$ percent of her current wealth. If the consumer evaluates lotteries using expected utility, the utility of this lottery will be

$$pu(xw) + (1 - p)u(yw).$$

Note that this multiplicative gamble has a different structure than the additive gambles analyzed above. Nevertheless, relative gambles of this sort often arise
in economic problems. For example, the return on investments is usually stated relative to the level of investment.

Just as before we can ask when one consumer will accept more small relative gambles than another at a given wealth level. Going through the same sort of analysis used above, we find that the appropriate measure turns out to be the Arrow-Pratt measure of relative risk aversion:

\[ \rho = \frac{-u''(w)w}{u'(w)}. \]

It is reasonable to ask how absolute and relative risk aversions might vary with wealth. It is quite plausible to assume that absolute risk aversion decreases with wealth: as you become more wealthy you would be willing to accept more gambles expressed in absolute dollars. The behavior of relative risk aversion is more problematic; as your wealth increases would you be more or less willing to risk losing a specific fraction of it? Assuming constant relative risk aversion is probably not too bad an assumption, at least for small changes in wealth.

**Example 5.4.4 (Mean-variance utility)** In general the expected utility of a gamble depends on the entire probability distribution of the outcomes. However, in some circumstances the expected utility of a gamble will only depend on certain summary statistics of the distribution. The most common example of this is a mean-variance utility function.

For example, suppose that the expected utility function is quadratic, so that \( u(w) = w - bw^2 \). Then expected utility is

\[ Eu(w) = Ew - bw^2 = \bar{w} - bw^2 - b\sigma_w^2. \]

Hence, the expected utility of a gamble is only a function of the mean and variance of wealth.

Unfortunately, the quadratic utility function has some undesirable properties: it is a decreasing function of wealth in some ranges, and it exhibits increasing absolute risk aversion.
A more useful case when mean-variance analysis is justified is the case when wealth is Normally distributed. It is well-known that the mean and variance completely characterize a Normal random variable; hence, choice among Normally distributed random variables reduces to a comparison on their means and variances. One particular case that is of special interest is when the consumer has a utility function of the form \( u(w) = -e^{-rw} \) which exhibits constant absolute risk aversion. Furthermore, when wealth is Normally distributed

\[
Eu(w) = -\int e^{-rw} f(z) dw = -e^{-r[\bar{w} - r\sigma_w^2/2]}.
\]

(To do the integration, either complete the square or else note that this is essentially the calculation that one does to find the moment generating function for the Normal distribution.) Note that expected utility is increasing in \( \bar{w} - r\sigma_w^2/2 \). This means that we can take a monotonic transformation of expected utility and evaluate distributions of wealth using the utility function \( u(\bar{w}, \sigma_w^2) = \bar{w} - \frac{r}{2}\sigma_w^2 \).

This utility function has the convenient property that it is linear in the mean and variance of wealth.

In our original analysis of choice under uncertainty, the prizes were simply abstract bundles of goods; later we specialized to lotteries with only monetary outcomes when considering the risk aversion issue. However, this is restrictive. After all, a complete description of the outcome of a dollar gamble should include not only the amount of money available in each outcome but also the prevailing prices in each outcome.

More generally, the usefulness of a good often depends on the circumstances or state of nature in which it becomes available. An umbrella when it is raining may appear very different to a consumer than an umbrella when it is not raining. These examples show that in some choice problems it is important to distinguish goods by the state of nature in which they are available.

For example, suppose that there are two states of nature, hot and cold, which we index by \( h \) and \( c \). Let \( x_h \) be the amount of ice cream delivered when it is hot and \( x_c \) the amount delivered when it is cold. Then if the probability of hot
weather is $p$, we may write a particular lottery as $pu(h, x_h) + (1 - p)u(c, x_c)$. Here the bundle of goods that is delivered in one state is “hot weather and $x_h$ units of ice cream,” and “cold weather and $x_c$, units of ice cream” in the other state.

A more serious example involves health insurance. The value of a dollar may well depend on one’s health – how much would a million dollars be worth to you if you were in a coma? In this case we might well write the utility function as $u(h, m_h)$ where $h$ is an indicator of health and $m$ is some amount of money. These are all examples of state-dependent utility functions. This simply means that the preferences among the goods under consideration depend on the state of nature under which they become available.

5.4.4 Anscombe-Aumann State Dependent Utility Theory

5.4.5 Model and Basic Axioms

5.4.6 State Dependent Utility Presentation Theorem

5.5 Savage Subjective Probability Theory

Rewrite this section.
5.5.1 Model and Basic Axioms

5.5.2 Subject Probability Utility Presentation Theorem

5.6 Further Extension of Choice Under Uncertainty

5.6.1 Two Classical Paradoxes on Choice Under Uncertainty

In the discussion of expected utility theory we have used “objective” probabilities — such as probabilities calculated on the basis of some observed frequencies and asked what axioms about a person’s choice behavior would imply the existence of an expected utility function that would represent that behavior.

However, many interesting choice problems involve subjective probabilities: a given agent’s perception of the likelihood of some event occurring. Similarly, we can ask what axioms about a person’s choice behavior can be used to infer the existence of subjective probabilities; i.e., that the person’s choice behavior can be viewed as if he were evaluating gambles according to their expected utility with respect to some subjective probability measures.

As it happens, such sets of axioms exist and are reasonably plausible. Subjective probabilities can be constructed in a way similar to the manner with which the expected utility function was constructed. Recall that the utility of some gamble $x$ was chosen to be that number $u(x)$ such that

$$x \sim u(x) \circ b \oplus (1 - u(x)) \circ w.$$ 

Suppose that we are trying to ascertain an individual’s subjective probability that it will rain on a certain date. Then we can ask at what probability $p$ will the individual be indifferent between the gamble $p \circ b \oplus (1 - p) \circ w$ and the gamble “Receive $b$ if it rains and $w$ otherwise.”

249
More formally, let $E$ be some event, and let $p(E)$ stand for the (subjective) probability that $E$ will occur. We define the subjective probability that $E$ occurs by the number $p(E)$ that satisfies $p(E) \circ b \oplus +(1 - p(E)) \circ w \sim \text{receive } b \text{ if } E \text{ occurs and } w \text{ otherwise.}$

It can be shown that under certain regularity assumptions the probabilities defined in this way have all of the properties of ordinary objective probabilities. In particular, they obey the usual rules for manipulation of conditional probabilities. This has a number of useful implications for economic behavior. We will briefly explore one such implication. Suppose that $p(H)$ is an individual’s subjective probability that a particular hypothesis is true, and that $E$ is an event that is offered as evidence that $H$ is true. How should a rational economic agent adjust his probability belief about $H$ in light of the evidence $E$? That is, what is the probability of $H$ being true, conditional on observing the evidence $E$?

We can write the joint probability of observing $E$ and $H$ being true as

$$p(H, E) = p(H|E)p(E) = p(H)p(E).$$

Rearranging the right-hand sides of this equation,

$$p(H|E) = \frac{p(E|H)p(H)}{p(E)}.$$ 

This is a form of Bayes’ law which relates the prior probability $p(H)$, the probability that the hypothesis is true before observing the evidence, to the posterior probability, the probability that the hypothesis is true after observing the evidence.

Bayes’ law follows directly from simple manipulations of conditional probabilities. If an individual’s behavior satisfies restrictions sufficient to ensure the existence of subjective probabilities, those probabilities must satisfy Bayes’ law. Bayes’ law is important since it shows how a rational individual should update his probabilities in the light of evidence, and hence serves as the basis for most models of rational learning behavior.
Thus, both the utility function and the subjective probabilities can be constructed from observed choice behavior, as long as the observed choice behavior follows certain intuitively plausible axioms. However, it should be emphasized that although the axioms are intuitively plausible it does not follow that they are accurate descriptions of how individuals actually behave. That determination must be based on empirical evidence.

Expected utility theory and subjective probability theory were motivated by considerations of rationality. The axioms underlying expected utility theory seem plausible, as does the construction that we used for subjective probabilities.

Unfortunately, real-life individual behavior appears to systematically violate some of the axioms. Here we present two famous examples.

**Example 5.6.1 (The Allais paradox)** You are asked to choose between the following two gambles:

**Gamble A.** A 100 percent chance of receiving 1 million.

**Gamble B.** A 10 percent chance of 5 million, an 89 percent chance of 1 million, and a 1 percent chance of nothing.

Before you read any further pick one of these gambles, and write it down.

Now consider the following two gambles.

**Gamble C.** An 11 percent chance of 1 million, and an 89 percent chance of nothing.

**Gamble D.** A 10 percent chance of 5 million, and a 90 percent chance of nothing.

Again, please pick one of these two gambles as your preferred choice and write it down. Many people prefer A to B and D to C. However, these choices violate the expected utility axioms! To see this, simply write the expected utility relationship implied by $A \succeq B$:

$$u(1) > .1u(5) + .89u(1) + .01u(0).$$

Rearranging this expression gives
and adding \(0.89u(0)\) to each side yields

\[
0.11u(1) + 0.89u(0) > 0.1u(5) + 0.90u(0),
\]

It follows that gamble C must be preferred to gamble D by an expected utility maximizer.

**Example 5.6.2 (The Ellsberg paradox)** The Ellsberg paradox concerns subjective probability theory. You are told that an urn contains 300 balls. One hundred of the balls are red and 200 are either blue or green.

- **Gamble A.** You receive $1,000 if the ball is red.
- **Gamble B.** You receive $1,000 if the ball is blue.
- Write down which of these two gambles you prefer. Now consider the following two gambles:
  - **Gamble C.** You receive $1,000 if the ball is not red.
  - **Gamble D.** You receive $1,000 if the ball is not blue.

It is common for people to strictly prefer A to B and C to D. But these preferences violate standard subjective probability theory. To see why, let \(R\) be the event that the ball is red, and \(\neg R\) be the event that the ball is not red, and define \(B\) and \(\neg B\) accordingly. By ordinary rules of probability,

\[
p(R) = 1 - p(\neg R) \tag{5.5}
\]
\[
p(B) = 1 - p(\neg B).
\]

Normalize \(u(0) = 0\) for convenience. Then if A is preferred to B, we must have \(p(R)u(1000) > p(B)u(1000)\), from which it follows that

\[
p(R) > p(B). \tag{5.6}
\]
If $C$ is preferred to $D$, we must have $p(\neg R)u(1000) > p(\neg B)u(1000)$, from which it follows that
\[ p(\neg R) > p(\neg B). \] (5.7)

However, it is clear that expressions (5.5), (5.6), and (5.7) are inconsistent.

The Ellsberg paradox seems to be due to the fact that people think that betting for or against $R$ is “safer” than betting for or against “blue.”

Opinions differ about the importance of the Allais paradox and the Ellsberg paradox. Some economists think that these anomalies require new models to describe people’s behavior. Others think that these paradoxes are akin to “optical illusions.” Even though people are poor at judging distances under some circumstances doesn’t mean that we need to invent a new concept of distance.

5.6.2 Prospect Theory
5.6.3 Max-Min Expected Utility Theory
5.6.4 Case Based Decision Theory

5.7 Reference

Books and Monographs:


Laffont, J. J. (1989). *The Economics of Uncertainty and Information*, MIT.


**Papers:**


256


Part III

Game Theory and Market Theory
We have restricted ourselves until now to the ideal situation (benchmark case) where the behavior of the others are summarized in non-individualized parameters – the prices of commodities, each individual makes decision independently by taking prices as given and individuals’ behavior are indirectly interacted through prices. This is clearly a very restricted assumption. In many cases, one individual’s action can directly affect the actions of the others, and also are affected by the others’s actions. Thus, it is very important to study the realistic case where individuals’ behavior are affected each other and they interact each other. The game theory is thus a powerful tool to study individuals’ cooperation and solving possible conflicts.

In this part, we will first discuss the game theory and then the various types of market structures where one individual’s decision may affect the decisions of the others. It is greatly acknowledged that the materials are largely drawn from the references in the end of each charter, especially from Varian (1992).
Chapter 6

Game Theory

To be written.

6.1 Introduction

Game theory is the study of interacting decision makers. In earlier chapters we studied the theory of optimal decision making by a single agent–firm or a consumer–in very simple environments. The strategic interactions of the agents were not very complicated. In this chapter we will lay the foundations for a deeper analysis of the behavior of economic agents in more complex environments.

There are many directions from which one could study interacting decision makers. One could examine behavior from the viewpoint of sociology, psychology, biology, etc. Each of these approaches is useful in certain contexts. Game theory emphasizes a study of cold-blooded “rational” decision making, since this is felt to be the most appropriate model for most economic behavior.

Game theory has been widely used in economics in the last two decade, and much progress has been made in clarifying the nature of strategic interaction in economic models. Indeed, most economic behavior can be viewed as special cases of game theory, and a sound understanding of game theory is a necessary component of any economist’s set of analytical tools.
6.2 Description of a game

There are several ways of describing a game. For our purposes, the strategic form and the extensive form will be sufficient. Roughly speaking the extensive form provides an “extended” description of a game while the strategic form provides a “reduced” summary of a game. We will first describe the strategic form, reserving the discussion of the extensive form for the section on sequential games.

6.2.1 Strategic Form

The strategic form of the game is defined by exhibiting a set of players $N = \{1, 2, \ldots, n\}$. Each player $i$ has a set of strategies $S_i$ from which he/she can choose an action $s_i \in S_i$ and a payoff function, $\phi_i(s)$, that indicate the utility that each player receives if a particular combination $s$ of strategies is chosen, where $s = (s_1, s_2, \ldots, s_n) \in S = \prod_{i=1}^{n} S_i$. For purposes of exposition, we will treat two-person games in this chapter. All of the concepts described below can be easily extended to multi-person contexts.

We assume that the description of the game – the payoffs and the strategies available to the players – are common knowledge. That is, each player knows his own payoffs and strategies, and the other player’s payoffs and strategies. Furthermore, each player knows that the other player knows this, and so on. We also assume that it is common knowledge that each player is “fully rational.” That is, each player can choose an action that maximizes his utility given his subjective beliefs, and that those beliefs are modified when new information arrives according to Bayes’ law.

Game theory, by this account, is a generalization of standard, one-person decision theory. How should a rational expected utility maximizer behave in a situation in which his payoff depends on the choices of another rational expected utility maximizer? Obviously, each player will have to consider the problem faced by the other player in order to make a sensible choice. We examine the outcome of this sort of consideration below.
Example 6.2.1 (Matching pennies) In this game, there are two players, Row and Column. Each player has a coin which he can arrange so that either the head side or the tail side is face-up. Thus, each player has two strategies which we abbreviate as Heads or Tails. Once the strategies are chosen there are payoffs to each player which depend on the choices that both players make.

These choices are made independently, and neither player knows the other’s choice when he makes his own choice. We suppose that if both players show heads or both show tails, then Row wins a dollar and Column loses a dollar. If, on the other hand, one player exhibits heads and the other exhibits tails, then Column wins a dollar and Row looses a dollar.

\[
\begin{array}{c|cc}
\text{Column} & \text{Heads} & \text{Tails} \\
\hline
\text{Row Heads} & (1, -1) & (-1, 1) \\
\text{Tails} & (-1, 1) & (1, -1) \\
\end{array}
\]

Table 6.1: Game Matrix of Matching Pennies

We can depict the strategic interactions in a game matrix. The entry in box (Head, Tails) indicates that player Row gets $-1$ and player Column gets $+1$ if this particular combination of strategies is chosen. Note that in each entry of this box, the payoff to player Row is just the negative of the payoff to player Column. In other words, this is a zero-sum game. In zero-sum games the interests of the players are diametrically opposed and are particularly simple to analyze. However, most games of interest to economists are not zero sum games.

Example 6.2.2 (The Prisoners Dilemma) Again we have two players, Row and Column, but now their interests are only partially in conflict. There are two strategies: to Cooperate or to Defect.

In the original story, Row and Column were two prisoners who jointly participated in a crime. They could cooperate with each other and refuse to give evidence (i.e., do not confess), or one could defect (i.e, confess) and implicate the
other. They are held in separate cells, and each is privately told that if he is the only one to confess, then he will be rewarded with a light sentence of 1 year while the recalcitrant prisoner will go to jail for 10 years. However, if the person is the only one not to confess, then it is the who will serve the 10-year sentence. If both confess, they will both be shown some mercy: they will each get 5 years. Finally, if neither confesses, it will still possible to convict both of a lesser crime that carries a sentence of 2 years. Each player wishes to minimize the time he spends in jail. The outcome can be shown in Table 6.2.

<table>
<thead>
<tr>
<th>Prisoner 2</th>
<th>Don’t Confess</th>
<th>Confess</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prisoner 1 Don’t Confess</td>
<td>(−2, −2)</td>
<td>(−10, −1)</td>
</tr>
<tr>
<td>Don’t confess</td>
<td>(−1, −10)</td>
<td>(−5, −5)</td>
</tr>
</tbody>
</table>

Table 6.2: The Prisoner’s Dilemma

The problem is that each party has an incentive to confess, regardless of what he or she believes the other party will do. In this prisoner’s dilemma, “confession” is the best strategy to each prisoner regardless the choice of the other.

An especially simple revised version of the above prisoner’s dilemma given by Aumann (1987) is the game in which each player can simply announce to a referee: “Give me $1,000,” or “Give the other player $3,000.” Note that the monetary payments come from a third party, not from either of the players; the Prisoner’s Dilemma is a variable-sum game.

The players can discuss the game in advance but the actual decisions must be independent. The Cooperate strategy is for each person to announce the $3,000 gift, while the Defect strategy is to take the $1,000 (and run!). Table 6.3 depicts the payoff matrix to the Aumann version of the Prisoner’s Dilemma, where the units of the payoff are thousands of dollars.

We will discuss this game in more detail below. Again, each party has an incentive to defect, regardless of what he or she believes the other party will do.
Table 6.3: A Revised Version of Prisoner’s Dilemma by Aumann

For if I believe that the other person will cooperate and give me a $3,000 gift, then I will get $4,000 in total by defecting. On the other hand, if I believe that the other person will defect and just take the $1,000, then I do better by taking $1,000 for myself.

In other applications, cooperate and defect could have different meanings. For example, in a duopoly situation, cooperate could mean “keep charging a high price” and defect could mean “cut your price and steal your competitor’s market.”

Example 6.2.3 (Cournot Duopoly) Consider a simple duopoly game, first analyzed by Cournot (1838). We suppose that there are two firms who produce an identical good with a marginal cost \( c \). Each firm must decide how much output to produce without knowing the production decision of the other duopolist. If the firms produce a total of \( x \) units of the good, the market price will be \( p(x) \); that is, \( p(x) \) is the inverse demand curve facing these two producers.

If \( x_i \) is the production level of firm \( i \), the market price will then be \( p(x_1 + x_2) \), and the profits of firm \( i \) are given by \( \pi_i(p(x_1 + x_2) - c)x_i \). In this game the strategy of firm \( i \) is its choice of production level and the payoff to firm \( i \) is its profits.

Example 6.2.4 (Bertrand duopoly) Consider the same setup as in the Cournot game, but now suppose that the strategy of each player is to announce the price at which he would be willing to supply an arbitrary amount of the good in question. In this case the payoff function takes a radically different form. It is plausible to suppose that the consumers will only purchase from the firm with the lowest price, and that they will split evenly between the two firms if they charge
the same price. Letting $x(p)$ represent the market demand function and $c$ the marginal cost, this leads to a payoff to firm 1 of the form:

$$
\pi_1(p_1, p_2) = \begin{cases} 
(p_1 - c)x(p_1) & \text{if } p_1 < p_2 \\
(p_1 - c)x(p_1)/2 & \text{if } p_1 = p_2 \\
0 & \text{if } p_1 > p_2
\end{cases}
$$

This game has a similar structure to that of the Prisoner’s Dilemma. If both players cooperate, they can charge the monopoly price and each reap half of the monopoly profits. But the temptation is always there for one player to cut its price slightly and thereby capture the entire market for itself. But if both players cut price, then they are both worse off.

Note that the Cournot game and the Bertrand game have a radically different structure, even though they purport to model the same economic phenomena – a duopoly. In the Cournot game, the payoff to each firm is a continuous function of its strategic choice; in the Bertrand game, the payoffs are discontinuous functions of the strategies. As might be expected, this leads to quite different equilibria. Which of these models is reasonable? The answer is that it depends on what you are trying to model. In most economic modelling, there is an art to choosing a representation of the strategy choices of the game that captures an element of the real strategic iterations, while at the same time leaving the game simple enough to analyze.

### 6.3 Solution Concepts

#### 6.3.1 Mixed Strategies and Pure Strategies

In many games the nature of the strategic interaction suggests that a player wants to choose a strategy that is not predictable in advance by the other player. Consider, for example, the Matching Pennies game described above. Here it is clear that neither player wants the other player to be able to predict his choice.
accurately. Thus, it is natural to consider a random strategy of playing heads with some probability $p_h$ and tails with some probability $p_t$. Such a strategy is called a \textbf{mixed strategy}. Strategies in which some choice is made with probability 1 are called \textbf{pure strategies}.

If $R$ is the set of pure strategies available to Row, the set of mixed strategies open to Row will be the set of all probability distributions over $R$, where the probability of playing strategy $r$ in $R$ is $p_r$. Similarly, $p_c$, will be the probability that Column plays some strategy $c$. In order to solve the game, we want to find a set of mixed strategies $(p_r, p_c)$ that are, in some sense, in equilibrium. It may be that some of the equilibrium mixed strategies assign probability 1 to some choices, in which case they are interpreted as pure strategies.

The natural starting point in a search for a solution concept is standard decision theory: we assume that each player has some probability beliefs about the strategies that the other player might choose and that each player chooses the strategy that maximizes his expected payoff.

Suppose for example that the payoff to Row is $u_r(r, c)$ if Row plays $r$ and Column plays $c$. We assume that Row has a subjective probability distribution over Column’s choices which we denote by $(\pi_c)$; see Chapter 5 for the fundamentals of the idea of subjective probability. Here $\pi_c$ is supposed to indicate the probability, as envisioned by Row, that Column will make the choice $c$. Similarly, Column has some beliefs about Row’s behavior that we can denote by $(\pi_r)$.

We allow each player to play a mixed strategy and denote Row’s actual mixed strategy by $(p_r)$ and Column’s actual mixed strategy by $(p_c)$. Since Row makes his choice without knowing Column’s choice, Row’s probability that a particular outcome $(r, c)$ will occur is $p_r \pi_c$. This is simply the (objective) probability that Row plays $r$ times Row’s (subjective) probability that Column plays $c$. Hence, Row’s objective is to choose a probability distribution $(p_r)$ that maximizes

Row’s expected payoff $= \sum_r \sum_c p_r \pi_c u_r(r, c).

Column, on the other hand, wishes to maximize

Column’s expected payoff $= \sum_c \sum_r p_c \pi_r u_c(r, c).$
So far we have simply applied a standard decision-theoretic model to this game—each player wants to maximize his or her expected utility given his or her beliefs. Given my beliefs about what the other player might do, I choose the strategy that maximizes my expected utility.

6.3.2 Nash equilibrium

In the expected payoff formulas given at the end of the last subsection, Row’s behavior—how likely he is to play each of his strategies represented by the probability distribution \( (p_r) \) and Column’s beliefs about Row’s behavior are represented by the (subjective) probability distribution \( (\pi_r) \).

A natural consistency requirement is that each player’s belief about the other player’s choices coincides with the actual choices the other player intends to make. Expectations that are consistent with actual frequencies are sometimes called rational expectations. A Nash equilibrium is a certain kind of rational expectations equilibrium. More formally:

**Definition 6.3.1 (Nash Equilibrium in Mixed Strategies.)** A Nash equilibrium in mixed strategies consists of probability beliefs \( (\pi_r, \pi_c) \) over strategies, and probability of choosing strategies \( (p_r, p_c) \), such that:

1. the beliefs are correct: \( p_r = \pi_r \) and \( p_c = \pi_c \) for all \( r \) and \( c \); and,
2. each player is choosing \( (p_r) \) and \( (p_c) \) so as to maximize his expected utility given his beliefs.

In this definition a Nash equilibrium in mixed strategies is an equilibrium in actions and beliefs. In equilibrium each player correctly foresees how likely the other player is to make various choices, and the beliefs of the two players are mutually consistent.

A more conventional definition of a Nash equilibrium in mixed strategies is that it is a pair of mixed strategies \( (p_r, p_c) \) such that each agent’s choice maximizes his expected utility, given the strategy of the other agent. This is equivalent to the
definition we use, but it is misleading since the distinction between the beliefs of the agents and the actions of the agents is blurred. We’ve tried to be very careful in distinguishing these two concepts.

One particularly interesting special case of a Nash equilibrium in mixed strategies is a Nash equilibrium in pure strategies, which is simply a Nash equilibrium in mixed strategies in which the probability of playing a particular strategy is 1 for each player. That is:

**Definition 6.3.2 (Nash equilibrium in Pure Strategies.)** A Nash equilibrium in pure strategies is a pair \((r^*, c^*)\) such that \(u_r(r^*, c^*) \geq u_r(r, c^*)\) for all Row strategies \(r\), and \(u_c(r^*, c^*) \geq u_c(r^*, c)\) for all Column strategies \(c\).

A Nash equilibrium is a minimal consistency requirement to put on a pair of strategies: if Row believes that Column will play \(c^*\), then Row’s best reply is \(r^*\) and similarly for Column. No player would find it in his or her interest to deviate unilaterally from a Nash equilibrium strategy.

If a set of strategies is not a Nash equilibrium then at least one player is not consistently thinking through the behavior of the other player. That is, one of the players must expect the other player not to act in his own self-interest – contradicting the original hypothesis of the analysis.

An equilibrium concept is often thought of as a “rest point” of some adjustment process. One interpretation of Nash equilibrium is that it is the adjustment process of “thinking through” the incentives of the other player. Row might think: “If I think that Column is going to play some strategy \(c_1\) then the best response for me is to play \(r_1\). But if Column thinks that I will play \(r_1\), then the best thing for him to do is to play some other strategy \(c_2\). But if Column is going to play \(c_2\), then my best response is to play \(r_2\)...” and so on.

**Example 6.3.1 (Nash equilibrium of Battle of the Sexes)** The following game is known as the “Battle of the Sexes.” The story behind the game goes something like this. Rhonda Row and Calvin Column are discussing whether to take microeconomics or macroeconomics this semester. Rhonda gets utility 2 and Calvin
gets utility 1 if they both take micro; the payoffs are reversed if they both take macro. If they take different courses, they both get utility 0.

Let us calculate all the Nash equilibria of this game. First, we look for the Nash equilibria in pure strategies. This simply involves a systematic examination of the best responses to various strategy choices. Suppose that Column thinks that Row will play Top. Column gets 1 from playing Left and 0 from playing Right, so Left is Column’s best response to Row playing Top. On the other hand, if Column plays Left, then it is easy to see that it is optimal for Row to play Top. This line of reasoning shows that (Top, Left) is a Nash equilibrium. A similar argument shows that (Bottom, Right) is a Nash equilibrium.

<table>
<thead>
<tr>
<th></th>
<th>Left (micro)</th>
<th>Right (macro)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top</td>
<td>(2, 1)</td>
<td>(0, 0)</td>
</tr>
<tr>
<td>Bottom</td>
<td>(0, 0)</td>
<td>(1, 2)</td>
</tr>
</tbody>
</table>

Table 6.4: Battle of the Sexes

We can also solve this game systematically by writing the maximization problem that each agent has to solve and examining the first-order conditions. Let \((p_t, p_b)\) be the probabilities with which Row plays Top and Bottom, and define \((p_l, p_r)\) in a similar manner. Then Row’s problem is

\[
\max_{(p_l, p_r)} p_t [p_l 2 + p_r 0] + p_b [p_l 0 + p_r 1]
\]

such that

\[
\begin{align*}
    p_t &\geq 0 \\
    p_b &\geq 0.
\end{align*}
\]

Let \(\lambda, \mu_t,\) and \(\mu_b\) be the Kuhn-Tucker multipliers on the constraints, so that the Lagrangian takes the form:

\[
\mathcal{L} = 2p_t p_l + p_b p_r - \lambda (p_t + p_b - 1) - \mu_t p_t - \mu_b p_b.
\]
Differentiating with respect to $p_t$ and $p_b$, we see that the Kuhn-Tucker conditions for Row are

\begin{align*}
2p_t &= \lambda + \mu_t \\
p_r &= \lambda + \mu_b \quad (6.1)
\end{align*}

Since we already know the pure strategy solutions, we only consider the case where $p_t > 0$ and $p_b > 0$. The complementary slackness conditions then imply that $\mu_t = \mu_b = 0$. Using the fact that $p_t + p_r = 1$, we easily see that Row will find it optimal to play a mixed strategy when $p_l = 1/3$ and $p_r = 2/3$.

Following the same procedure for Column, we find that $p_t = 2/3$ and $p_b = 1/3$. The expected payoff to each player from this mixed strategy can be easily computed by plugging these numbers into the objective function. In this case the expected payoff is 2/3 to each player. Note that each player would prefer either of the pure strategy equilibria to the mixed strategy since the payoffs are higher for each player.

**Remark 6.3.1** One disadvantage of the notion of a mixed strategy is that it is sometimes difficult to give a behavioral interpretation to the idea of a mixed strategy although a mixed strategies are the only sensible equilibrium for some games such as Matching Pennies. For example, a duopoly game – mixed strategies seem unrealistic.

### 6.3.3 Dominant strategies

Let $r_1$ and $r_2$ be two of Row’s strategies. We say that $r_1$ strictly dominates $r_2$ for Row if the payoff from strategy $r_1$ is strictly larger than the payoff for $r_2$ no matter what choice Column makes. The strategy $r_1$ weakly dominates $r_2$ if the payoff from $r_1$ is at least as large for all choices Column might make and strictly larger for some choice.

A **dominant strategy equilibrium** is a choice of strategies by each player such that each strategy (weakly) dominates every other strategy available to
that player. One particularly interesting game that has a dominant strategy equilibrium is the Prisoner’s Dilemma in which the dominant strategy equilibrium is (confess, confess). If I believe that the other agent will not confess, then it is to my advantage to confess; and if I believe that the other agent will confess, it is still to my advantage to confess.

Clearly, a dominant strategy equilibrium is a Nash equilibrium, but not all Nash equilibria are dominant strategy equilibria. A dominant strategy equilibrium, should one exist, is an especially compelling solution to the game, since there is a unique optimal choice for each player.

### 6.4 Repeated games

In many cases, it is not appropriate to expect that the outcome of a repeated game with the same players as simply being a repetition of the one-shot game. This is because the strategy space of the repeated game is much larger: each player can determine his or her choice at some point as a function of the entire history of the game up until that point. Since my opponent can modify his behavior based on my history of choices, I must take this influence into account when making my own choices.

Let us analyze this in the context of the simple Prisoner’s Dilemma game described earlier. Here it is in the “long-run” interest of both players to try to get to the (Cooperate, Cooperate) solution. So it might be sensible for one player to try to “signal” to the other that he is willing to “be nice” and play cooperate on the first move of the game. It is in the short-run interest of the other player to Defect, of course, but is this really in his long-run interest? He might reason that if he defects, the other player may lose patience and simply play Defect himself from then on. Thus, the second player might lose in the long run from playing the short-run optimal strategy. What lies behind this reasoning is the fact that a move that I make now may have repercussions in the future the other player’s future choices may depend on my current choices.
Let us ask whether the strategy of (Cooperate, Cooperate) can be a Nash equilibrium of the repeated Prisoner’s Dilemma. First we consider the case of where each player knows that the game will be repeated a fixed number of times. Consider the reasoning of the players just before the last round of play. Each reasons that, at this point, they are playing a one-shot game. Since there is no future left on the last move, the standard logic for Nash equilibrium applies and both parties Defect.

Now consider the move before the last. Here it seems that it might pay each of the players to cooperate in order to signal that they are “nice guys” who will cooperate again in the next and final move. But we’ve just seen that when the next move comes around, each player will want to play Defect. Hence there is no advantage to cooperating on the next to the last move — as long as both players believe that the other player will Defect on the final move, there is no advantage to try to influence future behavior by being nice on the penultimate move. The same logic of backwards induction works for two moves before the end, and so on. **In a repeated Prisoner’s Dilemma with a finite number of repetitions, the Nash equilibrium is still to Defect in every round.**

The situation is quite different in a repeated game with an infinite number of repetitions. In this case, at each stage it is known that the game will be repeated at least one more time and therefore there will be some (potential) benefits to cooperation. Let’s see how this works in the case of the Prisoner’s Dilemma.

Consider a game that consists of an infinite number of repetitions of the Prisoner’s Dilemma described earlier. The strategies in this repeated game are sequences of functions that indicate whether each player will Cooperate or Defect at a particular stage as a function of the history of the game up to that stage. The payoffs in the repeated game are the discounted sums of the payoffs at each stage; that is, if a player gets a payoff at time \( t \) of \( u_t \), his payoff in the repeated game is taken to be \( \sum_{t=0}^{\infty} u_t/(1 + r)^t \), where \( r \) is the discount rate.

Now we show that **as long as the discount rate is not too high there exists a Nash equilibrium pair of strategies such that each player finds**
it in his interest to cooperate at each stage. In fact, it is easy to exhibit an explicit example of such strategies. Consider the following strategy: “Cooperate on the current move unless the other player defected on the last move. If the other player defected on the last move, then Defect forever.” This is sometimes called a punishment strategy, for obvious reasons: if a player defects, he will be punished forever with a low payoff.

To show that a pair of punishment strategies constitutes a Nash equilibrium, we simply have to show that if one player plays the punishment strategy the other player can do no better than playing the punishment strategy. Suppose that the players have cooperated up until move \(T\) and consider what would happen if a player decided to Defect on this move. Using the numbers from the Prisoner’s Dilemma example, he would get an immediate payoff of 4, but he would also doom himself to an infinite stream of payments of 1. The discounted value of such a stream of payments is \(1/r\), so his total expected payoff from Defecting is \(4 + 1/r\).

On the other hand, his expected payoff from continuing to cooperate is \(3 + 3/r\). Continuing to cooperate is preferred as long as \(3 + 3/r > 4 + 1/r\), which reduces to requiring that \(r < 2\). As long as this condition is satisfied, the punishment strategy forms a Nash equilibrium: if one party plays the punishment strategy, the other party will also want to play it, and neither party can gain by unilaterally deviating from this choice.

This construction is quite robust. Essentially the same argument works for any payoffs that exceed the payoffs from (Defect, Defect). A famous result known as the Folk Theorem asserts precisely this: in a repeated Prisoner’s Dilemma any payoff larger than the payoff received if both parties consistently defect can be supported as a Nash equilibrium.

**Example 6.4.1 (Maintaining a Cartel)** Consider a simple repeated duopoly which yields profits \((\pi_c, \pi_c)\) if both firms choose to play a Cournot game and \((\pi_j, \pi_j)\) if both firms produce the level of output that maximizes their joint profits – that is, they act as a cartel. It is well-known that the levels of output that
maximize joint profits are typically not Nash equilibria in a single-period game — each producer has an incentive to dump extra output if he believes that the other producer will keep his output constant. However, as long as the discount rate is not too high, the joint profit-maximizing solution will be a Nash equilibrium of the repeated game. The appropriate punishment strategy is for each firm to produce the cartel output unless the other firm deviates, in which case it will produce the Cournot output forever. An argument similar to the Prisoner’s Dilemma argument shows that this is a Nash equilibrium.

6.5 Refinements of Nash equilibrium

The Nash equilibrium concept seems like a reasonable definition of an equilibrium of a game. As with any equilibrium concept, there are two questions of immediate interest: 1) will a Nash equilibrium generally exist; and 2) will the Nash equilibrium be unique?

Existence, luckily, is not a problem. Nash (1950) showed that with a finite number of agents and a finite number of pure strategies, an equilibrium will always exist. It may, of course, be an equilibrium involving mixed strategies. We will shown in Chapter 10 that it always exists a pure strategy Nash equilibrium if the strategy space is a compact and convex set and payoffs functions are continuous and quasi-concave.

Uniqueness, however, is very unlikely to occur in general. We have already seen that there may be several Nash equilibria to a game. Game theorists have invested a substantial amount of effort into discovering further criteria that can be used to choose among Nash equilibria. These criteria are known as refinements of the concept of Nash equilibrium, and we will investigate a few of them below.

6.5.1 Elimination of dominated strategies

When there is no dominant strategy equilibrium, we have to resort to the idea of a Nash equilibrium. But typically there will be more than one Nash equilibrium.
Our problem then is to try to eliminate some of the Nash equilibria as being “unreasonable.”

One sensible belief to have about players’ behavior is that it would be unreasonable for them to play strategies that are dominated by other strategies. This suggests that when given a game, we should first eliminate all strategies that are dominated and then calculate the Nash equilibria of the remaining game. This procedure is called elimination of dominated strategies; it can sometimes result in a significant reduction in the number of Nash equilibria.

For example, consider the game given in Table 6.5. Note that there are two pure strategy Nash equilibria, (Top, Left) and (Bottom, Right). However, the strategy Right weakly dominates the strategy Left for the Column player. If the Row agent assumes that Column will never play his dominated strategy, the only equilibrium for the game is (Bottom, Right).

Elimination of strictly dominated strategies is generally agreed to be an acceptable procedure to simplify the analysis of a game. Elimination of weakly dominated strategies is more problematic; there are examples in which eliminating weakly dominated strategies appears to change the strategic nature of the game in a significant way.

### 6.5.2 Sequential Games and Subgame Perfect Equilibrium

The games described so far in this chapter have all had a very simple dynamic structure: they were either one-shot games or a repeated sequence of one-shot
games. They also had a very simple information structure: each player in the game knew the other player’s payoffs and available strategies, but did not know in advance the other player’s actual choice of strategies. Another way to say this is that up until now we have restricted our attention to games with simultaneous moves.

But many games of interest do not have this structure. In many situations at least some of the choices are made sequentially, and one player may know the other player’s choice before he has to make his own choice. The analysis of such games is of considerable interest to economists since many economic games have this structure: a monopolist gets to observe consumer demand behavior before it produces output, or a duopolist may observe his opponent’s capital investment before making its own output decisions, etc. The analysis of such games requires some new concepts.

<table>
<thead>
<tr>
<th>Column</th>
<th>Left</th>
<th>Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row</td>
<td>Top</td>
<td>(1,9)</td>
</tr>
<tr>
<td></td>
<td>Bottom</td>
<td>(0,0)</td>
</tr>
</tbody>
</table>

Table 6.6: The Payoff Matrix of a Simultaneous-Move Game

Consider for example, the simple game depicted in Table 6.6. It is easy to verify that there are two pure strategy Nash equilibria in this game, (Top, Left) and (Bottom, Right). Implicit in this description of this game is the idea that both players make their choices simultaneously, without knowledge of the choice that the other player has made. But suppose that we consider the game in which Row must choose first, and Column gets to make his choice after observing Row’s behavior.

In order to describe such a sequential game it is necessary to introduce a new tool, the game tree. This is simply a diagram that indicates the choices that each player can make at each point in time. The payoffs to each player are indicated at
the “leaves” of the tree, as in Figure 6.1. This game tree is part of a description of the game in extensive form.

The nice thing about the tree diagram of the game is that it indicates the dynamic structure of the game — that some choices are made before others. A choice in the game corresponds to the choice of a branch of the tree. Once a choice has been made, the players are in a subgame consisting of the strategies and payoffs available to them from then on.

It is straightforward to calculate the Nash equilibria in each of the possible subgames, particularly in this case since the example is so simple. If Row chooses top, he effectively chooses the very simple subgame in which Column has the only remaining move. Column is indifferent between his two moves, so that Row will definitely end up with a payoff of 1 if he chooses Top.

If Row chooses Bottom, it will be optimal for Column to choose Right, which gives a payoff of 2 to Row. Since 2 is larger than 1, Row is clearly better off choosing Bottom than Top. Hence the sensible equilibrium for this game is (Bottom, Right). This is, of course, one of the Nash equilibria in the simultaneous-move game. If Column announces that he will choose Right, then Row’s optimal

Figure 6.1: A game tree. This illustrates the payoffs to the previous game where Row gets to move first.
response is Bottom, and if Row announces that he will choose Bottom then Column’s optimal response is Right.

But what happened to the other equilibrium, (Top, Left)? If Row believes that Column will choose Left, then his optimal choice is certainly to choose Top. But why should Row believe that Column will actually choose Left? Once Row chooses Bottom, the optimal choice in the resulting subgame is for Column to choose Right. A choice of Left at this point is not an equilibrium choice in the relevant subgame.

In this example, only one of the two Nash equilibria satisfies the condition that it is not only an overall equilibrium, but also an equilibrium in each of the subgames. A Nash equilibrium with this property is known as a subgame perfect equilibrium.

It is quite easy to calculate subgame-perfect equilibria, at least in the kind of games that we have been examining. One simply does a “backwards induction” starting at the last move of the game. The player who has the last move has a simple optimization problem, with no strategic ramifications, so this is an easy problem to solve. The player who makes the second to the last move can look ahead to see how the player with the last move will respond to his choices, and so on. The mode of analysis is similar to that of dynamic programming. Once the game has been understood through this backwards induction, the agents play it going forwards.

The extensive form of the game is also capable of modelling situations where some of the moves are sequential and some are simultaneous. The necessary concept is that of an information set. The information set of an agent is the set of all nodes of the tree that cannot be differentiated by the agent. For example, the simultaneous-move game depicted at the beginning of this section can be represented by the game tree in Figure 6.2. In this figure, the shaded area indicates that Column cannot differentiate which of these decisions Row made at the time when Column must make his own decision. Hence, it is just as if the choices are made simultaneously.
Figure 6.2: Information set. This is the extensive form to the original simultaneous-move game. The shaded information set indicates that column is not aware of which choice Row made when he makes his own decision.

Thus the extensive form of a game can be used to model everything in the strategic form plus information about the sequence of choices and information sets. In this sense the extensive form is a more powerful concept than the strategic form, since it contains more detailed information about the strategic interactions of the agents. It is the presence of this additional information that helps to eliminate some of the Nash equilibria as “unreasonable.”

Example 6.5.1 (A Simple Bargaining Model) Two players, A and B, have $1 to divide between them. They agree to spend at most three days negotiating over the division. The first day, A will make an offer, B either accepts or comes back with a counteroffer the next day, and on the third day A gets to make one final offer. If they cannot reach an agreement in three days, both players get zero.

A and B differ in their degree of impatience: A discounts payoffs in the future at a rate of $\alpha$ per day, and B discounts payoffs at a rate of $\beta$ per day. Finally, we assume that if a player is indifferent between two offers, he will accept the one that is most preferred by his opponent. This idea is that the opponent could offer some arbitrarily small amount that would make the player strictly prefer one
choice, and that this assumption allows us to approximate such an “arbitrarily small amount” by zero. It turns out that there is a unique subgame perfect equilibrium of this bargaining game.

As suggested above, we start our analysis at the end of the game, right before the last day. At this point A can make a take-it-or-leave-it offer to B. Clearly, the optimal thing for A to do at this point is to offer B the smallest possible amount that he would accept, which, by assumption, is zero. So if the game actually lasts three days, A would get 1 and B would get zero (i.e., an arbitrarily small amount).

Now go back to the previous move, when B gets to propose a division. At this point B should realize that A can guarantee himself 1 on the next move by simply rejecting B’s offer. A dollar next period is worth $\alpha$ to A this period, so any offer less than $\alpha$ would be sure to be rejected. B certainly prefers $1 - \alpha$ now to zero next period, so he should rationally offer $\alpha$ to A, which A will then accept. So if the game ends on the second move, A gets $\alpha$ and B gets $1 - \alpha$.

Now move to the first day. At this point A gets to make the offer and he realizes that B can get $1 - \alpha$ if he simply waits until the second day. Hence A must offer a payoff that has at least this present value to B in order to avoid delay. Thus he offers $\beta(1 - \alpha)$ to B. B finds this (just) acceptable and the game ends. The final outcome is that the game ends on the first move with A receiving $1 - \beta(1 - \alpha)$ and B receiving $\beta(1 - \alpha)$.

Figure 6.3 illustrates this process for the case where $\alpha = \beta < 1$. The outermost diagonal line shows the possible payoff patterns on the first day, namely all payoffs of the form $x_A + x_B = 1$. The next diagonal line moving towards the origin shows the present value of the payoffs if the game ends in the second period: $x_A + x_B = \alpha$. The diagonal line closest to the origin shows the present value of the payoffs if the game ends in the third period; this equation for this line is $x_A + x_B = \alpha^2$. The right angled path depicts the minimum acceptable divisions each period, leading up to the final subgame perfect equilibrium. Figure 6.3B shows how the same process looks with more stages in the negotiation.
Figure 6.3: A bargaining game. The heavy line connects together the equilibrium outcomes in the subgames. The point on the outer-most line is the subgame-perfect equilibrium.

It is natural to let the horizon go to infinity and ask what happens in the infinite game. It turns out that the subgame perfect equilibrium division is payoff to $A = \frac{1-\beta}{1-\alpha \beta}$ and payoff to $B = \frac{\beta(1-\alpha)}{1-\alpha \beta}$. Note that if $a = 1$ and $\beta < 1$, then player A receives the entire payoff, in accord with the principal expressed in the Gospels: “Let patience have her [subgame] perfect work.” (James 1:4).

6.5.3 Repeated games and subgame perfection

The idea of subgame perfection eliminates Nash equilibria that involve players threatening actions that are not credible - i.e., they are not in the interest of the players to carry out. For example, the Punishment Strategy described earlier is not a subgame perfect equilibrium. If one player actually deviates from the (Cooperate, Cooperate) path, then it is not necessarily in the interest of the other player to actually defect forever in response. It may seem reasonable to punish the other player for defection to some degree, but punishing forever seems extreme.

A somewhat less harsh strategy is known as Tit-for-Tat: start out cooperating on the first play and on subsequent plays do whatever your opponent did on the previous play. In this strategy, a player is punished for defection, but he is only punished once. In this sense Tit-for-Tat is a “forgiving” strategy.
Although the punishment strategy is not subgame perfect for the repeated Prisoner’s Dilemma, there are strategies that can support the cooperative solution that are subgame perfect. These strategies are not easy to describe, but they have the character of the West Point honor code: each player agrees to punish the other for defecting, and also punish the other for failing to punish another player for defecting, and so on. The fact that you will be punished if you don’t punish a defector is what makes it subgame perfect to carry out the punishments.

Unfortunately, the same sort of strategies can support many other outcomes in the repeated Prisoner’s Dilemma. The Folk Theorem asserts that essentially all distributions of utility in a repeated one-shot game can be equilibria of the repeated game.

This excess supply of equilibria is troubling. In general, the larger the strategy space, the more equilibria there will be, since there will be more ways for players to “threaten” retaliation for defecting from a given set of strategies. In order to eliminate the ”undesirable” equilibria, we need to find some criterion for eliminating strategies. A natural criterion is to eliminate strategies that are “too complex.” Although some progress has been made in this direction, the idea of complexity is an elusive one, and it has been hard to come up with an entirely satisfactory definition.

6.6 Games with incomplete information

6.6.1 Bayes-Nash Equilibrium

Up until now we have been investigating games of complete information. In particular, each agent has been assumed to know the payoffs of the other player, and each player knows that the other agent knows this, etc. In many situations, this is not an appropriate assumption. If one agent doesn’t know the payoffs of the other agent, then the Nash equilibrium doesn’t make much sense. However, there is a way of looking at games of incomplete information due to Harsanyi
(1967) that allows for a systematic analysis of their properties.

The key to the Harsanyi approach is to subsume all of the uncertainty that one agent may have about another into a variable known as the agent’s type. For example, one agent may be uncertain about another agent’s valuation of some good, about his or her risk aversion and so on. Each type of player is regarded as a different player and each agent has some prior probability distribution defined over the different types of agents.

A Bayes-Nash equilibrium of this game is then a set of strategies for each type of player that maximizes the expected value of each type of player, given the strategies pursued by the other players. This is essentially the same definition as in the definition of Nash equilibrium, except for the additional uncertainty involved about the type of the other player. Each player knows that the other player is chosen from a set of possible types, but doesn’t know exactly which one he is playing. Note in order to have a complete description of an equilibrium we must have a list of strategies for all types of players, not just the actual types in a particular situation, since each individual player doesn’t know the actual types of the other players and has to consider all possibilities.

In a simultaneous-move game, this definition of equilibrium is adequate. In a sequential game it is reasonable to allow the players to update their beliefs about the types of the other players based on the actions they have observed. Normally, we assume that this updating is done in a manner consistent with Bayes’ rule. Thus, if one player observes the other choosing some strategy $s$, the first player should revise his beliefs about what type the other player is by determining how likely it is that $s$ would be chosen by the various types.

Example 6.6.1 (A sealed-bid auction) Consider a simple sealed-bid auction for an item in which there are two bidders. Each player makes an independent bid without knowing the other player’s bid and the item will be awarded to the person with the highest bid. Each bidder knows his own valuation of the item being auctioned, $v$, but neither knows the other’s valuation. However, each player believes that the other person’s valuation of the item is uniformly distributed
between 0 and 1. (And each person knows that each person believes this, etc.)

In this game, the type of the player is simply his valuation. Therefore, a Bayes-Nash equilibrium to this game will be a function, $b(v)$, that indicates the optimal bid, $b$, for a player of type $v$. Given the symmetric nature of the game, we look for an equilibrium where each player follows an identical strategy.

It is natural to guess that the function $b(v)$ is strictly increasing; i.e., higher valuations lead to higher bids. Therefore, we can let $V(b)$ be its inverse function so that $V(b)$ gives us the valuation of someone who bids $b$. When one player bids some particular $b$, his probability of winning is the probability that the other player’s bid is less than $b$. But this is simply the probability that the other player’s valuation is less than $V(b)$. Since $v$ is uniformly distributed between 0 and 1, the probability that the other player’s valuation is less than $V(b)$ is $V(b)$.

Hence, if a player bids $b$ when his valuation is $v$, his expected payoff is

$$(v - b)V(b) + 0[1 - V(b)].$$

The first term is the expected consumer’s surplus if he has the highest bid; the second term is the zero surplus he receives if he is outbid. The optimal bid must maximize this expression, so

$$(v - b)V'(b) - V(b) = 0.$$ 

For each value of $v$, this equation determines the optimal bid for the player as a function of $v$. Since $V(b)$ is by hypothesis the function that describes the relationship between the optimal bid and the valuation, we must have

$$(V(b) - b)V'(b) \equiv V(b)$$

for all $b$.

The solution to this differential equation is

$$V(b) = b + \sqrt{b^2 + 2C},$$

285
where \( C \) is a constant of integration. In order to determine this constant of integration we note that when \( v = 0 \) we must have \( b = 0 \), since the optimal bid when the valuation is zero must be 0. Substituting this into the solution to the differential equation gives us

\[
0 = 0 + \sqrt{2C},
\]

which implies \( C = 0 \). It follows that \( V(b) = 2b \), or \( b = v/2 \), is a Bayes-Nash equilibrium for the simple auction. That is, it is a Bayes-Nash equilibrium for each player to bid half of his valuation.

The way that we arrived at the solution to this game is reasonably standard. Essentially, we guessed that the optimal bidding function was invertible and then derived the differential equation that it must satisfy. As it turned out, the resulting bid function had the desired property. One weakness of this approach is that it only exhibits one particular equilibrium to the Bayesian game — there could in principle be many others.

As it happens, in this particular game, the solution that we calculated is unique, but this need not happen in general. In particular, in games of incomplete information it may well pay for some players to try to hide their true type. For example, one type may try to play the same strategy as some other type. In this situation the function relating type to strategy is not invertible and the analysis is much more complex.

### 6.6.2 Discussion of Bayesian-Nash equilibrium

The idea of Bayesian-Nash equilibrium is an ingenious one, but perhaps it is too ingenious. The problem is that the reasoning involved in computing Bayesian-Nash equilibria is often very involved. Although it is perhaps not unreasonable that purely rational players would play according to the Bayesian-Nash theory, there is considerable doubt about whether real players are able to make the necessary calculations.
In addition, there is a problem with the predictions of the model. The choice that each player makes depends crucially on his beliefs about the distribution of various types in the population. Different beliefs about the frequency of different types leads to different optimal behavior. Since we generally don’t observe players beliefs about the prevalence of various types of players, we typically won’t be able to check the predictions of the model. Ledyard (1986) has shown that essentially any pattern of play is a Bayesian-Nash equilibrium for some pattern of beliefs.

Nash equilibrium, in its original formulation, puts a consistency requirement on the beliefs of the agents – only those beliefs compatible with maximizing behavior were allowed. But as soon as we allow there to be many types of players with different utility functions, this idea loses much of its force. Nearly any pattern of behavior can be consistent with some pattern of beliefs.

6.7 Reference

Books and Monographs:


**Papers:**


Chapter 7

Repeated Game Theory and Reputation

To be written.

7.1 Reference

Books and Monographs:

Chan, Jimmy (2012). *Lecture Notes*, SHUFE.


Papers:


Chapter 8

Cooperative Game Theory

8.1 Reference


**Papers:**


Chapter 9

Market Theory

9.1 Introduction

In previous chapters, we studied the behavior of individual consumers and firms, describing optimal behavior when markets prices were fixed and beyond the agent’s control. Here, we explore the consequences of that behavior when consumers and firms come together in markets. We will consider equilibrium price and quantity determination in a single market or group of closed related markets by the actions of the individual agents for different markets structures. This equilibrium analysis is called a partial equilibrium analysis because it focuses on a single market or group of closed related markets, implicitly assuming that changes in the markets under consideration do not change prices of other goods and upset the equilibrium that holds in those markets. We will treat all markets simultaneously in the general equilibrium theory.

We will concentrate on modelling the market behavior of the firm. How do firms determine the price at which they will sell their output or the prices at which they are willing to purchase inputs? We will see that in certain situations the “price-taking behavior” might be a reasonable approximation to optimal behavior, but in other situations we will have to explore models of the price-setting process. We will first consider the ideal (benchmark) case of perfect competi-
tion. We then turn to the study of settings in which some agents have market power. These settings include markets structures of pure monopoly, monopolistic competition, oligopoly, and monopsony.

The key insight of Adam Smith’s Wealth of Nations is simple: if an exchange between two parties is voluntary, it will not take place unless both believe they will benefit from it. How is this also true for any number of parties and for production case? The price system is the mechanism that performs this task very well without central direction.

Prices perform three functions in organizing economic activities in a free market economy:

(1) They transmit information about production and consumption. The price system transmit only the important information and only to the people who need to know. Thus, it transmits information in an efficiently way.

(2) They provide right incentives. One of the beauties of a free price system is that the prices that bring the information also provide an incentive to react on the information not only about the demand for output but also about the most efficient way to produce a product. They provide incentives to adopt those methods of production that are least costly and thereby use available resources for the most highly valued purposes.

(3) They determine the distribution of income. They determine who gets how much of the product. In general, one cannot use prices to transmit information and provide an incentive to act that information without using prices to affect the distribution of income. If what a person gets does not depend on the price he receives for the services of this resources, what incentive does he have to seek out information on prices or to act on the basis of that information?
9.2 Perfect Competition

Let us start to consider the case of pure competition in which there are a large number of independent sellers of some uniform product. In this situation, when each firm sets the price in which it sells its output, it will have to take into account not only the behavior of the consumers but also the behavior of the other producers.

The competitive markets are based on the following assumptions:

1. Large number of buyers and sellers — price-taking behavior
2. Unrestricted mobility of resources among industries: no artificial barrier or impediment to entry or to exit from market.
3. Homogeneous product: All the firms in an industry produce an identical production in the consumers’ eyes.
4. Passion of all relevant information (all relevant information are common knowledge): Firms and consumers have all the information necessary to make the correct economic decisions.

9.2.1 The Competitive Firm

A competitive firm is one that takes the market price of output as being given. Let $\bar{p}$ be the market price. Then the demand curve facing an ideal competitive firm takes the form

$$D(p) = \begin{cases} 
0 & \text{if } p > \bar{p} \\
\text{any amount} & \text{if } p = \bar{p} \\
\infty & \text{if } p < \bar{p}
\end{cases}$$

A competitive firm is free to set whatever price it wants and produce whatever quantity it is able to produce. However, if a firm is in a competitive market, it is clear that each firm that sells the product must sell it for the same price: for if any firm attempted to set its price at a level greater than the market price, it would immediately lose all of its customers. If any firm set its price at a level below
the market price, all of the consumers would immediately come to it. Thus, each firm must take the market price as given, exogenous variable when it determines its supply decision.

9.2.2 Competitive Firm’s Short-Run Supply Function

Since the competitive firm must take the market price as given, its profit maximization problem is simple. The firm only needs to choose output level \( y \) so as to solve

\[
\max_y p y - c(y)
\]

where \( y \) is the output produced by the firm, \( p \) is the price of the product, and \( c(y) \) is the cost function of production.

The first-order condition (in short, FOC) for interior solution gives:

\[
p = \frac{d}{dy} c(y) \equiv MC(y).
\]

The first order condition becomes a sufficient condition if the second-order condition (in short, SOC) is satisfied

\[
c''(y) > 0.
\]

Taken together, these two conditions determine the supply function of a competitive firm: at any price \( p \), the firm will supply an amount of output \( y(p) \) such that \( p = c'(y(p)) \). By \( p = c'(y(p)) \), we have

\[
1 = c''(y(p)) y'(p)
\]

and thus

\[
y'(p) > 0,
\]

which means the law of supply holds.

Recall that \( p = c'(y^*) \) is the first-order condition characterizing the optimum only if \( y^* > 0 \), that is, only if \( y^* \) is an interior optimum. It could happen that at
a low price a firm might very well decide to go out of business. For the short-run (in short, SR) case, 

$$c(y) = c_v(y) + F$$  
\[ (9.6) \]

The firm should produce if 

$$py(p) - c_v(y) - F \geq -F,$$  
\[ (9.7) \]

and thus we have 

$$p \geq \frac{c_v(y(p))}{y(p)} \equiv AVC.$$  
\[ (9.8) \]

That is, the necessary condition for the firm to produce a positive amount of output is that the price is greater than or equal to the average variable cost.

Thus, the supply curve for the competitive firm is in general given by: 

$$p = c'(y) \text{ if } p \geq \frac{c_v(y(p))}{y(p)} \text{ and } y = 0 \text{ if } p \leq \frac{c_v(y(p))}{y(p)}.$$  

That is, the supply curve coincides with the upward sloping portion of the marginal cost curve as long as the price covers average variable cost, and the supply curve is zero if price is less than average variable cost.

Figure 9.1: Firm’s supply curve, and AC, AVC, and MC Curves.

Suppose that we have \( j \) firms in the market. (For the competitive model to make sense, \( j \) should be rather large.) The industry supply function is simply
the sum of all individuals’ supply functions so that it is given by

\[ \hat{y}(p) = \sum_{j=1}^{J} y_j(p) \]  

(9.9)

where \( y_j(p) \) is the supply function of firm \( j \) for \( j = 1, \ldots, J \). Since each firm chooses a level of output where price equals marginal cost, each firm that produces a positive amount of output must have the same marginal cost. The industry supply function measures the relationship between industry output and the common cost of producing this output.

The aggregate (industry) demand function measures the total output demanded at any price which is given by

\[ \hat{x}(p) = \sum_{i=1}^{n} x_i(p) \]  

(9.10)

where \( x_i(p) \) is the demand function of consumer \( i \) for \( i = 1, \ldots, n \).

### 9.2.3 Single Commodity Market Equilibrium

How is the market price determined for a commodity? It is determined by the requirement that the total amount of output that the firms wish to supply will be equal to the total amount of output that the consumers wish to demand. Formerly, we have

A **partial equilibrium price** \( p^* \) is a price where the aggregate quantity demanded equals the aggregate quantity supplied. That is, it is the solution of the following equation:

\[ \sum_{i=1}^{n} x_i(p) = \sum_{j=1}^{J} y_j(p) \]  

(9.11)

Once this equilibrium price is determined, we can go back to look at the individual supply schedules of each firm and determine the firms level of output, its revenue, and its profits. In Figure 9.2, we have depicted cost curves for three firms. The first has positive profits, the second has zero profits, and the third has negative profits. Even though the third firm has negative profits, it may make sense for it to continue to produce as long as its revenues cover its variable costs.
Example 9.2.1 \( \hat{x}(p) = a - bp \) and \( c(y) = y^2 + 1 \). Since \( MC(y) = 2y \), we have

\[
y = \frac{p}{2}
\]

(9.12)

and thus the industry supply function is

\[
\hat{y}(p) = \frac{Jp}{2}
\]

(9.13)

Setting \( a - bp = \frac{Jp}{2} \), we have

\[
p^* = \frac{a}{b + J/2}.
\]

(9.14)

Now for general case of \( D(p) \) and \( S(p) \), what happens about the equilibrium price if the number of firms increases? From

\[
D(p(J)) = Jy(p(J))
\]

we have

\[
D'(p(J))p'(J) = y(p) + Jy'(p(J))p'(J)
\]

and thus

\[
p'(J) = \frac{y(p)}{X'(p) - Jy'(p)} < 0,
\]

which means the equilibrium price decreases when the number of firms increases.

9.2.4 Competitive Market and Increasing Returns to Scale

In progress.
9.2.5 Competitive in the Long Run

The long-run behavior of a competitive industry is determined by two sorts of effects. The first effect is the free entry and exit phenomena so that the profits made by all firms are zero. If some firm is making negative profits, we would expect that it would eventually have to change its behavior. Conversely, if a firm is making positive profits we would expect that this would eventually encourage entry to the industry. If we have an industry characterized by free entry and exit, it is clear that in the long run all firms must make the same level of profits. As a result, every firm makes zero profit at the long-run competitive equilibrium.

![Figure 9.3: Long-run adjustment with constant costs.](image)

The second influence on the long-run behavior of a competitive industry is that of technological adjustment. In the long run, firms will attempt to adjust their fixed factors so as to produce the equilibrium level of output in the cheapest way. Suppose for example we have a competitive firm with a long-run constant returns-to-scale technology that is operating at the position illustrated in Figure 9.3. Then in the long run it clearly pays the firm to change its fixed factors so as to operate at a point of minimum average cost. But, if every firm tries to do this, the equilibrium price will certainly change.

In the model of entry or exit, the equilibrium number of firms is the largest number of firms that can break even so the price must be chosen to minimum.
price.

**Example 9.2.2** \( c(y) = y^2 + 1 \). The break-even level of output can be found by setting

\[
AC(y) = MC(y)
\]

so that \( y = 1 \), and \( p = MC(y) = 2 \).

Suppose the demand is linear: \( X(p) = a - bp \). Then, the equilibrium price will be the smallest \( p^* \) that satisfies the conditions

\[
p^* = \frac{a}{b + J/2} \geq 2.
\]

As \( J \) increases, the equilibrium price must be closer and closer to 2.

### 9.2.6 Social Welfare under Perfect Competition

### 9.3 Pure Monopoly

#### 9.3.1 Monopoly in Product Market

**Profit Maximization Problem of Monopolist**

At the opposite pole from pure competition we have the case of pure monopoly. Here instead of a large number of independent sellers of some uniform product, we have only one seller. A monopolistic firm must make two sorts of decisions: how much output it should produce, and at what price it should sell this output. Of course, it cannot make these decisions unilaterally. The amount of output that the firm is able to sell will depend on the price that it sets. We summarize this relationship between demand and price in a market demand function for output, \( y(p) \). The market demand function tells how much output consumers will demand as a function of the price that the monopolist charges. It is often more convenient to consider the inverse demand function \( p(y) \), which indicates the price that consumers are willing to pay for \( y \) amount of output. We have provided the conditions under which the inverse demand function exists in Chapter 3. The
revenue that the firm receives will depend on the amount of output it chooses to supply. We write this revenue function as \( R(y) = p(y)y \).

The cost function of the firm also depends on the amount of output produced. This relationship was extensively studied in Chapter 4. Here we have the factor prices as constant so that the conditional cost function can be written only as a function of the level of output of the firm.

The profit maximization problem of the firm can then be written as:

\[
\max \ R(y) - c(y) = \max \ p(y)y - c(y)
\]

The first-order conditions for profit maximization are that marginal revenue equals marginal cost, or

\[
p(y^*) + p'(y^*)y^* = c'(y^*)
\]

The intuition behind this condition is fairly clear. If the monopolist considers producing one extra unit of output he will increase his revenue by \( p(y^*) \) dollars in the first instance. But this increased level of output will force the price down by \( p'(y^*) \), and he will lose this much revenue on unit of output sold. The sum of these two effects gives the marginal revenue. If the marginal revenue exceeds the marginal cost of production the monopolist will expand output. The expansion stops when the marginal revenue and the marginal cost balance out.

The first-order conditions for profit maximization can be expressed in a slightly different manner through the use of the price elasticity of demand. The price elasticity of demand is given by:

\[
\epsilon(y) = \frac{p}{y(p)} \frac{dy(p)}{dp}
\]

Note that this is always a negative number since \( dy(p)/dp \) is negative.

Simple algebra shows that the marginal revenue equals marginal cost condition can be written as:

\[
p(y^*) \left[ 1 + \frac{y^*}{p(y^*)} \frac{dp(y^*)}{dy} \right] = p(y^*) \left[ 1 + \frac{1}{\epsilon(y^*)} \right] = c'(y^*)
\]
that is, the price charged by a monopolist is a markup over marginal cost, with the level of the markup being given as a function of the price elasticity of demand.

There is also a nice graphical illustration of the profit maximization condition. Suppose for simplicity that we have a linear inverse demand curve: \( p(y) = a - by \). Then the revenue function is \( R(y) = ay - by^2 \), and the marginal revenue function is just \( R'(y) = a - 2by \). The marginal revenue curve has the same vertical intercept as the demand curve but is twice as steep. We have illustrated these two curves in Figure 9.4, along with the average cost and marginal cost curves of the firm in question.

![Figure 9.4: Determination of profit-maximizing monopolist’s price and output.](image)

The optimal level of output is located where the marginal revenue and the marginal cost curves intersect. This optimal level of output sells at a price \( p(y^*) \) so the monopolist gets an optimal revenue of \( p(y^*)y^* \). The cost of producing \( y^* \) is just \( y^* \) times the average cost of production at that level of output. The difference between these two areas gives us a measure of the monopolist’s profits.
Inefficiency of Monopoly

We say that a situation is Pareto efficient if there is no way to make one agent better off and the others are not worse off. Pareto efficiency will be a major theme in the discussion of welfare economics, but we can give a nice illustration of the concept here.

Let us consider the typical monopolistic configuration illustrated in Figure 9.5. It turns out a monopolist always operates in a Pareto inefficient manner. This means that there is some way to make the monopolist is better off and his customers are not worse off.

To see this let us think of the monopolist in Figure 9.5 after he has sold $y_m$ of output at the price $p_m$, and received his monopolist profit. Suppose that the monopolist were to produce a small unit of output $\Delta y$ more and offer to the public. How much would people be willing to pay for this extra unit? Clearly they would be willing to pay a price $p(y_m + \Delta y)$ dollars. How much would it cost to produce this extra output? Clearly, just the marginal cost $MC(y_m + \Delta y)$. Under this rearrangement the consumers are at least not worse off since they are freely purchasing the extra unit of output, and the monopolist is better off since he can sell some extra units at a price that exceeds the cost of its production. Here we are allowing the monopolist to discriminate in his pricing: he first sells

![Figure 9.5: Monopoly results in Pareto inefficient outcome.](image-url)
$y_m$ and then sells more output at some other price.

How long can this process be continued? Once the competitive level of output is reached, no further improvements are possible. The competitive level of price and output is Pareto efficient for this industry. We will investigate the concept of Pareto efficiency in general equilibrium theory.

**Monopoly in the Long Run**

We have seen how the long-run and the short-run behavior of a competitive industry may differ because of changes in technology and entry. There are similar effects in a monopolized industry. The technological effect is the simplest: the monopolist will choose the level of his fixed factors so as to maximize his long-run profits. Thus, he will operate where marginal revenue equals long-run marginal cost, and that is all that needs to be said.

The entry effect is a bit more subtle. Presumably, if the monopolist is earning positive profits, other firms would like to enter the industry. If the monopolist is to remain a monopolist, there must be some sort of barrier to entry so that a monopolist may make **positive profits** even in the long-run.

These barriers to entry may be of a legal sort, but often they are due to the fact that the monopolist owns some unique factor of production. For example, a firm might own a patent on a certain product, or might own a certain secret process for producing some item. If the monopoly power of the firm is due to a unique factor we have to be careful about we measure profits.

**9.3.2 Monopsony: Monopoly in Input Market**

To be revised

In our discussion up until now we have generally assumed that we have been dealing with the market for an output good. Output markets can be classified as “competitive” or “monopolistic” depending on whether firms take the market price as given or whether firms take the demand behavior of consumers as given.
There is similar classification for inputs markets. If firms take the factor prices as given, then we have competitive factor markets. If instead there is only one firm which demands some factor of production and it takes the supply behavior of its suppliers into account, then we say we have a monopsonistic factor market. The behavior of a monopsonist is analogous to that of a monopolist. Let us consider a simple example of a firm that is a competitor in its output market but is the sole purchaser of some input good. Let \( w(x) \) be the (inverse) supply curve of this factor of production. Then the profit maximization problem is:

\[
\max p f(x) - w(x)x
\]

The first-order condition is:

\[
p f'(x^*) - w(x^*) - w'(x^*)x^* = 0
\]

This just says that the marginal revenue product of an additional unit of the input good should be equal to its marginal cost. Another way to write the condition is:

\[
p \frac{\partial f(x^*)}{\partial x} = w(x^*)[1 + 1/\epsilon]
\]

where \( \epsilon \) is the price elasticity of supply. As \( \epsilon \) goes to infinity the behavior of a monopsonist approaches that of a pure competitor.

Recall that in Chapter 4, where we defined the cost function of a firm, we considered only the behavior of a firm with competitive factor markets. However, it is certainly possible to define a cost function for a monopsonistic firm. Suppose for example that \( x_i(w) \) is the supply curve for factor \( i \). Then we can define:

\[
c(y) = \min \sum w_i x_i(w) \quad s.t \quad f(x(w)) = y
\]

This just gives us the minimum cost of producing \( y \).

### 9.4 Monopolistic Competition Market

To be revised.
9.4.1 Long Run Monopolistic Competition Equilibrium

Recall that we assumed that the demand curve for the monopolist’s product depended only on the price set by the monopolist. However, this is an extreme case. Most commodities have some substitutes and the prices of those substitutes will affect the demand for the original commodity. The monopolist sets his price assuming all the producers of other commodities will maintain their prices, but of course, this will not be true. The prices set by the other firms will respond — perhaps indirectly — to the price set by the monopolist in question. In this section we will consider what happens when several monopolists “compete” in setting their prices and output levels.

We imagine a group of \( n \) “monopolists” who sell similar, but not identical products. The price that consumers are willing to pay for the output of firm \( i \) depends on the level of output of firm \( i \) but also on the levels of output of the other firms: we write this inverse demand function as \( p_i(y_i, y) \) where \( y = (y_1 \ldots y_n) \).

Each firm is interested in maximizing profits: that is, each firm \( i \) is wants to choose its level of output \( y_i \) so as to maximize:

\[
p_i(y_i, y)y_i - c_i(y_i)
\]

Unfortunately, the demand facing the \( i^{th} \) firm depends on what the other firms do. How is firm \( i \) supposed to forecast the other firms behavior?

We will adopt a very simple behavioral hypothesis: namely, that firm \( i \) assumes the other firms behavior will be constant. Thus, each firm \( i \) will choose its level of output \( y_i^* \) so as to satisfy:

\[
p_i(y_i^*, y) + \frac{\partial p_i(y_i^*, y)}{\partial y_i}y_i^* - c_i'(y_i^*) \leq 0 \quad \text{with equality if } y_i^* > 0
\]

For each combination of operating levels for the firms \( y_1, \ldots, y_n \), there will be some optimal operating level for firm \( i \). We will denote this optimal choice of output by \( Y_i(y_1, \ldots, y_n) \). (Of course the output of firm \( i \) is not an argument of this function but it seems unnecessary to devise a new notation just to reflect that fact.)
In order for the market to be in equilibrium, each firm’s forecast about the behavior of the other firms must be compatible with the other firms actually do. Thus if \((y_1^*, \ldots, y_n^*)\) is to be an equilibrium it must satisfy:

\[
\begin{align*}
    y_1^* &= Y_1(y_1^*, \ldots, y_n^*) \\
    &\vdots \\
    y_n^* &= Y_n(y_1^*, \ldots, y_n^*)
\end{align*}
\]

that is, \(y_i^*\) must be the optimal choice for firm \(i\) if it assumes the other firms are going to produce \(y_2^*, \ldots, y_n^*\), and so on. Thus a monopolistic competition equilibrium \((y_1^*, \ldots, y_n^*)\) must satisfy:

\[
p_i(y_i^*, y) + \frac{\partial p_i(y_i^*, y)}{\partial y_i} y_i^* - c_i'(y_i^*) \leq 0
\]

with equality if \(y_i^* > 0\) and \(i = 1, \ldots, n\).

![Figure 9.6: Short-run monopolistic competition equilibrium](image)

For each firm, its marginal revenue equals its marginal cost, given the actions of all the other firms. This is illustrated in Figure 9.6. Now, at the monopolistic competition equilibrium depicted in Figure 9.6, firm \(i\) is making positive profits. We would therefore expect others firms to enter the industry and share the market with the firm so the firm’s profit will decrease because of close substitute goods.
Thus, in the long run, firms would enter the industry until the profits of each firm were driven to zero. This means that firm $i$ must charge a price $p_i^*$ and produce an output $y_i^*$ such that:

$$p_i^* y_i^* - c_i(y^*) \leq 0 \quad \text{with equality if } y_i^* > 0$$

or

$$p_i^* - \frac{c_i(y^*)}{y_i^*} \leq 0 \quad \text{with equality if } y_i^* > 0 \quad i = 1, 2, \ldots, n.$$

Thus, the price must equal to average cost and on the demand curve facing the firm. As a result, as long as the demand curve facing each firm has some negative slope, each firm will produce at a point where average cost are greater than the minimum average costs. Thus, like a pure competitive firms, the profits made by each firm are zero and is very nearly the long run competitive equilibrium. On the other hand, like a pure monopolist, it still results in inefficient allocation as long as the demand curve facing the firm has a negative slope.

### 9.4.2 Social Welfare in Monopolistic Competition

### 9.4.3 Dixit-Stiglitz Model on Monopolistic Competition

### 9.5 Oligopoly

Oligopoly is the study of market interactions with a small number of firms. Such an industry usually does not exhibit the characteristics of perfect competition, since individual firms’ actions can in fact influence market price and the actions of other firms. The modern study of this subject is grounded almost entirely in the theory of games discussed in the last chapter. Many of the specifications of strategic market interactions have been clarified by the concepts of game theory. We now investigate oligopoly theory primarily from this perspective by introducing four models.
9.5.1 Price Competition: Bertrand Model

Another model of oligopoly of some interest is the so-called Bertrand model. The Cournot model and Stackelberg model take the firms’ strategy spaces as being quantities, but it seems equally natural to consider what happens if price is chosen as the relevant strategic variables. Almost 50 years after Cournot, another French economist, Joseph Bertrand (1883), offered a different view of firm under imperfect competition and is known as the Bertrand model of oligopoly. Bertrand argued that it is much more natural to think of firms competing in their choice of price, rather than quantity. This small difference completely change the character of market equilibrium. This model is striking, and it contrasts starkly with what occurs in the Cournot model: With just two firms in a market, we obtain a perfectly competitive outcome in the Bertrand model!

In a simple Bertrand duopoly, two firms produce a homogeneous product, each has identical marginal costs $c > 0$ and face a market demand curve of $D(p)$ which is continuous, strictly decreasing at all price such that $D(p) > 0$. The strategy of each player is to announce the price at which he would be willing to supply an arbitrary amount of the good in question. In this case the payoff function takes a radically different form. It is plausible to suppose that the consumers will only purchase from the firm with the lowest price, and that they will split evenly between the two firms if they charge the same price. This leads to a payoff to firm 1 of the form:

$$
\pi_1(p_1, p_2) = \begin{cases} 
(p_1 - c)x(p_1) & \text{if } p_1 < p_2 \\
(p_1 - c)x(p_1)/2 & \text{if } p_1 = p_2 \\
0 & \text{if } p_1 > p_2 
\end{cases}
$$

Note that the Cournot game and the Bertrand game have a radically different structure. In the Cournot game, the payoff to each firm is a continuous function of its strategic choice; in the Bertrand game, the payoffs are discontinuous functions of the strategies. What is the Nash equilibrium? It may be somewhat surprising, but in the unique Nash equilibrium, both firms charge a price equal to marginal cost, and both earn zero profit. Formerly, we have
Proposition 9.5.1 There is a unique Nash equilibrium \((p_1, p_2)\) in the Bertrand duopoly. In this equilibrium, both firms set their price equal to the marginal cost: 
\[ p_1 = p_2 = c \] and earn zero profit.

Proof. First note that both firms setting their prices equal to \(c\) is indeed a Nash equilibrium. Neither firm can gain by raising its price because it will then make no sales (thereby still earning zero); and by lowering its price below \(c\) a firm increase it sales but incurs losses. What remains is to show that there can be no other Nash equilibrium. Because each firm \(i\) chooses \(p_i \geq c\), it suffices to show that there are no equilibria in which \(p_i > c\) for some \(i\). So let \((p_1, p_2)\) be an equilibrium.

If \(p_1 > c\), then because \(p_2\) maximizes firm 2’s profits given firm 1’s price choice, we must have \(p_2 \in (c, p_1]\), because some such choice earns firm 2 strictly positive profits, whereas all other choices earns firm 2 zero profits. Moreover, \(p_2 \neq p_1\) because if firm 2 can earn positive profits by choosing \(p_2 = p_1\) and splitting the market, it can earn even higher profits by choosing \(p_2\) just slightly below \(p_1\) and supply the entire market at virtually the same price. Therefore, \(p_1 > c\) implies that \(p_2 > c\) and \(p_2 < p_1\). But by stitching the roles of firms 1 and 2, an analogous argument establishes that \(p_2 > c\) implies that \(p_1 > c\) and \(p_1 < p_2\). Consequently, if one firm’s price is above marginal cost, both prices must be above marginal cost and each firm must be strictly undercutting the other, which is impossible.

9.5.2 Price Competition with Production Capacity

9.5.3 Quantity Competition: Cournot Oligopoly

A fundamental model for the analysis of oligopoly was the Cournot oligopoly model that was proposed by Cournot, an French economist, in 1838. A Cournot equilibrium, already mentioned in the last chapter, is a special set of production levels that have the property that no individual firm has an incentive to change its own production level if other firms do not change theirs.
To formalize this equilibrium concept, suppose there are $J$ firms producing a single homogeneous product. If firm $j$ produces output level $q_j$, the firm’s cost is $c_j(q_j)$. There is a single market inverse demand function $p(\hat{q})$. The total supply is $\hat{q} = \sum_{j=1}^{J} q_j$. The profit to firm $j$ is

$$p(\hat{q})q_j - c_j(q_j)$$

**Definition 9.5.1 (Cournot Equilibrium)** A set of output levels $q_1, q_2, \ldots, q_J$ constitutes a Cournot equilibrium if for each $j = 1, 2, \ldots, J$ the profit to firm $j$ cannot be increased by changing $q_j$ alone.

Accordingly, the Cournot model can be regarded as one shot game: the profit of firm $j$ is its payoff, and the strategy space of firm $j$ is the set of outputs, and thus a Cournot equilibrium is just a pure strategy Nash equilibrium. Then the first-order conditions for the interior optimum are:

$$p'(\hat{q})q_j + p(\hat{q}) - c'_j(q_j) = 0 \quad j = 1, 2, \ldots, J.$$  

The first-order condition for firm determines firm $j$ optimal choice of output as a function of its beliefs about the sum of the other firms’ outputs, denoted by $\hat{q}_{-j}$, i.e., the FOC condition can be written as

$$p'(q_j + \hat{q}_{-j})q_j + p(\hat{q}) - c'_j(q_j) = 0 \quad j = 1, 2, \ldots, J.$$  

The solution to the above equation, denoted by $Q_j(\hat{q}_{-j})$, is called the reaction function to the total outputs produced by the other firms.

Reaction functions give a direct characterization of a Cournot equilibrium. A set of output levels $q_1, q_2, \ldots, q_J$ constitutes a Cournot equilibrium if for each reaction function given $q_j = Q_j(\hat{q}_{-j})$ $j = 1, 2, \ldots, J$.

An important special case is that of duopoly, an industry with just two firms. In this case, the reaction function of each firm is a function of just the other firm’s output. Thus, the two reaction functions have the form $Q_1(q_2)$ and $Q_2(q_1)$, which is shown in Figure 9.7. In the figure, if firm 1 selects a value $q_1$ on the horizontal axis, firm 2 will react by selecting the point on the vertical axis that
corresponds to the function $Q_2(q_1)$. Similarly, if firm 2 selects a value $q_2$ on the vertical axis, firm 1 will be reacted by selecting the point on the horizontal axis that corresponds to the curve $Q_1(q_2)$. The equilibrium point corresponds to the point of intersection of the two reaction functions.

Figure 9.7: Reaction functions.

9.5.4 Dynamic Explanation of Cournot Equilibrium

9.5.5 Sequential Quantity Competition: Stackelberg Model

There are alternative methods for characterizing the outcome of an oligopoly. One of the most popular of these is that of quantity leadership, also known as the Stackelberg model.

Consider the special case of a duopoly. In the Stackelberg formulation one firm, say firm 1, is considered to be the leader and the other, firm 2, is the follower. The leader may, for example, be the larger firm or may have better information. If there is a well-defined order for firms committing to an output decision, the leader commits first.

Given the committed production level $q_1$ of firm 1, firm 2, the follower, will select $q_2$ using the same reaction function as in the Cournot theory. That is, firm
2 finds $q_2$ to maximize

$$\pi_2 = p(q_1 + q_2)q_2 - c_2(q_2),$$

where $p(q_1 + q_2)$ is the industrywide inverse demand function. This yields the reaction function $Q_2(q_1)$.

Firm 1, the leader, accounts for the reaction of firm 2 when originally selecting $q_1$. In particular, firm 1 selects $q_1$ to maximize

$$\pi_1 = p(q_1 + Q_2(q_1))q_1 - c_1(q_1),$$

That is, firm 1 substitutes $Q_2(q_1)$ for $q_2$ in the profit expression.

Note that a Stackelberg equilibrium does not yield a system of equations that must be solved simultaneously. Once the reaction function of firm 2 is found, firm 1’s problem can be solved directly. Usually, the leader will do better in a Stackelberg equilibrium than in a Cournot equilibrium.

### 9.5.6 Dynamic Price Competition and Firms’ Collusion

All of the models described up until now are examples of non-cooperative games. Each firm maximizes its own profits and makes its decisions independently of the other firms. What happens if they coordinate their actions? An industry structure where the firms collude to some degree in setting their prices and outputs is called a cartel.

A natural model is to consider what happens if the two firms choose simultaneously choose $y_1$ and $y_2$ to maximize industry profits:

$$\max_{y_1, y_2} p(y_1 + y_2)(y_1 + y_2) - c_1(y_1) - c_2(y_2)$$

The first-order conditions are

$$p'(y_1^* + y_2^*)(y_1^* + y_2^*) + p(y_1^* + y_2^*) = c_1'(y_1^*)$$

$$p'(y_1^* + y_2^*)(y_1^* + y_2^*) + p(y_1^* + y_2^*) = c_2'(y_1^*)$$

It is easy to see from the above first-order conditions that profit maximization implies $c_1'(y_1^*) = c_2'(y_2^*)$. 

323
The problem with the cartel solution is that it is not “stable” unless they are completely merged. There is always a temptation to cheat: to produce more than the agreed-upon output. Indeed, when the other firm will hold its production constant, we have by rearranging the first-order condition of firm 1:

$$\frac{\partial \pi_1(y^*, y^*_2)}{\partial y_1} = p'(y_1^* + y_2^*)y_1^* + p(y_1^* + y_2^*) - c'_1(y_1^*) = -p'(y_1^* + y_2^*)y_2^* > 0$$

by noting the fact that the demand curves slope downward.

The strategic situation is similar to the Prisoner’s Dilemma: if you think that other firm will produce its quota, it pays you to defect to produce more than your quota. And if you think that the other firm will not produce at its quota, then it will in general be profitable for you to produce more than your quota.

In order to make the cartel outcome viable, some punishment mechanism should be provided for the cheat on the cartel agreement, say using a repeated game as discussed in the previous chapter.

9.5.7 Price Competition of Horizontal Product Differentiation: Hotelling Model

9.5.8 Vertical Product Differentiation Model

9.5.9 Market Entry Deterrence

9.5.10 Price Competition with Asymmetric Information

9.5.11 Limit Pricing with Asymmetric Information: Dynamic Market Structure

9.6 Reference

Books and Monographs:


**Papers:**


Part IV

General Equilibrium Theory and Social Welfare
Part IV is devoted to an examination of competitive market economies from a general equilibrium perspective at which all prices are variable and equilibrium requires that all markets clear.

The content of Part IV is organized into four chapters. Chapters 10 and 11 constitute the heart of the general equilibrium theory. Chapter 10 presents the formal structure of the equilibrium model, introduces the notion of competitive equilibrium (or called Walrasian equilibrium). The emphasis is on positive properties of the competitive equilibrium. We will discuss the existence, uniqueness, and stability of a competitive equilibrium. We will also discuss a more general setting of equilibrium analysis, namely the abstract economy which includes the general equilibrium model as a special case. Chapter 11 discusses the normative properties of the competitive equilibrium by introducing the notion of Pareto efficiency. We examine the relationship between the competitive equilibrium and Pareto optimality. The core is concerned with the proof of the two fundamental theorems of welfare economics. Chapter 12 explores extensions of the basic analysis presented in Chapters 10 and 11. Chapter 12 covers a number of topics whose origins lie in normative theory. We will study the important core equivalence theorem that takes the idea of Walrasian equilibria as the limit of cooperative equilibria as markets grow large, fairness of allocation, and social choice theory. Chapter 13 applies the general equilibrium framework developed in Chapters 10 to 12 to economic situations involving the exchange and allocation of resources under conditions of uncertainty.

It is greatly acknowledged that some materials are drawn from the references in the end of each charter.
Chapter 10

Positive Theory of Equilibrium: Existence, Uniqueness, and Stability

10.1 Introduction

One of the great achievements of economic theory in the last century is general equilibrium theory. General equilibrium theory is a branch of theoretical neoclassical economics. It seeks to explain the behavior of supply, demand and prices in a whole economy, by considering equilibrium in many markets simultaneously, unlike partial equilibrium theory which considers only one market at a time. Interaction between markets may result in a conclusion that is not obtained in a partial equilibrium framework. As such, it is a baseline theory to study market economy and so an abstraction from a real economy; it is proposed as being a useful model, both by considering equilibrium prices as long-term prices and by considering actual prices as deviations from equilibrium. The theory dates to the 1870s, particularly the work of French economist Léon Walras, and thus it is often called the Walrasian theory of market.

From a positive viewpoint, the general equilibrium theory is a theory of the
determination of equilibrium prices and quantities in a system of perfectly competitive markets. The proof of existence of a general equilibrium is generally considered one of the most important and robust results of economic theory. The mathematical model of a competitive economy of L. Walras (1874-77) was conceived as an attempt to explain the state of equilibrium reached by a large number of small agents interaction through markets. Walras himself perceived that the theory would be vacuous without a mathematical argument in support the existence of at least one equilibrium state. However, for more than half a century the equality of the number of equations and of the number of unknowns of the Walrasian model remained the only, unconvincing remark made in favor of the existence of a competitive equilibrium. Study of the existence problem began in the early 1930s when Neisser (1932), Stackelberg (1933), Zeuthen (1933), and Schesinger (1935) identified some of its basic features and when Wald (1935, 1936a, 1936b) obtained its first solutions. After an interruption of some twenty years, the questions of existence of an economic equilibrium was taken up again by Arrow and Debreu (1954), McKenzie (1954, 1955), Gale (1955), and many others.

The proof of the existence of a competitive equilibrium of was firstly given by Abraham Wald (1935, 1936a,b) in a production economy. Nevertheless, the essence of his contribution has not been fully recognized in the subsequent literature since its lacking of generality. von Neumann (1928, 1937) turned out to be of greater importance of the subject. He proved a topological lemma which, in its reformulation by Kakutani (1941) as a fixed-point theorem for a correspondence, becomes the most powerful tool for the proof of existence of an economic equilibrium.

A general equilibrium is defined as a state where the aggregate demand does not exceed the aggregate supply for all markets. Thus, equilibrium prices are endogenously determined. The general equilibrium approach has two central features:

(1) It views the economy as a closed and inter-related system in which
we must simultaneously determine the equilibrium values of all variables of interests (consider all markets together).

(2) It aims at reducing the set of variables taken as exogenous to a small number of physical realities.

It is to predict the final consumption and production in the market mechanism.

The general equilibrium theory consists of five components:

(1) Economic institutional environment (the fundamentals of the economy): economy that consists of consumption space, preferences, endowments of consumers, and production possibility sets of producers.

(2) Economic institutional arrangement: It is the price mechanism in which a price is quoted for every commodity.

(3) The behavior assumptions: price taking behavior for consumers and firms, utility maximization and profit maximization.

(4) Predicting outcomes: equilibrium analysis — positive analysis such as existence, uniqueness, and stability.

(5) Evaluating outcomes: normative analysis such as allocative efficiency of general equilibrium.

Questions to be answered in the general equilibrium theory.

(1) The existence and determination of a general equilibrium: What kinds of restrictions on economic environments (consumption sets, endowments, preferences, production sets) would guarantee the existence of a general equilibrium.

(2) Uniqueness of a general equilibrium: What kinds of restrictions on economic environments would guarantee a general equilibrium to be unique?
(3) Stability of a general equilibrium: What kinds of restrictions on economic environments would guarantee a competitive economy converge to equilibrium, and, if it does, how quickly does this happen?

(4) Welfare properties of a general equilibrium: What kinds of restrictions on consumption sets, endowments, preferences, production sets would ensure a general equilibrium to be socially optimal – Pareto efficient?

10.2 The Structure of General Equilibrium Model

10.2.1 Economic Environments

The fundamentals of the economy are economic institutional environments that are exogenously given and characterized by the following terms:

- $N = \{1, \ldots, n\}$: the set of consumers
- $J = \{1, \ldots, J\}$: the number of producers (firms)$^1$
- $L$: the number of (private) goods
- $X_i \subset \mathbb{R}^L$: the consumption space of consumer $i = 1, \ldots, n$, which specifies the boundary of consumptions, collection of all individually feasible consumptions of consumer $i$. Some components of an element may be negative such as a labor supply;
- $\succsim_i$: preferences ordering (or $u_i$ if a utility function exists) of consumer $i = 1, \ldots, n$;

$^1$With a slight abuse of notation, $J$ will represent the set as well as the number of producers.
• \( w_i \in X_i \): initial endowment vector of consumer \( i \).

• \( e_i = (X_i, \succ_i, w_i) \): the characteristic of consumer \( i \).

• \( Y_j \): production possibility set of firm \( j = 1, 2, \ldots, J \), which is the characteristic of producer \( j \).

• \( y_j \in Y_j \): a production plan, \( y_{jl} > 0 \) means \( y_{jl} \) is output and \( y_{jl} < 0 \) means \( y_{jl} \) is input. Most elements of \( y_j \) for a firm are zero, which means this good is neither produced nor used in production.

• \( e = (\{X_i, \succ_i, w_i\}_{i \in N}, \{Y_j\}_{j \in J}) \): an economy, or called an economic environment.

• \( X = X_1 \times X_2 \times \ldots \times X_n \): consumption space.

• \( Y = Y_1 \times Y_2 \times \ldots \times Y_J \): production space.

Remark 10.2.1 \( \succ_i \) is a preference ordering if it is reflexive \( (x_i \succ_i x_i) \), transitive \( (x_i \succ_i x'_i \text{ and } x'_i \succ_i x''_i \text{ implies } x_i \succ_i x''_i) \), and complete (for any pair \( x_i \) and \( x'_i \), either \( x_i \succ_i x'_i \) or \( x'_i \succ_i x_i \)). It can be shown that a preference ordering \( \succ_i \) can be represented by a continuous utility function if it is continuous.\(^2\) The existence of general equilibrium can be obtained even when preferences are weakened to be non-complete or non-transitive.

Remark 10.2.2 Recall that there can be three types of returns about production scales: diminishing returns to scale (i.e., \( y_j \in Y_j \) implies that \( \alpha y_j \in Y_j \) for all \( \alpha \in [0, 1] \)), increasing returns to scale (i.e., \( y_j \in Y_j \) implies that \( \alpha y_j \in Y_j \) for all \( \alpha \geq 1 \)), and constant returns to scale i.e., \( y_j \in Y_j \) implies that \( \alpha y_j \in Y_j \) for all \( \alpha \geq 0 \)). In other words, diminishing returns to scale implies any feasible input-output vector can be scaled down; increasing returns to scale implies any feasible input-output vector can be scaled up; constant returns to scale implies the

\(^{2}\)This result can be extended to the case in which both preference and utility function are upper hemicontinuous.
production set is the conjunction of increasing returns and diminishing returns. Geometrically, it is a cone.

Figure 10.1: Various Returns to Scale: IRS, DRS, and CRS.

Notice that, competitive equilibrium could exist and make sense only in the cases of constant and diminishing returns to scale. If the technology available exhibits increasing returns to scale, the firms may have unbounded input demands and output supplies for any given non-zero output prices. This is because of increasing marginal profitability from expanding the productive activity. Therefore, the existence of competitive equilibrium is not guaranteed. Convexity is a key assumption in general competitive analysis. In particular, convexity of production sets, or equivalently concavity (not merely monotonicity plus quasi-concavity) of production functions, is assumed and used in the existence proofs as well as in welfare theorems of Walrasian general equilibrium theory (Arrow and Debreu (1954), Debreu (1959)). This is also regarded as a main limitation of general equilibrium theory.

General equilibrium theory precludes the possibility of increasing returns to scale because the non-convexity of production set is essentially incompatible with a competitive market. There exist two main sources of non-convexity. It may arise
when the fixed cost of a firm outweights substantially its marginal cost. This is commonly observed in high-tech industries, where the input of R&D in advance is very large, and the subsequent marginal production cost is negligible small. In such industries, allowing the innovators to keep his monopoly position for a certain duration is a crucial method to protect innovation (say, the protection period of intellectual property in U.S.A lasts for 20 years, while the authorship property protection is 70 years after author’s death).

The non-convexity of production set may also arise in natural monopoly industry in which it is most efficient for production to be permanently concentrated in a single firm rather than contested competitively. Examples include public utilities such as water services and electricity. In these industries, allowing competition could only deprive the firms of minimum scale for operation and thus result in inefficient outcome. So only a unique or small number of firms are allowed to operate in such industries (not merely in developing countries as China, this measure is adopted by most developed market economies).

Either of these two cases leads to non-convexity of production set, and thus result in monopoly. In these cases, it is not reasonable to assume the firm to be a price taker. We will show in Section 6 of next Chapter, the allocative efficiency could not be achieved even under free competition. The profit maximizing rule is not feasible any more, some alternative ways of pricing, such as marginal cost pricing, may be adopted. The above cases show that non-convexity is essentially incompatible with competitive market, so it is very reasonable to exclude the possibility of increasing returns to scale.

10.2.2 Institutional Arrangement: Private Market Mechanism

- $p = (p^1, p^2, \ldots, p^L) \in \mathbb{R}_+^L$: a price vector;
- $p \cdot x_i$: the expenditure of consumer $i$ for $i = 1, \ldots, n$;
- $p \cdot y_j$: the profit of firm $j$ for $j = 1, \ldots, J$;
• \( p \cdot w_i \): the value of endowments of consumer \( i \) for \( i = 1, \ldots, n \);

• \( \theta_{ij} \in \mathbb{R}_+ \): the profit share of consumer \( i \) from firm \( j \), which specifies ownership (property rights) structures, so that \( \sum_{i=1}^{n} \theta_{ij} = 1 \) for \( j = 1, 2, \ldots, J \);

• \( \sum_{j=1}^{J} \theta_{ij} p \cdot y_j \): the total profit dividend received by consumer \( i \) from firms for \( i = 1, \ldots, n \).

For notional simplicity, in the following, we may also write the vector inner product without putting center dot between two vectors. For instance, we may write \( p \cdot x_i \) simply as \( px_i \).

For \( i = 1, 2, \ldots, n \), consumer \( i \)'s budget constraint is given by

\[
px_i \leq pw_i + \sum_{j=1}^{J} \theta_{ij} py_j
\]

and the budget set is given by

\[
B_i(p) = \{ x_i \in X_i : px_i \leq pw_i + \sum_{j=1}^{J} \theta_{ij} py_j \}. \tag{10.2}
\]

A private ownership economy then is referred to

\[
e = (\{ e_i \}_{i \in N} \{ Y_j \}_{j \in J}, \{ \theta_{ij} \}_{i \in N, j \in J}). \tag{10.3}
\]

The set of all such private ownership economies are denoted by \( E \).

10.2.3 Individual Behavior Assumptions:

We assume that all consumers and firms are rational, that is, consumers maximize their utilities subject to their budget constraints; firms maximizes their profits subject to their technology constraints; furthermore, the market is assumed to be competitive, that is, everyone is price taker.

(1) Perfect competitive markets: every player is a price-taker.
(2) Utility maximization: every consumer maximizes his preferences subject to $B_i(p)$. That is,

$$\max_{x_i} u_i(x_i)$$  \hfill (10.4)

s.t.

$$p \cdot x_i \leq p \cdot w_i + \sum_{j=1}^{J} \theta_{ij} p \cdot y_j$$  \hfill (10.5)

(3) Profit maximization: every firm maximizes its profit in $Y_j$. That is,

$$\max_{y_j \in Y_j} p \cdot y_j$$  \hfill (10.6)

for $j = 1, \ldots, J$.

### 10.2.4 Competitive Equilibrium

Before defining the notion of competitive equilibrium, we first give some notions on allocations which identify the set of possible outcomes in economy $e$. For notational convenience, “$\hat{a}$” will be used throughout the notes to denote the sum of elements of vectors $a_i$, i.e., $\hat{a} := \sum_{l=1}^{L} a_i^l$.

**Definition 10.2.1 (allocation)** An allocation $(x, y)$ is a specification of consumption vector $x = (x_1, \ldots, x_n)$ and production vector $y = (y_1, \ldots, y_J)$.

**Definition 10.2.2 (individually feasible allocation)** An allocation $(x, y)$ is individually feasible if $x_i \in X_i$ for all $i \in N$, $y_j \in Y_j$ for all $j = 1, \ldots, J$.

**Definition 10.2.3 (balanced allocation)** An allocation is weakly balanced if

$$\hat{x} \leq \hat{y} + \hat{w}$$  \hfill (10.7)

or specifically

$$\sum_{i=1}^{n} x_i \leq \sum_{j=1}^{J} y_j + \sum_{i=1}^{n} w_i$$  \hfill (10.8)

When inequality holds with equality, the allocation is called a balanced or attainable allocation.
Definition 10.2.4 (feasible allocation) An allocation \((x, y)\) is feasible if it is both individually feasible and (weakly) balanced.

Thus, an economic allocation is feasible if the total amount of each good consumed does not exceed the total amount available from both the initial endowment and production. Denote by \(A = \{(x, y) \in X \times Y : \hat{x} \leq \hat{y} + \hat{w}\}\) the set of all feasible allocations.

Aggregation:
\[
\begin{align*}
\hat{x} &= \sum_{i=1}^{n} x_i: \text{aggregation of consumption;} \\
\hat{y} &= \sum_{j=1}^{J} y_j: \text{aggregation of production;} \\
\hat{w} &= \sum_{i=1}^{n} w_i: \text{aggregation of endowments;}
\end{align*}
\]

Now we define the notion of competitive equilibrium.

Definition 10.2.5 (Competitive/Walrasian Equilibrium, CE) Given a private ownership economy, \(e = (\{e_i\}_{i \in N}, \{Y_j\}_{j \in J}, \{\theta_{ij}\}_{(i,j) \in N \times J})\), an allocation \((x, y) \in X \times Y\) and a price vector \(p \in \mathbb{R}^L\) consist of a competitive equilibrium if the following conditions are satisfied

(i) Utility maximization: \(x_i \succeq_i x'_i\) for all \(x'_i \in B_i(p)\) and \(x_i \in B_i(p)\) for \(i = 1, \ldots, n\).

(ii) Profit maximization: \(y_j \in Y_j\) and \(p \cdot y_j \geq p \cdot y'_j, \forall y'_j \in Y_j\).

(iii) Market Clear Condition: \(\hat{x} \leq \hat{w} + \hat{y}\).

\((x, y)\) is called an equilibrium allocation of CE, \(p\) is called equilibrium price vector of CE. We usually don’t require and is unable to guarantee a nonnegative price vector, unless some types of monotonicities of preferences are imposed.

Denote
\[
\begin{align*}
x_i(p) &= \{x_i \in B_i(p) : x_i \in B_i(p) \text{ and } x_i \succeq_i x'_i \text{ for all } x'_i \in B_i(p)\};
\end{align*}
\]
the demand correspondence of consumer \(i\) under utility maximization; it is called the demand function of consumer \(i\) if it is single-valued.
\[ y_j(p) = \{ y_j \in Y_j : p \cdot y_j \geq p \cdot y'_j \text{ for all } y'_j \in Y_j \} \]: the supply correspondence of the firm \( j \); it is called the supply function of firm \( j \) if it is single-valued.

- \( \hat{x}(p) = \sum_{i=1}^{n} x_i(p) \): the aggregate demand correspondence.
- \( \hat{y}(p) = \sum_{j=1}^{J} y_j(p) \): the aggregate supply correspondence.
- \( \hat{z}(p) = \hat{x}(p) - \hat{w} - \hat{y}(p) \): aggregate excess demand correspondence.

If \( \hat{z}(p) > 0 \), then good \( l \) is undersupplied; if \( \hat{z}(p) < 0 \) then good \( l \) is over-supplied. So an equivalent definition of competitive equilibrium is then given as follows:

**Definition 10.2.6** A price vector \( p^* \in \mathbb{R}_+^L \) is a competitive equilibrium price if there exists \( \hat{z} \in \hat{z}(p^*) \) such that \( \hat{z} \leq 0 \). If \( \hat{z}(p) \) is single-valued, \( \hat{z}(p^*) \leq 0 \) implies \( p^* \) is a competitive equilibrium price.

**Constrained Walrasian Equilibrium**

In the above definition of competitive equilibrium, a consumer maximizes its own utility subject only to his own budget constraint. If they maximize their utilities subject to the aggregate resources available, we get *constrained competitive equilibrium*.

Let

\[ C_i = \left\{ x_i \in X_i : x_i \leq \sum_{j=1}^{J} y_j + \sum_{i=1}^{n} w_i \right\}. \]

**Definition 10.2.7 (Constrained Competitive/Walrasian Equilibrium,CCE/CWE)** for a private ownership economy \( e = (\{e_i\}_{i \in N}, \{Y_j\}_{j \in J}, \{\theta_{ij}\}_{(i,j) \in N \times J}) \), an allocation \((x, y) \in X \times Y\) and price vector \( p \in \mathbb{R}^L\) constitutes a constrained competitive equilibrium if the following conditions hold:

(i) Utility-maximizing condition: \( \forall i = 1, \ldots, n, x_i \in B_i(p) \cap C_i(p) \)
   and for \( \forall x'_i \in B_i(p) \cap C_i(p) \), we have \( x_i \gg_i x'_i \);
(ii) Profit-maximizing condition: for \( \forall y_j' \in Y_j \), we have \( p y_j \geq p y_j' \), \( j = 1, \ldots, J \);

(iii) Market clearing condition: \( \hat{x} \leq \hat{w} + \hat{y} \)

Obviously, every CE is a CWE since \( x_i \in C_i \) for every CE \( (x, y) \). But the opposite is not necessarily true. This is because utility maximization for CWE is on the more restricted set (the intersection of \( B_i(p) \) and \( C_i \)) but not on \( B_i(p) \). However, when the preference is convex, (that is \( x'_i \succ_i x_i \) implies \( \lambda x'_i + (1 - \lambda)x_i \succ_i x_i, \ \forall \lambda \in (0, 1) \)), A CCE \( (x, y, p) \) with interior consumption \( x \in intX \) is also a CE.

**Proposition 10.2.1** For \( \forall i \in N, \succ_i \) is convex. Let \( (x, y, p) \) be a constrained competitive equilibrium. If \( x \in R^L_{++} \), then \( (x, y, p) \) is a competitive equilibrium.

**Proof.** We prove it by contradiction. Suppose that CWE \( (x, y, p) \) is not a CE. From the definition of CWE, \( (x, y, p) \) is feasible and \( y \) maximizes the profit of producers. Then, if \( (x, y, p) \) is not CE, there exists a consumer \( i \), whose utility is not maximized. That is, \( \exists x'_i \in X_i \) such that \( x'_i \succ_i x_i \) and \( x'_i \in B_i(p) \). Therefore, from the convexity of preference, we have \( \lambda x'_i + (1 - \lambda)x_i \succ_i x_i \) and \( \lambda x'_i + (1 - \lambda)x_i \in B_i(p) \). Now that \( x > 0 \), we have \( x_i < \sum_{j=1}^{J} y_j + \sum_{i=1}^{n} w_i \). Hence, when \( \lambda \) is sufficiently close to zero, \( \lambda x'_i + (1 - \lambda)x_i < \sum_{j=1}^{J} y_j + \sum_{i=1}^{n} w_i \), that is \( \lambda x'_i + (1 - \lambda)x_i \in C_i \). This contradicts the fact that \( (x, y, p) \) is CWE. □

CWE plays an important role in the mechanism design theory, it satisfies the necessary condition for Nash implementation, which is not satisfied by CE. For detailed discussion of CWE, see Tian (1988).

### 10.3 Some Examples of GE Models: Graphical Treatment

In most economies, there are three types of economic activities: production, consumption, and exchanges. Before formally stating the existence results on
competitive equilibrium, we will first give two simple examples of general equilibrium models: exchange economies and a production economy with only one consumer and one producer. If a unique person is both consumer and producer of an economy, we call this Robinson Crusoe economy. These examples introduce some of the questions, concepts, and common techniques that will occupy us for the rest of this part. In the rest part of this note, we focus on the case in which all goods have nonnegative price, unless otherwise stated.

10.3.1 Pure Exchange Economies

A pure exchange economy is an economy in which there is no production. This is a special case of general economy. In this case, economic activities only consist of trading and consumption.

The aggregate excess demand correspondence becomes \( \hat{z}(p) = \hat{x}(p) - \hat{w} \) so that we can define the individual excess demand by \( z_i(p) = x_i(p) - w_i \) for this special case. The simplest exchange economy with the possibility of mutual benefit exchange is the two-good and two-consumer exchange economy. As it turns out, this case is amenable to analysis by an extremely handy graphical device known as the Edgeworth Box.

Edgeworth Box.

Consider an exchange economy with two goods \((x_1, x_2)\) and two persons. The total endowment is \( \hat{w} = w_1 + w_2 \). For example, if \( w_1 = (1, 2) \), \( w_2 = (3, 1) \), then the total endowment is: \( \hat{w} = (4, 3) \). Note that the point, denoted by \( w \) in the Edgeworth Box, can be used to represent the initial endowments of two persons.
Figure 10.2: Edgeworth Box in which \( w_1 = (1, 2) \) and \( w_2 = (3, 1) \).

Advantage of the Edgeworth Box is that it gives all the possible (balanced) trading points. That is,

\[
x_1 + x_2 = w_1 + w_2
\]

for all points \( x = (x_1, x_2) \) in the box, where \( x_1 = (x_1^1, x_1^2) \) and \( x_2 = (x_2^1, x_2^2) \).

Thus, every point in the Edgeworth Box stands for an attainable allocation so that \( x_1 + x_2 = w_1 + w_2 \).

The shaded lens (portion) of the box in the above figure represents all the trading points that make both persons better off. Beyond the box, any point is not feasible.

Which point in the Edgeworth Box can be a competitive equilibrium?
Figure 10.3: In the Edgeworth Box, the point CE is a competitive equilibrium.

In the box, one person's budget line is also the budget line of the other person. They share the same budget line in the box.

Figure 10.4 below shows the market adjustment process to a competitive equilibrium. Originally, at price $p$, both persons want more good 2. This implies that the price of good 1, $p^1$, is too high so that consumers do not consume the total amounts of the good so that there is a surplus for $x_1$ and there is an excess demand for $x_2$, that is, $x_1 + x_2^1 < w_1^1 + w_2^1$ and $x_1^2 + x_2^2 > w_1^2 + w_2^2$. Thus, the market will adjust itself by decreasing $p^1$ to $p^1'$. As a result, the budget line will become flatter and flatter till it reaches the equilibrium where the aggregate demand equals the aggregate supply. In this interior equilibrium case, two indifference curves are tangent each other at a point that is on the budget line so that the marginal rates of substitutions for the two persons are the same that is equal to the price ratio.
Figure 10.4: This figure shows the market adjustment process.

What happens when indifference curves are linear?

Figure 10.5: A competitive equilibrium may still exist even if two persons’ indifference curves do not intersect.

In this case, there is no tangent point as long as the slopes of the indifference curves of the two persons are not the same. Even so, there still exists a competitive
equilibrium although the marginal rates of substitutions for the two persons are not the same.

**Offer Curve**: the locus of the optimal consumptions for the two goods when price varies. Note that, it consists of tangent points of the indifference curves and budget lines when price varies.

The offer curves of two persons are given in the following figure:

![Offer Curve Diagram](image)

Figure 10.6: The CE is characterized by the intersection of two persons’ offer curves.

The intersection of the two offer curves of the consumers can be used to check if there is a competitive equilibrium. By the above diagram, one can see that the one intersection of two consumers’ offer curves is always given at the endowment $w$. If there is another intersection of two consumers’ offer curves rather than the one at the endowment $w$, the intersection point must be the competitive equilibrium.

To enhance the understanding about the competitive equilibrium, you try to draw the competitive equilibrium in the following situations:
1. There are many equilibrium prices.

2. One person’s preference is such that two goods are perfect substitutes (i.e., indifference curves are linear).

3. Preferences of one person are the Leontief-type (perfect complement)

4. One person’s preferences are non-convex.

5. One person’s preferences are “thick”.

6. One person’s preferences are convex, but has a satiation point.

Note that a preference relation $\succ_i$ is convex if $x \succ_i x'$ implies $tx+(1-t)x' \succ_i x'$ for all $t \in (0,1)$ and all $x, x' \in X_i$. A preference relation $\succ_i$ has a satiation point $x$ if $x \succ_i x'$, for $\forall x' \in X_i$.

Cases in which there may be no Walrasian Equilibria

When the preference orderings are discontinuous or they are not strong monotonic and initial endowments are not interior point of consumption set, competitive equilibrium may fail to exist. To guarantee the existence of CE, some assumptions must be made to avoid these cases. We give two counterexamples as follows.

**Example 10.3.1 (Non-Convex Preferences)** Indifference curves (IC) are not convex. If the two offer curves are not intersected except for at the endowment points, then there may not exist a competitive equilibrium. This may be true when preferences are not convex.

This is only an extreme case. CE may exist even with convex preferences. But we cannot relax the assumption on convex preferences in order to obtain CE. This example shows that a graphical depiction is often very helpful for enlightening us to find the right answer.
Figure 10.7: A CE may not exist if indifference curves are not convex. In that case, offer curves of two consumers do not intersect.

**Example 10.3.2 (Non-Interior Endowment And Non-Strict Monotonic Preference)**

Competitive equilibrium may not exist when the initial endowment may not be an interior point of the consumption space, and preferences may not be strong monotonic, although the other nice properties are satisfied. Consider an exchange economy in which one person’s indifference curves put no values on one commodity, say, when Person 2’s indifference curves are vertical lines so that $u_2(x_1^2, x_2^2) = x_1^2$. Person 1’s utility function is regular one, say, which is given by a quasi-linear utility function, say, $u_1 = x_1^1 + x_1^2$. The initial endowments are given by $w_1 = (0, 1)$ and $w_2 = (1, 0)$. There is no competitive equilibrium. Why? There are two cases to be considered.

- If $p_1^1/p_2^2 > 0$, then

$$\begin{cases} x_2^1 = 1 \\ x_1^1 > 0, \end{cases}$$

Thus, $x_1^1(p) + x_2^1(p) > 1 = \hat{w}^1$, the aggregate demand exceeds the aggregate endowment, CE does not exist.
If \( p^1/p^2 = 0 \), then \( x_2^1 = \infty \), which violates the feasibility conditions. Thus, no CE exists.

Thus, when the initial endowment may not be an interior point of the consumption space and preferences may not be strong monotonic, there may not exist any competitive equilibrium. As such, in proving the existence of CE, the assumption either the initial endowment is an interior point of the consumption space or preferences is strong monotonic should be imposed.

Figure 10.8: A CE may not exist if an endowment is on the boundary.

10.3.2 The Economy With One Consumer And One Producer/Robinson Crusoe Economy

Now we introduce the possibility of production. To do so, in the simplest-possible setting in which there are only two price-taking economic agents.
Two agents: one producer so that $J = 1$ and one consumer so that $n = 1$.

Two goods: labor (leisure) and the consumption good produced by the firm. $w = (\bar{L}, 0)$: the endowment.

$\bar{L}$: the total units of leisure time.

$f(z) : R_+ \rightarrow R_+$: the production function that is strictly increasing, concave, and differentiable, where $z$ is labor input. To have an interior solution, we assume $f$ satisfies the Inada condition $f'(0) = +\infty$, and $\lim_{z \rightarrow 0} f'(z)z = 0$.

$(p, \omega)$: the price vector of the consumption good and labor.

$\theta = 1$: single person economy.

$u(x^1, x^2) : R^2_+ \rightarrow R^2_+$: is the utility function which is strictly quasi-concave, increasing, and differentiable. To have an interior solution, we assume $u$ satisfies the Inada condition $\frac{\partial u}{\partial x_i}(0) = +\infty$, and $\lim_{x_i \rightarrow 0} \frac{\partial u}{\partial x_i} x_i = 0$.

The firm’s problem is to choose the labor $z$ so as to solve

$\max_{z \geq 0} pf(z) - \omega z$ \hspace{1cm} (10.11)

FOC:

$pf'(z) = \omega$

or $f'(z) = \omega/p$

$(MRTS)_{z,q} = \text{Price ratio}$

which means the marginal rate of technique substitution of labor for the consumption good $q$ equals the price ratio of the labor input and the consumption good output.

Let
\( q(p, \omega) \) = the profit maximizing output for the consumption good.

\( z(p, \omega) \) = the profit maximizing input for the labor.

\( \pi(p, \omega) \) = the profit maximizing function.

---

Figure 10.9: Figures for the producer's problem, the consumer problem and the CE.
The consumer’s problem is to choose the leisure time and the consumption for the good so as to solve
\[
\max_{x^1, x^2} u(x^1, x^2) \\
\text{s.t. } px^2 \leq \omega(L - x^1) + \pi(p, \omega)
\]
where \(x^1\) is the leisure and \(x^2\) is the consumption of good.

(FOC:)
\[
\frac{\partial u}{\partial x^1} = \frac{\omega}{p}
\]
which means the marginal rate of substitution of the leisure consumption for the consumption good \(q\) equals the price ratio of the leisure and the consumption good, i.e., \(MRS_{x^1,x^2} = \omega/p\).

By (10.11) and (10.12)
\[
MRS_{x^1,x^2} = \frac{\omega}{p} = MRTS_{z,q}
\]
A competitive equilibrium for this economy involves a price vector \((p^*, \omega^*)\) at which
\[
x^2(p^*, \omega^*) = q(p^*, \omega^*); \\
x^1(p^*, \omega^*) + z(p^*, \omega^*) = L
\]
That is, the aggregate demand for the two goods equals the aggregate supply for the two goods. Figure 10.9 shows the problems of firm and consumer, and the competitive equilibrium, respectively.

10.4 Existence of Competitive Equilibrium

The proof of the existence of a competitive equilibrium is generally considered one of the most important and robust results of economic theory. In this section we will examine the existence of competitive equilibrium for the following four cases: (1) the single-valued aggregate excess demand function; (2) the aggregate
excess demand correspondence; (3) a general class of private ownership production economies. (4) pure exchange economy with non-complete and non-transitive preference.

The first two cases are based on excess demand (function or correspondence). In these cases, the existence of CEs are proved by showing that there is a price at which excess demand can be non-positive. Then some constraints are impose on characteristics of agents to guarantee the existence of CE. Methods in cases 3 and 4 are based directly on the underlying preference orderings and consumption and production sets.

Various methods could be adopted to prove the existence of CE in the above four cases, one could use Brouwer fixed point theorem, Kakutani fixed point theorem, KKM lemma, Michael fixed point theorem, introduced in Chapter 2 to prove the existence of CE. All these methods are essentially the same. They all states the existence of fixed points. The first attempt to model competitive equilibrium prices as solutions to a system of equations was made by Leon Walras in 1874. But, due to lacking of powerful mathematical weapons, the rigorous and formal proof was not put forth until 1950s.

### 10.4.1 The Existence of CE for Aggregate Excess Demand Function

The simplest case for the existence of a competitive equilibrium is the one when the aggregate excess demand correspondence is a single-valued function. Note that, when preference orderings and production sets are both strictly convex, we obtain excess demand functions rather than correspondences.

A very important property of excess demand function $\hat{z}(p)$ is Walras’ law, which can take one of the following three forms:

(1) strong Walras’ law given by

$$p \cdot \hat{z}(p) = 0 \text{ for all } p \in \mathbb{R}_+^L;$$

352
(2) weak Walras’ law given by

$$\mathbf{p} \cdot \hat{z}(\mathbf{p}) \leq 0 \text{ for all } \mathbf{p} \in \mathbb{R}^L_+;$$

(3) interior Walras’ law given by

$$\mathbf{p} \cdot \hat{z}(\mathbf{p}) = 0 \text{ for all } \mathbf{p} \in \mathbb{R}^L_{++}.$$

It is clear that strong Walras’ law implies weak Walras’ law, but converse may not be true. For the convention of discussion, if no confusion, we simply call strong Walras’ law and interior Walras’ law as Walras’ law.

Another important property of excess demand function is homogeneity of $\hat{z}(\mathbf{p})$: it is homogeneous of degree 0 in price $\hat{z}(\lambda \mathbf{p}) = \hat{z}(\mathbf{p})$ for any $\lambda > 0$. From this property, we can normalize prices.

Because of homogeneity, for example, we can normalize a price vector as follows:

1. $\mathbf{p}^l = p^l / p^1$, $l = 1, 2, \ldots, L$
2. $\mathbf{p}^l = p^l / \sum_{i=1}^L p^i$.

Thus, without loss of generality, we can restrict our attention to the unit simplex:

$$S^{L-1} = \{ \mathbf{p} \in \mathbb{R}^L_+ : \sum_{i=1}^L p^i = 1 \}. \quad (10.14)$$

Then, we have the following theorem on the existence of competitive equilibrium.

**Theorem 10.4.1 (The Existence Theorem I)** For a private ownership economy $e = (\{X_i, w_i, \succ_i\}_{i \in N}, \{Y_j\}_{j \in J}, \{\theta_{ij}\}_{(i,j) \in N \times J})$, if $\hat{z}(\mathbf{p})$ is a homogeneous of degree zero and continuous function and satisfies weak Walras’ Law, then there exists a competitive equilibrium, that is, there is $\mathbf{p}^* \in \mathbb{R}^L_+$ such that

$$\hat{z}(\mathbf{p}^*) \leq 0 \quad (10.15)$$
Proof. Define a continuous function \( g : S^{L-1} \to S^{L-1} \) by
\[
g^l(p) = \frac{p^l + \max \{0, \hat{z}^l(p)\}}{1 + \sum_{k=1}^{L} \max \{0, \hat{z}^k(p)\}}
\] for \( l = 1, 2, \ldots, L \).

First note that \( g \) is a continuous function since \( \max \{f(x), h(x)\} \) is continuous when \( f(x) \) and \( h(x) \) are continuous.

By Brouwer’s fixed point theorem, there exists a price vector \( p^* \) such that
\[
g(p^*) = p^*, \quad i.e.,
\]
\[
p^l = \frac{p^l + \max \{0, \hat{z}^l(p^*)\}}{1 + \sum_{k=1}^{L} \max \{0, \hat{z}^k(p^*)\}}
\] for \( l = 1, 2, \ldots, L \). (10.17)

We want to show \( p^* \) is in fact a competitive equilibrium price vector.

Cross multiplying \( 1 + \sum_{k=1}^{L} \max \{0, \hat{z}^k(p^*)\} \) on both sides of (10.17), we have
\[
\sum_{k=1}^{L} \max \{0, \hat{z}^k(p^*)\} = \max \{0, \hat{z}^l(p^*)\}.
\] (10.18)

Then, multiplying the above equation by \( \hat{z}^l(p^*) \) and making summation, we have
\[
\left[ \sum_{l=1}^{L} p^l \hat{z}^l(p^*) \right] \left[ \sum_{l=1}^{L} \max \{0, \hat{z}^l(p^*)\} \right] = \sum_{l=1}^{L} \hat{z}^l(p^*) \max \{0, \hat{z}^l(p^*)\}.
\] (10.19)

Then, by weak Walras’ Law, \( \sum_{l=1}^{L} p^l \hat{z}^l(p^*) \leq 0 \), and \( \sum_{l=1}^{L} \max \{0, \hat{z}^l(p^*)\} \geq 0 \), we must have
\[
\sum_{l=1}^{L} \hat{z}^l(p^*) \max \{0, \hat{z}^l(p^*)\} \leq 0.
\] (10.20)

Since each term of the summations is either zero or \( (\hat{z}^l(p^*))^2 > 0 \), to have the summation to be less than or equal to zero, we must have each term to be zero. That is, \( \hat{z}^l(p^*) \leq 0 \) for \( l = 1, \ldots, L \). The proof is completed. \( \blacksquare \)

The construction of function \( g(p) \) is depicted in the following figure 10.10 for the case \( L = 2 \).

Remark 10.4.1 Do not confuse competitive equilibrium of an aggregate excess demand function with strong Walras’ Law. Even though strong Walras’ Law
Figure 10.10: the fixed-point function $g(p)$.

holds, we may not have $\dot{z}(p) \leq 0$ for all $p$. Also, if $\dot{z}(p^*) \leq 0$ for some $p^*$, i.e., $p^*$ is a competitive equilibrium price vector, strong Walras’ Law may not hold unless some types of monotonicity are imposed such as local non-satiation.

**Fact 10.4.1 (free goods)** Under strong Walras’ Law, if $p^*$ is a competitive equilibrium price vector and $\dot{z}^l(p^*) < 0$, then $p^d = 0$.

**Proof.** Suppose not. Then $p^d > 0$. Thus, $p^d \dot{z}^l(p^*) < 0$, and so $p^* \cdot \dot{z}(p^*) < 0$, contradicting strong Walras’ Law.

**Definition 10.4.1 (Desirable goods)** we call good $l$ desirable good if $p^d = 0$ implies $\dot{z}^l(p^*) > 0$.

**Fact 10.4.2 (Equality of demand and supply)** If all goods are desirable and $p^*$ is a competitive equilibrium price vector, then $p^* > 0$ and $\dot{z}(p^*) = 0$.

**Proof.** Since goods are desirable, we must have the equilibrium $p^* > 0$. Now we show that $\dot{z}(p^*) = 0$. Suppose not. We then have $\dot{z}^l(p^*) < 0$ for some $l$. Then, by Fact 10.4.1, we have $p^d = 0$. Since good $l$ is desirable, we must have $\dot{z}^l(p^*) > 0$, a contradiction.
Remark 10.4.2 By strong Walras’ Law, if \( p > 0 \), and if \((L - 1)\) markets are in the equilibrium, the \(L\)-th market is also in the equilibrium. Thus, because of strong Walras’s Law, to verify that a price vector \( p > 0 \) clears all markets, it suffices check that it clear all markets but one.

The above existence theorem assume that the excess demand function is well defined for all prices in the closed unit simplex \( S^{L-1} \), including zero prices. However, when preferences are strictly monotone, excess demand functions are not well defined on the boundary of \( S^{L-1} \) so that the above existence theorem cannot be applied. Then, we need an existence theorem for strictly positive prices, which is given below.

Theorem 10.4.2 (The Existence Theorem II)) For a private ownership economy \( e = (\{X_i, w_i, \succ_i\}, \{Y_j\}, \\{\theta_{ij}\}) \), suppose the aggregate excess demand function \( \hat{\mathbf{z}}(\mathbf{p}) \) is defined for all strictly positive price vectors \( \mathbf{p} \in \mathbb{R}^{L}_{++} \), and satisfies the following conditions

(i) \( \hat{\mathbf{z}}(\cdot) \) is continuous;

(ii) \( \hat{\mathbf{z}}(\cdot) \) is homogeneous of degree zero;

(iii) \( \mathbf{p} \cdot \hat{\mathbf{z}}(\mathbf{p}) = 0 \) for all \( \mathbf{p} \in \mathbb{R}^{L}_{++} \) (interior Walras’ Law);

(iv) There is an \( s > 0 \) such that \( \hat{\mathbf{z}}^l(\mathbf{p}) > -s \) for every commodity \( l \) and all \( \mathbf{p} \in \mathbb{R}^{L}_{++} \);

(v) If \( \mathbf{p}_k \to \mathbf{p} \), where \( \mathbf{p} \neq 0 \) and \( p^l = 0 \) for some \( l \), then

\[
\max\{\hat{\mathbf{z}}^1(\mathbf{p}_k), \ldots, \hat{\mathbf{z}}^L(\mathbf{p}_k)\} \to \infty.
\]

Then there is \( \mathbf{p}^* \in \mathbb{R}^{L}_{++} \) such that

\[
\hat{\mathbf{z}}(\mathbf{p}^*) = 0 \quad (10.21)
\]

and thus \( \mathbf{p}^* \) is a competitive equilibrium.
**Proof.** Because of homogeneity of degree zero we can restrict our search for an equilibrium in the unit simplex $S$. Denote its interior by $\text{int}S$. We want to construct a correspondence $F$ from $S$ to $S$ such that any fixed point $p^*$ of $F$ is a competitive equilibrium, i.e., $p^* \in F(p^*)$ implies $\hat{z}(p^*) = 0$.

Define a correspondence $F : S \to 2^S$ by,

\[
F(p) = \begin{cases} 
\{ q \in S : \hat{z}(p) \cdot q \geq \hat{z}(p) \cdot q' \text{ for all } q' \in S \} & \text{if } p \in \text{int}S \\
\{ q \in S : p \cdot q = 0 \} & \text{if } p \text{ is on the boundary}
\end{cases}
\]

The construction of correspondence $F(\cdot)$ is depicted in the following figures 10.11 and 10.12. Note that, for $p \in \text{int}S$, $F(\cdot)$ means that, given the current "proposal" $p \in \text{int}S$, the "counterproposal" assigned by the correspondence $F(\cdot)$ is any price vector $q$ that maximizes the values of the excess demand vector among the permissible price vectors in $S$. Here $F(\cdot)$ can be thought as a rule that adjusts current prices in a direction that eliminates any excess demand, the correspondence $F(\cdot)$ assigns the highest prices to the commodities that are most in excess demand. In particular, we have

\[
F(p) = \{ q \in S : q' = 0 \text{ if } \hat{z}'(p^*_k) < \max\{\hat{z}^1(p^*_k), \ldots, \hat{z}^L(p^*_k)\} \}.
\]

Observe that if $\hat{z}(p) \neq 0$ for $p \in \text{int}S$, then because of interior Walras’ law we have $\hat{z}'(p) < 0$ for some $l$ and $\hat{z}''(p) > 0$ for some $l' \neq l$. Thus, for such a $p$,
any \( q \in F(p) \) has \( q'_l = 0 \) for some \( l \) (to maximize the values of excess demand vectors). Therefore, if \( \hat{z}(p) \neq 0 \) then \( F(p) \subset \text{Boundary } S = S \setminus \text{Int} \, S \). In contrast, \( \hat{z}(p) = 0 \) then \( F(p) = S \).

Now we want to show that the correspondence \( F \) is an upper hemi-continuous correspondence with nonempty, convex, and compact values. First, note that by the construction, the correspondence \( F \) is clearly compact and convex-valued. \( F(p) \) is also non-empty for all \( p \in S \). Indeed, when \( p \in \text{Int} \, S \), any price vector \( q \) that maximizes the value of \( q' \cdot \hat{z}(p) \) is in \( F(p) \) and so it is not empty. When \( p \) is on the boundary of \( S \), \( p'_l = 0 \) for at least some good \( l \) and thus there exists a \( q \in S \) such that \( p \cdot q = 0 \), which implies \( F(p) \) is also non-empty.

Now we show the correspondence \( F \) is upper hemi-continuous, or equivalently it has closed graph, i.e., for any sequences \( p_t \to p \) and \( q_t \to q \) with \( q_t \in F(p_t) \) for all \( t \), we have \( q \in F(p) \). There are two cases to consider.

Case 1. \( p \in \text{Int} \, S \). Then \( p_k \in \text{Int} \, S \) for \( k \) sufficiently large. From \( q_k \cdot \hat{z}(p_k) \geq q' \cdot \hat{z}(p_k) \) for all \( q' \in S \) and the continuity of \( \hat{z}(\cdot) \), we get \( q \cdot \hat{z}(p) \geq q' \cdot \hat{z}(p) \) for all \( q' \in S \), i.e., \( q \in F(p) \).

Case 2. \( p \) is a boundary point of \( S \). Take any \( l \) with \( p'_l > 0 \). We should argue that for \( k \) sufficiently large, we have \( q'_k = 0 \), and therefore it must be that \( q'_l = 0 \);
from this $q \in F(p)$ follows. Because $p^I > 0$, there is an $\epsilon > 0$ such that $p^I_k > \epsilon$
for $k$ sufficiently large. If, in addition, $p_k$ is on the boundary of $S$, then $q^I_k = 0$
by the definition of $F(p_k)$. If, instead, $p_k \in intS$, then by conditions (iv) and (v), for $k$
sufficiently large, we must have
\[
\hat{z}^I(p_k) < \max\{\hat{z}^1(p_k), \ldots, \hat{z}^L(p_k)\}
\]
and therefore that, again, $q^I_k = 0$. To prove the above inequality, note that by
condition (v) the right-hand side of the above expression goes to infinity with $k$
(because $p$ is a boundary point of $S$, some prices go to zero as $k \to \infty$). But the
left-hand side is bounded above because if it is positive then
\[
\hat{z}^I(p_k) \leq \frac{1}{\epsilon} p^I_k \hat{z}^I(p_k) = -\frac{1}{\epsilon} \sum_{l' \neq i} p^I_k \hat{z}^I(p_k) < \frac{1}{\epsilon} \sum_{l' \neq i} p^I_k < \frac{s}{\epsilon},
\]
where $s$ is the bound in excess supply given by condition (iv). Thus, we have
shown that for any $l$ with $p^I_l > 0$, when $k$ is sufficiently large, we have $q^I_k = 0$, and
therefore it must be that $q^I = 0$. But this guarantees $p \cdot q = 0$ and so $q \in F(p)$.
So $F$ must be upper hemi-continuous.

Thus the correspondence $F$ is an upper hemi-continuous correspondence with
nonempty, convex, and compact values. Therefore, by Kakutani’s fixed point
theorem (see section 1.2.8), we conclude that there is $p^* \in S$ with $p^* \in F(p^*)$.

Finally, we show that any fixed point $p^*$ of $F$ is a competitive equilibrium.
Suppose that $p^* \in F(p^*)$. Then $p^*$ cannot be a boundary point of $S$ because
$p \cdot p > 0$ and $p \cdot q = 0$ for all $q \in F(p)$ cannot occur simultaneously, and thus
$p^* \in intS$. If $\hat{z}(p^*) \neq 0$, then we already showed above that $F(p^*)$ must be a
subset of boundary points of $S$, which contradicts the fact that $p^*$ in an interior
point. Hence, if $p^* \in F(p^*)$, we must have $\hat{z}(p^*) = 0$. The proof is completed. \blacksquare

Notice that Theorem II applies to settings with strictly increasing preferences, but Theorem I does not.
### 10.4.2 The existence of CE under individual characteristics

All preassumptions in the above existence theorems are about aggregate excess demand functions. These conditions are less intuitive and are not easy to check. In this section we are trying to recover the individual behavioral properties from the analytical properties of aggregate excess demand functions. Under what conditions on individual characteristics will the budget constraint hold with equality and a single-valued and continuous aggregate excess demand correspondence is obtained. We will give these characteristics in this section.

In the above two theorems, (strong or interior) Walras’ Law is important to prove the existence of a competitive equilibrium. Under which conditions, is Walras’ Law held with equality?

- **Conditions for Walras’ Law to be held with equality**

When each consumer’s budget constraint holds with equality:

\[
p \cdot x_i(p) = p \cdot w_i + \sum_{j=1}^{J} \theta_{ij} p \cdot y_j(p)
\]

for all \(i\), we have

\[
\sum_{i=1}^{n} p \cdot x_i(p) = \sum_{i=1}^{n} p \cdot w_i + \sum_{i=1}^{n} \sum_{j=1}^{J} \theta_{ij} p \cdot y_j(p)
\]

\[
= \sum_{i=1}^{n} p \cdot w_i + \sum_{j=1}^{J} p \cdot y_j(p)
\]

which implies that

\[
p \left[ \sum_{i=1}^{n} x_i(p) - \sum_{i=1}^{n} w_i - \sum_{j=1}^{J} y_j(p) \right] = 0 \quad (10.22)
\]

so that

\[
p \cdot \hat{z}(p) = 0 \quad (Walras'/Law) \quad (10.23)
\]
Thus, as long as the budget line holds with equality and aggregate excess demand is well-defined on the domain of prices, Walras’ Law must hold.

The questions are under what conditions on economic environments a budget constraint holds with equality for all optimal consumption bundles, and the aggregate excess demand correspondence is single-valued or convex-valued? The following various types of monotonicities and convexities of preferences may be used to answer these questions.

- Types of monotonicity conditions

  1. **Strong monotonicity**: For any two consumption market bundles \((x \geq x')\) with \(x \neq x'\) implies \(x \succ_i x'\). A consumer obtains a larger utility level when he consumes a larger amount of at least one good.

  2. **Monotonicity**: if \(x > x'\) implies that \(x \succ_i x'\). A consumer obtains a larger utility level when he consumes a larger amount of all goods.

  3. **Weak Monotonicity**: if \(x \geq y\) implies \(x \succ y\). Every commodity is “good” rather than “bad”.

  4. **Local non-satiation**: For any point \(x\) and any neighborhood, \(N(x)\), there is \(x' \in N(x)\) such that \(x' \succ_i x\). For an arbitrary consumption bundle, a strictly better bundle will be found in an arbitrary neighbourhood of it.

  5. **Non-satiation**: For any \(x\), there exists \(x'\) such that \(x' \succ_i x\). A consumer could always strictly increase his utility through a change of his consumption bundle. Notice that local non-satiation properties says that a consumer could always strictly increase his utility through a minor change of his consumption bundle.
Remark 10.4.3 Strong monotonicity of preferences can be interpreted as individuals’ desires for goods: the more, the better. Local non-satiation means individuals’ desires are unlimited. Note that strong monotonicity implies monotonicity and weak monotonicity, monotonicity implies local non-satiation. In general, monotonicity does not imply weak monotonicity unless preferences are continuous.

- Types of convexities

Suppose $X$ is convex. The following four types of convexities are used in the textbooks and literature:

(i) **Strict convexity**: For any $x$ and $x'$ in $X$ with $x \succ_i x'$ and $x \neq x'$, $x_\lambda \equiv \lambda x + (1 - \lambda)x' \succ_i x'$ for $\lambda \in (0, 1)$.

![Figure 10.13: Strong convex indifference curves.](image)

(ii) **Strong convexity**: if $x'_i \sim_i x_i$ then $\lambda x_i + (1 - \lambda)x'_i \succ_i x_i, \forall \lambda \in (0, 1)$;

(iii) **Convexity**: If $x \succ_i x'$, then $x_\lambda = \lambda x + (1 - \lambda)x' \succ_i x'$ for $\lambda \in (0, 1)$. 

362
(iv) **Weak convexity**: If \( x \succeq_i x' \), then \( x_\lambda \succeq_i x', \forall \lambda \in (0, 1) \).

“Weak convexity” adopted my lecture notes, is usually referred as “Convexity” in most existing textbooks such as Varian (1992), Mas-colell, Whinston, and
Green (1995). The concepts “convexity” and “strong convexity” used in my lecture notes are both originated from Debreu (1959). When preference orderings are convex, the indifference curve may include segments of straight line, but strong convexity does not encompass this case. When preferences are weakly convex, the indifference curve may be “thick”, but this case is incompatible with “convexity” and “strong convexity”. Obviously, strict convexity implies strong convexity, convexity and weak convexity, but convexity does not necessarily imply weak convexity. However, under continuity of preferences, strong convexity implies strict convexity (and they are equivalent), they both implies convexity, which in turn implies weak convexity.

**Remark 10.4.4** The convexity of preferences implies that people want to diversify their consumptions, and thus, convexity can be viewed as the formal expression of basic measure of economic markets for diversification. Note that the strict convexity of $\succ_i$ implies the conventional diminishing marginal rates of substitution (MRS), and weak convexity of $\succeq_i$ is equivalent to the quasi-concavity of utility function $u_i$. Also notice that the continuity of $\succ_i$ is a sufficient condition for the continuous utility representations, that is, it guarantees the existence of continuous utility function $u_i(\cdot)$.

**Remark 10.4.5** Under the convexity of preferences $\succeq_i$, non-satiation implies local non-satiation. Why? The proof is left to readers.

Now we are ready to answer under which conditions Walras’s Law holds with equality, a demand correspondence can be convex-valued or a function. The following propositions answer the questions. The proofs are left to readers.

**Proposition 10.4.1** Under local non-satiation assumption, the budget constraint holds with equality for every optimal consumption bundle, and thus Walras’s Law holds with equality.

**Proposition 10.4.2** Under the strict convexity of $\succ_i$, $x_i(p)$ becomes a (single-valued) function.
Proposition 10.4.3 Under the weak convexity of preferences, demand correspondence \( x_i(p) \) is convex-valued.

We next explore the condition under which supply correspondence is single-valued.

Definition 10.4.2 (Strict convexity of production set) If \( y^1_j \in Y_j \) and \( y^2_j \in Y_j \), then the convex combination \( \lambda y^1_j + (1 - \lambda)y^2_j \in int Y_j \) for all \( 0 < \lambda < 1 \), where \( int Y_j \) denotes the interior points of \( Y_j \).

The proof of the following proposition is based on the maximum theorem.

Proposition 10.4.4 If \( Y_j \) is compact (i.e., closed and bounded) and strictly convex, then the supply correspondence \( y_j(p) \) is a well defined single-valued and continuous function.

Proof. By the maximum theorem, we know that \( y_j(p) \) is a non-empty valued upper hemi-continuous correspondence by the compactness of \( Y_j \) (by noting that \( 0 \in Y_j \)) for all \( p \in \mathbb{R}^L_+ \).

Now we show it is single-valued. Suppose not. \( y^1_j \) and \( y^2_j \) are two profit maximizing production plans for \( p \in \mathbb{R}^L_+ \), and thus \( p \cdot y^1_j = p \cdot y^2_j \). Then, by the strict convexity of \( Y_j \), we have \( \lambda y^1_j + (1 - \lambda)y^2_j \in int Y_j \) for all \( 0 < \lambda < 1 \). Therefore, there exists some \( t > 1 \) such that

\[
 t[\lambda y^1_j + (1 - \lambda)y^2_j] \in int Y_j. \tag{10.24}
\]

Then \( t[\lambda p \cdot y^1_j + (1 - \lambda)p y^2_j] = tp \cdot y^1_j > p \cdot y^1_j \) which contradicts the fact that \( y^1_j \) is a profit maximizing production plan.

So \( y_j(p) \) is a single-valued function. Thus, by the upper hemi-continuity of \( y_j(p) \), we know it is a single-valued and continuous function. ❑
Figure 10.16: The upper contour set $U_w(x_i)$ is given by all points above the indifference curve, and the lower contour set $L_w(x_i)$ is given by all points below the indifference curve in the figure.

Next, we answer under what conditions the Walras’ Law holds, and the aggregate excess demand correspondence is continuous and single-valued.

**Proposition 10.4.5** Suppose the following conditions hold for all consumers $i$

$(i) \ X_i = \mathbb{R}_+^L$;

$(ii) w_i \geq 0, \sum_{i \in N} w_i > 0$;

$(iii) \succ_i$ is continuous, strictly convex, locally non-satiated;

$(iv) Y_j$ is compact strictly convex setand $0 \in Y_j$ $j = 1, 2, \ldots, J$.

Then, for $\forall p \in \mathbb{R}_+^L$, we have

$(1)$ individual demand correspondence $x_i(p)$, and thus the aggregate excess demand correspondence is single-valued continuous function:

$(2)$ interior Walras’ Law hold
Proof. First, by strict convexity, we know \( x_i(p) \) is a single-valued function. Also, for all \( w_i \geq 0 \) and \( p > 0 \), the budget constrained set \( B_i(p) = \{ x_i \in X_i : p \cdot x_i \leq p \cdot w_i + \sum_{j=1}^{J} \theta_{ij} p y_j(p) \} \) is clearly non-empty and compact for all \( p \in \mathbb{R}_{++}^L \). In addition, since \( w_i \geq 0 \) and \( p > 0 \), \( B_i \) is a continuous correspondence. Indeed, \( B_i \) is clearly upper hemi-continuous. We only need to show that \( B_i \) is also lower hemi-continuous at every \( p \in \mathbb{R}_{++}^L \). Let \( p \in S \), \( x_i \in B_i(p) \), and let \( \{ p_t \} \) be a sequence such that \( p_t \to p \). We want to prove that there is a sequence \( \{ x_t \} \) such that \( x_t \to x \), and, for all \( t \), \( x_t \in B_i(p_t) \), i.e., \( x_t \in X_i \) and \( p_t x_t^i \leq p_t w_i + \sum_{j=1}^{J} \theta_{ij} p_t y_j(p_t) \). Two cases will be considered.

Case 1. \( p \cdot x_i < p \cdot w_i + \sum_{j=1}^{J} \theta_{ij} p y_j(p) \). Then, for any sequence \( \{ x_t \} \) such that \( x_t \to x \), we have \( p_t x_t^i < p_t w_i + \sum_{j=1}^{J} \theta_{ij} p_t y_j(p_t) \) for all \( t \) larger than a certain integer \( t' \), and thus \( x_t \in B_i(p_t) \).

Case 2. \( p \cdot x_i = p \cdot w_i + \sum_{j=1}^{J} \theta_{ij} p y_j(p) \). Let \( I_t^i = p_t w_i + \sum_{j=1}^{J} \theta_{ij} p_t y_j(p_t) \). Since \( y_j(p_t) \) is a profit maximizing production plan and also continuous \( y_j(p_t) \geq 0 \) and thus \( I_t^i > 0 \) for all \( t \) and is continuous in \( p_t \). Now, let \( x_t^i = \frac{I_t^i}{p_t x_i} x_i \). Since \( \frac{I_t^i}{p_t x_i} \to \frac{L}{p x_i} = 1 \), we have \( x_t^i \to x_i \). Also, since \( p_t x_t^i = \frac{I_t^i}{p_t x_i} p_t x_i = I_t^i \), we have \( x_t^i \in B_i(p_t) \). Therefore, the sequence \( \{ x_t \} \) has all the desired properties. Thus \( B_i \) is lower hemi-continuous at every \( p \in \mathbb{R}_{++}^L \). Hence, \( B_i \) is a continuous correspondence with non-empty and compact values. By assumption, \( \succ_i \) is also continuous. Then, by the maximum theorem, we know the demand correspondence \( x_i(p) \) is upper hemi-continuous. Furthermore, by the strict convexity of preferences, it is single-valued and continuous. Finally, by local non-satiation, we know the budget constraint holds with equality, and thus Walras’s Law is satisfied.

When \( w_i > 0 \), \( p \in \mathbb{R}_{++}^L \) can be weakened to \( p \in \mathbb{R}_+^L \) so that consumers income is still positive for \( \forall p \in \mathbb{R}_{++}^L \), and therefore the correspondence \( B_i \) is therefore nonempty and compact-valued. The aggregate excess demand correspondence \( z(p) \) is therefore a continuous function. From this propositions and the Theorem II above, we have the following existence theorem that provides sufficient
conditions directly based on the fundamentals of the economy.

**Theorem 10.4.3 (Existence Theorem III)** For a private ownership economy

\[ e = \left( \{ X_i, w_i, \succ_i \}_{i \in N}, \{ Y_j \}_{j \in J}, \{ \theta_{ij} \}_{(i,j) \in (N \times J)} \right), \]

there exists a competitive equilibrium if the following conditions hold

(i) \( X_i = \mathbb{R}_+^L \);

(ii) \( w_i \geq 0, \sum_{i \in N} w_i > 0 \);

(iii) \( \succ_i \) are continuous, strictly convex, and strongly monotone;

(iv) \( Y_j \) are compact, strictly convex, \( 0 \in Y_j \), \( j = 1, 2, \ldots, J \).

**Proof.** By the assumption imposed, we know that for \( \forall p \in \mathbb{R}_+^L \), \( x_i(p) \) and \( y_j(p) \) are continuous and single-valued. Thus the aggregate excess demand function is a continuous single-valued function and satisfies interior Walras’ Law by monotonicity of preferences. Thus, we only need to show conditions (iv) and (v) in Theorem II are also satisfied. The bound in (iv) follows from the nonnegativity of demand (i.e., the fact that \( X_i = \mathbb{R}_+^L \)) and bounded production sets, which implies that a consumer’s total net supply to the market of any good \( l \) can be no greater than the sum of his initial endowment and upper bound of production sets. Finally, we show that condition v is satisfied. As some prices go to zero, a consumer whose wealth tends to a strictly positive limit (note that, because \( p \hat{w} > 0 \), there must be at least one such consumer) and with strong monotonicity of preferences will demand an increasing large amount of some of the commodities whose prices go to zero. Hence, by Theorem II, there is a competitive equilibrium. The proof is completed.

Conditions of the above Theorem are more intuitive and easier to be check, since they are based on individual characteristics (preference, endowment, technology) rather on analytical properties of excess demand function (homogeneity, continuous, Walras’ Law, ect.)
In Theorem 10.4.5, it requires only local non-satiation to obtain the continuity of excess demand function. However, in Theorem (10.4.2), the strong monotonicity condition cannot weakened to be monotonicity or local non-satiation, which cannot guarantee condition (v) in Theorem II. Example 10.3.2 gives such an counterexample. Nevertheless, if all initial endowments are assumed to be interior so that $p \hat{w} > 0$, the strong monotonicity can be weakened to monotonicity. In this case, for every commodity, there is some consumer whose demand will go to infinity as its price goes to zero, and thus condition (v) is satisfied. We then have the proposition that is a slight variation of Existence Theorem III.

**Proposition 10.4.6 (Existence Theorem III')** For a private ownership economy

$$e = \{\{X_i, w_i, \succ_i\}_{i \in N}, \{Y_j\}_{j \in J}, \{\theta_{ij}\}_{(i,j) \in (N \times J)}\},$$

there exists a competitive equilibrium if the following conditions hold

(i) $X_i = \mathbb{R}^L_+$;

(ii) $w_i > 0$;

(iii) $\succ_i$ are continuous, strictly convex, and monotone;

(iv) $Y_j$ are compact, strictly convex, $0 \in Y_j$, $j = 1, 2, \ldots, J$.

**10.4.3 Examples of Computing CE**

Next, we give some examples of computing CE. It is worth mentioning that general equilibrium theory is based on competitive market, every individual (either producer or consumer) is required to be a price taker, which is obvious unrealistic for the cases with small number of agents. Therefore, the theory is only an approximation rather full and detailed depiction of the real world.

In the following examples, utility functions are assumed to be Cobb-Douglas, initial endowments are assumed to be on the boundary. From Walras’ Law, we need only to consider the clearance condition of one market. The price of good 1 is normalized to be one. For this example, it is easy to get the demand function
of good 1, then equilibrium price is obtained from market clearance condition of
good 1. This example shows that CE may exist even the endowment is on the
boundary of consumption space.

**Example 10.4.1** Consider an exchange economy with two consumers and two
goods with

\[
\begin{align*}
  w_1 &= (1, 0), & w_2 &= (0, 1) \\
  u_1(x_1) &= (x_1^1)^a(x_1^2)^{1-a}, & 0 < a < 1 \\
  u_2(x_2) &= (x_2^1)^b(x_2^2)^{1-b}, & 0 < b < 1
\end{align*}
\]

Let \( p = \frac{p_2}{p_1} \)

Consumer 1’s problem is to solve

\[
\max_{x_1} u_1(x_1)
\]

subject to

\[
x_1^1 + px_1^2 = 1.
\]

Since utility functions are Cobb-Douglas types of functions, the solution is then given by

\[
x_1^1(p) = \frac{a}{1} = a, \quad (10.28)
\]

\[
x_1^2(p) = \frac{1-a}{p}. \quad (10.29)
\]

Consumer 2’s problem is to solve

\[
\max_{x_2} u_2(x_2)
\]

subject to

\[
x_2^1 + px_2^2 = p.
\]

The solution is given by

\[
x_2^1(p) = \frac{b \cdot p}{1} = b \cdot p \quad (10.32)
\]

\[
x_2^2(p) = \frac{(1-b)p}{p} = (1-b). \quad (10.33)
\]

370
Then, by the market clearing condition,
\[ x_1^1(p) + x_1^2(p) = 1 \Rightarrow a + bp = 1, \quad (10.34) \]
and thus the competitive equilibrium is given by
\[ p = \frac{p^2}{p^1} = \frac{1 - a}{b}. \quad (10.35) \]
This is true because, by interior Walras’s Law, for \( L = 2 \), it is enough to show only one market clearing.

**Remark 10.4.6** Since the Cobb-Douglas utility function is widely used as an example of utility functions that have nice properties such as strong monotonicity on \( \mathbb{R}_{++}^L \), continuity, and strict quasi-concavity, it is useful to remember the functional form of the demand function derived from the Cobb-Douglas utility functions. It may be remarked that we can easily derive the demand function for the general function: \( u_i(x_i) = (x_1^i)^\alpha (x_2^i)^\beta \quad \alpha > 0, \beta > 0 \) by the a suitable monotonic transformation. Indeed, by invariant to monotonic transformation of utility function, we can rewrite the utility function as
\[ [(x_1^i)^\alpha (x_2^i)^\beta]\frac{1}{\alpha + \beta} = (x_1^i)^{\frac{\alpha}{\alpha + \beta}} (x_1^i)^{\frac{\beta}{\alpha + \beta}} \quad (10.36) \]
so that we have
\[ x_1^1(p) = \frac{\alpha}{\alpha + \beta} \frac{I}{p^1} \quad (10.37) \]
and
\[ x_1^2(p) = \frac{\beta}{\alpha + \beta} \frac{I}{p^2} \quad (10.38) \]
when the budget line is given by
\[ p^1 x_1^1 + p^2 x_1^2 = I. \quad (10.39) \]

We have give example that CE does not exist when endowment is located on the boundary. The above existence results are also obtained for interior endowment. Nevertheless, this example shows that interior endowment is not necessary for the existence of CE.

In the following example, the second consumer is assumed to have Leontief utility function. All other assumptions remain the same as the above example.
Example 10.4.2

\[ n = 2, \quad L = 2 \]
\[ w_1 = (1, 0), \quad w_2 = (0, 1) \]
\[ u_1(x_1) = (x_1^1)^a(x_1^2)^{1-a}, \quad 0 < a < 1 \]
\[ u_2(x_2) = \min\{x_2^1, bx_2^2\}, \quad \text{with } b > 0 \]

For consumer 1, we have already obtained
\[ x_1^1(p) = a, \quad x_1^2 = \frac{(1 - a)}{p}. \quad (10.41) \]

For consumer 2, his problem is to solve:
\[ \max_{x_2} u_2(x_2) \quad (10.42) \]
\[ \text{s.t.} \]
\[ x_2^1 + px_2^2 = p. \quad (10.43) \]

At the optimal consumption, we have
\[ x_2^1 = bx_2^2. \quad (10.44) \]

By substituting the solution into the budget equation, we have \( bx_2^2 + px_2^2 = p \) and thus \( x_2^2(p) = \frac{p}{b+bp} \) and \( x_1^2(p) = \frac{bp}{b+bp} \).

Then, by \( x_1^1(p) + x_2^1(p) = 1 \), we have
\[ a + \frac{bp}{b+p} = 1 \quad (10.45) \]
or
\[ (1 - a)(b + p) = bp \quad (10.46) \]
so that
\[ (a + b - 1)p = b(1 - a) \quad (10.47) \]

Thus,
\[ p^* = \frac{b(1 - a)}{a + b - 1}. \]

To make \( p^* \) be a competitive equilibrium price, we need to assume \( a + b > 1 \).
Here, the price must be nonnegative since the preference is monotonic. If we only consider the clearance of market 1, the price may be negative. To get a nonnegative price, the demand elasticities of two goods must sum up to one. This implies that for some ranges of parameters, CE may not exist. This is originated from the non-interior endowment and the fact that consumer 2’s utility is not strongly monotonic. Again, the assumption on interior endowments or strong monotonic preferences is indispensable. It is easy to see that this example violates condition (v) of Existence Theorem II.

When $p^1 \to 0$ (that is, $p^2 = \frac{p^2}{p^1} \to \infty$), we have

$$\lim_{p^1 \to 0} \hat{z}^1(p) = a + b - 1,$$

and

$$\lim_{p^1 \to 0} \hat{z}^2(p) = 0,$$

Both of them does not diverge to infinity, so condition (v) is violated.

**10.4.4 The existence of CE for discontinuous demand function**

The above result requires the aggregate excess demand function to be continuous. But in many situations, preferences/ production sets may not be convex, so that demand or supply functions are not continuous. By using the KKM lemma, we can prove the existence of competitive equilibrium by only assuming the aggregate excess demand function to be lower semi-continuous.

**Theorem 10.4.4 (The Existence Theorem IV)** For a private ownership economy $e = (\{X_i, w_i, \succ_i\}_{i \in N}, \{Y_j\}_{j \in J}, \{\theta_{ij}\}_{(i,j) \in N \times J})$, if the aggregate excess demand function $\hat{z}(p)$ is a lower semi-continuous function and satisfies weak Walras’ Law, then there exists a competitive equilibrium, that is, there is $p^* \in S^{k-1}$ such that

$$\hat{z}(p^*) \leq 0. \quad (10.48)$$
Proof. Define a correspondence $F : S^{L-1} \to 2^{S^{L-1}}$ by,

$$F(q) = \{ p \in S^{L-1} : q \cdot \hat{z}(p) \leq 0 \}$$

for all $q \in S^{L-1}$.

First note that $F(q)$ is nonempty for each $q \in S^{L-1}$ since $q \in F(q)$ by weak Walras’ Law. Since $p \geq 0$ and $\hat{z}(\cdot)$ is lower semi-continuous, the function defined by $\phi(q, p) \equiv q \cdot \hat{z}(p) = \sum_{l=1}^L q^l \hat{z}^l(p)$ is lower semi-continuous in $p$. Hence, the set $F(q)$ is closed for all $q \in S^{L-1}$. We now prove $F$ is FS-convex. Suppose, by way of contradiction, that there some $q_1, \ldots, q_m \in S^{L-1}$ and some convex combination $q_\lambda = \sum_{t=1}^m \lambda_t q_t$ such that $q_\lambda \notin \bigcup_{t=1}^m F(q_t)$. Then, $q_\lambda \notin F(q_t)$ and therefore $q_\lambda \cdot \hat{z}(q_\lambda) > 0$ for all $t = 1, \ldots, m$. Thus, $\sum_{t=1}^m \lambda_t q_\lambda \cdot \hat{z}(q_\lambda) = q_\lambda \cdot \hat{z}(q_\lambda) > 0$ which contradicts the fact that $\hat{z}$ satisfies weak Walras’ Law. So $F$ must be FS-convex. Therefore, by KKM lemma, we have

$$\bigcap_{q \in S} F(q) \neq \emptyset.$$  

Then there exists a $p^* \in S^{L-1}$ such that $p^* \in \bigcap_{q \in S^{L-1}} F(q)$, i.e., $p^* \in F(q)$ for all $q \in S^{L-1}$. Thus, $q \cdot \hat{z}(p^*) \leq 0$ for all $q \in S^{L-1}$. Now let $q_1 = (1, 0, \ldots, 0)$, $q_2 = (0, 1, 0, \ldots, 0)$, and $q_n = (0, \ldots, 0, 1)$. Then $q_t \in S^{L-1}$ and thus $q_t \cdot \hat{z}(p^*) = \hat{z}^t(p^*) \leq 0$ for all $t = 1, \ldots, L$. Thus we have $\hat{z}(p^*) \leq 0$, which means $p^*$ is a competitive equilibrium price vector. The proof is completed.  

The above existence theorems only give sufficient conditions. Recently, Tian (2016) provides a complete solution to the existence of competitive equilibrium in economies with general excess demand functions, in which commodities may be indivisible and excess demand functions may be discontinuous or do not have any structure except Walras’ law. 

Tian introduces a very weak condition, called recursive transfer lower semi-continuity, which is weaker than transfer lower semi-continuity and in turn weaker than lower semi-continuity. It is shown that the condition, together with Walras’s law, guarantees the existence of price equilibrium in economies with excess demand functions. The condition is also necessary, and thus our results generalize all the existing results on the existence of price equilibrium in economies where
excess demand is a function. For convenience of discussion, we introduce the following term.

We say that price system \( p \) upsets price system \( q \) if \( q \)'s excess demand is not affordable at price \( p \), i.e., \( p \cdot \hat{z}(q) > 0 \).

**Definition 10.4.3** An excess demand function \( \hat{z}(\cdot) : S^{L-1} \rightarrow \mathbb{R}^L \) is transfer lower semi-continuous if for all \( q, p \in S^{L-1} \), \( p \cdot \hat{z}(q) > 0 \) implies that there exists some point \( p' \in X \) and some neighborhood \( N(q) \) of \( q \) such that \( p' \cdot \hat{z}(q') > 0 \) for all \( q' \in N(q) \).

**Remark 10.4.7** The transfer lower semi-continuity of \( \hat{z}(\cdot) \) means that, if the aggregate excess demand \( \hat{z}(q) \) at price vector \( q \) is not affordable at price system \( p \), then there exists a price system \( p' \) such that \( \hat{z}(q') \) is also not affordable at price system \( p' \), provided \( q' \) is sufficiently close to \( q \). Note that, since \( p \geq 0 \), this condition is satisfied if \( \hat{z}(\cdot) \) is lower semi-continuous by letting \( p' = p \).

**Definition 10.4.4** (Recursive Upsetting) Let \( \hat{z}(\cdot) : S^{L-1} \rightarrow \mathbb{R}^L \) be an excess demand function. We say that a non-equilibrium price system \( p^0 \in S^{L-1} \) is recursively upset by \( p \in S^{L-1} \) if there exists a finite set of price systems \( \{p^1, p^2, \ldots, p\} \) such that \( p^1 \cdot \hat{z}(p^0) > 0 \), \( p^2 \cdot \hat{z}(p^1) > 0 \), \ldots, \( p \cdot \hat{z}(p^{m-1}) > 0 \).

In words, a non-equilibrium price system \( p^0 \) is recursively upset by \( p \) means that there exist finite upsetting price systems \( p^1, p^2, \ldots, p^m \) with \( p^m = p \) such that excess demand value under \( p^0 \) is not affordable at \( p^1 \), \( p^1 \)'s excess demand is not affordable at \( p^2 \), and \( p^{m-1} \)'s excess demand is not affordable at \( p^m \). When strong Walras’ law holds, this implies that \( p^0 \) is upset by \( p^1 \), \( p^1 \) is upset by \( p^2 \), \( p^2 \) is upset by \( p^3 \), and \( p^{m-1} \) is upset by \( p \).

For convenience, we say \( p^0 \) is directly upset by \( p \) when \( m = 1 \), and indirectly upset by \( p \) when \( m > 1 \). Recursive upsetting says that nonequilibrium price system \( p^0 \) can be directly or indirectly upset by a price system \( q \) through sequential upsetting price systems \( \{p^1, p^2, \ldots, p^{m-1}\} \) in a recursive way that \( p^0 \) is upset by \( p^1 \), \( p^1 \) is upset by \( p^2 \), \ldots, and \( p^{m-1} \) is upset by \( p \).
Definition 10.4.5 (Recursive Transfer Lower Semi-Continuity) An excess demand function \( \hat{z}(\cdot) : S^{L-1} \to \mathbb{R}^L \) is said to be recursively transfer lower semi-continuous on \( S^{L-1} \) if, whenever \( q \in S^{L-1} \) is not a competitive equilibrium price system, there exists some price system \( p^0 \in S^{L-1} \) (possibly \( p^0 = q \)) and a neighborhood \( V_q \) such that \( p \cdot \hat{z}(V_q) > 0 \) for any \( p \) that recursively upsets \( p^0 \), where \( p \cdot \hat{z}(V_q) > 0 \) means \( p \cdot \hat{z}(q') > 0 \) for all \( q' \in V_q \).

In the definition of recursive transfer lower semi-continuity, \( q \) is transferred to \( q^0 \) that could be any point in \( S^{L-1} \). Roughly speaking, recursive transfer lower semi-continuity of \( \hat{z}(\cdot) \) means that, whenever \( q \) is not an equilibrium price system, there exists another nonequilibrium price system \( p^0 \) such that all excess demands in some neighborhood of \( q \) are not affordable at any price system \( p \) that recursively upsets \( p^0 \). This implies that, if an excess demand function \( \hat{z}(\cdot) : S^{L-1} \to \mathbb{R}^L \) is not recursively transfer lower semi-continuous, then there is some non-equilibrium price system \( q \) such that for every other price system \( p^0 \) and every neighborhood of \( q \), excess demand of some price system \( q' \) in the neighborhood becomes affordable at price system \( p \) that recursively upsets \( p^0 \).

We then have the following theorem of competitive equilibrium in economies that have single-valued excess demand functions.

Theorem 10.4.5 (The Existence Theorem V)) Suppose an excess demand function \( \hat{z}(\cdot) : S^{L-1} \to \mathbb{R}^L \) satisfies weak Walras’ law. Then there exists a competitive price equilibrium \( p^* \in S^{L-1} \) if and only if \( \hat{z}(\cdot) \) is recursively transfer lower semi-continuous on \( S^{L-1} \).

Proof. Sufficiency \((\Leftarrow)\). Suppose, by way of contradiction, that there is no price equilibrium. Then, by recursive transfer lower semi-continuity of \( \hat{z}(\cdot) \), for each \( q \in S^{L-1} \), there exists \( p^0 \) and a neighborhood \( V_q \) such that \( p \cdot \hat{z}(V_q) > 0 \) whenever \( p^0 \in S^{L-1} \) is recursively upset by \( p \), i.e., for any sequence of recursive price systems \( \{p^1, \ldots, p^{m-1}, p\} \) with \( p \cdot \hat{z}(p^{m-1}) > 0 \), \( p^{m-1} \cdot \hat{z}(p^{m-2}) > 0 \), \ldots, \( p^1 \cdot \hat{z}(p^0) > 0 \) for \( m \geq 1 \), we have \( p \cdot \hat{z}(V_q) > 0 \). Since there is no price equilibrium
by the contrapositive hypothesis, \( p^0 \) is not a price equilibrium and thus, by recursive transfer lower semi-continuity, such a sequence of recursive price systems \( \{p^1, \ldots, p^{m-1}, p\} \) exists for some \( m \geq 1 \).

Since \( S^{L-1} \) is compact and \( S^{L-1} \subseteq \bigcup_{q \in S^{L-1}} \mathcal{V}_q \), there is a finite set \( \{q^1, \ldots, q^T\} \) such that \( S^{L-1} \subseteq \bigcup_{i=1}^T \mathcal{V}_{q_i} \). For each of such \( q^i \), the corresponding initial price system is denoted by \( p^{0i} \) so that \( p^i \cdot \hat{\mathbf{z}}(\mathcal{V}_{q^i}) > 0 \) whenever \( p^{0i} \) is recursively upset by \( p^i \).

Since there is no price equilibrium, for each of such \( p^{0i} \), there exists \( p^i \) such that \( p^i \cdot \hat{\mathbf{z}}(p^{0i}) > 0 \), and then, by 1-recursive transfer lower semi-continuity, we have \( p^i \cdot \hat{\mathbf{z}}(\mathcal{V}_{q^i}) > 0 \). Now consider the set of price systems \( \{p^1, \ldots, p^T\} \). Then, \( p^i \not\in \mathcal{V}_{q^i} \); otherwise, by \( p^i \cdot \hat{\mathbf{z}}(\mathcal{V}_{q^i}) > 0 \), we will have \( p^i \cdot \hat{\mathbf{z}}(p^i) > 0 \), contradicting to weak Walras’ law. So we must have \( p^i \not\in \mathcal{V}_{p^i} \).

Without loss of generality, we suppose \( p^1 \in \mathcal{V}_{p^1} \). Since \( p^2 \cdot \hat{\mathbf{z}}(p^1) > 0 \) by noting that \( p^1 \in \mathcal{V}_{p^2} \) and \( p^1 \cdot \hat{\mathbf{z}}(p^{01}) > 0 \), then, by 2-recursive transfer lower semi-continuity, we have \( p^2 \cdot \hat{\mathbf{z}}(\mathcal{V}_{q^2}) > 0 \). Also, \( q^2 \cdot \hat{\mathbf{z}}(\mathcal{V}_{q^2}) > 0 \). Thus \( p^2 \cdot \hat{\mathbf{z}}(\mathcal{V}_{q^1} \cup \mathcal{V}_{q^2}) > 0 \), and consequently \( p^2 \not\in \mathcal{V}_{q^1} \cup \mathcal{V}_{q^2} \).

Again, without loss of generality, we suppose \( p^2 \in \mathcal{V}_{q^2} \). Since \( p^3 \cdot \hat{\mathbf{z}}(p^2) > 0 \) by noting that \( p^2 \in \mathcal{V}_{p^3} \), \( p^2 \cdot \hat{\mathbf{z}}(p^1) > 0 \), and \( p^1 \cdot \hat{\mathbf{z}}(p^{01}) > 0 \), by 3-recursive transfer lower semi-continuity, we have \( p^3 \cdot \hat{\mathbf{z}}(\mathcal{V}_{q^1}) > 0 \). Also, since \( p^3 \cdot \hat{\mathbf{z}}(p^2) > 0 \) and \( p^3 \cdot \hat{\mathbf{z}}(p^{02}) > 0 \), by 2-recursive transfer lower semi-continuity, we have \( p^3 \cdot \hat{\mathbf{z}}(\mathcal{V}_{q^2}) > 0 \). Thus, we have \( p^3 \cdot \hat{\mathbf{z}}(\mathcal{V}_{q^1} \cup \mathcal{V}_{q^2} \cup \mathcal{V}_{q^2}) > 0 \), and consequently \( p^3 \not\in \mathcal{V}_{q^1} \cup \mathcal{V}_{q^2} \cup \mathcal{V}_{q^2} \).

With this process going on, we can show that \( p^k \not\in \mathcal{V}_{q^1} \cup \mathcal{V}_{q^2} \cup \ldots \cup \mathcal{V}_{q^T} \), i.e., \( p^k \) is not in the union of \( \mathcal{V}_{q^1}, \mathcal{V}_{q^2}, \ldots, \mathcal{V}_{q^T} \) for \( k = 1, 2, \ldots, T \). In particular, for \( k = T \), we have \( p^L \not\in \mathcal{V}_{q^1} \cup \mathcal{V}_{q^2} \cup \ldots \cup \mathcal{V}_{q^T} \) and so \( p^T \not\in S^{L-1} \subseteq \mathcal{V}_{q^1} \cup \mathcal{V}_{q^2} \ldots \cup \mathcal{V}_{q^T} \), a contradiction.

Thus, there exists \( p^* \in S^{L-1} \) such that \( p^* \cdot \hat{\mathbf{z}}(p^*) \leq 0 \) for all \( p \in S^{L-1} \). Letting \( p^1 = (1, 0, \ldots, 0) \), \( p^2 = (0, 1, 0, \ldots, 0) \), and \( p^L = (0, 0, 0, \ldots, 0, 1) \), we have \( \hat{\mathbf{z}}^l(p^*) \leq 0 \) for \( l = 1, \ldots, L \) and thus \( p^* \) is a price equilibrium.

*Necessity \((\Rightarrow)\).* Suppose \( p^* \) is a competitive price equilibrium and \( p^* \cdot \hat{\mathbf{z}}(q) > 0 \) for \( q, p \in S^{L-1} \). Let \( p^0 = p^* \) and \( \mathcal{N}(q) \) be a neighborhood of \( q \). Since \( p^* \cdot \hat{\mathbf{z}}(p^*) \leq 0 \)
for all $\mathbf{p} \in S^{L-1}$, it is impossible to find any sequence of finite price vectors 
$\{\mathbf{p}^1, \mathbf{p}^2, \ldots, \mathbf{p}^m\}$ such that $\mathbf{p}^1 \cdot \hat{z}(\mathbf{p}^0) > 0, \mathbf{p}^2 \cdot \hat{z}(\mathbf{p}^1) > 0, \ldots, \mathbf{p}^m \cdot \hat{z}(\mathbf{p}^{m-1}) > 0$.
Hence, the recursive transfer lower semi-continuity holds trivially. 

This theorem is useful to check the nonexistence of competitive equilibrium. The method of proof employed to obtain Theorem 10.4.5 is finite subcover theorem. While there are different ways of establishing the existence of competitive equilibrium, all the existing proofs essentially use the fixed-point-theorem related approaches. Moreover, a remarkable advantage of the above proof is that it is simple and elementary without using advanced math.

As shown by Tian (2016), Theorem 10.4.5 can also be extended to the case of any set, especially the positive price open set, of price systems for which excess demand is defined. To do so, we introduce the following version of recursive transfer lower semi-continuity.

**Definition 10.4.6** Let $D$ be a subset of $\text{int} S^{L-1}$. An excess demand function $\hat{z}(\cdot) : \text{int} S^{L-1} \rightarrow \mathbb{R}^L$ is said to be recursively transfer lower semi-continuous on $\text{int} S^{L-1}$ with respect to $D$ if, whenever $\mathbf{q} \in \text{int} S^{L-1}$ is not a competitive equilibrium price system, there exists some price system $\mathbf{p}^0 \in D$ (possibly $\mathbf{p}^0 = \mathbf{q}$) and a neighborhood $V_{\mathbf{q}}$ such that (1) whenever $\mathbf{p}^0$ is upset by a price system in $\text{int} S^{L-1} \setminus D$, it is upset by a price system in $D$, and (2) $\mathbf{p} \cdot \hat{z}(V_{\mathbf{q}}) > 0$ for any $\mathbf{p} \in D$ that recursively upsets $\mathbf{p}^0$.

We then have the following theorem that fully characterizes the existence of price equilibrium in economies with possibly indivisible commodity spaces and discontinuous excess demand functions.

**Theorem 10.4.6 (The Existence Theorem VI)** Suppose an excess demand function $\hat{z}(\cdot) : \text{int} S^{L-1} \rightarrow \mathbb{R}^L$ satisfies weak Walras’ law: $\mathbf{p} \cdot \hat{z}(\mathbf{p}) = 0$ for all $\mathbf{p} \in \text{int} S^{L-1}$. Then there is a competitive price equilibrium $\mathbf{p}^* \in \text{int} S^{L-1}$ if and only if there exists a compact subset $D \subseteq \text{int} S^{L-1}$ such that $\hat{z}(\cdot)$ is recursively transfer lower semi-continuous on $\text{int} S^{L-1}$ with respect to $D$. 

378
Proof. Sufficiency ($\Leftarrow$).

The proof of sufficiency is essentially the same as that of Theorem 10.4.5 and we just outline the proof here. We first show that there exists a price equilibrium $p^*$ on $D$ if it is recursively transfer lower semi-continuous on $S^{L-1}$ with respect to $D$. Suppose, by way of contradiction, that there is no competitive equilibrium in $D$. Then, since $\hat{z}$ is recursively diagonal transfer lower semi-continuous on $S^{L-1}$ with respect to $D$, for each $q \in D$, there exists $p^0$ and a neighborhood $\mathcal{V}_q$ such that (1) whenever $p^0$ is upset by a price system in $S^{L-1} \setminus D$, it is upset by a price system in $D$ and (2) $p \cdot \hat{z}(\mathcal{V}_q) > 0$ for any finite subset of price systems $\{p^1, \ldots, p^m\} \subset D$ with $p^m = p$ and $p \cdot \hat{z}(p^{m-1}) > 0$, $p^{m-1} \cdot \hat{z}(p^{m-2}) > 0$, ..., $p^1 \cdot \hat{z}(p^0) > 0$ for $m \geq 1$. Since there is no competitive equilibrium by the contrapositive hypothesis, $p^0$ is not a competitive equilibrium and thus, by recursive diagonal transfer lower semi-continuity on $S^{L-1}$ with respect to $D$, such a sequence of recursive securing price systems $\{p^1, \ldots, p^{m-1}, p\}$ exists for some $m \geq 1$.

Since $D$ is compact and $D \subseteq \bigcup_{q \in S^{L-1}} \mathcal{V}_q$, there is a finite set $\{q^1, \ldots, q^T\} \subseteq D$ such that $D \subseteq \bigcup_{i=1}^T \mathcal{V}_{q^i}$. For each of such $q^i$, the corresponding initial deviation price system is denoted by $p^{0i}$ so that $p^i \cdot \hat{z}(\mathcal{V}_{q^i}) > 0$ whenever $p^{0i}$ is recursively upset by $p^i$ through any finite subset of securing price systems $\{p^{i_1}, \ldots, p^{m_1}\} \subset D$ with $p^{m_1} = p^i$. Then, by the same argument as in the proof of Theorem 10.4.5, we will obtain that $z^k$ is not in the union of $\mathcal{V}_{q^1}, \mathcal{V}_{q^2}, \ldots, \mathcal{V}_{q^k}$ for $k = 1, 2, \ldots, T$. For $k = T$, we have $p^T \notin \mathcal{V}_{q^1} \cup \mathcal{V}_{q^2} \ldots \cup \mathcal{V}_{q^T}$ and so $p^T \notin D \subseteq \bigcup_{i=1}^T \mathcal{V}_{q^i}$, which contradicts that $p^T$ is a upsetting price in $D$.

We now show that $p \cdot \hat{z}(p^*) \leq 0$ for all $p \in \text{int} S^{L-1}$. Suppose not. $p^*$ would be upset by a price system in $\text{int} S^{L-1} \setminus D$, but it is upset by a price system in $D$, a contradiction. We want to show that $p^*$ in fact is a competitive price equilibrium. Note that $\text{int} S^{L-1}$ is open and $D$ is a compact subset of $\text{int} S^{L-1}$. One can always find a sequence of price vector $\{q^l_n\} \subseteq \text{int} S^{L-1} \setminus D$ such that $q^l_n \to p^l$ as $n \to \infty$, where $p^l = (0, \ldots, 0, 1, 0, \ldots, 0)$ is the unit vector that has only one argument - the $l$th argument - with value 1 and others with value 0.
Since $p \cdot \hat{z}(q)$ is continuous in $p$, we have $\hat{z}^l(p^*) \leq 0$ for $l = 1, \ldots, L$ and thus $p^*$ is a competitive price equilibrium.

**Necessity** ($\Rightarrow$). Suppose $p^*$ is a competitive equilibrium. Let $D = \{p^*\}$. Then, the set $D$ is clearly compact. Now, for any non-competitive equilibrium $q \in S^{L-1}$, let $p^0 = p^*$ and $V_q$ be a neighborhood of $q$. Since $p \cdot \hat{z}(p^*) \leq 0$ for all $p \in S^{L-1}$ and $p^0 = p^*$ is a unique element in $D$, there is no other upsetting price $p^1$ such that $p^1 \cdot \hat{z}(p^0) > 0$. Hence, the game is recursively diagonal transfer continuous on $S^{L-1}$ with respect to $D$.

Theorem 10.4.6 then strictly generalizes all the existing results on the existence of competitive equilibrium in economies with single-valued excess demand functions.

### 10.4.5 Existence of CE for Aggregate Excess Demand Correspondences

When preferences and/or production sets are not strictly convex, the demand correspondence and/or supply correspondence may not be single-valued, and consequently the aggregate excess demand correspondence may not be single-valued. As a result, one cannot use the above existence results to argue the existence of competitive equilibrium. Nevertheless, by using the KKM lemma, we can still prove the existence of competitive equilibrium when the aggregate excess demand correspondence satisfies certain conditions.

**Theorem 10.4.7 (The Existence Theorem VII)** For a private ownership economy $e = (\{X_i, w_i, \succeq_i\}_{i \in N}, \{Y_j\}_{j \in J}, \{\theta_{ij}\}_{(i,j) \in N \times J})$, if $\hat{z}(p)$ is a non-empty convex and compact-valued upper hemi-continuous correspondence and satisfies weak Walras’ Law, then there exists a competitive equilibrium, that is, there is a price vector $p^* \in S$ such that

$$\hat{z}(p^*) \cap \{-\mathbb{R}^L_+\} \neq \emptyset.$$  

(10.49)
Proof. The proof is analogous to the proof of Theorem III. Define a correspondence $F : S \to 2^S$ by,

$$ F(q) = \{ p \in S : q \cdot \hat{z} \leq 0 \text{ for some } \hat{z} \in \hat{z}(p) \}. $$

Since $\hat{z}(\cdot)$ is upper hemi-continuous, $F(q)$ is closed for each $q \in S$. We now prove $F$ is FS-convex. Suppose, by way of contradiction, that there are some $q_1, \ldots, q_m \in S$ and some convex combination $q_\lambda = \sum_{t=1}^m \lambda_t q_t$ such that $q_\lambda \not\in \bigcup_{t=1}^m F(q_t)$. Then, $q_\lambda \not\in F(q_t)$ for all $t = 1, \ldots, m$. Thus, for all $\hat{z} \in \hat{z}(q_\lambda)$, we have $q_t \cdot \hat{z} > 0$ for $t = 1, \ldots, m$. Hence, $\sum_{t=1}^m \lambda_t q_t \cdot \hat{z} > 0$ which contradicts the fact that $\hat{z}$ satisfies Walras’ Law. So $F$ must be FS-convex. Therefore, by KKM lemma, we have

$$ \bigcap_{q \in S} F(q) \neq \emptyset. $$

Then there exists a $p^* \in S$ such that $p^* \in \bigcap_{q \in S} F(q)$, i.e., $p^* \in F(q)$ for all $q \in S$. Thus, for each $q \in S$, there is $\hat{z}_q \in \hat{z}(p^*)$ such that

$$ q \cdot \hat{z}_q \leq 0. $$

We now prove $\hat{z}(p^*) \cap \{-\mathbb{R}^L_+\} \neq \emptyset$. Suppose not. Since $\hat{z}(p^*)$ is convex and compact and $-\mathbb{R}^L_+$ is convex and closed, by the Separating Hyperplane theorem, there exists a vector $q \in \mathbb{R}^L$ and some $c \in \mathbb{R}^L$ such that

$$ q \cdot (-\mathbb{R}^L_+) < c < q \cdot \hat{z}(p^*) $$

Since $(-\mathbb{R}^L_+)$ is a cone, we must have $c > 0$ and $q \cdot (-\mathbb{R}^L_+) \leq 0$. Thus, $q \in \mathbb{R}^L_+$ and $q \cdot \hat{z}(p^*) > 0$ for all $q$, a contradiction. The proof is completed.

Remark 10.4.8 The last part of the proof can be also shown by applying the following result: Let $K$ be a compact convex set. Then $K \cap \{-\mathbb{R}^L_+\} \neq \emptyset$ if and only if for any $p \in \mathbb{R}^L_+$, there exists $z \in K$ such that $p \cdot z \leq 0$. The proof of this result can be found, for example, in the book of K. Border (1985, p. 13).

Similarly, we have the following existence theorem that provides sufficient conditions directly based on economic environments by applying the Existence Theorem VII above.
Theorem 10.4.8 (The existence Theorem VIII) For a private ownership economy \( e = (\{X_i, w_i, \succ_i\}, \{Y_j\}, \{\theta_{ij}\}) \), there exists a competitive equilibrium if the following conditions hold

(i) \( X_i = \mathbb{R}_+^L \);
(ii) \( w_i \geq 0 \), and \( \sum_{i \in N} w_i > 0 \);
(iii) \( \succ_i \) are continuous, weakly convex and strong monotone;
(iv) \( Y_j \) are closed, convex, and \( 0 \in Y; j = 1, 2, \ldots, J \).

Similarly, we can weaken strong monotonicity to monotonicity when initial endowment is an interior point, and thus we have the following proposition.

Proposition 10.4.7 (The existence Theorem VIII') For a private ownership economy \( e = (\{X_i, w_i, \succ_i\}, \{Y_j\}, \{\theta_{ij}\}) \), there exists a competitive equilibrium if the following conditions hold

(i) \( X_i = \mathbb{R}_+^L \);
(ii) \( w_i \in \mathbb{R}_+^L \);
(iii) \( \succ_i \) are continuous, weakly convex and monotone;
(iv) \( Y_j \) are closed, convex, and \( 0 \in Y; j = 1, 2, \ldots, J \).

10.4.6 Existence of CE for General Production Economies

In the previous sections, we give the existence of CE for aggregate demand function. For a general private ownership production economy, Arrow and Debreu proved the existence of CE in 1954. This result was included in Debreu’s subsequent classical monograph “The Theory of Value”. This short book (roughly 100 pages) gives a clear exposition of the basic elements of axiomatic general equilibrium analysis and has changed the standards of mathematical rigor in economic theory.

For a general private ownership production economy,

\[
e = (\{X_i, \succ_i, w\}_{i \in N}, \{Y_j\}_{j \in J}, \{\theta_{ij}\}_{(i,j) \in N \times J}),
\]

(10.50)
We now state the following existence theorem for general production economies without proof. Since the proof is very complicated, we refer readers to the proof that can be found in Debreu (1959) who used the existence theorem on equilibrium of the abstract economy on Section 10.7.

**Theorem 10.4.9 (Existence Theorem IX, Debreu, 1959)** A competitive equilibrium for the private-ownership economy $e$ exists if the following conditions are satisfied:

1. $X_i$ is closed, convex and bounded from below;
2. $\succ_i$ are non-satiated;
3. $\succ_i$ are continuous;
4. $\succ_i$ are convex;
5. $w_i \in \text{int}X_i$;
6. $0 \in Y_j$ (possibility of inaction);
7. $Y_j$ are closed and convex (continuity and no IRS)
8. $Y_j \cap \{-Y_j\} = \{0\}$ (Irreversibility)
9. $\{-\Re_{+}^{L}\} \subseteq Y_j$ (free disposal)

There exist economic intuitions behind these mathematical conditions. Conditions (1) to (5) are about the properties of consumers; conditions (6) to (9) are about the properties of suppliers.

1. $X_i$ is bounded from below means that consumers could endure only a limited amount of “bads”. The convexity of consumption space implies the divisibility of goods. Closeness of consumption space is to guarantee that the optimal consumption choice is included in the it.

2. non-satiation of preference means that people in the economy is not too virtuous to abandon their self interests. The desire-scarcity conflict is a basic problem of economics. The function of economy is to allocate scarce resources among unlimited wants. Though market may sometimes fail, but it is by far
the best way to tackle this conflict. (we will show the desirability of market institution from different angles in the subsequent analysis.) Due to the moral limitation of people, some ideally high target cannot be accomplished in short run. That is why the policy maker usually has to take the second-best rather than the first-best objective.

(3) the assumption on continuity of preference guarantees the existence of consumer’s optimal choice within any budget set.

(4) convexity on preferences implies that people prefer diversity to extremity.

(5) interior endowment is a technical assumption to guarantee the continuity of budget correspondence and the existence of optimal demand choice.

(6) is a natural assumption that allows every producer to do nothing.

(7) the closeness of production set guarantees that the optimal production plan is within it. The convexity of production set eliminate the case of increasing returns to scale.

(8) Irreversibility says that the production process cannot be undone. That is, if \( y_j \in Y_j \) and \( y \neq 0 \), then \(-y_j \notin Y_j\). Actually, the laws of physics imply that all production processes are irreversible. You may be able to turn gold bars into jewelry and then jewelry back into gold bars, but in either case you use energy. So, this process is not really reversible.

(9) free disposal is a natural assumption means that nothing prevents the producer from being inefficient in the sense that it use more resources than necessary to produce the same amount of commodities.

These assumptions, though have limitations, are quite reasonable and easy to be met. Some implicit assumptions need to be considered in the above theorem. These assumptions include: everyone is price taker, information is complete, there is no public goods, the individual preferences are independent of others’ preferences. Without these assumptions, CE may fail to exist or is inefficient.
10.4.7 The Existence of CE Under Non-Ordered Preference

All the existence theorems mentioned above, however, are obtained by assuming that the preference are ordered, or even representable by an utility function. These assumptions are too restricted in many settings. Because the various environments in real world, the motivations for economists continually to be interested in setting forth conditions for the existence of equilibria under the weakest possible conditions. In this subsection, we explore the existence of CE of an abstract economy with non-total and non-transitive preference. The strict superior correspondence $U_{si} : X_i \to 2^{X_i}$ of individual $i$ is defined as follows:

$$U_{si}(x_i) = \{x_i' \in X_i : x_i' \succ_i x_i\}.$$

For simplicity of analysis, we discuss only the pure exchange economy.

**Definition 10.4.7 (CE under non-ordered preference)** Given a pure exchange economy $e = (e_1, \ldots, e_n)$, an allocation $x \in X$ and price vector $p \in \mathbb{R}_+^L$ constitute a competitive equilibrium provided the following conditions are met:

(i) maximization of preference: $x_i \in B_i(p)$, and $U_{si}(x) \cap B_i(p) = \emptyset, \forall i \in N$;

(ii) market clearance condition: $\hat{x} \leq \hat{w} + \hat{y}$.

Applying the existence Theorem 10.7.3 given in Section 10.7, we could prove the existence of CE under non-ordered preference.

**Theorem 10.4.10 (Existence theorem X, Tian (1992))** For private ownership economy $e = (\{X_i, w_i, \succ_i\})$, then an competitive equilibrium exists provided the following conditions hold:

(i) $X_i$ is compact subset of $\mathbb{R}_+^L$;

(ii) $w_i \geq 0, \sum_{i \in N} w_i > 0$;
(iii) $U_{s_i}$ is l.h.c., and the superior correspondence $U_{s_i}(x_i)$ is open;

(iv) $x_i \notin co U_{s_i}(x), \forall x \in Z$.

**Proof.** Since $w_i > 0$, it follows from the proof of proposition 10.4.5, budget correspondence $B_i$ is continuous. From the discussion of Section 10.7, we know that pure exchange economy could be regarded as an abstract economy, then all the conditions of Theorem 10.7.3 are met. Therefore, CE exists. ■

Notice that a sufficient condition for $P_i$ to be l.h.c and open is $P_i$ has an open graph.

### 10.5 Uniqueness of Competitive Equilibria

So now we know that a Walrasian equilibrium will exist under some regularity conditions such as continuity, monotonicity, and convexity of preferences and/or production possibility sets. We worry next about the other extreme possibility; for a given economy there are many equilibra so that we do not know what is the outcome of the market process. We can easily give examples in which there are multiple competitive equilibrium price vectors. When is there only one normalized price vector that clears all markets?

The free goods case is not of great interest here, so we will rule it out by means of the desirability assumption so that every equilibrium price of each good must be strictly positive. We want to also assume the continuous differentiability of the aggregate excess demand function. The reason is fairly clear; if indifference curves have kinks in them, we can find whole ranges of prices that are market equilibria. Not only are the equilibria not unique, they are not even locally unique. Thus, we answer this question for only considering the case of $p^* > 0$ and $\hat{z}(p)$ is differentiable.

The uniqueness and stability of competitive equilibrium were first studied by Arrow, Debreu, Block and Hurwicz in late 1950s. We will focus on these issues in this section. Firstly, we discuss the uniqueness of CE. The uniqueness property
requires that we cannot find two linearly independent equilibrium price vectors, otherwise, all the equilibrium prices may constitute a 2-dimensional subspace of \( \mathbb{R}^L \) since the zero homogeneity property hold.

**Theorem 10.5.1** Suppose all goods are desirable and gross substitute for all prices (i.e., \( \frac{\partial z^k(p)}{\partial p_l} > 0 \) for \( l \neq h \)). If \( p^* \) is a competitive equilibrium price vector and Walras’ Law holds, then it is the unique competitive equilibrium price vector.

**Proof.** By the desirability, \( p^* > 0 \). Suppose \( p \) is another competitive equilibrium price vector that is not proportional to \( p^* \). Let \( m = \max \frac{p_l}{p^*_l} = \frac{p_k}{p^*_k} \) for some \( k \).

By homogeneity and the definition of competitive equilibrium, we know that \( \hat{z}(p^*) = \hat{z}(mp^*) = 0 \). We know that \( m = \frac{p_k}{p^*_k} \geq \frac{p_l}{p^*_l} \) for all \( l = 1, \ldots, L \) and \( m > \frac{p_h}{p^*_h} \) for some \( h \). Then we have \( mp^*_l \geq p^*_l \) for all \( l \) and \( mp^*_h > p_h \) for some \( h \).

Thus, when the price of good \( k \) is fixed, the prices of the other goods are down. We must have the demand for good \( k \) down by the gross substitutes. Hence, we have \( \hat{z}^k(p) < 0 \), a contradiction. ■

When the aggregate demand function satisfies the Weak Axiom of Revealed Preference (WARP) and Walras’ Law holds, competitive equilibrium is also unique.

**The Weak Axiom of Revealed Preference (WARP) of the aggregate excess demand function:** If \( p \cdot \hat{z}(p) \geq p \cdot \hat{z}(p') \), then \( p' \cdot \hat{z}(p) > p' \cdot \hat{z}(p') \) for all \( p, p' \in \mathbb{R}^L_+ \).
Figure 10.17: The figure shows an aggregate demand function satisfies WARP.

WARP implies that, if \( \hat{z}(p') \) could have been bought at \( p \) where \( \hat{z}(p) \) was bought (so that \( \hat{z}(p) \succ \hat{z}(p') \) since \( \hat{z}(p) \) is the optimal choice), then at price \( p' \), \( \hat{z}(p) \) is outside the budget constraint (otherwise it contradicts to the fact that \( \hat{z}(p) \) is the optimal choice).
Figure 10.18: Both individual demand functions satisfy WARP.

Even if the individual demand functions satisfy WARP, the aggregate excess demand function does not necessarily satisfy the WARP in aggregate.

Figure 10.19: The aggregate excess demand function does not satisfy WARP.

The WARP is a weak restriction than the continuous concave utility function. However, the restriction on the aggregate excess demand function is not as weak as it may be seen. Even though two individuals satisfy the individual WARP, the aggregate excess demand function may not satisfy the aggregate WARP as shown in Figures 10.18 and 10.19.

**Lemma 10.5.1** Under the assumptions of Walras’ Law and WARP, we have \( p^* \tilde{z}(p) > 0 \) for all \( p \neq kp^* \) where \( p^* \) is a competitive equilibrium.

**Proof.** Suppose \( p^* \) is a competitive equilibrium:

\[
\tilde{z}(p^*) \leq 0. \quad (10.51)
\]
Also, by Walras’ Law, $p \cdot \hat{z}(p) = 0$. So we have $p \cdot \hat{z}(p) \geq p \cdot \hat{z}(p^*)$. Then, by WARP, $p^* \cdot \hat{z}(p) > p^* \cdot \hat{z}(p^*) = 0 \Rightarrow p^* \cdot \hat{z}(p) > 0$ for all $p \neq k p^*$.

**Theorem 10.5.2** Under the assumptions of Walras’ Law and WARP in aggregate excess demand function, the competitive equilibrium is unique.

**Proof.** By Lemma 10.5.1, for any $p \neq k p^*$, $p^* \cdot \hat{z}(p) > 0$ which means at least for some $l$, $\hat{z}^l > 0$.

The above results only provide sufficient conditions for the uniqueness of competitive equilibrium. Now we provide necessary and sufficient conditions for the uniqueness of competitive equilibrium. Mathematically, it is equivalent to that the excess demand function to be globally invertible, and further global invertibility of a continuous map means global homomorphism (i.e. the function is a one-to-one continuous mapping on the domain). To have the result, we first define the notion of proper map.

**Definition 10.5.1** (Proper Map) An excess demand function $\hat{z}(\cdot) : S^{L-1} \rightarrow \mathbb{R}^L$ is said to be proper map if its inverse $f^{-1}(K)$ is compact whenever $K$ is compact.

We then have following Browder’s Global Uniqueness Theorem.\(^3\)

**Lemma 10.5.2** (Browder Global Uniqueness Theorem) A function $F : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is homeomorphism if and only if it is a proper map and local homeomorphism.

Also, when $F$ is differentiable, it is well-known that local homeomorphism is equivalent to nonvanishing Jacobian, and then we have the following result on the uniqueness of competitive equilibrium.

**Theorem 10.5.3** Suppose that an excess demand function $\hat{z}(\cdot) : S^{L-1} \rightarrow \mathbb{R}^L$ is differentiable. Then, the competitive equilibrium is unique if and only if $\hat{z}(\cdot)$ has a nonvanishing Jacobian and is a proper map.

\(^3\)The proof can be found on page 29 in Lecture Notes in Mathematics Volume 977, 1983, pp 28-40; http://link.springer.com/chapter/10.1007/BFb0065570.
This result includes many existing results on the uniqueness of competitive equilibrium such as those in Gale and Nikaido (1965), Dierker (1972), Varian (1975), Mas-Colell (1979) and Mukherji (1997) as special cases. Since Varian’s result is relatively easier to verify, we introduce here as a corollary.

**Corollary 10.5.1** (Varian, 1975) Suppose that as the price of a good goes to zero, its excess demand becomes positive. If the Jacobian of the excess demand function $\dot{z}(\cdot)$ is positive at all Walras equilibria, there is exactly one equilibrium.

### 10.6 Stability of Competitive Equilibrium

The stability of the Walrasian economy became a central research focus in the years following the existence proofs (Arrow and Hurwicz 1958, 1959, 1960; Arrow, Block and Hurwicz 1959; Nikaido 1959; McKenzie 1960; Nikaido and Uzawa 1960). The concept of competitive equilibrium is a stationary concept. But, it has given no guarantee that the economy will actually operate at the equilibrium point. Or an economy may deviate from an equilibrium due to economic shocks. What forces exist that might tend to move prices to a market-clearing price? This is a topic about the stability on the price adjustment mechanism in a competitive equilibrium.

The reason for us to focus on stability of CE is the frequent fluctuations in economy. Whether and to what degree the competitive equilibrium is robust to some unanticipated disturbances such as turmoil originating from military, political, economic or social crises is a very important and realistic problem? For example, whether or not the world economy could recover from a financial crisis is the main concern of the whole world.

When the conditions guaranteeing the stability of CE are violated, an economy may vulnerable to shocks and diverge to an undesirable outcome. An central topic of development economics is whether or not the poorer economies’ per capita incomes will catch-up that of richer economies. Both “yes” and “no” answers to this question have found their historical evidences. The high-speed development
of China in the past three decades gives a “yes” answer. While Democratic
Republic of Congo, with the GDP per capita of only one hundred dollars is an
extreme example at the other side of the spectrum.

According to economic intuitions, it seems the stability of CE is automatically
guaranteed by the law of supply and demand. In fact, this intuition is not
always correct. We will see a counterexample given by Scarf in this section.

A paradoxical relationship between the idea of competition and price adjust-
ment is that: If all agents take prices as given, how can prices move? Who is
left to adjust prices? To solve this paradox, one introduces a “Walrasian auction-
eree” whose sole function is to seek for the market clearing prices. The Walrasian
auctioneer is supposed to call the prices and change the price mechanically re-
sponding to the aggregate excess demand till the market clears. Such a process
is called Tâtonnement adjustment process.

**Tâtonnement Adjustment Process** is defined, according to the laws of
demand and supply, by

\[
\frac{dp^l}{dt} = G^l(\hat{z}(p)) \quad l = 1, \ldots, L
\]

where \(G^l\) is a sign-preserving function of \(\hat{z}(p)\), i.e., \(G^l(x) > 0\) if \(x > 0\), \(G^l(x) = 0\)
if \(x = 0\), and \(G^l(x) < 0\) if \(x < 0\). The above equation implies that when the
aggregate excess demand is positive, we have a shortage and thus price should go
up by the laws of demand and supply.

As a special case of \(G^l\), \(G^l\) can be an identical mapping such that

\[
\dot{p}^l = \hat{z}^l(p) \quad (10.53)
\]

\[
\dot{p} = \hat{z}(p). \quad (10.54)
\]

Under Walras’ Law,

\[
\frac{d}{dt}(p\dot{p}) = \frac{d}{dt} \left[ \sum_{l=1}^{L} (p^l)^2 \right] = 2 \sum_{l=1}^{L} (p^l) \cdot \frac{dp^l}{dt}
\]

\[
= 2p\dot{p} = p \cdot \hat{z}(p) = 0
\]
which means that the sum of squares of the prices remain constant as the price adjusts. This is another price normalization. The path of the prices are restricted on the surface of a $L$-dimensional sphere.

Examples of price dynamics in Figure 10.20. The first and third figures show a stable equilibrium, the second and fourth figures show a unique unstable equilibrium. Formally, we have

**Definition 10.6.1** An equilibrium price $p^*$ is globally stable if

(i) $p^*$ is the unique competitive equilibrium,

(ii) for all $p_0$ there exists a unique price path $p = \phi(t, p_0)$ for $0 \leq t < \infty$ such that $\lim_{t \to \infty} \phi(t, p_0) = p^*$.

**Definition 10.6.2** An equilibrium price $p^*$ is locally stable if there is $\delta > 0$ and a unique price path $p = \phi(t, p_0)$ such that $\lim_{t \to \infty} \phi(t, p_0) = p^*$ whenever $|p^* - p_0| < \delta$. 

393
Figure 10.20: The first and third figures show the CEs are stable, and the second and fourth show they are not stable.

The local stability of a competitive equilibrium can be easily obtained from the standard result on the local stability of a differentiate equation.

**Theorem 10.6.1** A competitive equilibrium price \( p^* \) is locally stable if the Jacobean matrix defined by

\[
A = \left[ \frac{\partial z'(p^*)}{\partial p_k} \right]
\]

has all negative eigenvalues.

The global stability result can be obtained by Liaponov Theorem.
Liaponov’s function: For a differentiate equation system \( \dot{x} = f(x) \) with \( f(x^*) = 0 \), a function \( V \) is said to be a Liaponov function for the differentiate equation system if

1. there is a unique \( x^* \) such that \( V(x^*) = 0 \);
2. \( V(x) > 0 \) for all \( x \neq x^* \);
3. \( \frac{dV(x)}{dt} < 0 \).

Theorem 10.6.2 (Liaponov’s Theorem) If there exists a Liaponov function for \( \dot{x} = f(x) \), the unique stationary point \( x^* \) is globally stable.

Applying Liaponov’s theorem, Arrow, Block and Hurwicz (1959) give conditions under which CE found by the tâtonnement adjustment process is globally stable. It requires the excess demand function to satisfy WRPA. Debreu (1974) has shown essentially that any continuous function which satisfies the Walras’ Law is an aggregate demand function for some economy. Thus, utility maximization places no restriction on aggregate behavior. Thus, to get global stability, one has to make some additional assumptions such as gross substitutibility and the Weak Axiom of Revealed Preference (WARP) for aggregate excess demand functions.

Lemma 10.6.1 Under the assumptions of Walras’ Law and gross substitutability, we have \( p^* \cdot \dot{z}(p) > 0 \) for all \( p \neq k p^* \) where \( p^* > 0 \) is a competitive equilibrium.

Proof. The proof of this lemma is complicated. We illustrate the proof only for the case of two commodities by aid of the figure. The general proof of this lemma can be seen in Arrow and Hahn, 1971, p. 224.
Let $p^*$ be an equilibrium price vector and $\hat{x}^* = \hat{x}(p^*)$ be the aggregate equilibrium demand. Let $p \neq \alpha p^*$ for any $\alpha > 0$. Then we know $p$ is not an equilibrium price vector by the uniqueness of the competitive equilibrium under the gross substitutability. Since $p \cdot \hat{x}(p^*) = p \cdot \hat{w} = p \cdot \hat{x}(p^*)$ by Walras’ Law and $\hat{w} = \hat{x}(p^*)$, then the aggregate demand $\hat{x}(p)$ is on the line $AB$ which passes through the point $\hat{x}^*$ and whose slope is given by $-p_1/p_2$. Let $CD$ be the line which passes through the point $\hat{x}^*$ and whose slope is $p^*$. We assume that $p_1^*/p_2^* > p_1/p_2$ without loss of generality. We need to show that $\hat{x}_1^1(p) > \hat{x}_1^1$ and $\hat{x}_2(p) < \hat{x}_2^*$ so that the point $\hat{x}(p)$ must lie to the right of the point $\hat{x}^*$ in the figure.

By $p_1^*/p_2^* > p_1/p_2$, we have $p_1^*/p_1 > p_2^*/p_2 \equiv \mu$. Thus, $p_1^* > \mu p_1$ and $p_2^* = \mu p_2$. Therefore, we have $\hat{x}_1^x = \hat{x}_2^x(p^*) > \hat{x}_2(\mu p) = \hat{x}_2^x(p)$ by the gross substitutability and the homogeneity. Thus, by $p \cdot \hat{x}(p) = p \cdot \hat{x}(p^*)$, we have $\hat{x}_1^x(p) > \hat{x}_1^x(p^*) = \hat{x}_1^x$. Hence the point $\hat{x}(p)$ must lie to the right of the point $\hat{x}^*$ in the figure. Now draw a line parallel to $CD$ passing through the point $\hat{x}(p)$.
We see that $p^* \hat{x}(p) > p^* \hat{x}$. Adding $p^* \cdot \hat{w}$ both sides, we have $p^* \cdot \hat{z}(p) > 0$. The proof is completed. ■

Now we are ready to give the following theorem on the global stability of a competitive equilibrium.

**Theorem 10.6.3 (Arrow-Block-Hurwicz)** Under the assumptions of Walras’ Law, if $\hat{z}(p)$ satisfies either gross substitutability, or WARP, then the competitive equilibrium price is globally stable.

**Proof.** By the gross substitutability or WARP of the aggregate excess demand function, the price vector $p^*$ is unique. We now show it is globally stable. Define a Liaponov’s function by

$$V(p) = \sum_{l=1}^{L} (p^l(t) - p^*l)^2 = \|p - p^*\|^2$$

(10.55)

By the assumption, $V(p^*) = 0$ and $V(p) > 0, \forall p \neq p^*$. Also, since

$$\frac{dV}{dt} = 2(p(t) - p^*) \cdot \dot{p}(t) = 2(p(t) - p^*) \cdot \hat{z}(p) = -2p^* \cdot \hat{z}(p) < 0$$

by Walras’ Law and Lemma 10.5.1 and 10.6.1 for $p \neq k p^*$, we know $\dot{p} = \hat{z}(p)$ is globally stable by Leaponov’s theorem. ■

The above theorem establishes the global stability of a competitive equilibrium under Walras’ Law, homogeneity, and gross substitutability/WARP. It is natural to ask how far we can relax these assumptions, in particular gross substitutability. Since many goods in reality are complementary goods, can the global stability of a competitive equilibrium also hold for complementary goods? Scarf (1961) has constructed examples showing that a competitive equilibrium may not be globally stable in the presence of complementary goods.

**Example 10.6.1 (Scarf’s Counterexample on Global Instability)** Consider a pure exchange economy with three consumers and three commodities ($n=3$, $397$...
L=3). Suppose consumers’ endowments are given by $w_1 = (1, 0, 0)$, $w_2 = (0, 1, 0)$, $w_3 = (0, 0, 1)$ and their utility functions are given by

$$u_1(x_1) = \min\{x_1^1, x_1^2\}$$

$$u_2(x_2) = \min\{x_2^2, x_2^3\}$$

$$u_3(x_3) = \min\{x_3^1, x_3^3\}$$

so that they have $L$-shaped indifference curves. Then, the aggregate excess demand function is given by

$$\hat{z}_1^1(p) = -\frac{p^2}{p^1 + p^2} + \frac{p^3}{p^1 + p^3}$$

$$\hat{z}_2^2(p) = -\frac{p^3}{p^2 + p^3} + \frac{p^1}{p^1 + p^2}$$

$$\hat{z}_3^3(p) = -\frac{p^1}{p^1 + p^3} + \frac{p^2}{p^2 + p^3}$$

from which the only possible competitive equilibrium price system is given by $p^* = (1, 1, 1)$.

Then, the dynamic adjustment equation is given by

$$\dot{p} = \hat{z}(p).$$

We know that $\|p(t)\| = \text{constant}$ for all $t$ by Walras’ Law. Now we want to show that $\prod_{l=1}^3 p^l(t) = \text{constant}$ for all $t$. Indeed,

$$\frac{d}{dt}(\prod_{l=1}^3 p(t)) = \dot{p}^1 p^2 p^3 + \dot{p}^2 p^1 p^3 + \dot{p}^3 p^1 p^2$$

$$= \hat{z}_1^1 p^2 p^3 + \hat{z}_2^2 p^1 p^3 + \hat{z}_3^3 p^1 p^2 = 0.$$ 

Now, we show that the dynamic process is not globally stable. First choose the initial prices $p^l(0)$ such that $\sum_{l=1}^3 [p^l(0)]^2 = 3$ and $\prod_{l=1}^3 p^l(0) \neq 1$. Then, $\sum_{l=1}^3 [p^l(t)]^2 = 3$ and $\prod_{l=1}^3 p^l(t) \neq 1$ for all $t$. Since $\sum_{l=1}^3 [p^l(t)]^2 = 3$ and the only possible equilibrium prices are $p^* = (1, 1, 1)$, the solution of the above system of differential equations cannot converge to the equilibrium price $p^* = (1, 1, 1)$.
In this example, we may note the following facts. (i) there is no substitution
effect, (ii) the indifference curve is not strictly convex, and (iii) the difference
curve has a kink and hence is not differentiable. Scarf (1961) also provided the
examples of instability in which the substitution effect is present for the case of
Giffen’s goods. Thus, Scarf’s examples indicate that instability may occur in a
wide variety of cases.

10.7 Abstract Economy

The abstract economy defined by Debreu (Econometrica, 1952) generalizes the
notion of N-person Nash noncooperative game in that a player’s strategy set de-
pendson the strategy choices of all the other players and can be used to prove
the existence of competitive equilibrium since the market mechanism can be re-
garded as an abstract economy as shown by Arrow and Debreu (Econometrica
1954). Debreu (1952) proved the existence of equilibrium in abstract economies
with finitely many agents and finite dimensional strategy spaces by assuming
the existence of continuous utility functions. Since Debreu’s seminal work on ab-
stract economies, many existence results have been given in the literature. Shafer
and Sonnenschein (1975), and Tian (1992) extended Debreu’s results to abstract
economies without ordered preferences.

10.7.1 Equilibrium in Abstract Economy

Let $N$ be the set of agents which is any countable or uncountable set. Each agent
$i$ chooses a strategy $x_i$ in a set $X_i$ of $\mathbb{R}^L$. Denote by $X$ the (Cartesian) product
$\prod_{j \in N} X_j$ and $X_{-i}$ the product $\prod_{j \in N \setminus \{i\}} X_j$. Denote by $x$ and $x_{-i}$ an element of
$X$ and $X_{-i}$. Each agent $i$ has a payoff (utility) function $u_i : X \to \mathbb{R}$. Note that
agent $i$’s utility is not only dependent on his own choice, but also dependent on
the choice of the others. Given $x_{-i}$ (the strategies of others), the choice of the
$i$-th agent is restricted to a non-empty feasible choice set $F_i(x_{-i}) \subset X_i$, the $i$-th
agent chooses $x_i \in F_i(x_{-i})$ so as to maximize $u_i(x_{-i}, x_i)$ over $F_i(x_{-i})$. 
An abstract economy (or called generalized game) $\Gamma = (X_i, F_i, u_i)_{i \in N}$ is defined as a family of ordered triples $(X_i, F_i, P_i)$.

**Definition 10.7.1** A vector $x^* \in X$ is said to be an *equilibrium of an abstract economy* if $\forall i \in N$

(i) $x^*_i \in F_i(x^*_{-i})$ and

(ii) $x^*_i$ maximizes $u_i(x^*_{-i}, x_i)$ over $F_i(x^*_{-i})$.

If $F_i(x_{-i}) \equiv X_i, \forall i \in N$, the abstract economy reduces to the conventional game $\Gamma = (X_i, u_i)$ and the equilibrium is called a *Nash equilibrium*.

**Theorem 10.7.1 (Arrow-Debreu)** Let $X$ be a non-empty compact convex subset of $\mathbb{R}^{nL}$. Suppose that

i) the correspondence $F: X \to 2^X$ is a continuous correspondence with non-empty compact and convex values,

ii) $u_i : X \times X \to \mathbb{R}$ is continuous,

iii) $u_i : X \times X \to \mathbb{R}$ is either quasi-concave in $x_i$ or it has a unique maximum on $F_i(x_{-i})$ for all $x_{-i} \in X_{-i}$.

Then $\Gamma$ has an equilibrium.

**Proof.** For each $i \in N$, define the maximizing correspondence

$$M_i(x_{-i}) = \{x_i \in F_i(x_{-i}) : u_i(x_{-i}, x_i) \geq u_i(x_{-i}, z_i), \forall z_i \in F_i(x_{-i})\}.$$  

Then, the correspondence $M_i : X_{-i} \to 2^{X_i}$ is non-empty compact and convex valued because $u_i$ is continuous in $x$ and quasi-concave in $x_i$ and $F_i(x_{-i})$ is non-empty convex compact-valued. Also, by the Maximum Theorem, $M_i$ is an upper hemi-continuous correspondence. Therefore the correspondence

$$M(x) = \prod_{i \in N} M_i(x_{-i})$$
is an upper hemi-continuous correspondence with non-empty convex compact-values. Thus, by Kakutani’s Fixed Point Theorem, there exists $x^* \in X$ such that $x^* \in M(x^*)$ and $x^*$ is an equilibrium in the generalized game. ■

A competitive market mechanism can be regarded as an abstract economy. For simplicity, consider an exchange economy $e = (X_i, u_i, w_i)_{i \in \mathbb{N}}$. Define an abstract economy $\Gamma = (Z, F_i, u_{n+1})_{i \in \mathbb{N}}$ as follows. Let

$$
\begin{align*}
Z_i &= X_i \quad i = 1, \ldots, n \\
Z_{n+1} &= S^{l-1} \\
F_i(x_i, p) &= \{x_i \in X_i : p \cdot x_i \leq p \cdot w_i\} \quad i = 1, \ldots, n \\
F_{n+1} &= S^{l-1} \\
u_{n+1}(p, x) &= \sum_{i=1}^{n} p \cdot (x_i - w_i) \quad (10.59)
\end{align*}
$$

Here $N + 1$'s sole purpose is to change prices, and thus can be regarded as a Walrasian auctioneer. Then, we verify that the economy $e$ has a competitive equilibrium if the abstract economy defined above has an equilibrium by noting that $\sum_{i=1}^{n} x_i \leq \sum_{i=1}^{n} w_i$ at the equilibrium of the abstract equilibrium.

10.7.2 Existence of Equilibrium for General Preferences

The above theorem on the existence of an equilibrium in abstract economy has assumed the preference relation is an ordering and can be represented by a utility function. In this subsection, we consider the existence of equilibrium in an abstract economy where individuals' preferences $\succeq$ may be non total or not-transitive. Define agent $i$'s preference correspondence $P_i : X \rightarrow 2^{X_i}$ by

$$
P_i(x) = \{y_i \in X_i : (y_i, x_{-i}) \succ_i (x_i, x_{-i})\}
$$

We call $\Gamma = (X_i, F_i, P_i)_{i \in \mathbb{N}}$ an abstract economy.

A generalized game (or an abstract economy) $\Gamma = (X_i, F_i, P_i)_{i \in \mathbb{N}}$ is defined as a family of ordered triples $(X_i, F_i, P_i)$. An equilibrium for $\Gamma$ is an $x^* \in X$ such that $x^* \in F(x^*)$ and $P_i(x^*) \cap F_i(x^*) = \emptyset$ for each $i \in \mathbb{N}$.
Shafer and Sonnenschein (1975) proved the following theorem that generalizes the above theorem to an abstract economy with non-complete/non-transitive preferences.

**Theorem 10.7.2 (Shafer-Sonnenschein)** Let \( \Gamma = (X_i, F_i, P_i)_{i \in N} \) be an abstract economy satisfying the following conditions for each \( i \in N \):

(i) \( X_i \) is a non-empty, compact, and convex subset in \( \mathbb{R}^k \),

(ii) \( F_i : X \rightarrow 2^{X_i} \) is a continuous correspondence with non-empty, compact, and convex values,

(iii) \( P_i \) has open graph,

(iv) \( x_i \not\in \text{con} P_i(x) \) for all \( x \in Z \).

Then \( \Gamma \) has an equilibrium.

This theorem requires the preferences have open graph. Tian (International Journal of Game Theory, 1992) proved the following theorem that is more general and generalizes the results of Debreu (1952), Shafer and Sonnenschein (1975) by relaxing the openness of graphs or lower sections of preference correspondences. Before proceeding to the theorem, we state some technical lemmas which were due to Micheal (1956, Propositions 2.5, 2.6 and Theorem 3.1”).

**Lemma 10.7.1** Let \( X \subset \mathbb{R}^M \) and \( Y \subset \mathbb{R}^K \) be two convex subsets and \( \phi : X \rightarrow 2^Y, \psi : X \rightarrow 2^Y \) be correspondences such that

(i) \( \phi \) is lower hemi-continuous, convex valued, and has open upper sections,

(ii) \( \psi \) is lower hemi-continuous,

(iii) for all \( x \in X \), \( \phi(x) \cap \psi(x) \neq \emptyset \).

Then the correspondence \( \theta : X \rightarrow 2^Y \) defined by \( \theta(x) = \phi(x) \cap \psi(x) \) is lower hemi-continuous.
Lemma 10.7.2 Let $X \subset \mathbb{R}^M$ and $Y \subset \mathbb{R}^K$ be two convex subsets, and let $\phi : X \rightarrow 2^Y$ be lower hemi-continuous. Then the correspondence $\psi : X \rightarrow 2^Y$ defined by $\psi(x) = \text{con}\phi(x)$ is lower hemi-continuous.

Lemma 10.7.3 Let $X \subset \mathbb{R}^M$ and $Y \subset \mathbb{R}^K$ be two convex subsets. Suppose $F : X \rightarrow 2^Y$ is a lower hemi-continuous correspondence with non-empty and convex values. Then there exists a continuous function $f : X \rightarrow Y$ such that $f(x) \in F(x)$ for all $x \in X$.

Theorem 10.7.3 (Tian) Let $\Gamma = (X_i, F_i, P_i)_{i \in \mathbb{N}}$ be a generalized game satisfying for each $i \in \mathbb{N}$:

(i) $X_i$ is a non-empty, compact, and convex subset in $\mathbb{R}^k$,

(ii) $F_i$ is a continuous correspondence, and $F_i(x)$ is non-empty, compact, and convex for all $x \in X$,

(iii) $P_i$ is lower hemi-continuous and has open upper sections,

(iv) $x_i \notin \text{con}P_i(x)$ for all $x \in F$.

Then $\Gamma$ has an equilibrium.

Proof. For each $i \in \mathbb{N}$, define a correspondence $A_i : X \rightarrow 2^{X_i}$ by $A_i(x) = F_i(x) \cap \text{con}P_i(x)$. Let $U_i = \{x \in X : A_i(x) \neq \emptyset\}$. Since $F_i$ and $P_i$ are lower hemi-continuous in $X$, so are they in $U_i$. Then, by Lemma 10.7.2, $\text{con}P_i$ is lower hemi-continuous in $U_i$. Also since $P_i$ has open upper sections in $X$, so does $\text{con}P_i$ in $X$ and thus $\text{con}P_i$ in $U_i$. Further, $F_i(x) \cap \text{con}P_i(x) \neq \emptyset$ for all $x \in U_i$. Hence, by Lemma 10.7.1, the correspondence $A_i|U_i : U_i \rightarrow 2^{X_i}$ is lower hemi-continuous in $U_i$ and for all $x \in U_i$, $F(x)$ is non-empty and convex. Also $X_i$ is finite dimensional. Hence, by Lemma 10.7.3, there exists a continuous function $f_i : U_i \rightarrow X_i$ such that $f_i(x) \in A_i(x)$ for all $x \in U_i$. Note that $U_i$ is open since $A_i$ is lower hemi-continuous. Define a correspondence $G_i : X \rightarrow 2^{X_i}$ by

$$G_i(x) = \begin{cases}
\{f_i(x)\} & \text{if } x \in U_i \\
F_i(x) & \text{otherwise}
\end{cases} \quad (10.60)$$

403
Then $G_i$ is upper hemi-continuous. Thus the correspondence $G : X \to 2^X$ defined by $G(x) = \prod_{i \in N} G_i(x)$ is upper hemi-continuous and for all $x \in X$, $G(x)$ is non-empty, closed, and convex. Hence, by Kakutani’s Fixed Point Theorem, there exists a point $x^* \in X$ such that $x^* \in G(x^*)$. Note that for each $i \in N$, if $x^* \in U_i$, then $x^*_i = f_i(x^*) \in A(x^*) \subset \text{con } P_i(x^*)$, a contradiction to (iv). Hence, $x^* \not\in U_i$ and thus for all $i \in N$, $x^*_i \in F_i(x^*)$ and $F_i(x^*) \cap \text{con } P_i(x^*) = \emptyset$ which implies $F_i(x^*) \cap P_i(x^*) = \emptyset$. Thus $\Gamma$ has an equilibrium.

Note that a correspondence $p$ has open graph implies that it has upper and lower open sections; a correspondence $p$ has lower open sections implies $p$ is lower hemi-continuous. Thus, the above theorem is indeed weaker.

### 10.8 Reference

**Books and Monographs:**


**Papers:**


406

Wertlehre”, Ergebnisse eines Mathematischen Kolloquiums, No. 6, 10-11.

without Ordered Preferences”, Journal of Mathematical Economics, Vol. 2,
345-348.

Stackelberg, H. von (1933). “Zwei kritische Bemerkungen zur Preisthorie Gustav
Cassels”, Zeitschrift für Nationalökonomie, No. 4, 456-472.


tinuous Payoffs and Non-Compact Choice Spaces”, Journal of Mathematical
Economics, Vol. 21, No. 4, 379-388.

Spaces and Payoffs: A Full Characterization”, mimeo.


Wald, A. (1935). “Über die eindeutige positive Lösbarkeit der neuen Pro-
duktionsgleichungen”, Ergebnisse eines mathematischen Kolloquiums, No.
6, 12-20.


Chapter 11

Normative Theory of Equilibrium: Its Welfare Properties

11.1 Introduction

In the preceding chapter, we studied the conditions which would guarantee the existence, uniqueness, and stability of competitive equilibrium. In this chapter, we will study the welfare properties of competitive equilibrium, that is, how does the selfish individual behavior leads to optimality of social welfare. Normative theory is an important part of economics since both economists and policy makers are interested in “what should be” as well as “what”. Economists are interested not only in describing the world of the market economy, but also in evaluating it. Does a competitive market do a “good” job in allocating resources? Adam Smith’s “invisible hand” says that market economy is efficient. Then in what sense and under what conditions is a market efficient? The concept of efficiency is related to a concern with the well-being of those in the economy. The normative analysis can not only help us understand what advantage the market mechanism has, it can also help us to evaluate an economic system used in the real world,
as well as help us understand why China and East European countries want to change their economic institutions.

The term of economic efficiency consists of three requirements:

1. Exchange efficiency: goods are traded efficiently so that no further mutual beneficial trade can be obtained.
2. Production efficiency: there is no waste of resources in producing goods.
3. Product mix efficiency, i.e., the mix of goods produced by the economy reflects the preferences of those in the economy.

11.2 Pareto Efficiency of Allocation

When economists talk about the efficiency of allocation resources, it means Pareto efficiency. It provides a minimum criterion for efficiency of using resources.

Under the market institution, we want to know what is the relationship between a market equilibrium and Pareto efficiency. There are two basic questions: (1) If a market (not necessarily competitive) equilibrium exists, is it Pareto efficient? (2) Can any Pareto efficient allocation be obtained through the market mechanism by redistributing endowments?

The concept of Pareto efficiency is not just to study the efficiency of a market economy, but it also can be used to study the efficiency of any economic system. First let us define the notion of Pareto improvement.

**Definition 11.2.1 (Pareto Improvement)** An allocation can have a Pareto improvement if it is feasible and there is another feasible allocation such that one person would be better off and all other persons are not worse off.

**Definition 11.2.2 (Pareto Efficiency, also called Pareto Optimality)** An allocation is Pareto efficient or Pareto optimal (in short, P.O) if there is no Pareto improvement, that is, if it is feasible and there is no other feasible allocation such that one person would be better off and all other persons are not worse off.
More precisely, for exchange economies, a feasible allocation \( x \) is P.O. if there is no other allocation \( x' \) such that

(i) \( \sum_{i=1}^{n} x'_i \leq \sum_{i=1}^{n} w_i \)

(ii) \( x'_i \succeq_i x_i \) for all \( i \) and \( x'_k \succ_k x_k \) for some \( k = 1, \ldots, n \).

For production economy, \( (x, y) \) is Pareto optimal if and only if:

(1) \( \hat{x} \leq \hat{y} + \hat{w} \)

(2) there is no feasible allocation \( (x', y') \) s.t.

\[ x'_i \succeq x_i \text{ for all } i \]
\[ x'_k \succ_k x_k \text{ for some } k \]

A weaker concept about economic efficiency is the so-called weak Pareto efficiency.

**Definition 11.2.3 (Weak Pareto Efficiency)** An allocation is *weakly Pareto efficient* if it is feasible and there is no other feasible allocation such that all persons are strictly better off.

**Remark 11.2.1** Some textbooks such as Varian (1992) has used weak Pareto optimality as the definition of Pareto optimality. Under which conditions are they equivalent? It is clear that Pareto efficiency implies weak Pareto efficiency. But the converse may not be true. However, under the continuity and strong monotonicity of preferences, the converse is true.

**Proposition 11.2.1** Under the continuity and strong monotonicity of preferences, weak Pareto efficiency implies Pareto efficiency.

**Proof.** Suppose \( x \) is weakly Pareto efficient but not Pareto efficient. Then, there exists a feasible allocation \( x' \) such that \( x'_i \succeq_i x_i \) for all \( i \) and \( x'_k \succ_k x_k \) for some \( k \).
Define \( \bar{x} \) as follows
\[
\bar{x}_k = (1 - \theta)x'_k \\
\bar{x}_i = x'_i + \frac{\theta}{n-1}x'_k \text{ for } i \neq k
\]
Then we have
\[
\bar{x}_k + \sum_{i \neq k} \bar{x}_i = (1 - \theta)x'_k + \sum_{i \neq k} (x'_i + \frac{\theta}{n-1}x'_k) = \sum_{i=1}^{n} x'_i
\] (11.1)
which means \( \bar{x} \) is feasible. Furthermore, by the continuity of preferences, \( \bar{x}_k = (1 - \theta)x'_k \succ x_k \) when \( \theta \) is sufficiently close to zero, and \( \bar{x}_i \succ_i x_i \) for all \( i \neq k \) by the strong monotonicity of preferences. This contradicts the fact that \( x \) is weakly Pareto optimal. Thus, we must have every weak Pareto efficient allocation is Pareto efficient under the monotonicity and continuity of preferences. 

**Remark 11.2.2** The trick of taking a little from one person and then equally distribute it to the others will make every one better off is a useful one. We will use the same trick to prove the second theorem of welfare economics. Notice that, the strong monotonicity condition in the above theorem cannot is indispensable. Why? the readers may check it by themselves.

**Remark 11.2.3** The above proposition also depends on an implicit assumption that the goods under consideration are all private goods. Tian (Economics Letters, 1988) showed, by example that, if goods are public goods, we may not have such equivalence between Pareto efficiency and weak Pareto efficiency under the monotonicity and continuity of preferences.

The set of Pareto efficient allocations can be shown with the Edgeworth Box. Every point in the Edgeworth Box is attainable. A is the starting point. Is “A” a weak Pareto optimal? No. The point C, for example in the shaded area is better off to both persons. Is the point “B” a Pareto optimal? YES. Actually, all tangent points are Pareto efficient points where at least some agent does not have incentives to trade. The locus of all Pareto efficient is called the contract curve.
Remark 11.2.4 *Equity* and *Pareto efficiency* are two different concepts. The points like $O_A$ or $O_B$ are Pareto optimal points, but these are extremely unequal. To make an allocation be relatively equitable and also efficient, government needs to implement some institutional arrangements such as tax, subsidy to balance between equity and efficiency, but this is a value judgement and policy issue.

We will show that, when Pareto efficient points are given by tangent points of two persons’ indifference curves, it should satisfy the following conditions:

$$MRS^A_{x^1,x^2} = MRS^B_{x^1,x^2}$$

$$x_A + x_B = \tilde{w}.$$

When indifference curves of two agents are never tangent, we have the following cases.

Case 1. For linear indifference curves with non-zero (negative) slope, indifference curves of agents may not be tangent. How can we find the set of Pareto efficient allocations in this case? We can do it by comparing the steepness of indifference curves.
Figure 11.2: The set of Pareto efficient allocations is given by the upper and left edges of the box when indifference curves are linear and Person B’s indifference curves are steeper.

\[ MRS^A_{x_1,x_2} < MRS^B_{x_1,x_2} \]  \hspace{1cm} (11.2)

In this case, when indifference curves for B is given, say, by Line AA, then K is a Pareto efficient point. When indifferent curves for A is given, say, by Line BB, then p is a Pareto efficient point. Contract curve is then given by the upper and left edge of the box. As a usual method of finding all possible Pareto efficient points, wether or not difference curves have tangent points, taken one person’s any indifference as given, find another person’s utility maximizing point. The maximizing point must be Pareto efficient point.

Case 2. Suppose that indifference curves are given by

\[ u_A(x_A) = x_A^2 \]

and

\[ u_B(x_B) = x_B^1 \]
Figure 11.3: The only Pareto efficient allocation point is given by the upper and left corner of the box when individuals only care about one commodity.

Then, only Pareto efficient is the left upper corner point. But the set of weakly Pareto efficient is given by the upper and left edge of the box. Notice that utility functions in this example are continuous and monotonic, but a weak Pareto efficient allocation may not be Pareto efficient. This example shows that the strong monotonicity cannot be replaced by the monotonicity for the equivalence of Pareto efficiency and weak efficiency.

Case 3. Now suppose that indifference curves are perfect complementary. Then, utility functions are given by

\[ u_A(x_A) = \min\{ax_A^1, bx_A^2\} \]

and

\[ u_B(x_B) = \min\{cx_B^1, dx_B^2\} \]

A special case is the one where \(a = b = c = d = 1\).

Then, the set of Pareto optimal allocations is the area given by points between two 45° lines.
Case 4. One person’s indifference curves are “thick.” In this case, an weak Pareto efficient allocation may not be Pareto efficient.

Figure 11.4: The first figure shows that the contract curve may be the “thick” when indifference curves are perfect complementary. The second figure shows that a weak Pareo efficient allocation may not Pareto efficient when indifference curves are “thick.”
11.3 The First Fundamental Theorem of Welfare Economics

There is a well-known theorem, in fact, one of the most important theorems in economics, which characterizes a desirable nature of the competitive market institution. It claims that every competitive equilibrium allocation is Pareto efficient. A remarkable part of this theorem is that the theorem requires few assumptions, much fewer than those for the existence of competitive equilibrium. Some implicit assumptions in this section are that preferences are orderings, complete information, goods are divisible, and there are no public goods, or externalities. In the following, we will introduce three versions of the first theorems of welfare economics.

**Theorem 11.3.1 (The First Fundamental Theorem of Welfare Economics)**

If \((x, y, p)\) is a competitive equilibrium, then \((x, y)\) is weakly Pareto efficient, and further under local non-satiation, it is Pareto efficient.

Proof: Suppose \((x, y)\) is not weakly Pareto optimal, then there exists another feasible allocation \((x', y')\) such that \(x'_i \succ_i x_i\) for all \(i\). Thus, we must have \(p \cdot x'_i > p \cdot w_i + \sum_{j=1}^{J} \theta_{ij} p \cdot y_j\) for all \(i\). Therefore, by summation, we have

\[
\sum_{i=1}^{n} p \cdot x'_i > \sum_{i=1}^{n} p \cdot w_i + \sum_{j=1}^{J} p \cdot y_j.
\]

Since \(p \cdot y_j \geq p \cdot y'_j\) for all \(y'_j \in Y_j\) by profit maximization, we have

\[
\sum_{i=1}^{n} p \cdot x'_i > \sum_{i=1}^{n} p \cdot w_i + \sum_{j=1}^{J} p \cdot y_j.
\]

or

\[
p \left[ \sum_{i=1}^{n} x'_i - \sum_{i=1}^{n} w_i - \sum_{j=1}^{J} y'_j \right] > 0,
\]

which contradicts the fact that \((x', y')\) is feasible.
To show Pareto efficiency, suppose \((x, y)\) is not Pareto optimal. Then there exists another feasible allocation \((x', y')\) such that \(x'_i \succ_i x_i\) for all \(i\) and \(x'_k \succ_k x_k\) for some \(k\). Thus we have, by local non-satiation,

\[ p \cdot x'_i \geq p \cdot w_i + \sum_{j=1}^{J} \theta_{ij} p \cdot y_j \quad \forall i \]

and by \(x'_k \succ_k x_k\) (why? readers can prove this),

\[ p \cdot x'_k > p \cdot w_k + \sum_{j=1}^{J} \theta_{kj} p \cdot y_j \]

and thus

\[ \sum_{i=1}^{n} p \cdot x'_i > \sum_{i=1}^{n} p \cdot w_i + \sum_{j=1}^{J} p \cdot y_j \geq \sum_{i=1}^{n} p \cdot w_i + \sum_{j=1}^{J} p \cdot y'_j. \tag{11.5} \]

Again, it contradicts the fact that \((x', y')\) is feasible.

\[ \square \]

**Remark 11.3.1** If the local non-satiation condition is not satisfied, a competitive equilibrium allocation \(x\) may not be Pareto optimal, say, for the case of thick indifference curves.

The conditions of this theorem may be, at least approximately, satisfied by various practical settings. It worth mentioning that market does not always work, it may fail in many environments. Nevertheless, market institution is by far the most efficient and irreplaceable way for allocating resources. That is why market-oriented reforms have been widely adopted by China, Russia and many eastern European Countries who were used to be planned economies. Still, we need to focus on assumptions of this theorem. For example, the market may lead to inefficient outcome in the presence of incomplete information, imperfect competition, externality. We will discuss various settings under which market may fail and needs a revision in the last two parts of this note.
Figure 11.5: A CE allocation may not be Pareto efficient when the local non-satiation condition is not satisfied.

**Remark 11.3.2** Note that neither convexity of preferences nor convexity of production set is assumed in the theorem. The conditions required for Pareto efficiency of competitive equilibrium is much weaker than the conditions for the existence of competitive equilibrium.
Figure 11.6: A CE allocation may not be Pareto efficient when goods are indivisible.

**Remark 11.3.3** If goods are indivisible, then a competitive equilibrium allocation $x$ may not be Pareto optimal. The point $x$ is a competitive equilibrium allocation, but not Pareto optimal since $x'$ is preferred by person 1. Hence, the divisibility condition cannot be dropped for the First Fundamental Theorem of Welfare Theorem to be true.

For constrained competitive equilibrium, we could obtain an analogous result by replace local non-satiation with strong monotonicity.

**Theorem 11.3.2 (The First Fundamental Theorem Under CCE)** If $(x, y, p)$ is a constrained competitive equilibrium, then $(x, y)$ is weakly Pareto efficient. Moreover, if the preference are strongly monotone, then it is PO.

**Proof.** Using method analogous to the proof of Theorem 11.3.1, we can obtain the weak Pareto optimality of CCE. Next, we prove, by contradiction, that CCE
is also PO. Suppose that \((x, y)\) is not PO, then there exists another feasible allocation \((x', y')\), such that
\[
x'_i \succ_i x_i, \exists k, x'_k \succ_k x_k,
\]
for \(\forall i \in N\). Now, we prove that under strong monotonicity if \(x'_i \succ_i x_i\), we must have
\[
px'_i \geq pw_i + \sum_{j=1}^{J} \theta_{ij}py_j.
\]
Otherwise, we must have \(px'_i < pw_i + \sum_{j=1}^{J} \theta_{ij}py_j\). Therefore, there exists a good \(l\) such that
\[
x'_i < w_l + \sum_{j=1}^{J} \theta_{lj}y_j.
\]
Hence, there exists \(\hat{x}_i \in X_i\) satisfying
\[
\hat{x}_i \geq x'_i, \quad px'_i < pw_i + \sum_{j=1}^{J} \theta_{ij}py_j, \quad \hat{x}_i \leq w_i + \sum_{j=1}^{J} \theta_{ij}y_j.
\]
It follows from the strong monotonicity of preference
\[
\hat{x}_i \succ x'_i \succ_i x_i.
\]
This contradicts with the fact that \((x, y, p)\) is CCE. Therefore, for \(\forall i \in N\), we must have
\[
px'_i \geq pw_i + \sum_{j=1}^{J} \theta_{ij}py_j.
\]
For individual \(k\), it follows from \(x'_k \succ_k x_k\) that
\[
px'_k > pw_k + \sum_{j=1}^{J} \theta_{kj}py_j,
\]
therefore,
\[
\sum_{i=1}^{n} px'_i > \sum_{i=1}^{n} pw_i + \sum_{j=1}^{J} py_j \geq \sum_{i=1}^{n} pw_i + \sum_{j=1}^{J} py'_j.
\] (11.6)
This contradicts the fact that \((x', y')\) is feasible. ■
The strong monotonicity in the above theorem cannot be weakened by requiring instead merely local non-satiation.

*Competitive equilibrium with transfer payments* refers to the equilibrium obtained after redistribution of individual wealth.

**Definition 11.3.1 (CE with transfer payments)** For economy \( e = (e_1, \ldots, e_n, \{Y_j\}) \), an allocation \((x, y) \in X \times Y\) and price vector \(p \in \mathbb{R}_+^L\) constitutes a competitive equilibrium with transfer payments, if there exists a sequence of wealth \((I_1, \ldots, I_n)\) with \(\sum_i I_i = p \cdot \sum_i w_i + \sum_j p \cdot y_j\) such that the following conditions are met:

(i) for \(\forall i = 1, \ldots, n\), \(x' \succ_i x_i\) implies \(p \cdot x' > I_i\);

(ii) for \(\forall y' \in Y_j\), \(p \cdot y_j \geq p \cdot y'_j\);

(iii) \(\sum_i x_i \leq \sum_i w_i + \sum_j y_j\) (feasible condition).

Condition (i) is the utility-maximizing condition, that is, \(x_i\) is the optimal solution of \(\succ_i\) in budget set \(\{x'_i \in X_i : p \cdot x'_i \leq I_i\}\) Condition (ii) is profit-maximizing condition.

CE with transfers is a generalized version of CE, it depends only on the aggregate endowment rather than its initial allocation among individuals. The government could reduce income dispersion through redistributing the wealth among individuals and leave the other works to the market. This policy will not affect the efficiency of market institution.

**Definition 11.3.2 (Quasi Equilibrium with Transfer Payments)** Given an economy \( e = (e_1, \ldots, e_n, \{Y_j\}) \), an allocation \((x, y) \in X \times Y\) and price vector \(p \in \mathbb{R}_+^L\) constitute a quasi-CE with transfer payments if all conditions in definition 11.3.1 are satisfied except that condition (i) is replaced with

\[(i') \ x'_i \succ_i x_i \text{ implies } p \cdot x'_i \geq I_i, \forall i = 1, \ldots, n.\]

**Remark 11.3.4** We have several remarks on CE with transfers.
(1) It is obvious that condition $(i)$ implies $(i')$, CE with transfers is quasi-CE with transfers, but the opposite is not necessarily true.

(2) From $(i')$, we can see that if $x_i$ is the optimal consumption bundle within budget set $\{x_i' \in X_i : p \cdot x_i' \leq I_i \}$, then $x_i' \succ_i x_i$ and $p \cdot x_i' \leq I_i$ cannot hold simultaneously.

(3) If local non-satiation is satisfied, then $p \cdot x_i \geq I_i$. Indeed, if $p \cdot x_i < I_i$, then there exist $x_i'$ such that $x_i' \succ_i x_i$, and $p \cdot x_i' < I_i$, this contradicts condition $(i')$. Besides, feasible condition $(iii)$ implies $\sum_i p \cdot x_i \leq \sum_i I_i$. Therefore, under local non-satiation, we must have $p x_i = I_i$. Hence, condition $(i')$ in the definition of quasi-CE with transfers could be replaced by the following condition

$$(i'') x_i' \succ_i x_i \text{ implies } p x_i' \geq p x_i, \forall i = 1, \ldots, n.$$ 

Furthermore, under local non-satiation, $(i'')$ is equivalent to the condition that $x_i$ minimizes expenditure within set $\{x_i \in X_i : x_i' \succ_i x_i \}$.

(4) For CE with transfer payments, under local non-satiation condition, condition $(i)$ could be replaced by

$$(i'') x_i' \succ_i x_i \text{ implies } p x_i' > p x_i, \forall i = 1, \ldots, n.$$ 

Under what conditions will a quasi-CE with transfers be a CE with transfers? That is to say under what condition will $x_i' \succ_i x_i$ implies $p x_i' > I_i$. We have the next proposition.

**Proposition 11.3.1** Suppose that $X_i$ is a convex set, preferences are continuous. If $(x, y, p)$ is a quasi-CE with transfers, and there exists $\bar{x}_i \in X_i$ such that $p \bar{x}_i < I_i$, then $x_i' \succ_i x_i$ implies $p x_i' > I_i$, and therefore $(x, y, p)$ is a CE with transfer payments.

**Proof.** We prove it by contradiction. If the conclusion does not hold, then for $\forall x_i' \succ_i x_i$, we have $p x_i' = I_i$. Therefore, for an arbitrary $\alpha \in [0, 1)$, we have
\[ \alpha x'_i + (1 - \alpha)\bar{x}_i \in X_i \text{ and } p[\alpha x'_i + (1 - \alpha)\bar{x}_i] < I_i. \] It follows from the continuity of preferences that when \( \alpha \) is sufficiently close to 1, we have \( \alpha x'_i + (1 - \alpha)\bar{x}_i \succ_i x_i, \) a contradiction! Therefore, we must have \( x'_i \succ_i x_i \) implies \( px'_i \geq I_i. \) ■

A result may follows directly from the above proposition.

**Proposition 11.3.2** Suppose for all \( i, X_i \) is convex, \( 0 \in X_i, \) and \( \succeq_i \) is continuous. Then any quasi-CE with transfer payments under \( (I_1, \ldots, I_n) > 0 \) is also CE with transfer payments.

By an argument analogous to the proof the first fundamental theorem, we have the following theorem.

**Theorem 11.3.3 (The First Fundamental Theorem with Transfers)** Let \( (x, y, p) \) be CE with transfer payments, then \( x \) is weakly Pareto efficient. Moreover, if local non-satiation of preferences is also satisfied, \( x \) is Pareto efficient.

### 11.4 Calculations of Pareto Optimum by First-Order Conditions

We now discuss how to get the PO allocation. The basic method is to get it through solving a constrained optimization problem. In the \( 2 \times 2 \) pure exchange economy discussed above, we are able to obtain the Pareto optimal allocation by maximizing one consumer’s utility subject to the constraint that the other consumer’s utility is fixed. This approach could be adopted regardless of whether or not the two consumers’ indifference curves are tangent. The following proposition shows that this approach could be generalized to cases with multiple agents. More importantly, it transforms the problem of getting the Pareto optimal allocation to the constrained optimization problem of individuals. Therefore, the first-order conditions for Pareto efficiency may obtained from the first-order conditions of individual optimization when utility functions are differentiable.
11.4.1 Exchange Economies

**Proposition 11.4.1** A feasible allocation \( \mathbf{x}^* \) is Pareto efficient if and only if \( \mathbf{x}^* \) solves the following problem for all \( i = 1, 2, \ldots, n \)

\[
\max_x u_i(x_i) \\
\text{s.t. } \quad \sum x_k \leq \sum w_k \\
\quad u_k(x_k) \geq u_k(x_k^*) \quad \text{for } k \neq i
\]

**Proof.** Suppose \( \mathbf{x}^* \) solves all maximizing problems but \( \mathbf{x}^* \) is not Pareto efficient. This means that there is some allocation \( \mathbf{x}' \) where one consumer is better off, and the others are not worse off. But, then \( \mathbf{x}^* \) does not solve one of the maximizing problems, a contradiction. Conversely, suppose \( \mathbf{x}^* \) is Pareto efficient, but it does not solve one of the problems. Instead, let \( \mathbf{x}' \) solve that particular problem. Then \( \mathbf{x}' \) makes one of the agents better off without hurting any of the other agents, which contradicts the assumption that \( \mathbf{x}^* \) is Pareto efficient. The proof is completed.

If utility functions \( u_i(x_i) \) are differentiable, then we can define the Lagrangian function to get the optimal solution to the above problem:

\[
L = u_i(x_i) + q(\bar{w} - \bar{x}) + \sum_{k \neq i} t_k [u_k(x_k) - u_k(x_k^*)]
\]

The first order conditions are then given by

\[
\frac{\partial L}{\partial x^l_i} = \frac{\partial u_i(x_i)}{\partial x^l_i} - q^l \leq 0 \quad \text{with equality if } x^l_i > 0 \quad l = 1, \ldots, L, i = 1, \ldots (11.7)
\]

\[
\frac{\partial L}{\partial x^l_k} = t_k \frac{\partial u_k(x_k)}{\partial x^l_k} - q^l \leq 0 \quad \text{with equality if } x^l_k > 0 \quad l = 1, \ldots, L; k \neq l (11.8)
\]

By (11.7), when \( x^* \) is an interior solution, we have

\[
\frac{\partial u_i(x_i)}{\partial x^l_i} = \frac{q^l}{q^h} = \text{MRS}_{x^l_i, x^h_i} \quad (11.9)
\]

By (11.8)

\[
\frac{\partial u_k(x_k)}{\partial x^l_k} = \frac{q^l}{q^h} \quad (11.10)
\]
Thus, we have
\[
MRS_{x_i, x_i^l} = \cdots = MRS_{x_n, x_n^l} \quad l = 1, 2, \ldots, L; h = 1, 2, \ldots, L. \quad (11.11)
\]
which are the necessary conditions for the interior solutions to be Pareto efficient, which means that the MRS of any two goods are all equal for all agents. They become sufficient conditions when utility functions \( u_i(x_i) \) are differentiable and quasi-concave.

### 11.4.2 Production Economies

For simplicity, we assume there is only one firm. Let \( T(y) \leq 0 \) be the transformation frontier. Similarly, we can prove the following proposition.

**Proposition 11.4.2** A feasible allocation \((x^*, y^*)\) is Pareto efficient if and only if \((x^*, y^*)\) solves the following problem for all \( i = 1, 2, \ldots, n \)

\[
\max_x u_i(x_i)
\]
\[
s.t. \quad \sum_{k \in N} x_k = \sum_{k \in N} w_k + y
\]
\[
\quad u_k(x_k) \geq u_k(x_k^*) \quad \text{for} \quad k \neq i
\]
\[
T(y) \leq 0.
\]

If utility functions \( u_i(x_i) \) are differentiable, then we can define the Lagrangian function to get the first order conditions:

\[
L = u_i(x_i) - \lambda T(\hat{x} - \hat{w}) + \sum_{k \neq i} t_k [u_k(x_k) - u_k(x_k^*)]
\]

FOC:
\[
\frac{\partial L}{\partial x_i^l} = \frac{\partial u_i(x_i)}{x_i^l} - \lambda \frac{\partial T(y)}{\partial x_i^l} \leq 0 \quad \text{with equality if} \; x_i^l > 0 \quad l = 1, \ldots, L, i = 1, \ldots, n \quad (11.12)
\]
\[
\frac{\partial L}{\partial x_k^l} = t_k \frac{\partial u_k(x_k)}{x_k^l} - \lambda \frac{\partial T(y)}{\partial x_i^l} \leq 0 \quad \text{with equality if} \; x_k^l > 0 \quad l = 1, \ldots, L, k \neq l \quad (11.13)
\]

When \( x^* \) is an interior solution, we have by (11.12)
\[
\frac{\partial u_i(x_i)}{\partial x_i^l} = \frac{\partial T(y)}{\partial y^l},
\]
\[
\frac{\partial u_k(x_k)}{\partial x_k^l} = \frac{\partial T(y)}{\partial y^l}, \quad (11.14)
\]

426
and by (11.13)
\[
\frac{\partial u_k(x_k)}{\partial x^l_k} = \frac{\partial T(y)}{\partial y^l}.
\]

(11.15)

Thus, we have

\[MRS_{x^l_1, x^h_1} = \cdots = MRS_{x^l_n, x^h_n} = MRTS_{y^l, y^h} \quad l = 1, 2, \ldots, L; h = 1, 2, \ldots, L\]

which are the necessary condition for the interior solutions to be Pareto efficient, which means that the \(MRS\) of any two goods for all agents equals the \(MRTS\). They become sufficient conditions when utility functions \(u_i(x_i)\) are differentiable and quasi-concave and the production functions are concave.

### 11.5 The Second Fundamental Theorem of Welfare Economics

Now we can assert a converse of the First Fundamental Theorem of Welfare Economics. The Second Fundamental Theorem of Welfare Economics gives conditions under which a Pareto optimum allocation can be “supported” by a competitive equilibrium if we allow some redistribution of endowments. It tells us, under some regularity assumptions, including essential condition of convexity of preferences and production sets, that any desired Pareto optimal allocation can be achieved as a market-based equilibrium with transfers.

This is one of the most important theorems in modern economics, and the theorem is also one of the theorems in microeconomic theory whose proof is complicated. It implies that, despite its dysfunctionalities in various environments, one does not need to seek any alternative economic institution to reach Pareto efficient allocations. In order to get an efficient and equitable allocation, the government needs only to redistribute the social wealth through a proper tax-transfer system and leave the other works to the market, then Laissez-faire may produce a Pareto optimal outcome. In this section, we will give a mathemat-
ically rigorous representation and proof of the second fundamental theorem of welfare economics. Combining the first and second fundamental theorems, we could find that when certain regular conditions are satisfied, the set of Pareto optimal allocations coincides with the set of competitive equilibrium with transfer payments.

**Theorem 11.5.1 (The Second Fundamental Theorem of Welfare Economics I)**

Suppose \((x^*, y^*)\) with \(x^* > 0\) is Pareto optimal, suppose \(\succ_i\) are continuous, convex and strongly monotonic, and suppose that \(Y_j\) are closed and convex. Then, there is a price vector \(p \geq 0\) such that \((x^*, y^*, p)\) is a competitive equilibrium with transfer payments, i.e.,

1. if \(x'_i \succ_i x^*_i\), then \(p \cdot x'_i > p \cdot x^*_i\) for \(i = 1, \ldots, n\).
2. \(p \cdot y'_j \geq p \cdot y_j\) for all \(y'_j \in Y_j\) and \(j = 1, \ldots, J\).

**Proof.** Let

\[
P(x^*_i) = \{x_i \in X_i : x_i \succ_i x^*_i\}
\]  

(11.17)

be the strict upper contour set and let

\[
P(x^*) = \sum_{i=1}^{n} P(x^*_i).
\]

(11.18)

By the convexity of \(\succ_i\), we know that \(P(x^*_i)\) and thus \(P(x^*)\) are convex.
Figure 11.7: $P(x_i^*)$ is the set of all points strictly above the indifference curve through $x_i^*$.  

Let $W = \{\hat{w}\} + \sum_{j=1}^{J} Y_j$ which is closed and convex. Then $W \cap P(x^*) = \emptyset$ by Pareto optimality of $(x^*, y^*)$, and thus, by the Separating Hyperplane Theorem in Chapter 2, there is a $p \neq 0$ such that

$$p\hat{z} \geq p\hat{\sigma} \text{ for all } \hat{z} \in P(x^*) \text{ and } \hat{\sigma} \in W$$

(11.19)

Now we show that $p$ is a competitive equilibrium price vector by the following four steps.

1. $p \geq 0$
To see this, let $e^l = (0, \ldots, 1, 0, \ldots, 0)$ with the $l$-th component one and other places zero. Let

$$\hat{z} = \hat{\sigma} + e^l \text{ for some } \hat{\sigma} \in W.$$ 

Then $\hat{z} = \hat{\sigma} + e^l \in P(x^*)$ by strong monotonicity and redistribution of $e^l$. Thus, we have by (11.19).

$$p(\hat{\sigma} + e^l) \geq p\hat{\sigma} \quad (11.20)$$

and thus

$$pe^l \geq 0 \quad (11.21)$$

which means

$$p^l \geq 0 \quad \text{for } l = 1, 2, \ldots, L. \quad (11.22)$$

2. $py_j^* \geq py_j$ for all $y_j \in Y_j$, $j = 1, \ldots, J$.

Since $\hat{x}^* = \hat{y}^* + \hat{w}$ by noting $(x^*, y^*)$ is a Pareto efficient allocation and preference orderings are strongly monotonic, we have $p\hat{x}^* = p(\hat{w} + \hat{y}^*)$.

Thus, by $p\hat{z} \geq p(\hat{w} + \hat{y})$ in (11.19) and $p\hat{w} = p\hat{x}^* - p\hat{y}^*$, we have

$$p(\hat{z} - \hat{x}^*) \geq p(\hat{y} - \hat{y}^*) \quad \forall \hat{y} \in \hat{Y}.$$ 

Letting $\hat{z} \rightarrow x^*$, we have

$$p(\hat{y} - \hat{y}^*) \leq 0 \quad \forall \hat{y} \in \hat{Y}.$$ 

Letting $y_k = y_k^*$ for $k \neq j$, we have from the above equation,

$$p \cdot y_j^* \leq p \cdot y_j \quad \forall y_j \in Y_j.$$ 

3. If $x_i \succ_i x_i^*$, then

$$p \cdot x_i \geq p \cdot x_i^*. \quad (11.23)$$

To see this, let

$$x_i' = (1 - \theta)x_i \quad 0 < \theta < 1$$

$$x_k' = x_k^* + \frac{\theta}{n-1}x_i \quad \text{for } k \neq i$$
Then, by the continuity of $\succeq_i$ and the strong monotonicity of $\succeq_k$, we have $x'_i \succ_i x_i^*$ for all $i \in N$, and thus

$$x' \in P(x^*) \quad (11.24)$$

if $\theta$ is sufficiently small. By (11.19), we have

$$p \cdot (x'_i + \sum_{k \neq i} x'_k) = p \cdot [(1 - \theta)x_i + \sum_{k \neq i} (x_k^* + \frac{\theta}{n-1}x_i)] \geq p \cdot \sum_{k=1}^n x_k^* \quad (11.25)$$

and thus we have

$$p \cdot x_i \geq p \cdot x_i^* \quad (11.26)$$

4. If $x_i \succ_i x_i^*$, we must have $p \cdot x_i > p \cdot x_i^*$. To show this, suppose by way of contradiction, that

$$p \cdot x_i = p \cdot x_i^* \quad (11.27)$$

Since $x_i \succ_i x_i^*$, then $\lambda x_i \succ_i x_i^*$ for $\lambda$ sufficiently close to one by the continuity of preferences for $0 < \lambda < 1$. By step 3, we know $\lambda p \cdot x_i \geq p \cdot x_i^* = px_i$ so that $\lambda \geq 1$ by $p \cdot x_i = p \cdot x_i^* > 0$, which contradicts the fact that $\lambda < 1$.

The strong monotonicity condition of preferences can be weakened to locally non-satiation. But the nonnegative price (which fits goods resulting in negative utilities such as labor, economic bads) cannot be guaranteed and the Pareto optimal allocations must satisfy $\sum_i x_i = \sum_i w_i + \sum_j y_j$.

**Definition 11.5.1 (The Second Fundamental Theorem of Welfare Economics II)**

Given an economy $e = (e_1, \ldots, e_n, \{Y_j\})$, suppose $\succeq_i$ is continuous, convex, and local non-satiated, $Y_j$ is convex and close. Then for any Pareto optimal allocation $(x^*, y^*)$, there exists $p \neq 0$ such that $(x^*, y^*, p)$ is a quai-equilibrium with transfer payments. That is, there exist a set of wealth $(I_1, \ldots, I_n)$ with $\sum_i I_i = p \cdot \sum_i w_i + \sum_j p \cdot y_j^*$ such that

1. if $x'_i \succ_i x_i^*$, then $px'_i \geq px_i^*$, $i = 1, \ldots, n$;

2. for an arbitrary $y'_j \in Y_j$, $j = 1, \ldots, J$, we have $py'_j \geq py_j^*$. 

431
Furthermaore, if \( \forall i, 0 \in X_i \) and \( px_i^* > I_i \), then \( (x^*, y^*, p) \) is a CE with transfer payments.

**Proof.** Analogous to the proof of above theorems, we define the upper section set \( P(x^*) \) and set of feasible allocations \( W \), then applying the separating theorem of hyperplane, we get that there exists \( p \neq 0 \) such that

\[
 p\hat{z} \geq p\hat{\sigma}, \quad \forall \hat{z} \in P(x^*), \forall \hat{\sigma} \in W. \tag{11.28}
\]

The next prove \( p \) is a quasi-equilibrium with transfer payments in three steps.

1. for all \( i \), if \( x_i \succ x_i^* \), then \( \sum_i px_i \geq \sum_j px_j^* \).

   For all \( i \) and \( x_i \succ x_i^* \), local non-satiation implies there exists a \( x_i' \) close enough to \( x_i \), such that

   \[ x_i' \succ_i x_i \succ_i x_i^*. \]

   Therefore, \( \sum_i x_i' \in P(x^*) \). Meanwhile,

   \[
   \sum_i x_i^* = \sum_i w_i + \sum_j y_j^* \in W.
   \]

   Then, from (11.28), we have \( \sum_i px_i' \geq \sum_j px_j^* \). Let \( x_i' \to x_i \), we have \( \sum_i px_i \geq \sum_j px_j^* \).

2. For all \( y_j \in Y_j, j = 1, \ldots, J \), we have \( py_j^* \geq py_j \).

   Notice that the Pareto optimal allocation is attainable, that is \( \hat{x}^* = \hat{y}^* + \hat{w} \).

   Therefore, \( \sum_i px_i^* = \sum_i pw_i + \sum_j py_j^* \). The remaining analysis is similar to the proof of the above theorem.

3. for any \( i \), if \( x_i \succ_i x_i^* \), then \( px_i \geq px_i^* \). Let

   \[
   x_i' = x_i, \\
   x_k' = x_k^*, \quad k \neq i.
   \]

   then it follows from step 1 that \( px_i + \sum_{k \neq i} px_k^* \geq \sum_j px_j^* \). Consequently,

   \[ px_i \geq px_i^*. \]
Therefore, \((x^*, y^*, p)\) is a quasi-equilibrium with transfers are satisfied.

Finally, from proposition 11.3.2, we know if for all \(i\), 0 \(\in\) \(X_i\) and \(px_i^* > I_i\), then quasi-equilibrium with transfers \((x^*, y^*, p)\) is also equilibrium with transfers, the proof is finished. ■

The Second Fundamental Theorem of Welfare Economics establishes that under some regular conditions, the market mechanism, modified by the addition of lump-sum transfers, can achieve virtually any desired optimal distribution.

A main attack of market institution is that it ignores distribution. Laissez-faire may produce a Pareto optimal outcome, but an intolerable degree of inequality may also arise. There are many different Pareto optima, and some are fairer than others. Some people are endowed with resources that make them extremely rich, while others, through no fault of their own, are extremely poor. The objections of market-oriented reforms in China, the former Soviet Union, and other eastern European countries came mainly from this concern.

One approach to rectifying the distributional inequities of laissez-faire is the command economy: a central bureaucracy makes detailed decisions about the consumption decisions of all individuals and production decisions of all producers. The main theoretical problem with the command approach is that it fails to create appropriate incentives for individuals and firms. On the empirical side, the experiences of the late Soviet and Maoist command economies are proved to be unsuccessful and even disastrous.

The second approach to solve distribution problems is originated from the second fundamental theorem: to transfer income or purchasing power among individuals, and then to let the market work. The lump-sum money transfers will not cause incentive-related losses. This approach allows a limited degree of governmental intervention and admits the main role of market. The evidences from transitional economies show that this approach is by far the most successful remedy for distribution problems.

For exchange economies, the competitive equilibrium with transfers is the same as a regular competitive equilibrium when \(w_i = x_i^*\). Then we have the
following corollary.

**Corollary 11.5.1** Suppose $x^*>0$ is Pareto optimal, suppose $\succ_i$ are continuous, convex and strongly monotonic. Then, $x^*$ is a competitive equilibrium for the initial endowment $w_i = x_i^*$.

**Remark 11.5.1** If $\succ_i$ can be represented by a concave and differentiable utility function, then the proof of the Second Fundamental Theorem of Welfare Economics can be much simpler. A sufficient condition for concavity is that the Hessian matrix is negative definite. Also note that monotonic transformation does not change preferences so that we may be able to transform a quasi-concave utility function to a concave utility function as follows, for example

$$u(x, y) = xy$$

which is concave after monotonic transformation.

**Differentiation Version of the Second Fundamental Theorem of Welfare Economics for Exchange Economies**

Proof: If $x^*>0$ is Pareto Optimal, then we have

$$Du_i(x_i) = q \quad i = 1, 2, \ldots, n. \quad (11.29)$$

We want to show $q$ is a competitive equilibrium price vector. To do so, we only need to show that each consumer maximizes his utility s.t. $B(q) = \{x_i \in X_i : qx_i \leq qx_i^*\}$. Indeed, by concavity of $u_i$

$$u_i(x_i) \leq u_i(x_i^*) + Du_i(x_i^*) (x_i - x_i^*)$$

$$\quad = u_i(x_i^*) + q(x_i - x_i^*)/t_i$$

$$\leq u_i(x_i^*)$$

The reason the inequality holds for a concave function is because that, from Figure 11.8, we have

$$\frac{u(x) - u(x^*)}{x - x^*} \leq u'(x^*). \quad (11.30)$$
Thus, we have \( u_i(x_i) \leq u(x_i^*) + Du_i(x_i^*) (x_i - x_i^*) \).

### 11.6 Non-Convex Production Technologies and Marginal Cost Pricing

The indispensability of convexity for the second welfare theorem can be observed in Figure 11.9(a). There, the allocation \( x^* \) maximizes the welfare of the
Figure 11.9: Figure (a) shows failure of the second welfare theorem with a non-convex technology. Figure (b) shows the first welfare theorem applies even with a non-convex technology.

consumer, but for the only value of relative prices that could support $x^*$ as a utility-maximizing bundle, the firm does not maximize profits even locally (i.e., at the relative prices $w/p$, there are productions arbitrarily close to $x^*$ yielding higher profits). In contrast, the first welfare theorem remains applicable even in the presence of non-convexities. As Figure 11.9(b) suggests, any Walrasian equilibrium maximizes the well-being of the consumer in the feasible production set.

The second welfare theorem runs into difficulties in the presence of non-convex production sets (here we do not question the assumption of convexity on the consumption side). In the first place, large non-convexities caused by the presence of fixed costs or extensive increasing returns lead to a world of a small number of large firms, making the assumption of price taking less plausible. Yet, even if price taking can somehow be relied on, it may still be impossible to support a given Pareto optimal allocation. Examples are provided by Figures 11.9(a) and 11.10. In Figure 11.10, at the only relative prices that could support the production $y^*$ locally, the firm sustains a loss and would rather avoid it by shutting down. In Figure 11.9(a), on the other hand, not even local profit maximization can be guaranteed.

For the industries exhibiting increasing return to scale and requiring innovations, what measures should be taken to provide incentives to innovation as well as obtain efficiency? Laissez-faire is not applicable in this case, the government or a third party need to price the goods. Then how to set the price to optimize the allocation of resources? In the presence of convexity or increasing return to scale, profit-maximizing rule is not feasible any more, one should adopt some alternative pricing methods, such as average cost pricing, marginal cost pricing, loss-free pricing, etc.
Analogous to CE and CE with transfer payments, we now introduce marginal cost pricing equilibrium and marginal cost pricing equilibrium with transfers.

Figure 11.10: The firm incurs a loss at the prices that locally support be Pareto optimal allocation.

**Definition 11.6.1 (Marginal Cost Pricing Equilibrium)** For private ownership economy $e = (e_1, \ldots, e_n, \{Y_j\}, \{\theta_{ij}\})$, an allocation $(x, y) \in X \times Y$ and price vector $p \in \mathbb{R}_+^L$ constitute an equilibrium of marginal cost pricing if the following conditions are satisfied:

(i) for $\forall j$,
$$p = \gamma_j \nabla T_j(y),$$

(ii) for $\forall i$, $x_i$ maximize preference $\succ_i$ within budget set
$$\{x_i' \in X_i : p \cdot x_i' \leq p \cdot w_i\};$$

(iii) $\sum_i x_i \leq \sum_i w_i + \sum_j y_j^*$

The motivation for this concept comes from the one-output, one-input case, it is easy to find that the price in this equilibrium equals to marginal cost.

**Definition 11.6.2 (Marginal Cost Pricing Equilibrium with Transfers)**

For an economy $e = (e_1, \ldots, e_n, \{Y_j\})$, an allocation $(x, y) \in X \times Y$ and price vector $p \in \mathbb{R}_+^L$ constitute a marginal cost pricing equilibrium with transfers if there exists a sequence of redistributed wealth levels $(I_1, \ldots, I_n)$ such that $
\sum_i I_i = p \cdot \sum_i w_i + \sum_j p \cdot y_j^*$ and the following conditions are satisfied:
(i) for $\forall j$, we have
\[ p = \gamma_j \nabla T_j(y), \exists \gamma_j > 0; \]

(ii) for $\forall i$, $x_i$ maximizes $\succ_i$ with the budget set
\[ \{x'_i \in X_i : p \cdot x'_i \leq I_i\}; \]

(iii) $\sum x_i \leq \sum w_i + \sum y_j$.  

For marginal cost pricing under non-convexity, the first fundamental theorem of welfare economics does not hold. It hold only in very restricted class of economic environments (see Quinzii (1992)). Although non-convexities may prevent us from supporting the Pareto optimal production allocation as a profit-maximizing choice, under the differentiability assumptions on pricing rules of firms, we can use the first-order necessary conditions derived there to formulate a weaker result that parallels the second welfare theorem.

**Theorem 11.6.1 (The Second Welfare Theorem for Non-Convex Production Sets)**

*Suppose that firm’s $j$ transformation function $T_j(y)$ are smooth, and all consumers have continuous, monotonic, and convex preferences. If $(x^*, y^*)$ is Pareto optimal, then there exists a price vector $p$ and wealth levels $(I_1, \ldots, I_n)$ with $\sum I_i = p \cdot \sum w_i + \sum p \cdot y^*_j$ such that:

(i) For any firm $j$, we have
\[ p = \gamma_j \nabla T_j(y^*) \text{ for some } \gamma_j > 0. \]

(ii) For any $i$, $x^*_i$ is maximal for $\succ_i$ in the budget set
\[ \{x_i \in X'_i : p \cdot x_i \leq I_i\}. \]

(iii) $\sum x^*_i = \sum w_i + \sum y^*_j$.  


Condition (i) also implies that

\[
\frac{p^j}{p^k} = \frac{\partial T^j(y)}{\partial y^j} = \cdots = \frac{\partial T^j(y)}{\partial y^k},
\tag{11.31}
\]

and thus marginal rates of technic substitutions, \(MRTS\), of two goods are equal for all firms. As we have noted, although condition (i) is necessary, but not sufficient, for profit maximization, and in fact, it does not imply that the \((y^*_1, \ldots, y^*_J)\) are profit-maximizing production plans for price-taking firms. The condition says only that small changes in production plans have no first-order effect on profit. But small changes may still have positive second-order effects (as in Figure 11.9(a), where at a marginal cost price equilibrium the firm actually chooses the production that minimizes profits among the efficient productions) and, at any rate, large changes may increase profits (as in Figure 11.10). Thus, to achieve allocation \((x^*, y^*)\) may require that a regulatory agency prevent the managers of non-convex firms from attempting to maximize profits at the given prices \(p\). Recently, Tian (2010) presented such an incentive compatible mechanism that implements marginal cost price equilibrium allocations with transfers in Nash equilibrium. Tian (2009) also consider implementation of other equilibrium allocations when firms pricing rule is given by such as average-cost pricing, loss-free pricing, etc.

It should be noted that the converse result to the above theorem, which would assert that every marginal cost price equilibrium is Pareto optimal, is not true. In Figure 11.11, for example, we show a one-consumer economy with a non-convex production set. In the figure, \(x^*\) is a marginal cost price equilibrium with transfers for the price system \(p = (1, 1)\). Yet, allocation \(x'\) yields the consumer a higher utility. Informally, this occurs because marginal cost pricing neglects second-order conditions and it may therefore happen that, as at allocation \(x^*\), the second-order conditions for the social utility maximization problem are not. A marginal cost pricing equilibrium need not be Pareto optimal allocation. As a result, satisfaction of the first-order marginal optimality conditions (which in the case of Figure 11.11 amounts simply to the tangency of the indifference curve and
the production surface) does not ensure that the allocation is Pareto optimal.

Figure 11.11: The firm incurs a loss at the prices that locally support be Pareto optimal allocation.

See Quinzii (1992) for extensive background and discussion on the material presented in this section.

11.7 Pareto Optimality and Social Welfare Maximization

Pareto efficiency is only concerned with efficiency of allocations and has nothing to say about distribution of welfare. Even if we agree with Pareto optimality, we still do not know which one we should be at. One way to solve the problem is to assume the existence of a social welfare function.

**Definition 11.7.1 (Social Welfare Function)** A social welfare function $W : X \rightarrow \mathbb{R}$ is defined by $W(u_1(x_1), \ldots, u_n(x_n))$, where we assume that $W(\cdot)$ is increasing.
The idea behind a social welfare function $W(u_1, \ldots, u_n)$ is that it accurately expresses society’s judgments on how individual utilities have to be compared to produce an ordering of possible social outcomes. Two common examples of $W$ are the utilitarian social welfare function and the Rawlsian social welfare function defined respectively as follows.

**Definition 11.7.2 (Utilitarian Social Welfare Function)** The utilitarian social welfare function is given by

$$W(u_1, \ldots, u_n) = \sum_{i=1}^{n} a_i u_i(x_i)$$

with $\sum a_i = 1, a_i \geq 0$. Under a utilitarian rule, social states are ranked according to the linear sum of utilities. The utilitarian form is by far the most common and widely applied social welfare function in economics.

**Definition 11.7.3 (Rawlsian Social Welfare Function)** The Rawlsian social welfare function is defined by

$$W(\cdot) = \min\{u_1(x_1), u_2(x_2), \ldots, u_n(x_n)\}.$$  

So the utility function is not strongly monotonic increasing. The Rawlsian form gives priority to the interests of the worst off members, and it is used in the ethical system proposed by Rawls (1971).

**11.7.1 Social Welfare Maximization for Exchange Economies**

We suppose that a society should operate at a point that maximizes social welfare; that is, we should choose an allocation $x^*$ such that $x^*$ solves

$$\max W(u_1(x_1), \ldots, u_n(x_n))$$

subject to

$$\sum_{i=1}^{n} x_i \leq \sum_{i=1}^{n} w_i.$$
How do the allocations that maximize this welfare function compare to Pareto efficient allocations? The following is a trivial consequence if the strong monotonicity assumption is imposed.

**Proposition 11.7.1** Under strong monotonicity of the social welfare function, $W(\cdot)$ if $x^*$ maximizes a social welfare function, then $x^*$ must be Pareto Optimal.

**Proof.** If $x^*$ is not Pareto Optimal, then there is another feasible allocation $x'$ such that $u_i(x'_i) \geq u_i(x_i)$ for all $i$ and $u_k(x'_k) > u_k(x_k)$ for some $k$. Then, by strong monotonicity of $W(\cdot)$, we have $W(u_1(x'_1), \ldots, u_n(x'_n)) > W(u_1(x_1), \ldots, u_n(x_n))$ and thus it does not maximizes the social welfare function. ■

Thus, every social welfare maximum is Pareto efficient. Is the converse necessarily true? By the Second Fundamental Theorem of Welfare Economics, we know that every Pareto efficient allocation is a competitive equilibrium allocation by redistributing endowments. This gives us a further implication of competitive prices, which are the multipliers for the welfare maximization. Thus, the competitive prices really measure the (marginal) social value of a good. Now we state the following proposition that shows every Pareto efficient allocation is a social welfare maximum for the social welfare function with a suitable weighted sum of utilities.

**Proposition 11.7.2** Let $x^* > 0$ be a Pareto optimal allocation. Suppose $u_i$ is concave, differentiable and strongly monotonic. Then, there exists some choice of weights $a^*_i$ such that $x^*$ maximizes the welfare functions

$$W(u_1, \ldots, u_n) = \sum_{i=1}^{n} a_i u_i(x_i) \quad (11.32)$$

Furthermore, $a^*_i = \frac{1}{\lambda_i}$ with $\lambda_i = \frac{\partial V_i(p,I_i)}{\partial I_i}$

where $V_i(\cdot)$ is the indirect utility function of consumer $i$.

**Proof:** Since $x^*$ is Pareto optimal, it is a competitive equilibrium allocation with $w_i = x_i^*$ by the second theorem of welfare economies. So we have

$$D_i u_i(x_i^*) = \lambda p \quad (11.33)$$
by the first order condition, where \( \mathbf{p} \) is a competitive equilibrium price vector.

Now for the welfare maximization problem

\[
\max \sum_{i=1}^{n} a_i u_i(x_i)
\]

\[
s.t. \sum x_i \leq \sum x_i^*
\]

since \( u_i \) is concave, \( x^* \) solves the problem if the first order condition

\[
a_i \frac{\partial u_i(x_i)}{\partial x_i} = q \quad i = 1, \ldots, n
\]

is satisfied for some \( q \). Thus, if we let \( p = q \), then \( a_i^* = \frac{1}{x_i} \). We know \( x^* \) also maximizes the welfare function \( \sum_{i=1}^{n} a_i^* u_i(x_i^*) \).

Thus, the price is the Lagrangian multiples of the welfare maximization, and this measures the marginal social value of goods.

### 11.7.2 Welfare Maximization in Production Economy

Define a choice set by the transformation function

\[
T(\hat{x}) = 0 \quad \text{with} \quad \hat{x} = \sum_{i=1}^{n} x_i.
\]

The social welfare maximization problem for the production economy is

\[
\max W(u_1(x_1), u_2(x_2), \ldots, u_n(u_n))
\]
subject to

\[ T(\hat{x}) = 0. \]

Define the Lagrangian function

\[ L = W(u_1(x_1), \ldots, u_n(u_n)) - \lambda T(\hat{x}). \quad (11.37) \]

The first order condition is then given by

\[ W'(-) \frac{\partial u_i(x_i)}{\partial x_i^j} - \lambda \frac{\partial T(\hat{x})}{\partial x_i^j} \leq 0 \quad \text{with equality if } x_i^j > 0, \quad (11.38) \]

and thus when \( x \) is an interior point, we have

\[ \frac{\partial u_i(x_i)}{\partial x_i^j} = \frac{\partial T(\hat{x})}{\partial x_i^j} \quad (11.39) \]

That is,

\[ MRS_{x_i^j, x_i^k} = MRTS_{x_i^j, x_i^k}. \quad (11.40) \]

The conditions characterizing welfare maximization require that the marginal rate of substitution between each pair of commodities must be equal to the marginal rate of transformation between the two commodities for all agents.

### 11.8 Political Overtones

1. By the First Fundamental Theorem of Welfare Economics, implication is that what the government should do is to secure the competitive environment in an economy and give people full economic freedom. So, as long as the market system works well, there should be no subsidizing, no price floor, no price ceiling, stop a rent control, no regulations, lift the tax and the import-export barriers.

2. Even if we want to reach a preferred Pareto optimal outcome which may be different from a competitive equilibrium from a given endowment, you might do so by adjusting the initial endowment but not disturbing prices, imposing...
taxes or regulations. That is, if a derived Pareto optimal is not “fair”, all the
government has to do is to make a lump-sum transfer payments to the poor first,
keeping the competitive environments intact. We can adjust initial endowments
to obtain a desired competitive equilibrium by the Second Fundamental Theorem
of Welfare Economics.

3. Of course, when we reach the above conclusions, you should note that there
are conditions on the results. In many cases, we have market failures, in the sense
that either a competitive equilibrium does not exist or a competitive equilibrium
may not be Pareto optimal so that the First or Second Fundamental Theorem of
Welfare Economics cannot be applied.

The conditions for the existence of a competitive equilibrium are: (i) convex-
ity (diversification of consumption and no IRS), (ii) monotonicity (self-interest),
and (iii) continuity, (iv) divisibility, (v) perfect competition, (vi) complete infor-
modation, etc. If these conditions are not satisfied, we may not obtain the existence
of a competitive equilibrium. The conditions for the First Fundamental Theo-
rem of Welfare Economics are: (i) local non-satiation (unlimited desirability),
(ii) divisibility, (iii) no externalities, (iv) perfect competition, (v) complete in-
formation etc. If these conditions are not satisfied, we may not guarantee that
every competitive equilibrium allocation is Pareto efficient. The conditions for
the Second Fundamental Theorem of Welfare Economics are: (i) the convexity
of preferences and production sets, (ii) monotonicity (self-interest), and (iii) con-
tinuity, (iv) divisibility, (v) perfect competition, (vi) complete information, etc.
If these conditions are not satisfied, we may not guarantee every Pareto efficient
allocation can be supported by a competitive equilibrium with transfers.

Thus, as a general notice, before making an economic statement, one should
pay attention to the assumptions which are implicit and/or explicit involved. As
for a conclusion from the general equilibrium theory, one should notice conditions
such as divisibility, no externalities, no increasing returns to scale, perfect com-
petition, complete information. If these assumptions are relaxed, a competitive
equilibrium may not exist or may not be Pareto efficient, or a Pareto efficient
allocation may not be supported by a competitive equilibrium with transfer payments. Only in this case of a market failure, we may adopt another economic institution. We will discuss the market failure and how to solve the market failure problem in Part 4 and Part 5.

11.9 Reference

Books and Monographs:


Papers:


Chapter 12
Economic Core, Fair Allocations, and Social Choice Theory

12.1 Introduction

In this chapter we briefly discuss some topics in the framework of general equilibrium theory, namely economic core, fair allocations, and social choice theory. The theory of core is important because it gives an insight into how a competitive equilibrium is achieved as a result of individual strategic behavior instead of results of an auctioneer and the Walrasian tâtonnement mechanism. It shows the necessity of adopting a market institution as long as individuals behavior self-interestedly.

We have also seen that Pareto optimality may be too weak a criterion to be meaningful. It does not address any question about income distribution and equity of allocations. Fairness is a notion to overcome this difficulty. This is one way to restrict a set of Pareto optimum.

In a slightly different framework, suppose that a society is deciding the social priority among finite alternatives. Alternatives may be different from Pareto optimal allocations. Let us think of a social “rule” to construct the social ordering (social welfare function) from many individual orderings of different alternatives. The question is: Is it possible to construct a rule satisfying several desirable
both “fairness” and “social welfare function” address a question of social justice.

12.2 The Core of Exchange Economies

The use of a competitive (market) system is just one way to allocate resources. What if we use some other social institution? Would we still end up with an allocation that was “close” to a competitive equilibrium allocation? The answer will be that, if we allow agents to form coalitions, the resulting allocation can only be a competitive equilibrium allocation when the economy becomes large. Such an allocation is called a core allocation and was originally considered by Edgeworth (1881).

The core is a concept in which every individual and every group agree to accept an allocation instead of moving away from the social coalition.

There is some reason to think that the core is a meaningful political concept. If a group of people find themselves able, using their own resources to achieve a better life, it is not unreasonable to suppose that they will try to enforce this threat against the rest of community. They may find themselves frustrated if the rest of the community resorts to violence or force to prevent them from withdrawing.

The theory of the core is distinguished by its parsimony. Its conceptual apparatus does not appeal to any specific trading mechanism nor does it assume any particular institutional setup. Informally, the notion of competition that the theory explores is one in which traders are well informed of the economic characteristics of other traders, and in which the members of any group of traders can bind themselves to any mutually advantageous agreement.

For simplicity, we consider exchange economies. We say two agents are of the same type if they have the same preferences and endowments.

The $r$-replication of the original economy: There are $r$ times as many agents of each type in the original economy.
A coalition is a group of agents, and thus it is a subset of $n$ agents.

**Definition 12.2.1 (Blocking Coalition)** A group of agents $S$ (a coalition) is said to block (improve upon) a given allocation $x$ if there is a Pareto improvement with their own resources, that is, if there is some allocation $x'$ such that

1. it is feasible for $S$, i.e., $\sum_{i \in S} x'_i \leq \sum_{i \in S} w_i$,
2. $x'_i \succ_i x_i$ for all $i \in S$ and $x'_k \succ_k x_k$ for some $k \in S$.

**Definition 12.2.2 (Core)** A feasible allocation $x$ is said to have the core property if it cannot be improved upon for any coalition. The core of an economy is the set of all allocations that have the core property.

**Remark 12.2.1** Every allocation in the core is Pareto optimal (coalition by whole people).

**Definition 12.2.3 (Individual Rationality)** An allocation $x$ is individually rational if $x_i \succ_i w_i$ for all $i = 1, 2, \ldots, n$.

The individual rationality condition is also called the participation condition which means that a person will not participate the economic activity if he is worse off than at the initial endowment.

**Remark 12.2.2** Every allocation in the core must be individually rational.

**Remark 12.2.3** When $n = 2$ and preference relations are weakly monotonic, an allocation is in the core if and only if it is Pareto optimal and individually rational.

**Remark 12.2.4** Many textbooks and papers use the weak Pareto improvement to define a blocking coalition and consequently the core property. An allocation $x$ is said to have a weak Pareto improvement if it is feasible and there is no other feasible allocation $x'$ such that $x'_i \succ_i x_i$ for all $i$. If so, we then need to impose continuity and strong monotonicity to get most results in this section such as Pareto optimality of an allocation in the core, the results on Equal Treatment in the Core, Shrinking Core Theorem.
Figure 12.1: The set of allocations in the core is simply given by the set of Pareto efficient and individually rational allocations when $n = 2$.

Remark 12.2.5 Even though a Pareto optimal allocation is independent of individual endowments, an allocation in the core depends on individual endowments.

What is the relationship between core allocations and competitive equilibrium allocations?

Theorem 12.2.1 Under local non-satiation, if $(x, p)$ is a competitive equilibrium, then $x$ has the core property.

Proof: Suppose $x$ is not an allocation in the core. Then there is a coalition $S$ and a feasible allocation $x'$ such that

$$\sum_{i \in S} x'_i \leq \sum_{i \in S} w_i \quad (12.1)$$
and $x'_i \succ_i x_i$ for all $i \in S$, $x'_k \succ_k x_k$ for some $k \in S$. Then, by local non-satiation, we have

$$px'_i \geq px_i \text{ for all } i \in S \text{ and }$$

$$px'_k > px_k \text{ for some } k$$

Therefore, we have

$$\sum_{i \in S} px'_i > \sum_{i \in S} px_i = \sum_{i \in S} pw_i \quad (12.2)$$

a contradiction. Therefore, the competitive equilibrium must be an allocation in the core.

We refer to the allocations in which consumers of the same type get the same consumption bundles as equal-treatment allocations. It can be shown that any allocation in the core must be an equal-treatment allocation.

**Proposition 12.2.1 (Equal Treatment in the Core)** Suppose agents’ preferences are strictly convex. Then if $\mathbf{x}$ is an allocation in the r-core of a given economy, then any two agents of the same type must receive the same bundle.

Proof: Let $\mathbf{x}$ be an allocation in the core and index the $2r$ agents using subscripts $A_1, \ldots, Ar$ and $B_1, \ldots, Br$. If all agents of the same type do not get the same allocation, there will be one agent of each type who is most poorly treated. We will call these two agents the “type-A underdog” and the “type-B underdog.” If there are ties, select any of the tied agents.

Let $\bar{x}_A = \frac{1}{r} \sum_{j=1}^{r} x_{A_j}$ and $\bar{x}_B = \frac{1}{r} \sum_{j=1}^{r} x_{B_j}$ be the average bundle of the type-A and type-B agents. Since the allocation $\mathbf{x}$ is feasible, we have

$$\frac{1}{r} \sum_{j=1}^{r} x_{A_j} + \frac{1}{r} \sum_{j=1}^{r} x_{B_j} = \frac{1}{r} \sum_{j=1}^{r} \omega_{A_j} + \frac{1}{r} \sum_{j=1}^{r} \omega_{B_j} = \frac{1}{r} r \omega_A + \frac{1}{r} r \omega_B.$$

It follows that

$$\bar{x}_A + \bar{x}_B = \omega_A + \omega_B,$$

so that $(\bar{x}_A, \bar{x}_B)$ is feasible for the coalition consisting of the two underdogs. We are assuming that at least for one type, say type A, two of the type-A agents
receive different bundles. Hence, the A underdog will strictly prefer $\bar{x}_A$ to his present allocation by strict convexity of preferences (since it is a weighted average of bundles that are at least as good as $x_A$), and the B underdog will think $\bar{x}_B$ is at least as good as his present bundle, thus forming a coalition that can improve upon the allocation. The proof is completed.

Since any allocation in the core must award agents of the same type with the same bundle, we can examine the cores of replicated two-agent economies by use of the Edgeworth box diagram. Instead of a point $x$ in the core representing how much A gets and how much B gets, we think of $x$ as telling us how much each agent of type A gets and how much each agent of type B gets. The above proposition tells us that all points in the $r$-core can be represented in this manner.

The following theorem is a converse of Theorem 12.2.1 and shows that any allocation that is not a market equilibrium allocation must eventually not be in the $r$-core of the economy. This means that core allocations in large economies look just like Walrasian equilibria.

**Theorem 12.2.2 (Shrinking Core Theorem)** Suppose $\succ_i$ are strictly convex and continuous. Suppose $x^*$ is a unique competitive equilibrium allocation. Then, if $y$ is not a competitive equilibrium, there is some replication $V$ such that $y$ is not in the $V$-core.

**Proof:** We prove the theorem with the aid of Figure 12.2. From the figure, one can see that $y$ is not a competitive equilibrium. We want to show that there is a coalition such that the point $y$ can be improved upon for $V$-replication. Since $y$ is not a competitive equilibrium, the line segment through $y$ and $w$ must cut at least one agent’s, say agent A’s, indifference curve through $y$. Then, by strict convexity and continuity of $\succ_i$, there are integers $V$ and $T$ with $0 < T < V$ such that

$$g_A = \frac{T}{V} w_A + (1 - \frac{T}{V}) y_A \succ_A y_A.$$
We can do so since any real number can be approached by a rational number that consists of the ratio of two suitable integers \( T \).

![Figure 12.2: The shrinking core. A point like \( y \) will eventually not be in the core.](image)

Now, form a coalition consisting of \( V \) consumers of type A and \( V - T \) consumers of type B and consider the allocation \( x = (g_A, \ldots, g_A, y_B, \ldots, y_B) \) (in which there are \( V \) Type A and \( V - T \) Type B). We want to show \( x \) is feasible for this coalition.

\[
V g_A + (V - T)y_B = V \left[ \frac{T}{V} w_A + \left( 1 - \frac{T}{V} y_A \right) \right] + (V - T)y_B \\
= Tw_A + (V - T)y_A + (V - T)y_B \\
= Tw_A + (V - T)(y_A + y_B) \\
= Tw_A + (V - T)(w_A + w_B) \\
= V w_A + (V - T)w_B
\]

by noting \( y_A + y_B = w_A + w_B \). Thus, \( x \) is feasible in the coalition and \( g_A >_A y_A \) for all agents in type A and \( y_B \sim_B y_B \) for all agents in type B which means \( y \) is not in the \( V \)-core for the \( V \)-replication of the economy. The proof is completed.
Remark 12.2.6 The shrinking core theorem then shows that the only allocations that are in the core of a large economy are market equilibrium allocations, and thus Walrasian equilibria are robust: even very weak equilibrium concepts, like that of core, tend to yield allocations that are close to Walrasian equilibria for larger economies. Thus, this theorem shows the essential importance of competition and fully economic freedom.

Remark 12.2.7 Many of the restrictive assumptions in this proposition can be relaxed such as strong monotonicity, convexity, uniqueness of competitive equilibrium, and two types of agents.

From the above discussion, we have the following limit theorem.

Theorem 12.2.3 (Limit Theorem on the Core) Under the strict convexity and continuity, the core of a replicated two person economy shrinks when the number of agents for each type increases, and the core coincides with the competitive equilibrium allocation if the number of agents goes to infinity.

This result means that any allocation which is not a competitive equilibrium allocation is not in the core for some $r$-replication.

12.3 Fairness of Allocation

Pareto efficiency gives a criterion of how the goods are allocated efficiently, but it may be too weak a criterion to be meaningful. It does not address any questions about income distribution, and does not give any “equity” implication. Fairness is a notion that may overcome this difficulty. This is one way to restrict the whole set of Pareto efficient outcomes to a small set of Pareto efficient outcomes that satisfy the other properties.

What is the equitable allocation?

How can we define the notion of equitable allocation?
Definition 12.3.1 (Envy) An agent $i$ is said to envy agent $k$ if agent $i$ prefers agent $k$’s consumption. i.e., $x_k \succ_i x_i$.

Definition 12.3.2 An allocation $x$ is equitable if no one envies anyone else, i.e., for each $i \in N$, $x_i \succ_i x_k$ for all $k \in N$.

Definition 12.3.3 (Fairness) An allocation $x$ is said to be fair if it is both Pareto optimal and equitable.

![Figure 12.3: A fair allocation.](image)

Remark 12.3.1 By the definition, a set of fair allocations is a subset of Pareto efficient allocations. Therefore, fairness restricts the size of Pareto optimal allocations.

The following strict fairness concept is due to Lin Zhou (JET, 1992, 57: 158-175).

An agent $i$ envies a coalition $S$ ($i \notin S$) at an allocation $x$ if $\bar{x}_S \succ_i x_i$, where $\bar{x}_S = \frac{1}{|S|} \sum_{j \in S} x_j$.
**Definition 12.3.4** An allocation $x$ is strictly equitable or strictly envy-free if no one envies any other coalitions.

**Definition 12.3.5 (Strict Fairness)** An allocation $x$ is said to be strictly fair if it is both Pareto optimal and strictly equitable.

**Remark 12.3.2** The set of strictly fair allocations are a subset of Pareto optimal allocations.

**Remark 12.3.3** For a two person exchange economy, if $x$ is Pareto optimal, it is impossible for two persons to envy each other.

**Remark 12.3.4** It is clear every strictly fair allocation is a fair allocation, but the converse may not be true. However, when $n = 2$, a fair allocation is a strictly fair allocation.

The following figure shows that $x$ is Pareto efficient, but not equitable.

![Figure 12.4: x is Pareto efficient, but not equitable.](image)

The figure below shows that $x$ is equitable, but it is not Pareto efficient.
How to test a fair allocation?

**Graphical Procedures for Testing Fairness:**

Let us restrict an economy to a two-person economy. An easy way for agent $A$ to compare his own allocation $x_A$ with agent $B$’s allocation $x_B$ in the Edgeworth Box is to find a point symmetric of $x_A$ against th center of the Box. That is, draw a line from $x_A$ to the center of the box and extrapolate it to the other side by the same length to find $x'_A$, and then make the comparison. If the indifference curve through $x_A$ cuts “below” $x'_A$, then $A$ envies $B$. Then we have the following way to test whether an allocation is a fair allocation:

**Step 1:** Is it Pareto optimality? If the answer is “yes”, go to step 2; if no, stop.

**Step 2:** Construct a reflection point $(x_B, x_A)$. (Note that $\frac{x_A + x_B}{2}$ is the center of the Edgeworth box.)

**Step 3:** Compare $x_B$ with $x_A$ for person $A$ to see if $x_B \succ_A x_A$ and compare $x_A$ with $x_B$ for person $B$ to see if $x_A \succ_B x_B$. If the answer is “no” for both persons, it is a fair allocation.
We have given some desirable properties of “fair” allocations. A question is whether it exists at all. The following theorem provides one sufficient condition to guarantee the existence of fairness.

**Theorem 12.3.1 (Fairness Theorem)** Let \((x^*, p^*)\) be a competitive equilibrium. Under local non-satiation, if all individuals’ income is the same, i.e., \(p^*w_1 = p^*w_2 = \ldots = p^*w_n\), then \(x^*\) is a strictly fair allocation.

**Proof:** By local non-satiation, \(x^*\) is Pareto optimal by the First Fundamental Theorem of Welfare Economics. We only need to show \(x^*\) is strictly equitable. Suppose not. There is \(i\) and a coalition \(S\) with \(i \not\in S\) such that

\[
\pi_S^* = \frac{1}{|S|} \sum_{k \in S} x^*_k \succ_i x^*_i
\]  

(12.3)

Then, we have \(p^*\pi_S^* > p^*x^*_i = p^*w_i\). But this contradicts the fact that

\[
p^*\pi_S^* = \frac{1}{|S|} \sum_{k \in S} p^*x^*_k = \frac{1}{|S|} \sum_{k \in S} p^*w_k = p^*w_i
\]  

(12.4)
by noting that \( p^*w_1 = p^*w_2 = \ldots = p^*w_n \). Therefore, it must be a strictly fair allocation.

**Definition 12.3.6** An allocation \( x \in \mathbb{R}^n_{+} \) is an equal income Walrasian allocation if there exists a price vector such that

1. \( px_i \leq p\bar{w}, \) where \( \bar{w} = \frac{1}{n} \sum_{k=1}^{n} w_k \) : average endowment.
2. \( x'_i \succ_i x_i \) implies \( px'_i > p\bar{w} \).
3. \( \sum x_i \leq \sum w_i \).

Notice that every equal Walrasian allocation \( x \) is a competitive equilibrium allocation with \( w_i = \bar{w} \) for all \( i \).

**Corollary 12.3.1** Under local non-satiation, every equal income Walrasian allocation is strictly fair allocation.

**Remark 12.3.5** An “equal” division of resource itself does not give “fairness,” but trading from “equal” position will result in “fair” allocation. This “divide-and-choose” recommendation implies that if the center of the box is chosen as the initial endowment point, the competitive equilibrium allocation is fair. A political implication of this remark is straightforward. Consumption of equal bundle is not Pareto optimum, if preferences are different. However, an equal division of endowment plus competitive markets result in fair allocations.

**Remark 12.3.6** A competitive equilibrium from an equitable (but not “equal” division) endowment is not necessarily fair.

**Remark 12.3.7** Fair allocations are defined without reference to initial endowments. Since we are dealing with optimality concepts, initial endowments can be redistributed among agents in a society.

In general, there is no relationship between an allocation in the core and a fair allocation. However, when the social endowments are divided equally among two persons, we have the following theorem.
Theorem 12.3.2  In a two-person exchange economy, if $\succeq_i$ are convex, and if the total endowments are equally divided among individuals, then the set of allocations in the core is a subset of (strictly) fair allocation.

Proof: We know that an allocation in the core $x$ is Pareto optimal. We only need to show that $x$ is equitable. Since $x$ is a core allocation, then $x$ is individually rational (everyone prefers initial endowments). If $x$ is not equitable, there is some agent $i$, say, agent $A$, such that

$$x_B \succ_A x_A \succ_A w_A = \frac{1}{2}(w_A + w_B)$$

by noting that $w_A = w_B$ and $x$ is feasible. Thus, $x_A \succ_A \frac{1}{2}(x_A + x_B)$. But, on the other hand, since $\succeq_A$ is convex, $x_B \succ_A x_A$ implies that $\frac{1}{2}(x_A + x_B) \succ_A x_A$, a contradiction.

12.4 Social Choice Theory

12.4.1 Introduction

In this section, we present a vary brief summary and introduction of social choice theory. We analyze the extent to which individual preferences can be aggregated into social preferences, or more directly into social decisions, in a “satisfactory” manner (in a manner compatible with the fulfilment of a variety of desirable conditions).

As was shown in the discussion of “fairness,” it is difficult to come up with a criterion (or a constitution) that determines that society’s choice. Social choice theory aims at constructing such a rule which could be allied with not only Pareto efficient allocations, but also any alternative that a society faces. We will give some fundamental results of social choice theory: Arrow Impossibility Theorem, which states there does not exist any non-dictatorial social welfare function satisfying a number of “reasonable” assumptions. Gibbard-Satterthwaite
theorem states that no social choice mechanism exists which is non-dictatorial and can never be advantageously manipulated by some agents.

12.4.2 Basic Settings

To simplify the exposition, we will return to the notation used in pre-orders of preferences. Thus $aP_i b$ means that $i$ strictly prefers $a$ to $b$. We will also assume that the individual preferences are all strict, in other words, that there is never a situation of indifference. For all $i$, $a$ and $b$, one gets preference $aP_i b$ or $bP_i a$ (this hypothesis is hardly restrictive if $A$ is finite).

$$N = \{1, 2, \ldots, n\} : \text{the set of individuals;}$$

$$X = \{x_1, x_2, \ldots, x_m\} : (m \geq 3) : \text{the set of alternatives (outcomes);}$$

$$P_i = (\text{or } \succ_i) : \text{strict preference orderings of agent } i;$$

$\mathcal{P}_i$ the class of allowed individual orderings;

$$P = (P_1, P_2, \ldots, P_n) : \text{a preference ordering profile;}$$

$\mathcal{P}$: the set of all profiles of individuals orderings;

$$S(X) : \text{the class of allowed social orderings.}$$

Arrow’s social welfare function:

$$F : \mathcal{P} \rightarrow S(X) \quad (12.5)$$

which is a mapping from individual ordering profiles to social orderings.

Gibbard-Satterthwaite’s social choice function (SCF) is a mapping from individual preference orderings to the alternatives

$$f : \mathcal{P} \rightarrow X \quad (12.6)$$

Note that even though individuals’ preference orderings are transitive, a social preference ordering may not be transitive. To see this, consider the following example.
Example 12.4.1 (The Condorcet Paradox) Suppose a social choice is determined by the majority voting rule. Does this determine a well-defined social welfare function? The answer is in general no by the well-known Condorcet paradox. Consider a society with three agents and three alternatives: \( x, y, z \). Suppose each person’s preference is given by

\[
x \succ_1 y \succ_1 z \quad \text{(by person 1)}
\]
\[
y \succ_2 z \succ_2 x \quad \text{(by person 2)}
\]
\[
z \succ_3 x \succ_3 y \quad \text{(by person 3)}
\]

By the majority rule,

For \( x \) and \( y \), \( xFy \) (by social preference)

For \( y \) and \( z \), \( yFz \) (by social preference)

For \( x \) and \( z \), \( zFx \) (by social preference)

Then, pairwise majority voting tells us that \( x \) must be socially preferred to \( y \), \( y \) must be socially preferred to \( z \), and \( z \) must be socially preferred to \( x \). This cyclic pattern means that social preference is not transitive.

The number of preference profiles can increase very rapidly with increase of number of alternatives.

Example 12.4.2 \( X = \{x, y, z\}, n = 3 \)

\[
x \succ y \succ z
\]
\[
x \succ z \succ y
\]
\[
y \succ x \succ z
\]
\[
y \succ z \succ x
\]
\[
z \succ x \succ y
\]
\[
z \succ y \succ x
\]

Thus, there are six possible individual orderings, i.e., \( |P_i| = 6 \), and therefore there are \( |P_1| \times |P_2| \times |P_3| = 6^3 = 216 \) possible combinations of 3-individual preference orderings on three alternatives. The social welfare function is a mapping from each of these 216 entries (cases) to one particular social ordering (among six possible social orderings of three alternatives). The social choice function is
a mapping from each of these 216 cases to one particular choice (among three alternatives). A question we will investigate is what kinds of desirable conditions should be imposed on these social welfare or choice functions.

You may think of a hypothetical case that you are sending a letter of listing your preference orderings, $P_i$, on announced national projects (alternatives), such as reducing deficits, expanding the medical program, reducing society security program, increasing national defence budget, to your Congressional representatives. The Congress convenes with a huge stake of letters $P$ and try to come up with national priorities $F(P)$. You want the Congress to make a certain rule (the Constitution) to form national priorities out of individual preference orderings. This is a question addressed in social choice theory.

12.4.3 Arrow’s Impossibility Theorem

Definition 12.4.1 Unrestricted Domain (UD): A class of allowed individual orderings $\mathcal{P}_i$ consists of all possible orderings defined on $X$.

Definition 12.4.2 Pareto Principle (P): if for $x, y \in X$, $x P_i y$ for all $i \in N$, then $xF(P)y$ (social preferences).

Definition 12.4.3 Independence of Irrelevant Alternatives (IIA): For any two alternatives $x, y, \in X$ and any two preference profiles $P, P' \in \mathcal{P}^n$, $xF(P)y$ and \( \{i : xP_i y\} = \{i : xP'_i y\} \) for all $i$ implies that $xF(P')y$. That is, the social preference between any two alternatives depends only on the profile of individual preferences over the same alternatives.

Remark 12.4.1 IIA means that the ranking between $x$ and $y$ for any agent is equivalent in terms of $P$ and $P'$ implies that the social ranking between $x$ and $y$ by $F(P)$ and $F(P')$ is the same. In other words, if two different preference profiles that are the same on $x$ and $y$, the social order must be also the same on $x$ and $y$. 

464
Remark 12.4.2 By IIA, any change in preference ordering other than the ordering of \( x \) and \( y \) should not affect social ordering between \( x \) and \( y \).

Example 12.4.3 Suppose \( x P_i y \) and \( x P'_i z \) and \( x P'_i z P'_i y \).

By IIA, if \( xF(P)y \), then \( xF(P')y \).

Definition 12.4.4 (Dictator) There is some agent \( i \in N \) such that \( F(P) = P_i \) for all \( P \in \mathcal{P}^n \), and agent \( i \) is called a dictator.

Theorem 12.4.1 (Arrow’s Impossibility Theorem) Any social welfare function that satisfies \( m \geq 3 \), UD, \( P \), IIA conditions is dictatorial.

Proof: The idea for the following proof dates back to Vickrey (1960); it comprises three lemmas.

Lemma 12.4.1 (Neutrality) Given a division \( N = M \cup I \) and \( (a, b, x, y) \in A^4 \) such that

\[
\forall i \in M, xP_i y \text{ and } aP'_i b \\
\forall i \in I, yP_i x \text{ and } bP'_i a
\]

then

\[
x \hat{P} y \Leftrightarrow a \hat{P}' b \quad (12.7)
\]

\[
x \hat{I} y \Leftrightarrow a \hat{I}' b \quad (12.8)
\]

Proof: The interpretation of lemma 12.4.1 is simple: If \( x \) and \( y \) are ordered by each individual in \( P \) as \( a \) and \( b \) in \( P' \), then this must also be true of social preference; that is, \( x \) and \( y \) must be ordered by \( \hat{P} \) as \( a \) and \( b \) by \( \hat{P}' \). If this weren’t the case, then the procedure of aggregation would treat the pairs \( (a, b) \) and \( (x, y) \) in a non-neutral manner, hence the name of the lemma.
Let us first prove (12.7). Suppose that \( x \hat{\prec} P' y \) and that \((a, b, x, y)\) are distinct two by two. Let preferences \( P'' \) be such that
\[
\forall i \in M, a P''_i x P''_i y P''_i b \\
\forall i \in I, y P''_i b P''_i a P''_i x
\]
(such preferences exist because the domain is universal). By IIA, since \( x \) and \( y \) are ordered individually in \( P'' \) as in \( P \), we have \( x \hat{\prec} P' y \). By the Pareto principle, \( a \hat{\prec} P'' b \). By transitivity, we get \( a \hat{\prec} P'' b \). Finally, in reapplying IIA, we find that \( a \hat{\prec} P' b \). The cases where \( a \) or \( b \) coincide with \( x \) or \( y \) are treated similarly.

Part (12.8) is obtained directly. If \( x \hat{\succ} y \) but, for example \( a \hat{\succ} P' b \), we find the contradiction \( x \hat{\succ} P' y \) by (12.7) in reversing the roles of \((a, b)\) and \((x, y)\).

Before stating Lemma 12.4.2, two terms must be defined. We say that a set of agents \( M \) is almost decisive on \((x, y)\) if for every \( \hat{P} \),
\[
(\forall i \in M, x P_i y \text{ and } \forall i \notin M, y P_i x) \Rightarrow x \hat{\prec} P y
\]
We say that \( M \) is decisive on \((x, y)\) if for every \( \hat{P} \),
\[
(\forall i \in M, x P_i y) \Rightarrow x \hat{\prec} P y
\]

**Lemma 12.4.2** If \( M \) is almost decisive on \((x, y)\), then it is decisive on \((x, y)\).

Proof: Suppose that \( \forall i \in M, x P_i y \). By IIA, only individual preferences count on \((x, y)\); the others can be changed. Assume therefore, with no loss of generality, that \( z \) exists such that \( x P_i z P_i y \) if \( i \in M \) and \( z \) is preferred to \( x \) and \( y \) by the other agents. Neutrality imposes that \( x \hat{\prec} P z \) because the individual preferences are oriented as on \((x, y)\) and \( M \) is almost decisive on \((x, y)\). Finally, the Pareto principle implies that \( z \hat{\prec} P y \) and transitivity implies the conclusion that \( x \hat{\prec} P y \).

Note that neutrality implies that if \( M \) is decisive on \((x, y)\), it is decisive on every other pair. We will therefore just say that \( M \) is decisive.

**Lemma 12.4.3** If \( M \) is decisive and contains at least two agents, then a strict subset of \( M \) exists that is decisive.
Proof: Divide $M = M_1 \cup M_2$, and choose a $\hat{P}$ such that

- on $M_1$, $x P_i y P_i z$
- on $M_2$, $y P_i z P_i x$
- outside $M$, $z P_i x P_i y$

As $M$ is decisive, we have $y \hat{P} z$.

One of two things results:

- Either $y \hat{P} x$, and (since $z$ doesn’t count by IIA) $M_2$ is almost decisive on $(x, y)$, and therefore decisive by lemma 12.4.2.
- Or $x \hat{P} y$, and by transitivity $x \hat{P} z$; then ($y$ doesn’t count by IIA) $M_1$ is almost decisive on $(x, y)$, and therefore is decisive by lemma refsoc2.

The proof is concluded by noting that by the Pareto principle, $I$ as a whole is decisive. In apply Lemma 12.4.3, the results is an individual $i$ that is decisive, and this is therefore the sought-after dictator. The attentive reader will have noted that lemma 12.4.1 remains valid when there are only two alternatives, contrary to Lemmas 12.4.2 and 12.4.3. The proof of the theorem is completed.

The impact of the Arrow possibility theorem has been quite substantial. Obviously, Arrow’s impossibility result is a disappointment. The most pessimistic reaction to it is to conclude that there is just no acceptable way to aggregate individual preferences, and hence no theoretical basis for treating welfare issues. A more moderate reaction, however, is to examine each of the assumptions of the theorem to see which might be given up. Conditions imposed on social welfare functions may be too restrictive. Indeed, when some conditions are relaxed, then the results could be positive, e.g., UD is usually relaxed.

12.4.4 Some Positive Result: Restricted Domain

When some of the assumptions imposed in Arrow’s impossibility theorem is removed, the result may be positive. For instance, if alternatives have certain
characteristics which could be placed in a spectrum, preferences may show some patterns and may not exhaust all possibilities orderings on $X$, thus violating (UD). In the following, we consider the case of restricted domain. A famous example is a class of “single-peaked” preferences. We show that under the assumption of single-peaked preferences, non-dictatorial aggregation is possible.

**Definition 12.4.5** A binary relation $\geq$ on $X$ is a *linear order* if $\geq$ is reflexive ($x \geq x$), transitive ($x \geq y \geq z$ implies $x \geq z$), and total (for distinct $x, y \in X$, either $x \geq y$ or $y \geq x$, but not both).

Example: $X = \mathbb{R}$ and $x \geq y$.

**Definition 12.4.6** $\succ_i$ is said to be *single-peaked* with respect to the linear order $\geq$ on $X$, if there is an alternative $x \in X$ such that $\succ_i$ is increasing with respect to $\geq$ on the lower contour set $L(x) = \{y \in X : y \leq x\}$ and decreasing with respect to $\geq$ on the upper contour set $U(x) = \{y \in X : y \geq x\}$. That is,

1. $x \geq z > y$ implies $z \succ_i y$
2. $y > z \geq x$ implies $z \succ_i y$

In words, there is an alternative that represents a peak of satisfaction and, moreover, satisfaction increases as we approach this peak so that, in particular, there cannot be any other peak of satisfaction.
Figure 12.7: $u$ in the left figure is single-peaked, $u$ in the right figure is not single-peaked.

Figure 12.8: Five Agents who have single-peaked preferences.

Given a profile of preference $(\succsim_1, \ldots, \succsim_n)$, let $x_i$ be the maximal alternative for $\succsim_i$ (we will say that $x_i$ is “individual $i$’s peak”).

**Definition 12.4.7** Agent $h \in N$ is a median agent for the profile $(\succsim_1, \ldots, \succsim_n)$ if $\# \{i \in N : x_i \succeq x_h\} \geq \frac{n}{2}$ and $\# \{i \in N : x_h \succeq x_i\} \geq \frac{n}{2}$.

**Proposition 12.4.1** Suppose that $\succeq$ is a linear order on $X$, $\succsim_i$ is single-peaked. Let $h \in N$ be a median agent, then the majority rule $\tilde{F}(\succeq)$ is aggregatable:

$$x_h \tilde{F}(\succeq) y \quad \forall y \in X.$$  

That is, the peak $x_h$ of the median agent is socially optimal (cannot be defeated by any other alternative) by majority voting. Any alternative having this property
is called a Condorect winner. Therefore, a Condorect winner exists whenever the preferences of all agents are single-peaked with respect to the same linear order.

Proof. Take any $y \in X$ and suppose that $x_h > y$ (the argument is the same for $y > x_h$). We need to show that

$$\#\{i \in N : x_h \succ_i y\} \geq \#\{i \in N : y \succ_i x_h\}.$$ 

Consider the set of agents $S \subseteq N$ that have peaks larger than or equal to $x_h$, that is, $S = \{i \in N : x_i \geq x_h\}$. Then $x_i \geq x_h > y$ for every $i \in S$. Hence, by single-peakedness of $\succ_i$ with respect to $\succeq$, we get $x_h \succ_i y$ for every $i \in S$ and thus $\#S \leq \#\{i \in N : x_h \succ_i y\}$. Hence, $\{i \in N : y \succ_i x_h\} \subseteq (N \setminus S)$, and then $\#\{i \in N : y \succ_i x_h\} \leq \#(N \setminus S)$. On the other hand, because agent $h$ is a median agent, we have that $\#S \geq n/2$ and therefore $\#\{i \in N : y \succ_i x_h\} \leq \#(N \setminus S) \leq n/2 \leq \#S \leq \#\{i \in N : x_h \succ_i y\}$.

### 12.4.5 Gibbard-Satterthwaite Impossibility Theorem

The task we set ourselves to accomplish in the previous sections was how to aggregate profiles of individual preference relations into a rational social preference order. Presumably, this social preference order is then used to make decisions. In this section we focus directly on social decision and pose the aggregation question as one analyzing how profiles of individual preferences turn into social decisions.

The main result we obtain again yields a dictatorship conclusion. The result amounts in a sense, to a translation of the Arrow’s impossibility theorem into the language of choice functions. It provides a link towards the incentive-based analysis in the mechanism design we will discuss in Part VI.

**Definition 12.4.8** A social choice function (SCF) is manipulable at $P \in \mathcal{P}^n$ if there exists $P'_i \in \mathcal{P}$ such that

$$f(P_{-i}, P'_i) P_i f(P_{-i}, P)$$

(12.9)

where $P_{-i} = (P_1, \ldots, P_{i-1}, P_{i+1}, \ldots, P_n)$. 

470
**Definition 12.4.9** A SCF is strongly individually incentive compatible (SIIC) if there exists no preference ordering profile at which it is manipulable. In other words, the truth telling is a dominant strategy equilibrium:

\[ f(P'_{-i}, P_i) P_i f(P'_{-i}, P'_i) \text{ for all } P' \in \mathcal{P}^n \]  

(12.10)

**Definition 12.4.10** A SCF is dictatorial if there exists an agent whose optimal choice is the social optimal.

**Theorem 12.4.2 (Gibbard-Satterthwaite Theorem)** If \( X \) has at least 3 alternatives, a SCF which is SIIC and UD is dictatorial.

Proof. There are several the ways to prove the Gibbard-Satterthwaite’s Impossibility Theorem. The proof adopted here is due to Schmeidler-Sonnenschein (1978). It has the advantage of bringing to light a close relationship to Arrow’s Impossibility Theorem.

We want to prove that if such a mechanism is manipulable, then \( f \) is dictatorial. The demonstration comprises two lemmas; it consists of starting from an implementable SCF \( f \) and of constructing from it an SWF \( F \) that verifies Arrow’s conditions. One concludes from this that \( F \) is dictatorial, which implies that \( f \) also is.

**Lemma 12.4.4** Suppose that \( f(P) = a_1 \) and \( f(P'_i, P_{-i}) = a_2 \), where \( a_2 \neq a_1 \). Then

1. \( f \) is manipulable by \( i \) in \( (P'_i, P_{-i}) \) if \( a_1 P'_i a_2 \)
2. \( f \) is manipulable by \( i \) in \( P \) if \( a_2 P_i a_1 \)

*Proof* In both cases it is sufficient to write the definition of manipulability.

We will need the notation \( P'_i \), which, for a profile \( P \) and given agents \( i < j \), will represent the vector \( (P_i, \ldots, P_j) \).

**Lemma 2** Let \( B \) be a subset of the image of \( f \) and \( P \) a profile such that
• \( \forall a_1 \in B, \forall a_2 \notin B, \forall i = 1, \ldots, n, a_1 P_i, a_2 \)

Then \( f(P) \in B \).

**Proof**  
This can be shown by contradiction. Let \( a_2 = f(P) \), and suppose that \( a_2 \notin B \). Let \( P' \) be a profile such that \( f(P') = a_1 \in B \) (such a profile does exist, since \( B \) is included in the image of \( f \) and given the universal domain hypothesis). Now construct a sequence \( (a_3^i)_{i=0,\ldots,n} \) by

- \( a_3^0 = a_2 \notin B \)
- for \( i = 1, \ldots, n - 1 \), \( a_3^i = f(P_1^i, P_{i+1}^n) \)
- \( a_3^n = a_1 \in B \)

Let \( j \) be the first integer such that \( a_3^j \in B \). We then get

- \( f(P_1^j, P_{j+1}^n) = a_3^j \in B \)
- \( f(P_1^{j-1}, P_j^n) = a_3^{j-1} \notin B \)

and by the hypothesis of the lemma, \( a_3^j P_j a_3^{j-1} \). Lemma 1 then implies that \( f \) is manipulable.

Now construct an SWF \( F \). Let \( P \) be any profile and \( a_1, a_2 \) two choices in \( A \). Define a new profile (using UD) \( \tilde{P} \) such that for each \( i \),

- \( \tilde{P}_i \), coincides with \( P_i \), on \( \{a_1, a_2\} \)
- \( \tilde{P}_i \), coincides with \( P_i \), on \( A - \{a_1, a_2\} \)
- \( \{a_1, a_2\} \) is placed at the top of the preferences \( \tilde{P}_i \)

(Strictly speaking, \( \tilde{P} \) of course depends on \( a_1 \) and \( a_2 \), and the notation should reflect this.)

Lemma 2 implies that \( f(\tilde{P}) \in \{a_1, a_2\} \) (taking \( B = \{a_1, a_2\} \) and replacing \( P \) by \( \tilde{P} \) in the statement of the lemma). \( F \) can therefore be defined by
Now we can verify Arrow’s conditions: There are surely at least three choices.

- $a_1 F(P) a_2 \iff f(\tilde{P}) = a_1$

Now we can verify Arrow’s conditions: There are surely at least three choices.

- $F$ is, by construction, of universal domain.

- $F$ satisfies the Pareto principle: if for every $i$, $a_1 P a_2$, then $a_1$ is at the top of all preferences $\tilde{P}_i$. By taking $B = \{a_1, a_2\}$ in the statement of lemma 2, we indeed get $f(\tilde{P}) = a_1$.

- $F$ satisfies IIA: if this were not the case, there would exist $P, P', a_1$ and $a_2$ such that

  for every $i$, $a_1 P a_2 \iff a_1 P' a_2$

- $a_1 F(P) a_2$ and $a_2 F(P) a_1$

Now define a sequence $(a_i^3)_{i=0,\ldots,n}$ by

- $a_3^n = a_1$
  
  for $i = 1, \ldots, n - 1, a_i^3 = f(\tilde{P}_1^i, \tilde{P}_i^{n+1})$

- $a_3^n = a_2$

  Lemma 2 implies that $a_3^i \in \{a_1, a_2\}$ for every $i$. Therefore let $j$ be the first integer such that $a_3^j = a_2$. This gives $f(\tilde{P}_1^j, \tilde{P}_j^n) = a_2$ and $f(\tilde{P}_1^{j-1}, \tilde{P}_j^n) = a_1$.

Now one of two things can result:

- $a_1 P_j a_2$

  This implies $a_1 P'_j a_2$ and therefore $a_1 \tilde{P}_j' a_2$, so lemma 1 implies that $f$ is manipulable.

- $a_2 P_j a_1$
This implies $a_2 \tilde{P}_j a_1$, so lemma 1 again implies that $f$ is manipulable.

But there is contradiction in both results: for every $P, F(P)$ is clearly a complete and asymmetrical binary relation. What remains for us is to verify that it is transitive.

Take the opposite case so that we have a cycle on a triplet \{a_1, a_2, a_3\}. For every $i$, let $P'_i$, which coincides with $P_i$, on \{a_1, a_2, a_3\} and on $A - \{a_1, a_2, a_3\}$ be such that \{a_1, a_2, a_3\} is at the top of $P'_i$ (using UD). Lemma 2 implies that $f(P') \in \{a_1, a_2, a_3\}$; without any loss of generality, we can assume that $f(P') = a_1$. Since $F(P)$ has a cycle on \{a_1, a_2, a_3\}, we necessarily get $a_2 F(P) a_1$ or $a_3 F(P) a_1$. Here again, without loss of generality, we can assume that $a_3 F(P) a_1$. Now modify $P'$ in $P''$ by making $a_2$ move into third place in each individual preference ($P''$ is admissible by UD). Note that $a_3 P_i a_1$ if and only if $a_3 P''_i a_1$; in applying IIA (which we have just shown is satisfied), we get $a_3 F(P'') a_1$, which again implies $a_3 = f(P'')$.

At the risk of seeming redundant, we now define a sequence $(a^i_4)_{i=0, \ldots, n}$ by

- $a^0_4 = a_1$
- For $i = 1, \ldots, n - 1$, $a^i_4 = f(\tilde{P}^m_{i-1}, \tilde{P}^n_{i+1})$
- $a^n_4 = a_3$

Lemma 2 implies that $a^i_4 \in \{a_1, a_2, a_3\}$ for every $i$. Therefore let $j$ be the first integer such that $a^j_4 \neq a_1$. One of two things results:

- $a^j_4 = a_2$

but $a_1 P''_j a$, since $a_2$ is only in third position in $P''_j$. Therefore $f(\tilde{P}^m_{j-1}, \tilde{P}^n_{j}) P''_j f(\tilde{P}^m_{j}, \tilde{P}^n_{j+1})$, so $f$ is manipulable.

- $a^j_4 = a_3$

Now, if $a_1 P''_j a_3$, we also have $a_1 P''_j a_3$. Therefore $f(\tilde{P}^m_{j-1}, \tilde{P}^n_{j}) P''_j f(\tilde{P}^m_{j}, \tilde{P}^n_{j+1})$, and $f$ is manipulable. If $a_3 P'_j a_1$, we directly get $f(\tilde{P}^m_{j}, \tilde{P}^n_{j+1}) P''_j f(\tilde{P}^m_{j-1}, \tilde{P}^n_{j})$, and
$f$ is still manipulable. We are led therefore to a contradiction in every case, which shows that $F(P)$ is transitive.

Since $F$ verifies all of Arrow’s conditions, $F$ must be dictatorial; let $i$ be the dictator. Let $P$ be any profile and arrange the choices in such a way that $a_1P_1a_2P_2\ldots$ Since $i$ is the dictator, more precisely we have $a_1F(P)a_2$ and therefore $f(\tilde{P}) = a_1$. But, by construction, $\tilde{P}$ coincides with $P$ and $f(P)$ is therefore $a_1$, the preferred choice of $i$, which concludes the proof showing that $i$ is also a dictator for $f$.

The readers who are interested in the other ways of proofs are referred to Mas-Colell, Whinston, and Green (1995).

12.5 Reference

Books and Monographs:


Papers:


Chapter 13

General Equilibrium Under Uncertainty

13.1 Introduction

In this chapter, we apply the general equilibrium framework developed in Chapters 10 to 12 to economic situations involving the exchange and allocation of resources under conditions of uncertainty.

We begin by formalizing uncertainty by means of states of the world and then introduce the key idea of a contingent commodity: a commodity whose delivery is conditional on the realized state of the world. We then use these tools to define the concept of an Arrow-Debreu equilibrium. This is simply a Walrasian equilibrium in which contingent commodities are traded. It follows from the general theory of Chapter 11 that an Arrow-Debreu equilibrium results in a Pareto optimal allocation of risk.

In Section 13.4, we provide an important reinterpretation of the concept of Arrow-Debreu equilibrium. We show that, under the assumptions of self-fulfilling, rational expectations, Arrow-Debreu equilibria can be implemented by combining trade in a certain restricted set of contingent commodities with spot trade that occurs after the resolution of uncertainty. This results in a significant reduction
in the number of ex ante (i.e., before uncertainty) markets that must operate. In Section 13.5 we briefly illustrate some of the welfare difficulties raised by the possibility of incomplete markets, that is, by the possibility of there being too few asset markets to guarantee a fully Pareto optimal allocation of risk.

13.2 A Market Economy with Contingent Commodities

As in our previous chapters, we contemplate an environment with $L$ physical commodities, $n$ consumers, and $J$ firms. The new element is that technologies, endowments, and preferences are now uncertain.

Throughout this chapter, we represent uncertainty by assuming that technologies, endowments, and preferences depend on the state of the world. A state of the world is to be understood as a complete description of a possible outcome of uncertainty. For simplicity we take $S$ to be a finite set with (abusing notation slightly) $S$ elements. A typical element is denoted $s = 1, \ldots, S$.

We state the key concepts of a (state-)contingent commodity and a (state-)contingent commodity vector. Using these concepts we shall then be able to express the dependence of technologies, endowments, and preferences on the realized states of the world.

**Definition 13.2.1** For every physical commodity $\ell = 1, \ldots, L$ and state $s = 1, \ldots, S$, a unit of (state-)contingent commodity $\ell s$ is a title to receive a unit of the physical good $\ell$ if and only if $s$ occurs. Accordingly, a (state-)contingent commodity vector is specified by

$$x = (x_{11}, \ldots, x_{L1}, \ldots, x_1s, \ldots, x_{LS}) \in \mathbb{R}^{LS},$$

and is understood as an entitlement to receive the commodity vector $(x_1s, \ldots, x_{LS})$ if state $s$ occurs.
We can also view a contingent commodity vector as a collection of $L$ random variables, the $l$th random variable being $(x_{l1}, \ldots, x_{ls})$.

With the help of the concept of contingent commodity vectors, we can now describe how the characteristics of economic agents depend on the state of the world. To begin, we let the endowments of consumer $i = 1, \ldots, n$ be a contingent commodity vector:

$$w_i = (w_{1i}, \ldots, w_{Li}, \ldots, w_{1si}, \ldots, w_{Lsi}) \in \mathbb{R}^{LS}.$$ 

The meaning of this is that if state $s$ occurs then consumer $i$ has endowment vector $(w_{1si}, \ldots, w_{Lsi}) \in \mathbb{R}^L$.

The preferences of consumer $i$ may also depend on the state of the world (e.g., the consumer's enjoyment of wine may well depend on the state of his health). We represent this dependence formally by defining the consumer's preferences over contingent commodity vectors. That is, we let the preferences of consumer $i$ be specified by a rational preference relation $\succ_i$, defined on a consumption set $X_i \subset \mathbb{R}^{LS}$.

The consumer evaluates contingent commodity vectors by first assigning to state $s$ a probability $\pi_{si}$ (which could have an objective or a subjective character), then evaluating the physical commodity vectors at state $s$ according to a Bernoulli state-dependent utility function $u_{si}(x_{1si}, \ldots, x_{Lsi})$, and finally computing the expected utility. That is, the preferences of consumer $i$ over two contingent commodity vectors $x_i, x'_i \in X_i \subset \mathbb{R}^{LS}$ satisfy

$$x_i \succ_i x'_i \text{ if and only if } \sum_s \pi_{si} u_{si}(x_{1si}, \ldots, x_{Lsi}) \geq \sum_s \pi_{si} u_{si}(x'_{1si}, \ldots, x'_{Lsi}).$$

It should be emphasized that the preferences $\succ_i$ are in the nature of ex ante preferences: the random variables describing possible consumptions are evaluated before the resolution of uncertainty.

Similarly, the technological possibilities of firm $j$ are represented by a production set $Y_j \subset \mathbb{R}^{LS}$. The interpretation is that a (state-)contingent production plan $y_j \in \mathbb{R}^{LS}$ is a member of $Y_j$ if for every $s$ the input–output vector $(y_{1sj}, \ldots, y_{Lsj})$ of physical commodities is feasible for firm $j$ when state $s$ occurs.
Example 13.2.1 Suppose there are two states, $s_1$ and $s_2$, representing good and bad weather. There are two physical commodities: seeds ($\ell = 1$) and crops ($\ell = 2$). In this case, the elements of $Y_j$ are four-dimensional vectors. Assume that seeds must be planted before the resolution of the uncertainty about the weather and that a unit of seeds produces a unit of crops if and only if the weather is good. Then

$$y_j = (y_{11j}, y_{21j}, y_{12j}, y_{22j}) = (-1, 1, -1, 0)$$

is a feasible plan. Note that since the weather is unknown when the seeds are planted, the plan $(-1, 1, 0, 0)$ is not feasible: the seeds, if planted, are planted in both states. Thus, in this manner we can imbed into the structure of $Y_j$ constraints on production related to the timing of the resolution of uncertainty.

To complete the description of a private market economy it only remains to specify ownership shares for every consumer $i$ and firm $j$. In principle, these shares could also be state-contingent. It will be simpler, however, to let $\theta_{ij} \geq 0$ be the share of firm $j$ owned by consumer $i$ whatever the state. Of course $\sum_j \theta_{ij} = 1$ for every $i$.

A private market economy with uncertain thus can be written

$$e = \left( S, \{X_i, w_i, \pi_i\}_{i=1}^n, \{Y_j\}_{j=1}^n, \{\theta_{ij}\}_{i,j=1}^{n,n} \right).$$

Information and the Resolution of Uncertainty

In the setting just described, time plays no explicit formal role. In reality, however, states of the world unfold over time. Figure 13.1 captures the simplest example. In the figure, we have a period 0 in which there is no information whatsoever on the true state of the world and a period 1 in which this information has been completely revealed.
Figure 13.1: Two periods. Perfect information at $t = 1$.

The same methodology can be used to incorporate into the formalism a much more general temporal structure. Suppose we have $T+1$ dates $t = 0, 1, \ldots, T$ and, as before, $S$ states, but assume that the states emerge gradually through a tree, as in Figure 13.2. Here final nodes stand for the possible states realized by time $t = T$, that is, for complete histories of the uncertain environment. When the path through the tree coincides for two states, $s$ and $s'$, up to time $t$, this means that in all periods up to and including period $t$, $s$ and $s'$ cannot be distinguished.

Figure 13.2: An information tree: gradual release of information.

Subsets of $S$ are called events. A collection of events $\mathcal{L}$ is an information structure if it is a partition, that is, if for every state $s$ there is $E \in \mathcal{L}$ with $s \in E$.
and for any two $E, E' \in \mathcal{L}, E \neq E'$, we have $E \cap E' = \emptyset$. The interpretations is that if $s$ and $s'$ belong to the same event in $\mathcal{L}$ then $s$ and $s'$ cannot be distinguished in the information structure $\mathcal{L}$.

To capture formally a situation with sequential revelation of information we look at a family of information structures: $(\mathcal{L}_0, \ldots, \mathcal{L}_t, \ldots, \mathcal{L}_T)$. The process of information revelation makes the $\mathcal{L}_t$ increasingly fine: once one has information sufficient to distinguish between two states, the information is not forgotten.

**Example 13.2.2** Consider the tree in Figure 13.2. We have

\[
\begin{align*}
\mathcal{L}_0 &= \{1, 2, 3, 4, 5, 6\}, \\
\mathcal{L}_1 &= \{1, 2\}, \{3\}, \{4, 5, 6\} \\
\mathcal{L}_2 &= \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}.
\end{align*}
\]

The partitions could in principle be different across individuals. However, we shall assume that the information structure is the same for all consumers.

A pair $(t, E)$ where $t$ is a date and $E \in \mathcal{L}_t$ is called a date-event. Date-events are associated with the nodes of the tree. Each date-event except the first has a unique predecessor, and each date-event not at the end of the tree has one or more successors.

With this temporal modeling it is now necessary to be explicit about the time at which a physical commodity is available. Suppose there is a number $H$ of basic physical commodities (bread, leisure, etc.). We will use the double index $ht$ to indicate the time at which a commodity $h$ is produced, appears as endowment, or is available for consumption. Then $x_{hts}$ stands for an amount of the physical commodity $h$ available at time $t$ along the path of state $s$.

Fortunately, this multi-period model can be formally reduced to the timeless structure introduced above. To see this, we define a new set of $L = H(T + 1)$ physical commodities, each of them being one of these double-indexed (i.e., $ht$) commodities. We then say that a vector $z \in \mathbb{R}^{LS}$ is measurable with respect to the family of information partitions $(\mathcal{L}_0, \ldots, \mathcal{L}_T)$ if, for every $hts$ and $hts'$, we have
that \( z_{hts} = z_{hts'} \) whenever \( s, s' \) belong to the same element of the partition \( \mathcal{L}_t \).
That is, whenever \( s \) and \( s' \) cannot be distinguished at time \( t \), the amounts assigned to the two states cannot be different. Finally, we impose on endowments \( w_i \in \mathbb{R}^{LS} \), consumption sets \( X_i \subset \mathbb{R}^{LS} \) and production sets \( Y_j \subset \mathbb{R}^{LS} \) the restriction that all their elements be measurable with respect to the family of information partitions. With this, we have reduced the multi-period structure to our original formulation.

### 13.3 Arrow-Debreu Equilibrium

We postulate the existence of a market for every contingent commodity \( l_s \). These markets open before the resolution of uncertainty. The price of the commodity is denoted by \( p_{ls} \). Notice that for this market to be well defined it is indispensable that all economic agents be able to recognize the occurrence of \( s \). That is, information should be symmetric across economic agents.

Formally, the market economy just described is nothing but a particular case of the economies we have studied in previous chapters. We can, therefore, apply to our market economy the concept of Walrasian equilibrium and, with it, all the theory developed so far. When dealing with contingent commodities it is customary to call the Walrasian equilibrium an Arrow-Debreu equilibrium.

**Definition 13.3.1** An allocation

\[
(x_1^*, \ldots, x_I^*, y_1^*, \ldots, Y_J^*) \in X_1 \times \cdots \times X_I \times Y_1 \times \cdots \times Y_J \subset \mathbb{R}^{LS(I+J)}
\]

and a system of prices for the contingent commodities \( p = (p_{11}, \ldots, P_{LS}) \in \mathbb{R}^{LS} \) constitute an Arrow-Debreu equilibrium if:

(i) For every \( j \), \( Y_j^* \) satisfies \( p \cdot y_j^* \geq p \cdot y_j \) for all \( y_j \in Y_j \).

(ii) For every \( i \), \( x_i^* \) is maximal for \( \succeq_i \) in the budget set \( \{x_i \in X_i: p \cdot x_i \leq p \cdot w + \sum_ j \theta_{ij} p \cdot y_j^* \} \).

(iii) \( \sum_i x_i^* = \sum_j y_j^* + \sum_i w_i \).
The positive and welfare theorems of Chapters 10 and 11 apply without modification to the Arrow-Debreu equilibrium. Recall that, in the present context, the convexity assumption takes on an interpretation in terms of risk aversion. For example, in the expected utility setting, the preference relation \( \succ_i \) is convex if the Bernoulli utilities \( u_{si}(x_{si}) \) are concave.

The Pareto optimality implication of Arrow-Debreu equilibrium says, effectively, that the possibility of trading in contingent commodities leads, at equilibrium, to an efficient allocation of risk.

It is important to realize that at any production plan the profit of a firm, \( p \cdot y_j \), is a nonrandom amount of dollars. Productions and deliveries of goods do, of course, depend on the state of the world, but the firm is active in all the contingent markets and manages, so to speak, to insure completely. This has important implications for the justification of profit maximization as the objective of the firm.

**Example 13.3.1** Consider an exchange economy with \( n = 2, L = 1, \) and \( S = 2 \). This lends itself to an Edgeworth box representation because there are precisely two contingent commodities. In Figures xx(a) and xx(b) we have \( w_1 = (1, 0), w_2 = (0, 1) \), and utility functions of the form \( \pi_{1i}u_1(x_{1i}) + \pi_{2i}u_2(x_{2i}) \), where \( (\pi_{1i}, \pi_{2i}) \) are the subjective probabilities of consumer \( i \) for the two states. Since \( w_1 + w_2 = (1, 1) \) there is no aggregate uncertainty, and the state of the world determines only which consumer receives the endowment of the consumption good. Recall that for this model [in which the \( u_i(\cdot) \) do not depend on \( s \)], the marginal rate of substitution of consumer \( i \) at any point where the consumption is the same in the two states equals the probability ratio \( \pi_{1i}/\pi_{2i} \).

In Figure 13.3(a) the subjective probabilities are the same for the two consumers (i.e., \( \pi_{11} = \pi_{12} \)) and therefore the Pareto set coincides with the diagonal of the box (the box is a square and so the diagonal coincides with the 45-degree line, where the marginal rates of substitution for the two consumers are equal: \( \pi_{11}/\pi_{21} = \pi_{12}/\pi_{22} \)). Hence, at equilibrium, the two consumers insure completely; that is, consumer \( i \)'s equilibrium consumption does not vary across the two
states. In Figure 13.3(b) the consumers' subjective probabilities are different. In particular, $\pi_{11} < \pi_{12}$ (i.e., the second consumer gives more probability to state 1). In this case, each consumer’s equilibrium consumption is higher in the state he thinks comparatively more likely (relative to the beliefs of the other consumer).

Figure 13.3: (a) No aggregate risk: some probability assessments. (b) No aggregate risk: different probability assessments.

13.4 Sequential Trade

The Arrow-Debreu framework provides a remarkable illustration of the power of general equilibrium theory. Yet, it is hardly realistic. Indeed, at an Arrow-Debreu equilibrium all trade takes place simultaneously and before the uncertainty is resolved. Trade is a one-shot affair. In reality, however, trade takes place to a large extent sequentially over time, and frequently as a consequence of information disclosures. The aim of this section is to introduce a first model of sequential trade and show that Arrow-Debreu equilibria can be reinterpreted by means of trading processes that actually unfold through time.
To be as simple as possible we consider only exchange economies. In addition, we take \( X_i = \mathbb{R}_{+}^{LS} \) for every \( i \). To begin with, we assume that there are two dates, \( t = 0 \) and \( t = 1 \), that there is no information whatsoever at \( t = 0 \), and that the uncertainty has resolved completely at \( t = 1 \). Thus, the date-event tree is as in Figure 13.3. Again for simplicity, we assume that there is no consumption at \( t = 0 \).

Suppose that markets for the \( LS \) possible contingent commodities are set up at \( t = 0 \), and that \((x_1^*, \ldots, x_n^*) \in \mathbb{R}^{LSn}\) is an Arrow-Debreu equilibrium allocation with prices \((p_{11}, \ldots, p_{LS}) \in \mathbb{R}^{LS}\). Recall that these markets are for delivery of goods at \( t = 1 \) (they are commonly called forward markets). When period \( t = 1 \) arrives, a state of the world \( s \) is revealed, contracts are executed, and every consumer \( i \) receives \( x_{si}^* = (x_{1si}^*, \ldots, x_{LSi}^*) \in \mathbb{R}^{L} \). Imagine now that, after this but before the actual consumption of \( x_{si}^* \), markets for the \( L \) physical goods were to open at \( t = 1 \) (these are called spot markets). Would there be any incentive to trade in these markets? The answer is “no.” To see why, suppose that there were potential gains from trade among the consumers, that is, that there were \( x_{si} = (x_{1si}, \ldots, x_{LSi}) \in \mathbb{R}^{L} \) for \( i = 1, \ldots, n \), such that \( \sum_i x_{si} \leq \sum_i w_{si} \) and \((x_{1i}^*, \ldots, x_{si}^*, \ldots, x_{LSi}^*) \succ_i (x_{1i}, \ldots, x_{si}, \ldots, x_{LSi})\) for all \( i \), with at least one preference strict. It then follows from the definition of Pareto optimality that the Arrow-Debreu equilibrium allocation \((x_1^*, \ldots, x_n^*) \in \mathbb{R}^{LSn}\) is not Pareto optimal, contradicting the conclusion of the first welfare theorem. In summary at \( t = 0 \) the consumers can trade directly to an overall Pareto optimal allocation; hence there is no reason for further trade to take place. In other words, ex ante pareto optimality implies ex post Pareto optimality and thus no ex post trade.

Matters are different if not all the \( LS \) contingent commodity markets are available at \( t = 0 \). Then the initial trade to a Pareto optimal allocation may not be feasible and it is quite possible that ex post (i.e., after the revelation of the state \( s \)) the resulting consumption allocation is not Pareto optimal. There would then be an incentive to reopen the markets and retrade.

A most interesting possibility, first observed by Arrow (1953), is that, even
if not all the contingent commodities are available at \( t = 0 \), it may still be the case under some conditions that the retrading possibilities at \( t = 1 \) guarantee that Pareto optimality is reached, nevertheless. That is, the possibility of ex post trade can make up for an absence of some ex ante markets. In what follows, we shall verify that this is the case whenever at least one physical commodity can be traded contingently at \( t = 0 \) if, in addition, spot markets occur at \( t = 1 \) and the spot equilibrium prices are correctly anticipated at \( t = 0 \). The intuition for this result is reasonably straightforward: if spot trade can occur within each state, then the only task remaining at \( t = 0 \) is to transfer the consumer’s overall purchasing power efficiently across states. This can be accomplished using contingent trade in a single commodity. By such a procedure we are able to reduce the number of required forward markets for \( LS \) to \( S \).

Let us be more specific. At \( t = 0 \) consumers have *expectations* regarding the spot prices prevailing at \( t = 1 \) for each possible state \( s \in S \). Denote the price vector expected to prevail in state \( s \) spot market by \( p_S \in \mathbb{R}^L \); and the overall expectation vector by \( p = (p_1, \ldots, p_S) \in \mathbb{R}^{LS} \). Suppose that, in addition, at date \( t = 0 \) there is trade in the \( S \) contingent commodities denoted by 11 to 1S; that is, there is contingent trade only in the physical good with the label 1. We denote the vector of prices for these contingent commodities traded at \( t = 0 \) by \( q = (q_1, \ldots, q_S) \in \mathbb{R}^S \).

Faced with prices \( q \in \mathbb{R}^S \) at \( t = 0 \) and expected spot prices \( (p_1, \ldots, p_S) \in \mathbb{R}^{LS} \) at \( t = 1 \), every consumer \( i \) formulates a consumption, or trading, plan \((z_{1i}, \ldots, z_{si}) \in \mathbb{R}^S\) for contingent commodities at \( t = 0 \), as well as a set of spot market consumption plans \((x_{1i}, \ldots, x_{si}) \in \mathbb{R}^{LS}\) for the different states that may occur at \( t = 1 \). Of course, these plans must satisfy a budget constraint. Let \( U_i(\cdot) \) be a utility function for \( \succsim_i \). Then the problem of consumer \( i \) can be expressed...
formally as

\[
\begin{align*}
\max_{(x_{1i}, \ldots, x_{si}) \in \mathbb{R}^{LS}, (z_{1i}, \ldots, z_{si}) \in \mathbb{R}^S} & \quad U_i(x_{1i}, \ldots, x_{si}) \\
\text{s.t.} & \quad (i) \sum_s q_s z_{si} \leq 0, \\
& \quad (ii) p_s \cdot x_{si} \leq p_s \cdot w_{si} + p_{1s} z_{si} \quad \text{for every } s.
\end{align*}
\]

Restriction (i) is the budget constraint corresponding to trade at \( t = 0 \). The family of restrictions (ii) are the budget constraints for the different spot markets. Note that the value of wealth at a state \( s \) is composed of two parts: the market value of the initial endowments, \( p_s \cdot w_{si} \), and the market value of the amounts \( z_{si} \) of good 1 bought or sold forward at \( t = 0 \). Observe that we are not imposing any restriction on the sign or the magnitude of \( z_{si} \). If \( z_{si}, < -w_{1si} \) then one says that at \( t = 0 \) consumer \( i \) is selling good 1 short. This is because he is selling at \( t = 0 \), contingent on state \( s \) occurring, more than he has at \( t = 1 \) if \( s \) occurs. Hence, if \( s \) occurs he will actually have to buy in the spot market the extra amount of the first good required for the fulfillment of his commitments. The possibility of selling short is, however, indirectly limited by the fact that consumption, and therefore ex post wealth, must be nonnegative for every \( s \).

To define an appropriate notion of sequential trade we shall impose a key condition: Consumers’ expectations must be self-fulfilled, or rational; that is, we require that consumers’ expectations of the prices that will clear the spot markets for the different states \( s \) do actually clear them once date \( t = 1 \) has arrived and a state \( s \) is revealed.

**Definition 13.4.1** A collection formed by a price vector \( q = (q_1, \ldots, q_S) \in \mathbb{R}^S \)

\(^1\)Observe also that we have taken the wealth at \( t = 0 \) to be zero (that is, there are no initial endowments of the contingent commodities). This is simply a convention. Suppose, for example, that we regard \( w_{1si} \), the amount of good 1 available at \( t = 1 \) in state \( s \), as the amount of the \( s \) contingent commodity that \( i \) owns at \( t = 0 \) (to avoid double counting, the initial endowment of commodity 1 in the spot market \( s \) at \( z = 1 \) should simultaneously be put to zero). The budget constraints are then: (i) \( \sum_s q_s (z'_{si} - w_{1si}) \leq 0 \) and (ii) \( p_s \cdot x_{si} \leq \sum_{l \neq 1} p_{ls} w_{l1} + p_{1s} z'_{si} \), for every \( s \). But letting \( z'_{si} = z_{si} + w_{1si} \), we see that these are exactly the constraints of (13.1).
for contingent first good commodities at \( t = 0 \), a spot price vector 

\[
p_s = (p_{1s}, \ldots, p_{LS}) \in \mathbb{R}^L
\]

for every \( s \), and, for every consumer \( i \), consumption plans \( z^*_i = (z^*_{1i}, \ldots, z^*_{si}) \in \mathbb{R}^S \) at \( t = 0 \) and \( x^*_i = (x^*_{1i}, \ldots, x^*_{si}) \in \mathbb{R}^{LS} \) at \( t = 1 \) constitutes a Radner equilibrium [see Radner (1982)] if:

(i) For every \( i \), the consumption plans \( z^*_i, x^*_i \) solve problem (13.1).

(ii) \( \sum_i z^*_{si} \leq 0 \) and \( \sum_i x^*_i \leq \sum_i w_{si} \) for every \( s \).

At a Radner equilibrium, trade takes place through time and, in contrast to the Arrow-Debreu setting, economic agents face a sequence of budget sets, one at each date-state (more generally, at every date-event).

We can see from an examination of problem (13.1) that all the budget constraints are homogeneous of degree zero with respect to prices. It is natural to choose the first commodity and to put \( p_{1s} = 1 \) for every \( s \), so that a unit of the \( s \) contingent commodity then pays off 1 dollar in state \( s \). Note that this still leaves one degree of freedom, that corresponding to the forward trades at date 0 (so we could put \( q_1 = 1 \), or perhaps \( \sum_s q_s = 1 \)).

In the following proposition, which is the key result of this section, we show that for this model the set of Arrow-Debreu equilibrium allocations (induced by the arrangement of one-shot trade in \( LS \) contingent commodities) and the set of Radner equilibrium allocations (induced by contingent trade in only one commodity, sequentially followed by spot trade) are identical.

**Proposition 13.4.1** We have:

(i) If the allocation \( x^* \in \mathbb{R}^{LSu} \) and the contingent commodities price vector \( (p_1, \ldots, p_{LS}) \in \mathbb{R}^{LS}_{++} \) constitute an Arrow-Debreu equilibrium, then there are prices \( q \in \mathbb{R}^S_{++} \) for contingent first good commodities and consumption plans for these commodities \( z^* = (z^*_1, \ldots, z^*_u) \in \mathbb{R}^u \).
\( \mathbb{R}^S \) such that the consumptions plans \( x^*, z^* \), the prices \( q \), and the spot prices \((p_1, \ldots, p_S)\) constitute a Radner equilibrium.

(ii) Conversely, if the consumption plans \( x^* \in \mathbb{R}^{LS}, z^* \in \mathbb{R}^S \) and prices \( q \in \mathbb{R}^{S_+}, (p_1, \ldots, p_S) \in \mathbb{R}^{LS}_+ \) constitute a Radner equilibrium, then there are multipliers \((\mu_1, \ldots, \mu_S) \in \mathbb{R}^{S_+} \) such that the allocation \( x^* \) and the contingent commodities price vector \((\mu_1 p_1, \ldots, \mu_S p_S) \in \mathbb{R}^{LS}_+ \) constitute an Arrow-Debreu equilibrium. (The multiplier \( \mu_s \) is interpreted as the value, at \( t = 0 \), of a dollar at \( t = 1 \) and state \( s \)).

**Proof:**

i. It is natural to let \( q_s = p_{1s} \) for every \( s \). With this we claim that, for every consumer \( i \), the budget set of the Arrow-Debreu problem,

\[
B_i^{AD} = \{ (x_{1i}, \ldots, x_{si}) \in \mathbb{R}^{LS}_+ : \sum_s p_s (x_{si} - w_{si}) \leq 0 \},
\]

is identical to the budget set of the Radner problem,

\[
B_i^{AD} = \{ (x_{1i}, \ldots, x_{si}) \in \mathbb{R}^{LS}_+ : \text{there are } (z_{1i}, \ldots, z_{si}) \text{ such that } \sum_s p_s z_{si} \leq 0 \text{ and } p_s (x_{si} - w_{si}) \leq p_{1s} z_{si} \text{ for every } s \}.
\]

To see this, suppose that \( x_i = (x_{1i}, \ldots, x_{si}) \in B_i^{AD} \). For every \( s \), denote \( z_{si} = (1/p_{1s}) p_s (x_{si} - w_{si}) \). Then \( \sum_s p_s z_{si} = \sum_s p_{1s} z_{si} = \sum_s p_s (x_{si} - w_{si}) \leq 0 \) and \( p_s (x_{si} - w_{si}) = p_{1s} z_{si} \) for every \( s \). Hence, \( x_i \in B_i^{R} \). Conversely, suppose that \( x_i = (x_{1i}, \ldots, x_{si}) \in B_i^{R} \); that is, for some \((z_{1i}, \ldots, z_{si})\) we have \( \sum_s q_s z_{si} \leq 0 \) and \( p_s (x_{si} - w_{si}) = p_{1s} z_{si} \), for every \( s \). Summing over \( s \) gives \( \sum_s p_s (x_{si} - w_{si}) \leq \sum_s p_{1s} z_{si} = \sum_s q_s z_{si} \leq 0 \). Hence, \( x_i \in B_i^{AD} \).

We conclude that our Arrow-Debreu equilibrium allocation is also a Radner equilibrium allocation supported by \( q = (p_{11}, \ldots, p_{1S}) \in \mathbb{R}^S \), the spot prices \((p_1, \ldots, p_S)\), and the contingent trades \((z^*_{1i}, \ldots, z^*_{si}) \in \mathbb{R}^S \) defined by \( z^*_{si} = (1/p_{1s}) p_s (x^*_{si} - w_{si}) \). Note that the contingent markets clear since, for every \( s \), \( \sum_i z^*_{si} = (1/p_{1s}) p_s [\sum_i (x^*_{si} - w_{si})] \leq 0 \).
ii. Choose \( \mu_s \) so that \( \mu_s p_{1s} = q_s \). Then we can rewrite the Radner budget set of every consumer \( i \) as \( B_i^R = \{(x_{1i}, \ldots, x_{si}) \in \mathbb{R}^{LS} : \text{there are } (z_{1i}, \ldots, z_{si}) \text{ such that } \sum_s q_s z_{si} \leq 0 \text{ and } \mu_s p_s \cdot (x_{si} - w_{si}) \leq q_s z_{si} \text{ for every } s \} \). But from this we can proceed as we did in part (i) and rewrite the constraints, and therefore the budget set, in the Arrow-Debreu form:

\[
B_i^R = B_i^{AD} = \{(x_{1i}, \ldots, x_{si}) \in \mathbb{R}^{LS} : \sum_s \mu_s p_s \cdot (x_{si} - w_{si}) \leq 0 \}.
\]

Hence, the consumption plan \( x_i^* \) is also preference maximizing in the budget set \( B_i^{AD} \). Since this is true for every consumer \( i \), we conclude that the price vector \( (\mu_1 p_1, \ldots, \mu_S p_S) \in \mathbb{R}^{LS} \) clears the markets for the \( LS \) contingent commodities.

Example 13.4.1 : Consider a two-good, two-state, two-consumer pure exchange economy. Suppose that the two states are equally likely and that every consumer has the same, state-independent, Bernoulli utility function \( u(x_{si}) \). The consumers differ only in their initial endowments. The aggregate endowment vectors in the two states are the same; however, endowments are distributed so that consumer 1 gets everything in state 1 and consumer 2 gets everything in state 2. (See Figure 13.4.)

By the symmetry of the problem, at an Arrow-Debreu equilibrium each consumer gets, in each state, half of the total endowment of each good. In Figure 491
19.D.1, we indicate how these consumptions will be reached by means of contingent trade in the first commodity and spot markets. The spot prices will be the same in the two states. The first consumer will sell an amount $\alpha$ of the first good contingent on the occurrence of the first state and will in exchange buy an amount $\beta$ of the same good contingent on the second state.

**Remark 13.4.1** It is important to emphasize that, although the concept of Radner equilibrium cuts down the number of contingent commodities required to attain optimality (from $LS$ to $S$), this reduction is not obtained free of charge. With the smaller number of forward contracts, the correct anticipation of future spot prices becomes crucial.

### 13.5 Incomplete Markets

In this section we explore the implications of having fewer than $S$ assets, that is, of having an asset structure that is necessarily incomplete. We pursue this point in the two-period framework of the previous sections.

We begin by observing that when $K < S$ a Radner equilibrium need not be Pareto optimal. This is not surprising: if the possibilities of transferring wealth across states are limited, then there may well be a welfare loss due to the inability to diversify risks to the extent that would be convenient. Just consider the extreme case where there are no assets at all. The following example provides another interesting illustration of this type of failure.

**Example 13.5.1** [Sunspots]. Suppose that preferences admit an expected utility representation and that the set of states $S$ is such that, first, the probability estimates for the different states are the same across consumers (i.e., $\pi_{si} = \pi'_{si} = \pi_s$ for all $i, i', \text{ and } s$) and second, that the states do not affect the fundamentals
of the economy; that is, the Bernoulli utility functions and the endowments of every consumer \(i\) are uniform across states [i.e., \(u_{si} (\cdot) = u_i (\cdot)\) and \(w_{si}, = w_i\), for all \(s\)]. Such a set of states is called a \textit{sunspot} set. The question we shall address is whether in these circumstances the Radner equilibrium allocations can assign varying consumptions across states. An equilibrium where this happens is called a \textit{sunspot equilibrium}.

Under the assumption that consumers are strictly risk averse, so that the utility functions \(u_i (\cdot)\) are strictly concave, \textit{any Pareto optimal allocation} \((x_1, \ldots, x_n) \in \mathbb{R}^{LS^n}\) \textit{must be uniform across states} (or \textit{state independent}); that is, for every \(i\) we must have \(x_{1i} = x_{2i} = \cdots = x_{si} = \cdots = x_{si}\). To see this, suppose that, for every \(i\) and \(s\), we replace the consumption bundle of consumer \(i\) in state \(s\), \(x_{si} \in \mathbb{R}^{L}\), by the expected consumption bundle of this consumer: \(\bar{x}_i = \sum_s \pi_s x_{si} \in \mathbb{R}^{L}\). The new allocation is state independent, and it is also feasible because

\[
\sum_i \bar{x}_i = \sum_i \sum_s \pi_s x_{si} = \sum_s \pi_s \left( \sum_i x_{si} \right) \leq \sum_s \pi_s \left( \sum_i w_i \right) = \sum_i w_i.
\]

By the concavity of \(u_i (\cdot)\) it follows that no consumer is worse off:

\[
\sum_s \pi_s u_i (\bar{x}_i) = u_i (\bar{x}_i) = u_i \left( \sum_s \pi_s x_{si} \right) \geq \sum_s \pi_s u_i (x_{si}) \text{ for every } i.
\]

Because of the Pareto optimality of \((x_1, \ldots, x_n)\), the above weak inequalities must in fact be equalities; that is, \(u_i (\bar{x}_i) = \sum_s \pi_s u_i (x_{si})\) for every \(i\). But, if so, then the strict concavity of \(u_i (\cdot)\) yields \(x_{si} = \bar{x}_i\) for every \(s\). In summary: the Pareto optimal allocation \((x_1, \ldots, x_n) \in \mathbb{R}^{LS^n}\) is state independent.

From the state independence of Pareto optimal allocations and the first welfare theorem we reach the important conclusion that if a system of complete markets over the states \(S\) can be organized, then the equilibria are \textit{sunspot free}, that is, consumption is uniform across states. In effect, traders wish to insure completely and have instruments to do so.

It turns out, however, that if there is not a complete set of insurance opportunities, then the above conclusion does not hold true. Sunspot-free, Pareto optimal
equilibria always exist (just make the market “not pay attention” to the sunspot. But it is now possible for the consumption allocation of some Radner equilibria to depend on the state, and consequently to fail the Pareto optimality test. In such an equilibrium consumers expect different prices in different states, and their expectations end up being self-fulfilling. The simplest, and most trivial, example is when there are no assets whatsoever \((K = 0)\). Then a system of spot prices \((p_1, \ldots, p_s) \in \mathbb{R}^{LS} \) is a Radner equilibrium if and only if every \(p_s\) is a Walrasian equilibrium price vector for the spot economy defined by \(\{(u_i(\cdot), w_i)\}_{i=1}^n\). If, as is perfectly possible, this economy admits several distinct Walrasian equilibria, then by selecting different equilibrium price vectors for different states, we obtain a sunspot equilibrium, and hence a Pareto inefficient Radner equilibrium.

We have seen that Radner equilibrium allocations need not be Pareto optimal, and so, in principle, there may exist reallocations of consumption that make all consumers at least as well off, and at least one consumer strictly better off. It is important to recognize, however, that this need not imply that a welfare authority who is “as constrained in interstate transfers as the market is” can achieve a Pareto optimum. An allocation that cannot be Pareto improved by such an authority is called a constrained Pareto optimum. A more significant and reasonable welfare question to ask is, therefore, whether Radner equilibrium allocations are constrained Pareto optimal. We now address this matter. This is a typical instance of a second-best welfare issue.

To proceed with the analysis we need a precise description of the constrained feasible set and of the corresponding notion of constrained Pareto optimality. This is most simply done in the context where there is a single commodity per state, that is, \(L = 1\). The important implication of this assumption is that then the amount of consumption good that any consumer \(i\) gets in the different states is entirely determined by the portfolio \(z_i\). Indeed, \(x_{si} = \sum_k z_k r_{sk} + w_{si}\). Hence, we can let

\[
U^*_i(z_i) = U^*_i(z_{1i}, \ldots, z_{Ki}) = U_i\left(\sum_k z_k r_{1k} + w_{1i}, \ldots, \sum_k z_k r_{sk} + w_{si}\right)
\]
denote the utility induced by the portfolio $z_i$. The definition of constrained Pareto optimality is then quite natural.

**Definition 13.5.1** The asset allocation $(z_1, \ldots, z_n) \in \mathbb{R}^{Kn}$ is constrained Pareto optimal if it is feasible (i.e., $\sum_i z_i \leq 0$) and if there is no other feasible asset allocation $(z'_1, \ldots, z'_n) \in \mathbb{R}^{Kn}$ such that

$$U^*_i(z'_1, \ldots, z'_n) \geq U^*_i(z_1, \ldots, z_n)$$

for every $i$, with at least one inequality strict.

In this $L = 1$ context the utility maximization problem of consumer $i$ becomes

$$\max_{z_i \in \mathbb{R}^K} U^*_i(z_{1i}, \ldots, z_{Ki})$$

$$\text{s.t. } q \cdot z_i \leq 0.$$  

Suppose that $z_i \in \mathbb{R}^K$ for $i = 1, \ldots, n$, is a family of solutions to these individual problems, for the asset price vector $q \in \mathbb{R}^K$. Then $q \in \mathbb{R}^K$ is a Radner equilibrium price if and only if $\sum_i z^*_i \leq 0$.

To it we can apply the first welfare theorem and reach the conclusion of the following proposition.

**Proposition 13.5.1** Suppose that there two periods and only one consumption good in the second period. Then any Radner equilibrium is constrained Pareto optimal in the sense that there is no possible redistribution of assets in the first period that leaves every consumer as well off and at least one consumer strictly better off.

The situation considered in Proposition 13.5.1 is very particular in that once the initial asset portfolio of a consumer is determined, his overall consumption is fully determined: with only one consumption good, there are no possibilities for trade once the state occurs. In particular, second-period relative prices do

---

2Recall that, given $z_i$ the consumptions in every state are determined. Also, the price of consumption good in every state is formally fixed to be 1.
not matter, simply because there are no such prices. Things change if there is more than one consumption good in the second period, or if there are more than two periods. such an example can be found in Mas-Colell, Whinston, and Green (1995, p. 711).

13.6 Reference

Books and Monographs:


Papers:


Part V

Externalities and Public Goods
In Chapters 10 and 11, we have introduced the notions of competitive equilibrium and Pareto optimality, respectively. The concept of competitive equilibrium provides us with an appropriate notion of market equilibrium for competitive market economies. The concept of Pareto optimality offers a minimal and uncontroversial test that any social optimal economic outcome should pass since it is a formulation of the idea that there is no further improvement in society, and it conveniently separates the issue of economic efficiency from more controversial (and political) questions regarding the ideal distribution of well-being across individuals.

The important results and insights obtained in Chapters 10 and 11 are the First and Second Fundamental Theorems of Welfare Economics. The first welfare theorem provides a set of conditions under which a market economy will achieve a Pareto optimal; it is, in a sense, the formal expression of Adam Smith’s claim about the “invisible hand” of the market. The second welfare theorem goes even further. It states that under the same set of conditions as the first welfare theorem plus convexity and continuity conditions, all Pareto optimal outcomes can in principle be implemented through the market mechanism by appropriately redistributing wealth and then “letting the market work.”

Thus, in an important sense, the general equilibrium theory establishes the perfectly competitive case as a benchmark for thinking about outcomes in market economies. In particular, any inefficiency that arise in a market economy, and hence any role for Pareto-improving market intervention, must be traceable to a violation of at least one of these assumptions of the first welfare theorem. The remainder of the notes, can be viewed as a development of this theme, and will study a number of ways in which actual markets may depart from this perfectly competitive ideal and where, as a result, market equilibria fail to be Pareto optimal, a situation known market failure.

In the current part, we will study externalities and public goods in Chapter 14 and Chapter 15, respectively. In both cases, the actions of one agent directly affect the utility or production of other agents in the economy. We will see
these nonmarketed “goods” or “bads” lead to a non-Pareto optimal outcome in general; thus a market failure. It turns out that private markets are often not a very good mechanism in the presence of externalities and public goods. We will consider situations of incomplete information which also result in non-Pareto optimal outcomes in general in Part VI.

It is greatly acknowledged that the materials are largely drawn from the references in the end of each charter, especially from Varian (1992).
Chapter 14

Externalities

14.1 Introduction

In this chapter we deal with the case of externalities so that a market equilibrium may lead to non-Pareto efficient allocations in general, and thus there is a market failure. The reason is that there are things that people care about are not priced. Externality can happen in both cases of consumption and production.

*Consumption Externality:*

\[ u_i(x_i) : \text{without preference externality} \]
\[ u_i(x_1, ..., x_n) : \text{with preference externality} \]

in which other individuals' consumptions effect an individual’s utility.

**Example 14.1.1**

(i) One person’s quiet environment is disturbed by another person’s local stereo.

(ii) Mr. A hates Mr. D smoking next to him.

(ii) Mr. A’s satisfaction decreases as Mr. C’s consumption level increases, because Mr. A envies Mr. C’s lifestyle.

*Production Externality:*

502
A firm’s production includes arguments other than its own inputs.

For example, downstream fishing is adversely affected by pollutants emitted from an upstream chemical plant.

This leads to an examination of various suggestions for alternative ways to allocate resources that may lead to efficient outcomes. Achieving an efficient allocation in the presence of externalities essentially involves making sure that agents face the correct prices for their actions. Ways of solving externality problem include taxation, regulation, property rights, merges, etc.

14.2 Consumption Externalities

When there are no consumption externalities, agent $i$’s utility function is a function of only his own consumption:

$$u_i(x_i)$$ (14.1)

In this case, the first order conditions for the competitive equilibrium are given by

$$MRS_{xy}^A = \frac{p_x}{p_y} = MRS_{xy}^B$$

and the first order conditions for Pareto efficiency are given by:

$$MRS_{xy}^A = MRS_{xy}^B.$$

Thus, because of price-taking behavior, every competitive equilibrium implies Pareto efficiency if utility functions are quasi-concave.

The main purpose of this section is to show that a competitive equilibrium allocation is not in general Pareto efficient when there exists an externality in consumption. We show this by examining that the first order conditions for a competitive equilibrium is not in general the same as the first order conditions for Pareto efficient allocations in the presence of consumption externalities. The following materials are mainly absorbed from Tian and Yang (2009).
Consider the following simple two-person and two-good exchange economy.

\[ u_A(x_A, x_B, y_A) \] (14.2)

\[ u_B(x_A, x_B, y_B) \] (14.3)

which are assumed to be strictly increasing in his own goods consumption, quasi-concave, and satisfies the Inada condition \( \frac{\partial u_i}{\partial x_i}(0) = +\infty \) and \( \lim_{x_i \to 0} \frac{\partial u_i}{\partial x_i} x_i = 0 \) so it results in interior solutions. We further assume that the gradient of \( u_i(\cdot) \) is nonzero at Pareto efficient allocations. Here good \( x \) results in consumption externalities.

The first order conditions for the competitive equilibrium are the same as before:

\[ MRS^A_{xy} = \frac{p_x}{p_y} = MRS^B_{xy}. \]

We now find the first order conditions for Pareto efficient allocations in exchange economies with externalities. Thus Pareto efficient allocations \( x^* \) can be completely determined by the FOCs of the following problem.

\[
\begin{aligned}
\text{Max}_{x \in \mathbb{R}^4_+} & \quad u_B(x_A, x_B, y_B) \\
\text{s.t.} & \quad x_A + x_B \leq w_x \\
& \quad y_A + y_B \leq w_y \\
& \quad u_A(x_A, x_B, y_A) \geq u_A(x^*_A, x^*_B, y^*_A) \\
\end{aligned}
\]

The first order conditions are

\[
\begin{aligned}
x_A : & \quad \frac{\partial u_B}{\partial x_A} - \lambda_x + \mu \frac{\partial u_A}{\partial x_A} = 0 \quad (14.4) \\
y_A : & \quad -\lambda_y + \mu \frac{\partial u_A}{\partial y_A} = 0 \quad (14.5) \\
x_B : & \quad \frac{\partial u_B}{\partial x_B} - \lambda_x + \mu \frac{\partial u_A}{\partial x_B} = 0 \quad (14.6) \\
y_B : & \quad \frac{\partial u_B}{\partial y_B} - \lambda_y = 0 \quad (14.7) \\
\lambda_x : & \quad w_x - x_A - x_B \geq 0, \lambda_x \geq 0, \lambda_x (w_x - x_A - x_B) = 0 \quad (14.8) \\
\lambda_y : & \quad w_y - y_A - y_B \geq 0, \lambda_y \geq 0, \lambda_y (w_y - y_A - y_B) = 0 \quad (14.9) \\
\mu : & \quad u_A - u^*_A \geq 0, \mu \geq 0, \mu (u_A - u^*_A) = 0 \quad (14.10)
\end{aligned}
\]
By (14.7), $\lambda_y = \frac{\partial u_B}{\partial y_B} > 0$, and thus by (14.9),

$$y_A + y_B = w_y$$  \hspace{1cm} (14.11)

which means there is never destruction of the good which does not exhibit a negative externality. Also, by (14.5) and (14.7), we have

$$\mu = \frac{\partial u_B}{\partial y_B} \frac{\partial u_A}{\partial y_A}.$$

Then, by (14.4) and (14.5), we have

$$\frac{\lambda_x}{\lambda_y} = \left[ \frac{\partial u_A}{\partial x_A} + \frac{\partial u_B}{\partial x_A} \frac{\partial u_A}{\partial y_A} \right]$$  \hspace{1cm} (14.13)

and by (14.6) and (14.7), we have

$$\frac{\lambda_x}{\lambda_y} = \left[ \frac{\partial u_B}{\partial x_B} + \frac{\partial u_A}{\partial x_B} \frac{\partial u_B}{\partial y_B} \right]$$  \hspace{1cm} (14.14)

Thus, by (14.13) and (14.14), we have

$$\frac{\partial u_A}{\partial y_A} + \frac{\partial u_B}{\partial y_B} = \frac{\partial u_B}{\partial y_B} + \frac{\partial u_A}{\partial y_A},$$

which expresses the equality of the social marginal rates of substitution for the two consumers at Pareto efficient points. Thus, we immediately have the following conclusion:

A competitive equilibrium allocations may not be Pareto optimal because the first order conditions for competitive equilibrium and Pareto optimality are not the same.

From the above marginal equality condition, we know that, in order to evaluate the relevant marginal rates of substitution for the optimality conditions, we must take into account both the direct and indirect effects of consumption activities in the presence of externalities. That is, to achieve Pareto optimality, when one consumer increases the consumption of good $x$, not only does the consumer’s...
consumption of good \( y \) need to change, the other consumer’s consumption of good \( y \) must also be changed. Therefore the social marginal rate of substitution of good \( x \) for good \( y \) by consumer \( i \) equals \( \frac{\partial u_i}{\partial x_i} + \frac{\partial u_j}{\partial y_j} \).

Solving (14.4) and (14.6) for \( \mu \) and \( \lambda_x \), we have

\[
\mu = \frac{\partial u_B}{\partial x_B} - \frac{\partial u_A}{\partial x_A} > 0 \quad (14.16)
\]

and

\[
\lambda_x = \frac{\partial u_A}{\partial x_A} \frac{\partial u_B}{\partial x_B} - \frac{\partial u_A}{\partial x_A} \frac{\partial u_B}{\partial x_B}.
\quad (14.17)
\]

When the consumption externality is positive, from (14.13) or (14.14), we can easily see that \( \lambda_x \) is always positive since \( \lambda_y = \frac{\partial u_A}{\partial y_B} > 0 \). Also, when no externality or a one-sided externality\(^1\) exists, by either (14.13) or (14.14), \( \lambda_x \) is positive. Thus, the marginal equality condition (14.15) and the balanced conditions, completely determine all Pareto efficient allocations for these cases. However, when there are negative externalities for both consumers, the Kuhn-Tucker multiplier \( \lambda_x \) directly given by (14.17) or indirectly given by (14.13) or (14.14) is the sum of a negative and positive term, and thus the sign of \( \lambda_x \) may be indeterminate. However, unlike the claim in some textbooks such as Varian (1992, 438), the marginal equality condition, (14.15), and the balanced conditions may not guarantee finding Pareto efficient allocations correctly.

To guarantee that an allocation is Pareto efficient in the presence of negative externalities, we must require \( \lambda_x \geq 0 \) at efficient points, which in turn requires that social marginal rates of substitution be nonnegative, that is,

\[
\frac{\partial u_A}{\partial x_A} + \frac{\partial u_B}{\partial x_B} = \frac{\partial u_A}{\partial y_B} + \frac{\partial u_B}{\partial y_A} \geq 0, \quad (14.18)
\]

or equivalently requires both (14.15) and

\[
\frac{\partial u_A}{\partial x_A} \frac{\partial u_B}{\partial x_B} \geq \frac{\partial u_A}{\partial y_B} \frac{\partial u_B}{\partial x_A} \quad (14.19)
\]

\( \lambda_x \)

\(^1\)Only one consumer imposes an externality on another consumer.
for all Pareto efficient points.

We can interpret the term in the left-hand side of (14.19), $\frac{\partial u_A}{\partial x_A} \frac{\partial u_B}{\partial x_B}$, as the joint marginal benefit of consuming good $x$, and the term in the right-hand side, $\frac{\partial u_A}{\partial x_B} \frac{\partial u_B}{\partial x_A}$, as the joint marginal cost of consuming good $x$ because the negative externality hurts the consumers. To consume the goods efficiently, a necessary condition is that the joint marginal benefit of consuming good $x$ should not be less than the joint cost of consuming good $x$.

Thus, the following conditions

$$\begin{cases} \frac{\partial u_A}{\partial x_A} + \frac{\partial u_B}{\partial x_B} = \frac{\partial u_B}{\partial x_B} + \frac{\partial u_A}{\partial x_A} \geq 0 \\ y_A + y_B = w_y \\ x_A + x_B \leq w_x \\ \left(\frac{\partial u_A}{\partial x_A} \frac{\partial u_B}{\partial x_B} - \frac{\partial u_A}{\partial x_B} \frac{\partial u_B}{\partial x_A}\right) (w_x - x_A - x_B) = 0 \end{cases}$$

constitute a system (PO) from which all Pareto efficient allocations can be obtained. We can do so by considering three cases.

Case 1. When $\frac{\partial u_A}{\partial x_A} \frac{\partial u_B}{\partial x_B} > \frac{\partial u_A}{\partial x_B} \frac{\partial u_B}{\partial x_A}$, or equivalently $\frac{\partial u_A}{\partial x_A} + \frac{\partial u_B}{\partial x_A} = \frac{\partial u_B}{\partial x_B} + \frac{\partial u_A}{\partial x_B} > 0$, $\lambda_x > 0$ and thus the last two conditions in the above system (PO) reduce to $x_A + x_B = w_x$. In this case, there is no destruction. Substituting $x_A + x_B = w_x$ and $y_A + y_B = w_y$ into the marginal equality condition (14.15), it would give us a relationship between $x_A$ and $y_A$, which exactly defines the Pareto efficient allocations.

Case 2. When the joint marginal benefit equals the joint marginal cost:

$$\frac{\partial u_A}{\partial x_A} \frac{\partial u_B}{\partial x_B} = \frac{\partial u_A}{\partial x_B} \frac{\partial u_B}{\partial x_A};$$

then

$$\frac{\partial u_A}{\partial x_A} + \frac{\partial u_B}{\partial x_A} = \frac{\partial u_B}{\partial x_B} + \frac{\partial u_A}{\partial x_B} = 0$$

and thus $\lambda_x = 0$. In this case, when $x_A + x_B \leq w_x$, the necessity of destruction is indeterminant. However, even when destruction is necessary, we can still
determine the set of Pareto efficient allocations by using \( y_A + y_B = w_y \) and the zero social marginal equality conditions (14.21). Indeed, after substituting \( y_A + y_B = w_y \) into (14.21), we can solve for \( x_A \) in terms \( y_A \).

Case 3. When \( \frac{\partial u_A}{\partial x_A} \frac{\partial u_B}{\partial x_B} < \frac{\partial u_A}{\partial x_B} \frac{\partial u_B}{\partial x_A} \) for any allocations that satisfy \( x_A + x_B = w_x \), \( y_A + y_B = w_y \), and the marginal equality condition (14.15), the social marginal rates of substitution must be negative. Hence, the allocation will not be Pareto efficient. In this case, there must be a destruction for good \( x \) for Pareto efficiency, and a Pareto efficient allocation satisfies (14.21).

Summarizing, we have the following proposition that provides two categories of sufficiency conditions for characterizing whether or not there should be destruction of endowment \( w_x \) in achieving Pareto efficient allocations.

**Proposition 14.2.1** For \( 2 \times 2 \) pure exchange economies, suppose that utility functions \( u_i(x_A, x_B, y_i) \) are continuously differentiable, strictly quasi-concave, and \( \frac{\partial u_i(x_A, x_B, y_i)}{\partial x_i} > 0 \) for \( i = A, B \).

1. If the social marginal rates of substitution are positive at a Pareto efficient allocation \( x^* \), then there is no destruction of \( w_x \) in achieving Pareto efficient allocation \( x^* \).

2. If the social marginal rates of substitution are negative for any allocation \( (x_A, x_B) \) satisfying \( x_A + x_B = w_x \), \( y_A + y_B = w_y \), and the marginal equality condition (14.15), then there is destruction of \( w_x \) in achieving any Pareto efficient allocation \( x^* \). That is, \( x^*_A + x^*_B < w_x \) and \( x^* \) is determined by \( y_A + y_B = w_y \) and (14.21).

Thus, from the above proposition, we know that a sufficient condition for destruction is that for any allocation \( (x_A, y_A, x_B, y_B) \),

\[
\begin{align*}
  \frac{\partial u_A}{\partial x_A} + \frac{\partial u_B}{\partial x_A} &= \frac{\partial u_B}{\partial x_B} + \frac{\partial u_A}{\partial x_B} \\
  x_A + x_B &= w_x \\
  y_A + y_B &= w_y
\end{align*}
\]

\[
\Rightarrow \frac{\partial u_A}{\partial x_A} \frac{\partial u_B}{\partial x_B} < \frac{\partial u_A}{\partial x_B} \frac{\partial u_B}{\partial x_A}.
\]

\[\text{(14.22)}\]

\(^2\)As we discussed above, this is true if the consumption externality is positive, or there is no externality or only one side externality.
A sufficient condition for all Pareto efficient allocations \((x_A, y_A, x_B, y_B)\) for no destruction is,
\[
\begin{aligned}
\frac{\partial u_A}{\partial x_A} \frac{\partial u_B}{\partial x_B} + \frac{\partial u_B}{\partial y_A} \frac{\partial u_A}{\partial y_B} \\
x_A + x_B \leq w_x \quad y_A + y_B = w_y
\end{aligned}
\] 
\Rightarrow \frac{\partial u_A}{\partial x_A} \frac{\partial u_B}{\partial x_B} > \frac{\partial u_A}{\partial x_B} \frac{\partial u_B}{\partial x_A} \tag{14.23}
\]

**Example 14.2.1** Consider the following utility function:
\[
u_i(x_A, x_B, y_i) = \sqrt{x_i y_i} - x_j, \quad i \in \{A, B\}, j \in \{A, B\}, j \neq i
\]

By the marginal equality condition (14.15), we have
\[
\left(\frac{\sqrt{y_A}}{x_A} + 1\right)^2 = \left(\frac{\sqrt{y_B}}{x_B} + 1\right)^2 \tag{14.24}
\]
and thus
\[
\frac{y_A}{x_A} = \frac{y_B}{x_B}. \quad \tag{14.25}
\]
Let \(x_A + x_B \equiv \bar{x}\). Substituting \(x_A + x_B = \bar{x}\) and \(y_A + y_B = w_y\) into (14.25), we have
\[
\frac{y_A}{x_A} = \frac{w_y}{\bar{x}}. \quad \tag{14.26}
\]
Then, by (14.25) and (14.26), we have
\[
\frac{\partial u_A}{\partial x_A} \frac{\partial u_B}{\partial x_B} = \frac{1}{4} \frac{y_A}{x_A} \sqrt{\frac{y_B}{x_B}} = \frac{y_A}{x_A} \frac{w_y}{4\bar{x}} \tag{14.27}
\]
and
\[
\frac{\partial u_A}{\partial x_B} \frac{\partial u_B}{\partial x_A} = 1. \quad \tag{14.28}
\]
Thus, \(\bar{x} = w_y/4\) is the critical point that makes \(\frac{\partial u_A}{\partial x_A} \frac{\partial u_B}{\partial x_B} - \frac{\partial u_A}{\partial x_B} \frac{\partial u_B}{\partial x_A} = 0\), or equivalently \(\frac{\partial u_A}{\partial x_A} \frac{\partial u_B}{\partial x_B} = \frac{\partial u_A}{\partial x_B} \frac{\partial u_B}{\partial x_A} = 0\). Hence, if \(w_x > \frac{w_y}{4}\), then \(\frac{\partial u_A}{\partial x_A} \frac{\partial u_B}{\partial x_B} - \frac{\partial u_A}{\partial x_B} \frac{\partial u_B}{\partial x_A} < 0\), and thus, by (14.22), there is destruction in any Pareto efficient allocation. If \(w_x < \frac{w_y}{4}\), then \(\frac{\partial u_A}{\partial x_A} \frac{\partial u_B}{\partial x_B} - \frac{\partial u_A}{\partial x_B} \frac{\partial u_B}{\partial x_A} > 0\), and, by (14.23), no Pareto optimal allocation requires destruction. Finally, when \(w_x = \frac{w_y}{4}\), any allocation that satisfies the marginal equality condition (14.15) and the balanced conditions \(x_A + x_B = w_x\) and \(y_A + y_B = w_y\) also satisfies (14.19), and thus it is a Pareto efficient allocation with no destruction.
Now, let us set out the sufficient conditions for destruction and non-destruction in detail. Note that if the sufficient condition for non-destruction holds, then \( x_A + x_B = w_x \) would also hold as an implication. Thus, to apply both the sufficient conditions for non-destruction and destruction, one should use the following three conditions

\[
\begin{aligned}
\frac{\partial u_A}{\partial x_A} + \frac{\partial u_B}{\partial x_B} &= \frac{\partial u_B}{\partial y_B} + \frac{\partial u_A}{\partial y_A} \\
 x_A + x_B &= w_x \\
y_A + y_B &= w_y
\end{aligned}
\]

If an allocation also satisfies the condition \( \frac{\partial u_A}{\partial x_A} \frac{\partial u_B}{\partial x_B} \geq \frac{\partial u_A}{\partial x_B} \frac{\partial u_B}{\partial x_A} \), then the allocation is Pareto efficient, and further there is no destruction. Otherwise, it is not Pareto efficient. If this is true for all such non-Pareto allocations, all Pareto efficient condition must hold with destruction.

Note that, since \( \frac{\partial u_A}{\partial x_A} \) and \( \frac{\partial u_B}{\partial x_B} \) represent marginal benefit, they are usually diminishing in consumption in good \( x \). Since \( \frac{\partial u_A}{\partial x_B} \) and \( \frac{\partial u_B}{\partial x_A} \) are in the form of a marginal cost, their absolute values would be typically increasing in the consumption of good \( x \). Hence, when total endowment \( w_x \) is small, the social marginal benefit would exceed the social marginal cost so that there is no destruction. As the total endowment of \( w_x \) increases, the social marginal cost will ultimately outweigh the marginal social benefit, which results in the destruction of the endowment of \( w_x \).

Alternatively, we can get the same result by using social marginal rates of substitution. When utility functions are strictly quasi-concave, marginal rates of substitution are diminishing. Therefore, in the presence of negative consumption externalities, social marginal rates of substitution may become negative when the consumption of good \( x \) becomes sufficiently large. When this occurs, it is better to destroy some resources for good \( x \). As the destruction of good \( x \) increases, which will, in turn, decrease the consumption of good \( x \), social marginal rates of substitution will increase. Eventually they will become nonnegative.

The destruction issue is not only important in theory, but also relevant to
reality. It can be used to explain a well-known puzzle of the happiness-income relationship in the economic and psychology literatures: happiness rises with income up to a point, but not beyond it. For example, well-being has declined over the last quarter of a century in the U.S., and life satisfaction has run approximately flat across the same time in Britain. If we interpret income as a good, and if consumers envy each other in terms of consumption levels, by our result, when the income reaches a certain level, one may have to freely dispose of wealth to achieve Pareto efficient allocations; otherwise the resulting allocations will be Pareto inefficient. Therefore, economic growth does not raise well-being indexed by any social welfare function once the critical income level is achieved. For detailed discussion, see Tian and Yang (2007, 2009).

14.3 Production Externality

We now show that allocation of resources may not be efficient also for the case of externality in production. To show this, consider a simple economy with two firms. Firm 1 produces an output $x$ which will be sold in a competitive market. However, production of $x$ imposes an externality cost denoted by $e(x)$ to firm 2, which is assumed to be convex and strictly increasing.

Let $y$ be the output produced by firm 2, which is sold in competitive market.

Let $c_x(x)$ and $c_y(y)$ be the cost functions of firms 1 and 2 which are both convex and strictly increasing.

The profits of the two firms:

$$
\pi_1 = p_x x - c_x(x) \quad (14.29)
$$

$$
\pi_2 = p_y y - c_y(y) - e(x) \quad (14.30)
$$

where $p_x$ and $p_y$ are the prices of $x$ and $y$, respectively. Then, by the first order conditions, we have for positive amounts of outputs:

$$
p_x = c'_x(x) \quad (14.31)
$$

$$
p_y = c'_y(y) \quad (14.32)
$$
However, the profit maximizing output $x_c$ from the first order condition is too large from a social point of view. The first firm only takes account of the private cost – the cost that is imposed on itself– but it ignores the social cost – the private cost plus the cost it imposes on the other firm.

What’s the social efficient output?

The social profit, $\pi_1 + \pi_2$, is not maximized at $x_c$ and $y_c$ which satisfy (14.31) and (14.32). If the two firms merged so as to internalize the externality

$$\max_{x,y} p_x x + p_y y - c_x(x) - e(x) - c_y(y)$$  \hfill (14.33)

which gives the first order conditions:

$$p_x = c'_x(x^*) + e'(x^*)$$
$$p_y = c'_y(y^*)$$

where $x^*$ is an efficient amount of output; it is characterized by price being equal to the marginal social cost. Thus, production of $x^*$ is less than the competitive output in the externality case by the convexity of $e(x)$ and $c_x(x)$.

Figure 14.1: The efficient output $x^*$ is less than the competitive output $x_c$. 

512
14.4 Solutions to Externalities

From the above discussion, we know that a competitive market in general may not result in Pareto efficient outcome in the presence of externalities, and one needs to seek some other alternative mechanisms to solve the market failure problem. In this section, we now introduce some remedies to this market failure of externality such as:

1. Pigovian taxes
2. Voluntary negotiation (Coase Approach)
3. Compensatory tax/subsidy
4. Creating a missing markets with property rights
5. Direct intervention
6. Merges
7. Incentives mechanism design

Any of the above solution may result in Pareto efficient outcomes, but may lead to different income distributions. Also, it is important to know what kind of information are required to implement a solution listed above.

Most of the above proposed solutions need to make the following assumptions:

1. The source and degree of the externality is identifiable.
2. The recipients of the externality is identifiable.
3. The causal relationship of the externality can be established objectively.
4. The cost of preventing (by different methods) an externality are perfectly known to everyone.
5. The cost of implementing taxes and subsides is negligible.
6. The cost of voluntary negotiation is negligible.
14.4.1 Pigovian Tax

Set a tax rate, \( t \), such that \( t = e'(x^*) \). This tax rate to firm 1 would internalize the externality.

\[
\pi_1 = p_x \cdot x - c_x(x) - t \cdot x
\]  

(14.34)

The first order condition is:

\[
p_x = c'_x(x) + t = c'_x(x) + e'(x^*),
\]  

(14.35)

which is the same as the one for social optimality. That is, when firm 1 faces the wrong price of its action, and a tax \( t = e'(x^*) \) should be imposed for each unit of firm 1’s production. This will lead to a social optimal outcome that is less than that of competitive equilibrium outcome. Such correction taxes are called Pigovian taxes.

The problem with this solution is that it requires that the taxing authority knows the externality cost \( e(x) \). But, how does the authority know the externality and how do they estimate the value of externality in real world? If the authority knows this information, it would work such as imposing a Pigovian tax on gasoline since emission of automobile can be relatively determined. But, in most cases, it will not work well.

14.4.2 Coase’ Voluntary Negotiation Property Rights Approach

A different approach to the externality problem relies on the parties to negotiate a solution to the problem themselves.

The greatest novelty of Nobel laureate Ronald Coase’s contribution was the systematic treatment of trade in property rights. To solve the externality problem, Coase in a famous article, “The Problem of Social Cost”, in 1960 emphasized that the amount of damage that one party causes to another typically depends on the actions of both parties and argues that, regardless of the way that the law assigns liability, if the perpetrator and recipient are able to bargain freely, they
are likely to reach an efficient outcome. As such, the success of such a system depends on making sure that property rights are clearly assigned.

Coase’s paper consists of a series of examples and insightful discussions. He made no claims of a formal theorem based on explicit assumptions. The term “Coase Theorem” originated with George Stigler, who explained Coase’s ideas in his textbook, “Theory of Price”, pp 110-114. Stigler asserted that the Coase theorem actually contains two claims.

Claim 1 (Coase Efficiency Theorem): In the absence of transaction costs, voluntary negotiations over externalities will lead to a Pareto-optimal outcome, i.e., private bargaining in the absence of transaction costs will lead to a Pareto optimal level of externality-producing actions. In other words, under perfect competition, private and social costs will be equal.

Claim 2 (Coase Neutrality Theorem or Independence Theorem): The level of the externality will be the same regardless of the assignment of property rights.

Stigler’s Claim 2 would follow from Claim 1 if it were true that every Pareto optimal allocation has the same level of externality-producing activities, regardless of the way that private goods are distributed. Thus, the so-called Coase’s Theorem assesses that as long as property rights are clearly assigned, the two parties will negotiate in such a way that the optimal level of the externality-producing activity is implemented. As a policy implication, a government should simply rearrange property rights with appropriately designed property rights. Market then could take care of externalities without direct government intervention.

Coase shows his claims mainly by examples with only two agents and detrimental externality and insightful discussions. Coase made an observation that in the presence of externalities, the victim has an incentive to pay the firm to stop the production if the victim can compensate the firm by paying $p_x - c'_x(x^*)$. The following example captures Coase’s arguments.
Example 14.4.1 Two firms: One is chemical factory that discharges chemicals into a river and the other is the fisherman. Suppose the river can produce a value of $50,000. If the chemicals pollute the river, the fish cannot be eaten. How does one solve the externality? Coase’s method states that as long as the property rights of the river are clearly assigned, it results in efficient outcomes. That is, the government should give the ownership of the lake either to the chemical firm or to the fisherman, then it will yield an efficient output. To see this, assume that:

The cost of the filter is denoted by $c_f$.

**Case 1:** The lake is given to the factory.

i) $c_f < $50,000. The fisherman is willing to buy a filter for the factory. The fisherman will pay for the filter so that the chemical cannot pollute the lake.

ii) $c_f > $50,000 – The chemical is discharged into the lake. The fisherman does not want to install any filter.

**Case 2:** The lake is given to the fisherman, and the firm’s net product revenue is greater than $50,000.

i) $c_f < $50,000 – The factory buys the filter so that the chemical cannot pollute the lake.

ii) $c_f > $50,000 – The firm pays $50,000 to the fisherman then the chemical is discharged into the lake.

Like the above example, Cases’s own examples supporting his claims do with negotiations between firms or business rather than those between individuals. This difference is important since firms maximize profits rather than utility, and act as fiduciaries for individuals. We have to make some restricted assumptions on consumers’ utility functions to make Coase’s Theorem to be held.

Now consider an economy with two consumers with $L$ goods. Further, consumer $i$ has initial wealth $w_i$. Each consumer has preferences over both the
commodities he consumes and over some action $h$ that is taken by consumer 1. That is,

$$u_i(x^1_i, \ldots, x^L_i, h).$$

Activity $h$ is something that has no direct monetary cost for person 1. For example, $h$ is the quantity of loud music played by person 1. In order to play it, the consumer must purchase electricity, but electricity can be captured as one of the components of $x_i$. From the point of view of consumer 2, $h$ represents an external effect of consumer 1’s action. In the model, we assume that

$$\frac{\partial u_2}{\partial h} \neq 0.$$

Thus the externality in this model lies in the fact that $h$ affects consumer 2’s utility, but it is not priced by the market. Let $v_i(p, w_i, h)$ be consumer i’s indirect utility function:

$$v_i(w_i, h) = \max_{x_i} u_i(x_i, h) \quad \text{s.t.} \quad px_i \leq w_i.$$

We assume that preferences are quasi-linear with respect to some numeraire commodity. Thus, the consumer’s indirect utility function takes the form:

$$v_i(w_i, h) = \phi_i(h) + w_i.$$

We further assume that utility is strictly concave in $h$: $\phi''_i(h) < 0$. Again, the competitive equilibrium outcome in general is not Pareto optimal. In order to maximize utility, the consumer 1 should choose $h$ in order to maximize $v_1$ so that the interior solution satisfies $\phi'_1(h^*) = 0$. Even though consumer 2’s utility depends on $h$, it cannot affect the choice of $h$.

On the other hand, the socially optimal level of $h$ will maximize the sum of the consumers’ utilities:

$$\max_h \phi_1(h) + \phi_2(h).$$
The first-order condition for an interior maximum is:

$$\phi_1'(h^{**}) + \phi_2'(h^{**}) = 0,$$

where $h^{**}$ is the Pareto optimal amount of $h$. Thus, the social optimum is where the sum of the marginal benefit of the two consumers equals zero. In the case where the externality is bad for consumer 2 (loud music), the level of $h^* > h^{**}$. That is, too much $h$ is produced. In the case where the externality is good for consumer 2 (baking bread smell or yard beautification), too little will be provided, $h^* < h^{**}$.

Now we show that, as long as property rights are clearly assigned, the two parties will negotiate in such a way that the optimal level of the externality-producing activity is implemented. We first consider the case where consumer 2 has the right to prohibit consumer 1 from undertaking activity $h$. But, this right is contractible. Consumer 2 can sell consumer 1 the right to undertake $h_2$ units of activity $h$ in exchange for some transfer, $T_2$. The two consumers will bargain both over the size of the transfer $T_2$ and over the number of units of the externality good produced, $h_2$.

In order to determine the outcome of the bargaining, we first specify the bargaining mechanism as follows:

1. Consumer 2 offers consumer 1 a take-it-or-leave-it contract specifying a payment $T_2$ and an activity level $h_2$.

2. If consumer 1 accepts the offer, that outcome is implemented. If consumer 1 does not accept the offer, consumer 1 cannot produce any of the externality good, i.e., $h = 0$.

To analyze this, begin by considering which offers $(h, T)$ will be accepted by consumer 1. Since in the absence of agreement, consumer 1 must produce $h = 0$, consumer 1 will accept $(h_2, T_2)$ if and only if it offers higher utility than $h = 0$. That is, consumer 1 accepts if and only if:

$$\phi_1(h_2) - T_2 \geq \phi_1(0).$$
Given this constraint on the set of acceptable offers, consumer 2 will choose \((h_2, T_2)\) in order to solve the following problem:

\[
\begin{align*}
\max_{h_2, T_2} & \quad \phi_2(h_2) + T_2 \\
\text{s.t.} & \quad \phi_1(h_2) - T_2 \geq \phi_1(0).
\end{align*}
\]

Since consumer 2 prefers higher \(T_2\), the constraint will bind at the optimum. Thus the problem becomes:

\[
\max_{h_2} \phi_1(h_2) + \phi_2(h_2) - \phi_1(0).
\]

The first-order condition for this problem is given by:

\[
\phi_1'(h_2) + \phi_2'(h_2) = 0.
\]

But, this is the same condition that defines the socially optimal level of \(h_2\). Thus consumer 2 chooses \(h_2 = h^{**}\), and, using the constraint, \(T_2 = \phi_1(h^{**}) - \phi_1(0)\). And, the offer \((h_2, T_2)\) is accepted by consumer 1. Thus this bargaining process implements the social optimum.

Now we consider the case where consumer 1 has the right to produce as much of the externality as she wants. We maintain the same bargaining mechanism. Consumer 2 makes consumer 1 a take-it-or-leave-it offer \((h_1, T_1)\), where the subscript indicates that consumer 1 has the property right in this situation. However, now, in the event that 1 rejects the offer, consumer 1 can choose to produce as much of the externality as she wants, which means that she will choose to produce \(h^*\). Thus the only change between this situation and the first case is what happens in the event that no agreement is reached. In this case, consumer 2’s problem is:

\[
\begin{align*}
\max_{h_1, T_1} & \quad \phi_2(h_1) - T_1 \\
\text{s.t.} & \quad \phi_1(h_1) + T_1 \geq \phi_1(h^*)
\end{align*}
\]

Again, we know that the constraint will bind, and so consumer 2 chooses \(h_1\) and \(T_1\) in order to maximize

\[
\max \phi_1(h_1) + \phi_2(h_1) - \phi_1(h^*)
\]
which is also maximized at $h_1 = h^{**}$, since the first-order condition is the same. The only difference is in the transfer. Here $T_1 = \phi_1(h^*) - \phi_1(h^{**})$.

While both property-rights allocations implement $h^{**}$, they have different distributional consequences. The payment of the transfer is positive in the case where consumer 2 has the property rights, while it is negative when consumer 1 has the property rights. The reason for this is that consumer 2 is in a better bargaining position when the non-bargaining outcome is that consumer 1 is forced to produce 0 units of the externality good.

However, note that in the quasilinear framework, redistribution of the numeraire commodity has no effect on social welfare. The fact that regardless of how the property rights are allocated, bargaining leads to a Pareto optimal allocation is an example of the Coase Theorem: If trade of the externality can occur, then bargaining will lead to an efficient outcome no matter how property rights are allocated (as long as they are clearly allocated). Note that well-defined, enforceable property rights are essential for bargaining to work. If there is a dispute over who has the right to pollute (or not pollute), then bargaining may not lead to efficiency. An additional requirement for efficiency is that the bargaining process itself is costless. Note that the government doesn’t need to know about individual consumers here - it only needs to define property rights. However, it is critical that it do so clearly. Thus the Coase Theorem provides an argument in favor of having clear laws and well-developed courts.

However, Hurwicz (Japan and the World Economy 7, 1995, pp. 49-74) argued that, even when the transaction cost is zero, absence of income effects in the demand for the good with externality is not only sufficient (which is well known) but also necessary for Coase Neutrality Theorem to be true, i.e., when the transaction cost is negligible, the level of pollution will be independent of the assignments of property rights if and only if preferences of the consumers are quasi-linear with respect to the private good, leading to absence of income effects in the demand for the good with externality.

Unfortunately, as shown by Chipman and Tian (2012, Economic Theory), the
proof of Hurwicz’s claim on the necessity of parallel preferences for “Coase’s conjecture” is incorrect. To see this, consider the following class of utility functions $U_i(x_i, h)$ that have the functional form:

$$U_i(x_i, h) = x_i e^{-h} + \phi_i(h), \quad i = 1, 2$$

(14.36)

where

$$\phi_i(h) = \int e^{-b_i(h)} dh.$$  

(14.37)

$U_i(x_i, h)$ is then clearly not quasi-linear in $x_i$. It is further assumed that for all $h \in (0, \eta]$, $b_1(h) > \xi$, $b_2(h) < 0$, $b'_i(h) < 0 \ (i = 1, 2)$, $b_1(0) + b_2(0) \geq \xi$, and $b_1(\eta) + b_2(\eta) \leq \xi$.

We then have

$$\frac{\partial U_1}{\partial x_i} = e^{-h} > 0, \quad i = 1, 2,$$

$$\frac{\partial U_1}{\partial h} = -x_1 e^{-h} + b_1(h)e^{-h} > e^{-h}[\xi - x_1] \geq 0,$$

$$\frac{\partial U_2}{\partial h} = -x_2 e^{-h} + b_2(h)e^{-h} < 0$$

for $(x_i, h) \in (0, \xi) \times (0, \eta), \ i = 1, 2$. Thus, by the mutual tangency equality condition for Pareto efficiency, we have

$$0 = \frac{\partial U_1}{\partial h} \left/ \frac{\partial U_1}{\partial x_1} \right/ \frac{\partial U_2}{\partial h} \left/ \frac{\partial U_2}{\partial x_2} = -x_1 - x_2 + b_1(h) + b_2(h) = b_1(h) + b_2(h) - \xi,$$

(14.38)

which is independent of $x_i$. Hence, if $(x_1, x_2, h)$ is Pareto optimal, so is $(x'_1, x'_2, h)$ provided $x_1 + x_2 = x'_1 + x'_2 = \xi$. Also, note that $b'_i(h) < 0 \ (i = 1, 2)$, $b_1(0) + b_2(0) \geq \xi$, and $b_1(\eta) + b_2(\eta) \leq \xi$. Then $b_1(h) + b_2(h)$ is strictly monotone and thus there is a unique $h \in [0, \eta]$, satisfying (14.38). Thus, the contract curve is horizontal even though individuals’ preferences need not be parallel.

**Example 14.4.2** Suppose $b_1(h) = (1 + h)^\alpha \eta^n + \xi$ with $\alpha < 0$, and $b_2(h) = -h^n$. Then, for all $h \in (0, \eta]$, $b_1(h) > \xi$, $b_2(h) < 0$, $b'_i(h) < 0 \ (i = 1, 2)$, $b_1(0) + b_2(0) > \xi$, and $b_1(\eta) + b_2(\eta) < \xi$. Thus, $\phi_i(h) = \int e^{-b_i(h)} dh$ is concave, and $U_i(x_i, h) = x_i e^{-h} + \int e^{-b_i(h)} dh$ is quasi-concave, $\partial U_i/\partial x_i > 0$ and $\partial U_i/\partial h > 0$, and $\partial U_2/\partial h < 0$ for $(x_i, h) \in (0, \xi) \times (0, \eta), \ i = 1, 2$, but it is not quasi-linear in $x_i$. 

521
Chipman and Tian (2012) then investigate the necessity for the “Coase conjecture” that the level of pollution is independent of the assignments of property rights. This reduces to developing the necessary and sufficient conditions that guarantee that the contract curve is horizontal so that the set of Pareto optima for the utility functions is \( h \)-constant. This in turn reduces to finding the class of utility functions such that the mutual tangency (first-order) condition does not contain \( x_i \) and consequently it is a function, denoted by \( g(h) \), of \( h \) only:

\[
\frac{\partial U_1}{\partial h} \frac{\partial U_1}{\partial x_1} + \frac{\partial U_2}{\partial h} \frac{\partial U_2}{\partial x_2} = g(h) = 0. \tag{14.39}
\]

Let \( F_i(x_i, h) = \frac{\partial U_i}{\partial h} / \frac{\partial U_i}{\partial x_i} \) \( (i = 1, 2) \), which can be generally expressed as

\[ F_i(x_i, h) = x_i \psi_i(h) + f_i(x_i, h) + b_i(h), \]

where \( f_i(x_i, h) \) are nonseparable and nonlinear in \( x_i \). \( \psi_i(h) \), \( b_i(h) \), and \( f_i(x_i, h) \) will be further specified below.

Let \( F(x, h) = F_1(x, h) + F_2(\xi - x, h) \). Then the mutual tangency equality condition can be rewritten as

\[ F(x, h) = 0. \tag{14.40} \]

Thus, the contract curve, i.e., the locus of Pareto-optimal allocations, can be expressed by a function \( h = f(x) \) that is implicitly defined by (14.40).

Then, the Coase Neutrality Theorem, which is characterized by the condition that the set of Pareto optima (the contract curve) in the \((x, h)\) space for \( x_i > 0 \) is a horizontal line \( h = \text{constant} \), implies that

\[ h = f(x) = \bar{h} \]

with \( \bar{h} \) constant, and thus we have

\[ \frac{dh}{dx} = -\frac{F_x}{F_h} = 0 \]

for all \( x \in [0, \xi] \) and \( F_h \neq 0 \), which means that the function \( F(x, h) \) is independent of \( x \). Then, for all \( x \in [0, \xi] \),

\[ F(x, h) = x \psi_1(h) + (\xi - x) \psi_2(h) + f_1(x, h) + f_2(\xi - x, h) + b_1(h) + b_2(h) \equiv g(h). \tag{14.41} \]
Since the utility functions $U_1$ and $U_2$ are functionally independent, and $x$ disappears in (14.41), we must have $\psi_1(h) = \psi_2(h) \equiv \psi(h)$ and $f_1(x, h) = -f_2(\xi - x, h) = 0$ for all $x \in [0, \xi]$. Therefore,

$$F(x, h) = \xi \psi(h) + b_1(h) + b_2(h) \equiv g(h), \quad (14.42)$$

and

$$\frac{\partial U_i}{\partial h} / \frac{\partial U_i}{\partial x_i} = F_i(x_i, h) = x_i \psi(h) + b_i(h) \quad (14.43)$$

which is a first-order linear partial differential equation. Then, from Polyanin, Zaitsev, and Moussiaux (2002), we know that the principal integral $U_i(x_i, h)$ of (14.43) is given by

$$U_i(x_i, h) = x_i e^{\int \psi(h) dh} + \phi_i(h), \quad i = 1, 2 \quad (14.44)$$

with

$$\phi_i(h) = \int e^{\int \psi(h) dh} b_i(h) dh. \quad (14.45)$$

The general solution of (14.43) is then given by $\bar{U}_i(x, y) = \psi(U_i)$, where $\psi$ is an arbitrary function. Since a monotonic transformation preserves orderings of preferences, we can regard the principal solution $U_i(x_i, h)$ as a general functional form of utility functions that is fully characterized by (14.43).

Note that (14.44) is a general utility function that contains quasi-linear utility in $x_i$ and the utility function given in (14.36) as special cases. Indeed, it represents parallel preferences when $\psi(h) \equiv 0$ and also reduces to the utility function given by (14.36) when $\psi(h) = -1$.

To make the the mutual tangency (first-order) condition (14.39) be also sufficient for the contract curve to be horizontal in a pollution economy, we assume that for all $h \in (0, \eta]$, $x_1 \psi(h) + b_1(h) > 0$, $x_2 \psi(h) + b_2(h) < 0$, $\psi'(h) \leq 0$, $b_i'(h) < 0$ ($i = 1, 2$), $\xi \psi(0) + b_1(0) + b_2(0) \geq 0$, and $\xi \psi(\eta) + b_1(\eta) + b_2(\eta) \leq 0$.

\textsuperscript{3}It can be also seen from http://eqworld.ipmnet.ru/en/solutions/fpde/fpde1104.pdf.
We then have for \((x_i, h) \in (0, \xi) \times (0, \eta), i = 1, 2,\)

\[
\frac{\partial U_i}{\partial x_i} = e^{\int \psi(h)} > 0, \quad i = 1, 2, \\
\frac{\partial U_1}{\partial h} = e^{\int \psi(h)} [x_1 \psi(h) + b_1(h)] > 0, \\
\frac{\partial U_2}{\partial h} = e^{\int \psi(h)} [x_2 \psi(h) + b_2(h)] < 0,
\]

and thus

\[
0 = \frac{\partial U_1}{\partial h} \left( \frac{\partial U_1}{\partial x_1} + \frac{\partial U_2}{\partial h} \right) \frac{\partial U_2}{\partial x_2} = (x_1 + x_2) \psi(h) + b_1(h) + b_2(h) = \xi \psi(h) + b_1(h) + b_2(h),
\]

(14.46)

which does not contain \(x_i\). Hence, if \((x_1, x_2, h)\) is Pareto optimal, so is \((x'_1, x'_2, h)\)
provided \(x_1 + x_2 = x'_1 + x'_2 = \xi\). Also, note that \(\psi'(h) \leq 0, b'_i(h) < 0 \ (i = 1, 2), \)
\(\xi \psi(0) + b_1(0) + b_2(0) \geq 0, \) and \(\xi \psi(h) + b_1(h) + b_2(h) \leq 0\). Then \(\xi \psi(h) + b_1(h) + b_2(h)\)
is strictly monotone and thus there is a unique \(h \in [0, \eta]\) that satisfies (14.46).

Thus, the contract curve is horizontal even though individuals’ preferences need
not be parallel.

The formal statement of Coase Neutrality Theorem obtained by Chipman and
Tian (2012) thus can be set forth as follows:

**Proposition 14.4.1 (Coase Neutrality Theorem)** In a pollution economy
considered in the chapter, suppose that the transaction cost equals zero, and that
the utility functions \(U_i(x_i, h)\) are differentiable and such that \(\partial U_i/\partial x_i > 0, \) and
\(\partial U_1/\partial h > 0 \) but \(\partial U_2/\partial h < 0 \) for \((x_i, h) \in (0, \xi) \times (0, \eta), i = 1, 2.\) Then, the level
of pollution is independent of the assignments of property rights if and only if the
utility functions \(U_i(x, y)\), up to a monotonic transformation, have a functional
form given by

\[
U_i(x_i, h) = x_i e^{\int \psi(h)} + \int e^{\int \psi(h) dh} b_i(h) dh, \quad (14.47)
\]

where \(h\) and \(b_i\) are arbitrary functions such that the \(U_i(x_i, h)\) are differentiable,
\(\partial U_i/\partial x_i > 0, \) and \(\partial U_1/\partial h > 0 \) but \(\partial U_2/\partial h < 0 \) for \((x_i, h) \in (0, \xi) \times (0, \eta), i = 1, 2.\)

Although the above Coase neutrality theorem covers a wider class of preferences
than quasi-linear utility functions, it still puts a significant restriction
on the domain of its validity due to the special functional forms of the utility functions with respect to the private good. Thus, the Coase Neutrality Theorem may be best applied to production externalities between firms, but rather than consumption externalities.

The problem of this Coase theorem is that, costs of negotiation and organization, in general, are not negligible, and the income effect may not be zero. Thus, a privatization is optimal only in case of zero transaction cost, no income effect, and perfect economic environments. Tian (2000, 2001) shown that the state ownership and the collective ownership can also be optimal when an economic environment is sufficient imperfect or somewhere in the middle.

The problem of the Coase Efficiency Theorem is more serious. First, as Arrow (1979, p. 24) pointed out, the basic postulate underlying Coase’s theory appears to be that the process of negotiation over property rights can be modelled as a cooperative game, and this requires the assumption that each player know the preferences or production functions of each of the other players. When information is not complete or asymmetric, in general we do not have a Pareto optimal outcome. For instance, when there is one polluter and there are many pollutees, a “free-rider” problem arises and there is an an incentive for pollutees to misrepresent their preferences. Whether the polluter is liable or not, the pollutees may be expected overstate the amount they require to compensate for the externality. Thus, we may need to design an incentive mechanism to solve the free-rider problem.

Secondly, even if the information is complete, there are several circumstances that have led a number of authors to questions the conclusion in the Coase Efficiency Theorem:

(1) The core may be empty, and hence no Pareto optimum exists. An example of this for a three-agent model was presented by Aivazian and Callen (1981).

(2) There may be a fundamental non-convexity that prevents a Pareto optimum from being supported by a competitive equilibrium. S-
tarrettt (1972) showed that externalities are characterized by “fundamental non-convexities” that may preclude existence of competitive equilibrium.

(3) When an agent possesses the right to pollute, there is a built-in incentive for extortion. As Andel (1966) has pointed out, anyone with the right to pollute has an incentive to extract payment from potential pullutees, e.g., threat to blow a bugle in the middle of the night.

Thus, the hypothesis that negotiations over externalities will mimic trades in a competitive equilibrium is, as Coase himself has conceded, not one that can be logically derived from his assumptions, but must be regarded as an empirical conjecture that may or may not be confirmed from the data. A lot of theoretical work therefore still remains in order to provide Coasian economics with the rigorous underpinning.

14.4.3 Missing Market

We can regard externality as a lack of a market for an “externality.” For the above example in Pigovian taxes, a missing market is a market for pollution. Adding a market for firm 2 to express its demand for pollution - or for a reduction of pollution - will provide a mechanism for efficient allocations. By adding this market, firm 1 can decide how much pollution it wants to sell, and firm 2 can decide how much pollution it wants to buy.

Let $r$ be the price of pollution.

$x_1 =$ the units of pollution that firm 1 wants to sell;

$x_2 =$ the units of pollution for firm 2 wants to buy.

Normalize the output of firm 1 to $x_1$. 
The profit maximization problems become:

\[ \pi_1 = p_x x_1 + r x_1 - c_1(x_1) \]
\[ \pi_2 = p_y y - r x_2 - e_2(x_2) - c_y(y) \]

The first order conditions are:

\[ p_x + r = c'_1(x_1) \quad \text{for Firm 1} \]
\[ p_y = c'_y(y) \quad \text{for Firm 2} \]
\[ -r = e'(x_2) \quad \text{for Firm 2}. \]

At the market equilibrium, \( x_1^* = x_2^* = x^* \), we have

\[ p_x = c'_1(x^*) + e'(x^*) \quad (14.48) \]

which results in a social optimal outcome.

14.4.4 The Compensation Mechanism

The Pigovian taxes were not adequate in general to solve externalities due to the information problem: the tax authority cannot know the cost imposed by the externality. How can one solve this incomplete information problem?

Varian (AER 1994) proposed an incentive mechanism which encourages the firms to correctly reveal the costs they impose on the other. Here, we discuss this mechanism. In brief, a mechanism consists of a message space and an outcome function (rules of game). We will introduce in detail the mechanism design theory in Part VI.

Strategy Space (Message Space): \( M = M_1 \times M_2 \) with \( M_1 = \{(t_1, x_1)\} \), where \( t_1 \) is interpreted as a Pigovian tax proposed by firm 1 and \( x_1 \) is the proposed level of output by firm 1, and \( t_2 \) is interpreted as a Pigovian tax proposed by firm 2 and \( y_2 \) is the proposed level of output by firm 2.

The mechanism has two stages:

Stage 1: (Announcement stage): Firms 1 and 2 name Pigovian tax rates, \( t_i \), \( i = 1, 2 \), which may or may not be the efficient level of such a tax rate.
Stage 2: (Choice stage): If firm 1 produces $x$ units of pollution, firm 1 must pay $t_2x$ to firm 2. Thus, each firm takes the tax rate as given. Firm 2 receives $t_1x$ units as compensation. Each firm pays a penalty, $(t_1 - t_2)^2$, if they announce different tax rates.

Thus, the payoffs of two firms are:

$$
\pi_1^* = \max_x px - c_x(x) - t_2x - (t_1 - t_2)^2
$$

$$
\pi_2^* = \max_y py - c_y(y) + t_1x - e(x) - (t_1 - t_2)^2.
$$

Because this is a two-stage game, we may use the subgame perfect equilibrium, i.e., an equilibrium in which each firm takes into account the repercussions of its first-stage choice on the outcomes in the second stage. As usual, we solve this game by looking at stage 2 first.

At stage 2, firm 1 will choose $x(t_2)$ to satisfy the first order condition:

$$
p_x - c'_x(x) - t_2 = 0 \quad (14.49)
$$

Note that, by the convexity of $c_x$, i.e., $c''_x(x) > 0$, we have

$$
x'(t_2) = -\frac{1}{c''_x(x)} < 0. \quad (14.50)
$$

Firm 2 will choose $y$ to satisfy $p_y = c'_y(y)$.

Stage 1: Each firm will choose the tax rate $t_1$ and $t_2$ to maximize their payoffs.

For Firm 1,

$$
\max_{t_1} px - c_x(x) - t_2x(t_2) - (t_1 - t_2)^2 \quad (14.51)
$$

which gives us the first order condition:

$$
2(t_1 - t_2) = 0
$$

so the optimal solution is

$$
t_1^* = t_2. \quad (14.52)
$$

For Firm 2,

$$
\max_{t_2} py - c_y(y) + t_1x(t_2) - e(x(t_2)) - (t_1 - t_2)^2 \quad (14.53)
$$

528
so that the first order condition is
\[ t_1 x'(t_2) - e'(x(t_2)) x'(t_2) + 2(t_1 - t_2) = 0 \]
and thus
\[ [t_1 - e'(x(t_2))] x'(t_2) + 2(t_1 - t_2) = 0. \] (14.54)

By (14.50),(14.52) and (14.54), we have
\[ t^* = e'(x(t^*)) \quad \text{with} \quad t^* = t_{1}^* = t_{2}^*. \] (14.55)

Substituting the equilibrium tax rate, \( t^* = e'(x(t^*)) \), into (14.49) we have
\[ p_x = c'_x (x^*) + e'(x^*) \] (14.56)
which is the condition for social efficiency of production.

**Remark 14.4.1** This mechanism works by setting opposing incentives for two agents. Firm 1 always has an incentive to match the announcement of firm 2. But consider firm 2’s incentive. If firm 2 thinks that firm 1 will propose a large compensation rate \( t_1 \) for him, he wants firm 1 to be taxed as little as possible so that firm 1 will produce as much as possible. On the other hand, if firm 2 thinks firm 1 will propose a small \( t_1 \), it wants firm 1 to be taxed as much as possible. Thus, the only point where firm 2 is indifferent about the level of production of firm 1 is where firm 2 is exactly compensated for the cost of the externality.

In general, the individual’s objective is different from the social goal. However, we may be able to construct an appropriated mechanism so that the individual’s profit maximizing goal is consistent with the social goal such as efficient allocations. Tian (2003) also gave the solution to the consumption externalities by giving the incentive mechanism that results in Pareto efficient allocations. Tian (2004) study the informational efficiency problem of the mechanisms that results in Pareto efficient allocations for consumption externalities.
14.5 Reference

Books and Monographs:


Papers:


Chapter 15

Public Goods

15.1 Introduction

In the previous Chapter, we discussed the environments with externalities. It is very important to consider this issue when analysing the efficiency of a market (or even doing everything). The possibility of externality determines whether or not market failure may occur, whether or not intervening and supervisory measures are needed. In the presence of externalities, the market may fail to achieve efficient allocation even under perfect competition and freedom of choice. Therefore, some remedies need to be adopted. These measures include: Pigouvian tax, Coasian’s approach, building a market for the transactions of emission rights, and designing a proper incentive mechanism, etc.

The presence of public goods is another main reason for market failure. Once public goods are present in an economy, externalities and thus market failure may occur. It is well-known that financing a public project via voluntary donation is very difficult. This is because that public good is essentially different from externality. Two main differences are the non-exclusivity and non-rivalry. A good is *excludable* if people can be excluded from consuming it. A good is *non-rival* if one person’s consumption does not reduce the amount available to other consumers.
A pure public good is a good in which consuming one unit of the good by an individual in no way prevents others from consuming the same unit of the good. Thus, the good is nonexcludable and non-rival. Examples of public goods include street lights, policemen, fire protection, highway system, national defence, flood-control project, public television and radio broadcast, public parks, and a public project. The most representable public good is national defense. It protects all citizens from aggression.

Local Public Goods: when there is a location restriction for the service of a public good.

The non-exclusivity of public good may result in free-rider problem. For example, every citizen wants to get benefits from but don’t want to contribute to a public project. The inefficiency of some state-owned enterprises also originated from the free-rider problem. These enterprises lack proper incentive mechanisms, therefore, everyone wants enjoy the efforts provided by others. Even if the competitive market is an efficient social institution for allocating private goods in an efficient manner, it turns out that a private market is not a very good mechanism for allocating public goods.

15.2 Notations and Basic Settings

In a general setting of public goods economy that includes consumers, producers, private goods, and public goods.

Let

\( n \): the number of consumers.

\( L \): the number of private goods.

\( K \): the number of public goods.

\( Z_i \subseteq \mathbb{R}_+^L \times \mathbb{R}_+^K \): the consumption space of consumer \( i \).

\( Z \subseteq \mathbb{R}_+^{nL} \times \mathbb{R}_+^{nK} \): consumption space.

\( x_i \in \mathbb{R}_+^L \): a consumption of private goods by consumer \( i \).
\( y \in \mathbb{R}_+^K \): a consumption/production of public goods.

\( w_i \in \mathbb{R}_+^L \): the initial endowment of private goods for consumer \( i \). For simplicity, it is assumed that there is no public goods endowment, but they can be produced from private goods by a firm.

\( Y \subseteq \mathbb{R}^{L+K} \): the set of production possibilities of the firm. For simplicity, we assume there is only one firm to produce the public goods.

\( y \in Y \): a production plan, where the positive elements are outputs of public goods and negative elements are inputs.

\( f: \mathbb{R}_+^L \rightarrow \mathbb{R}_+^K \): production function with \( y = f(v) \), where \( v \in \mathbb{R}_+^L \): the private goods input.

\( \theta_i \): the profit share of consumer \( i \) from the production.

\( (x_i, y) \in Z_i \).

\( (x, y) = (x_1, ..., x_n, y) \in Z \): an allocation.

\( \succ_i \) (or \( u_i \) if exists) is a preference ordering.

\( e_i = (Z_i, \succ_i, w_i, \theta_i) \): the characteristic of consumer \( i \).

An allocation \( z \equiv (x, y) \) is feasible if

\[
\sum_{i=1}^n x_i + v \leq \sum_{i=1}^n w_i \tag{15.1}
\]

and

\[
y = f(v) \tag{15.2}
\]

\( e = (e_1, ..., e_n, f) \): a public goods economy.

A simple class of economic environments with public goods is depicted above. Analogous to the discussion in general equilibrium problem, the economic environments could be generalized to allow for production possibility sets of general form, an arbitrary number of firms, either public or private goods as input or output, see Foley (1970) and Milleron (1972) as detailed discussion.
Definition 15.2.1 Allocation \( z \equiv (x, y) = (x_1, \ldots, x_n, y) \in Z \) is feasible, if 
\[
(\sum_{i=1}^n x_i - \hat{w}, y) \in Y,
\]
where \( \hat{w} = \sum_{i=1}^n w_i \). If technology could be presented by function \( y = f(v) \), feasibility condition could be written as:
\[
\sum_{i=1}^n x_i + v \leq \sum_{i=1}^n w_i \tag{15.3}
\]
and
\[
y = f(v). \tag{15.4}
\]

Definition 15.2.2 An allocation \((x, y)\) is Pareto efficient for a public goods economy \( e \) if it is feasible and there is no other feasible allocation \((x', y')\) such that \((x'_i, y'_i) \succeq_i (x_i, y)\) for all consumers \( i \) and \((x'_k, y'_k) \succ_k (x_k, y)\) for some \( k \).

Definition 15.2.3 An allocation \((x, y)\) is weakly Pareto efficient for the public goods economy \( e \) if it is feasible and there is no other feasible allocation \((x', y')\) such that \((x'_i, y'_i) \succ_i (x_i, y)\) for all consumers \( i \).

Remark 15.2.1 Unlike private goods economies, even though under the assumptions of continuity and strong monotonicity, a weakly Pareto efficient allocation may not be Pareto efficient for the public goods economies. The following proposition is due to Tian (Economics Letters, 1988).

Proposition 15.2.1 For the public goods economies, a weakly Pareto efficient allocation may not be Pareto efficient even if preferences satisfy strong monotonicity and continuity.

Proof. The proof is by way of an example. Consider an economy with \((n, L, K) = (3, 1, 1)\), constant returns in producing \( y \) from \( x \) (the input-output coefficient normalized to one), and the following endowments and utility functions: \( w_A = w_B = w_3 = 1, u_1(x_1, y) = x_1 + y, \) and \( u_i(x_i, y) = x_i + 2y \) for \( i = 2, 3 \). Then \( z = (x, y) \) with \( x = (0.5, 0, 0) \) and \( y = 2.5 \) is weakly Pareto efficient but not Pareto efficient because \( z' = (x', y') = (0, 0, 0, 3) \) Pareto-dominates \( z \) by consumers 2 and 3. ■

However, under an additional condition of strict convexity, they are equivalent. The proof is left to readers.
15.3 Discrete Public Goods

15.3.1 Efficient Provision of Public Goods

For simplicity, consider a public good economy with $n$ consumers and two goods: one private good and one public good.

Discrete public good, also called public project, is indivisible. Its quantities may either be 1 (the project is constructed) or 0 (it is not constructed).

Let $g_i$ be the contribution made by consumer $i$, so that

$$x_i + g_i = w_i$$

$$\sum_{i=1}^{n} g_i = v$$

Assume $u_i(x_i, y)$ is strongly monotonic increasing and continuous.

Let $c$ be the cost of producing the public project so that the production technology is given by

$$y = \begin{cases} 
1 & \text{if } \sum_{i=1}^{n} g_i \geq c \\
0 & \text{otherwise}
\end{cases}$$

We first want to know under what conditions providing the public good will be Pareto dominate to not producing it, i.e., there exists $(g_1, \ldots, g_n)$ such that

$$\sum_{i=1}^{n} g_i \geq c$$

and

$$u_i(x_i - g_i, 1) > u_i(x_i, 0) \quad \forall i.$$  \hspace{1cm} (15.5)

Let $r_i$ be the maximum willingness-to-pay (reservation price) of consumer $i$, i.e., $r_i$ must satisfy

$$u_i(x_i - r_i, 1) = u_i(x_i, 0).$$  \hspace{1cm} (15.5)

If producing the public project Pareto dominates not producing the public project, we have

$$u_i(x_i - g_i, 1) > u_i(x_i, 0) = u_i(x_i - r_i, 1) \quad \text{for all } i$$  \hspace{1cm} (15.6)

By monotonicity of $u_i$, we have

$$x_i - g_i > x_i - r_i \quad \text{for } i$$  \hspace{1cm} (15.7)
Then, we have

\[ r_i > g_i \quad \text{(15.8)} \]

and thus

\[ \sum_{i=1}^{n} r_i > \sum_{i=1}^{n} g_i \geq c \quad \text{(15.9)} \]

That is, the sum of the willingness-to-pay for the public good must exceed the cost of providing it. This condition is necessary. In fact, this condition is also sufficient. In summary, we have the following proposition.

**Proposition 15.3.1** Providing a public good Pareto dominates not producing the public good if and only if

\[ \sum_{i=1}^{n} r_i > \sum_{i=1}^{n} g_i \geq c. \]

The question is information related to individual preferences and utility functions is unknown to the central planner. So the issue of mechanism design and rent extraction may arise.

### 15.3.2 Free-Rider Problem

How effective is a private market at providing public goods? The answer as shown below is that we cannot expect that purely independent decision will necessarily result in an efficient amount of the public good being produced. To see this, suppose

\[
\begin{align*}
  r_i &= 100 \quad i = 1, 2 \\
  c &= 150 \text{ (total cost)} \\
  g_i &= \begin{cases} 
    150/2 = 75 & \text{if both agents make contributions} \\
    150 & \text{if only agent } i \text{ makes contribution}
  \end{cases}
\end{align*}
\]

Each person decides independently whether or not to buy the public good. As a result, each one has an incentive to be a free-rider on the other as shown the following payoff matrix.
Table 15.3.2: Private Provision of a Discrete Public Good

Note that net payoffs are defined by \( r_i - g_i \). Thus, it is given by 100 - 150/2 = 25 when both consumers are willing to produce the public project, and 100 - 150 = -50 when only one person wants to buy, but the other person does not.

This is the prisoner’s dilemma. The dominant strategy equilibrium in this game is (doesn’t buy, doesn’t buy). Thus, no body wants to share the cost of producing the public project, but wants to free-ride on the other consumer. As a result, the public good is not provided at all even thought it would be efficient to do so. Thus, voluntary contribution in general does not result in the efficient level of the public good.

15.3.3 Voting for a Discrete Public Good

The amount of a public good is often determined by a voting. Will this generally results in an efficient provision? The answer is no.

Voting does not result in efficient provision. Consider the following example.

Example 15.3.1

\[
\begin{align*}
c &= 99 \\
r_1 &= 90, \quad r_2 = 30, \quad r_3 = 30
\end{align*}
\]

Clearly, \( r_1 + r_2 + r_3 > c \). \( g_i = 99/3 = 33 \). So the efficient provision of the public good should be yes. However, under the majority rule, only consumer 1 votes “yes” since she receives a positive net benefit if the good is provided. The 2nd and 3rd persons vote “no” to produce public good, and therefore, the public good
will not be provided so that we have inefficient provision of the public good. The problem with the majority rule is that it only measures the net benefit for the public good, whereas the efficient condition requires a comparison of willingness-to-pay.

This example reflects the relationship between democracy and efficiency. Since the voters are all driven by their own interests, a democratic voting rule may thus lead to an inefficient provision of public goods. To overcome this problem, the election of high-rank officials who are responsible for the designing of major strategies of an organization should be conducted under a democratic voting procedure. Choosing a right person needs the full respect for the will of all voters. However, once a right person is elected, his authority conferred by all voters ought to be fully respected. Applying the simple majority rule in every specific issues may often lead to inefficient outcome. Therefore, even in democratic system, the major leader of an organization (say, presidents of a country of an university) usually has the power to nominate his deputy and the whole team of leadership. Of course, the leader himself will be responsible to all the voters and pursue a chance of being reelected via good performance. An example is the professors’ committee in universities. Its duty is to evaluate the academic performance and promotion of faculties rather than getting involved in the daily executive works. If every professor has voting right to support his own field of speciality, then the inefficient outcome as described in the above example may arise.

The above analysis shows that neither market nor democratic voting procedure could lead to efficient provision of public goods. The solution to this problem depends on designing of proper mechanism. We will discuss the VCG (Groves – Clark – Vickrey) mechanism in Chapter 18, which may elicit efficient provision of public goods and truthtelling of voters.
15.4 Continuous Public Goods

15.4.1 Efficient Provision of Public Goods

Again, for simplicity, we assume there is only one public good and one private
good that may be regarded as money, and \( y = f(v) \).

The welfare maximization approach shows that Pareto efficient allocations
can be characterized by

\[
\max_{(x, y)} \sum_{i=1}^{n} a_i u_i(x_i, y)
\]

\( \text{s.t.} \quad \sum_{i=1}^{n} x_i + v \leq \sum_{i=1}^{n} w_i \\
y \leq f(v) \)

Define the Lagrangian function:

\[
L = \sum_{i=1}^{n} a_i u_i(x_i, y) + \lambda \left( \sum_{i=1}^{n} w_i - \sum_{i=1}^{n} x_i - v \right) + \mu (f(v) - y). \quad (15.10)
\]

When \( u_i \) is strictly quasi-concave and differentiable and \( f(v) \) is concave and d-
derivative, the set of Pareto optimal allocations are characterized by the first
order condition:

\[
\frac{\partial L}{\partial x_i} = 0 \quad a_i \frac{\partial u_i}{\partial x_i} = \lambda \quad (15.11)
\]

\[
\frac{\partial L}{\partial v} = 0 \quad \mu f'(v) = \lambda \quad (15.12)
\]

\[
\frac{\partial L}{\partial y} = 0 \quad \sum_{i=1}^{n} a_i \frac{\partial u_i}{\partial y} = \mu. \quad (15.13)
\]

So at an interior solution, by (15.11) and (15.12)

\[
\frac{a_i}{\mu} = \frac{f'(v)}{\frac{\partial u_i}{\partial x_i}} \quad (15.14)
\]

Substituting (15.14) into (15.13),

\[
\sum_{i=1}^{n} \frac{\partial u_i}{\partial y} \frac{\partial u_i}{\partial x_i} = \frac{1}{f'(v)} \quad (15.15)
\]
Thus, we obtain the well-known Lindahl-Samuelson condition.

In conclusion, the conditions for Pareto efficiency are given by

\[
\begin{cases}
\sum_{i=1}^{n} MRS_{y|x_i} = MRTS_{yv} \\
\sum x_i + v \leq \sum_{i=1}^{n} w_i \\
y = f(v)
\end{cases}
\]  
(15.16)

**Example 15.4.1**

\[
\begin{align*}
u_i &= a_i \ln y + \ln x_i \\
y &= v
\end{align*}
\]

the Lindahl-Samuelson condition is

\[
\sum_{i=1}^{n} \frac{\partial u_i}{\partial y} = 1
\]  
(15.17)

and thus

\[
\sum_{i=1}^{n} \frac{a_i y}{x_i} = \sum_{i=1}^{n} \frac{a_i x_i}{y} = 1 \Rightarrow \sum a_i x_i = y
\]  
(15.18)

which implies the level of the public good is not uniquely determined.

Thus, in general, the marginal willingness-to-pay for a public good depends on the amount of private goods consumption, and therefor, the efficient level of \( y \) depends on \( x_i \). However, in the case of quasi-linear utility functions,

\[
u_i(x_i, y) = x_i + u_i(y)\]  
(15.19)

the Lindahl-Samuelson condition becomes

\[
\sum_{i=1}^{n} u'_i(y) = \frac{1}{f'(v)} \equiv c'(y)
\]  
(15.20)

and thus \( y \) is uniquely determined.

**Example 15.4.2**

\[
\begin{align*}
u_i &= a_i \ln y + x_i \\
y &= v
\end{align*}
\]
the Lindahl-Samuelson condition is
\[
\sum_{i=1}^{n} \frac{\partial u_i}{\partial y} = 1 \quad (15.21)
\]
and thus
\[
\sum_{i=1}^{n} \frac{a_i}{y} = \sum_{i=1}^{n} \frac{a_i}{y} = 1 \Rightarrow \sum a_i = y \quad (15.22)
\]
which implies the level of the public good is uniquely determined.

### 15.4.2 Lindahl Equilibrium

We have given the conditions for Pareto efficiency in the presence of public goods. The next problem is how to achieve a Pareto efficient allocation in a decentralized way. In private-goods-only economies, any competitive equilibrium is Pareto optimal. However, with public goods, a competitive mechanism does not help. For instance, if we tried the competitive solution with a public good and two consumers, the utility maximization would equalize the MRS and the relative price, e.g.,
\[
\frac{\text{MRS}}{A_{yx}} = \frac{\text{MRS}}{B_{yx}} = \frac{p_y}{p_x}.
\]
This is an immediate violation to the Samuelson-Lindahl optimal condition.

Lindahl suggested to use a tax method to provide a public good. Each person is signed a specific “personalized price” for the public good. The Lindahl solution is a way to mimic the competitive solution in the presence of public goods. Suppose we devise a mechanism to allocate the production cost of a public good between consumers, then we can achieve the Samuelson-Lindahl condition. To this end, we apply different prices of a public good to different consumers. The idea of the Lindahl solution is that the consumption level of a public goods is the same to all consumers, but the price of the public good is personalized among consumers in the way that the price ratio of two goods for each person being equal the marginal rates of substitutions of two goods.

To see this, consider a public goods economy $e$ with $x_i \in R^L_+$ (private goods) and $y \in R^K_+$ (public goods). For simplicity, we assume the CRS for $y = f(v)$. A
feasible allocation
\[
\sum_{i=1}^{n} x_i + v \leq \sum_{i=1}^{n} w_i
\]  \hspace{1cm} (15.23)

Let \( q_i \in R^K_+ \) be the personalized price vector of consumer \( i \) for consuming the public goods.

Let \( q = \sum_{i=1}^{n} q_i \) : the market price vector of \( y \).

Let \( p \in R^L_+ \) be the price vector of private goods.

The profit is defined as \( \pi = qy - pv \) with \( y = f(v) \).

**Definition 15.4.1 (Lindahl Equilibrium)** An allocation \((x^*, y^*) \in R^{nL+K}_+\) is a Lindahl equilibrium allocation if it is feasible and there exists a price vector \( p^* \in R^L_+ \) and personalized price vectors \( q_i^* \in R^K_+ \), one for each individual \( i \), such that

(i) \( p^*x_i^* + q_i^*y^* \leq p^*w_i \);

(ii) \( (x_i, y) \succ_i (x_i^*, y^*) \) implies \( p^*x_i + q_i^*y > p^*w_i \);

(iii) \( q^*y^* - p^*v^* = 0 \),

where \( v^* = \sum_{i=1}^{n} w_i - \sum_{i=1}^{n} x_i^* \) and \( \sum_{i=1}^{n} q_i^* = q^* \).

We call \((x^*, y^*, p^*, q_1^*, \ldots, q_n^*)\) a Lindahl equilibrium.

**Remark 15.4.1** Because of CRS, the maximum profit is zero at the Lindahl equilibrium. That is, \( \hat{q}^*y^* - p^*v^* = 0 \), therefore

\[
\sum_{i=1}^{n} p^*x_i^* = \sum_{i=1}^{n} w_i + \hat{q}^*y^*.
\]

then the budget constraint (i) holds with equality at equilibrium for every consumer.

We may regard a Walrasian equilibrium as a special case of a Lindahl equilibrium when there are no public goods. In fact, the concept of Lindahl equilibrium in economies with public goods is, in many ways, a natural generalization of the
Walrasian equilibrium notion in private goods economies, with attention to the well-known duality that reverses the role of prices and quantities between private and public goods, and between Walrasian and Lindahl allocations. In the Walrasian case, prices must be equalized while quantities are individualized; in the Lindahl case the quantities of the public good must be the same for everyone, while prices charged for public goods are individualized. In addition, the concepts of Walrasian and Lindahl equilibria are both relevant to private-ownership economies. Furthermore, they are characterized by purely price-taking behavior on the part of agents. It is essentially this property that one can consider the Lindahl solution as an informationally decentralized process.

Similar to Walrasian Equilibrium, Lindahl equilibrium has various good properties. In fact, an economy with public goods might be regarded as an economy with only private goods through redefining the consumption space. The Lindahl equilibrium could be regarded as Walrasian equilibrium under this transformation. This method is broadly adopted by existing literature to prove the existence of Lindahl equilibrium. The analogous fundamental I and II theorems may hold.

**Theorem 15.4.1 (Existence Theorem on Lindahl Equilibrium)** For economy

\[
 e = \left( \{X_i, w_i, \succ_i\} , \{Y_j\} , \{\theta_{ij}\} \right),
\]

there exist a Lindahl equilibrium if the following conditions hold:

(i) \( Z_i = \mathbb{R}_+^L \times \mathbb{R}_+^K \);

(ii) \( w_i \geq 0, \sum_{i \in N} w_i > 0 \);

(iii) \( \succ_i \) is continuous, strictly convex, and strongly monotone;

(iv) \( Y \) is a closed and convex cone, \( 0 \in Y , (-\mathbb{R}_+^L, 0) \subseteq Y \) (free disposal property).

**Proof.** We prove this theorem by constructing a economy with only private goods to which the existence theorem of CE is applicable. Treating the valuations of different consumers as different private goods, then the consumption
space of consumer $i$ is $\tilde{Z}_i = (Z_i, \{0\}) \subseteq \mathbb{R}^{L+K} \times \mathbb{R}^{(n-1)K}$, here $0$ is null element of $(n-1)K$-dimensional space. The consumption bundle if $i$ is $z_i = (x_i, y_i, 0, \ldots, 0)$. In addition to conditions of this theorem, the consumption space constructed above satisfy all requirements of Theorem 10.4.3 (Existence Theorem III for competitive equilibriums), the existence of CE is guaranteed. Therefore, a Lindahl equilibrium exists. ■

Similarly, we can weaken strong monotonicity to monotonicity when initial endowment is an interior point. Thus we have the following proposition.

**Proposition 15.4.1** For economy

\[ e = \{\{X_i, w_i, \succ_i\}, \{Y_j\}, \{\theta_{ij}\}\}, \]

there exist a Lindahl equilibrium if the following conditions hold:

(i) $Z_i = \mathbb{R}_+^{L} \times \mathbb{R}_+^{K}$;

(ii) $w_i \in \mathbb{R}_+^{L}$;

(iii) $\succ_i$ is continuous, strictly convex, and strongly monotone;

(iv) $Y$ is a closed and convex cone, $0 \in Y$, $(-\mathbb{R}_+^{L}, 0) \subseteq Y$ (free disposal property).

For a public economy with one private good and one public good $y = \frac{1}{q}v$, the definition of Lindahl equilibrium becomes much simpler.

**Definition 15.4.2** An allocation $(x^*, y^*)$ is a Lindahl Allocation if $(x^*, y^*)$ is feasible (i.e., $\sum_{i=1}^n x_i^* + qy^* \leq \sum w_i$) and there exists $q_i^*$, $i = 1, \ldots, n$ such that

(i) $x_i^* + q_i y^* \leq w_i$

(ii) $(x_i, y) \succ_i (x_i^*, y^*)$ implies $x_i + q_i y > w_i$

(iii) $\sum_{i=1}^n q_i = q$

In fact, the feasibility condition is automatically satisfied when the budget constraints (i) is satisfied.
If \((x^*, y^*)\) is an interior Lindahl equilibrium allocation, from the utility maximization, we can have the first order condition:

\[
\frac{\partial u_i}{\partial y} = \frac{q_i}{1} \tag{15.24}
\]

which means the Lindahl-Samuelson condition holds:

\[
\sum_{i=1}^{n} MRS_{yx_i} = q_i
\]

which is the necessary condition for Pareto efficiency.

**Example 15.4.3**

\[
u_i(x_i, y) = x_i^{\alpha_i} y^{(1-\alpha_i)} \quad \text{for} \quad 0 < \alpha_i < 1
\]

\[
y = \frac{1}{q} v
\]

The budget constraint is:

\[
x_i + q_i y = w_i.
\]

The demand functions for \(x_i\) and \(y_i\) of each \(i\) are given by

\[
x_i = \alpha_i w_i \tag{15.25}
\]

\[
y_i = \frac{(1 - \alpha_i) w_i}{q_i} \tag{15.26}
\]

Since \(y_1 = y_2 = \ldots y_n = y^*\) at the equilibrium, we have by (15.26)

\[
q_i y^* = (1 - \alpha_i) w_i. \tag{15.27}
\]

Making summation, we have

\[
q y^* = \sum_{i=1}^{n} (1 - \alpha_i) w_i.
\]

Then, we have

\[
y^* = \frac{\sum_{i=1}^{n} (1 - \alpha_i) w_i}{q}
\]

and thus, by (15.27), we have

\[
q_i = \frac{(1 - \alpha_i) w_i}{y^*} = \frac{q(1 - \alpha_i) w_i}{\sum_{i=1}^{n} (1 - \alpha_i) w_i}. \tag{15.28}
\]

If we want to find a Lindahl equilibrium, we must know the preferences or MRS of each consumer. But because of the free-rider problem, it is very difficult for consumers to report their preferences truthfully.
15.4.3 The First Fundamental Theorem

Similarly, we have the following First Fundamental Theorem of Welfare Economics for public goods economies.

**Theorem 15.4.2**: Every Lindahl allocation \((x^*, y^*)\) with the price system \((p^*, q^1_1, \ldots, q^*_n)\) is weakly Pareto efficient, and further under local non-satiation, it is Pareto efficient.

**Proof.** We only prove the second part. The first part is the simple. Suppose not. There exists another feasible allocation \((x_i, y)\) such that \((x_i, y) \succ_i (x_i^*, y^*)\) for all \(i\) and \((x_j, y) \succ_j (x_j^*, y^*)\) for some \(j\). Then, by local-non-satiation of \(\succ_i\), we have

\[
p^* x_i + q^*_i y \geq p^* w_i \quad \text{for all } i = 1, 2, \ldots, n
\]

\[
p^* x_j + q^*_j y > p^* w_j \quad \text{for some } j.
\]

Thus

\[
\sum_{i=1}^{n} p^* x_i + \sum_{i=1}^{n} q^*_i y > \sum_{i=1}^{n} p^* w_i \quad (15.29)
\]

So

\[
p^* \sum_{i=1}^{n} x_i + q^* y > \sum_{i=1}^{n} p^* w_i
\]

or

\[
p^* \sum_{i=1}^{n} x_i + p^* v > \sum_{i=1}^{n} p^* w_i
\]

by noting that \(q^* y - p^* v \leq q^* y^* - p^* v^* = 0\). Hence,

\[
p^* \left[ \sum_{i=1}^{n} (x_i - w_i) + v \right] > 0
\]

which contradicts the fact that \((x, y)\) is feasible. ■

Similarly, we can define Lindahl equilibrium (LE) with transfer payments.

**Definition 15.4.3 (Lindahl Equilibrium with Transfer Payments)** For an economy with public goods \(e = (e_1, \ldots, e_n, Y)\), an allocation \((x^*, y^*) \in Z\), a price
vector of private goods \( \mathbf{p}^* \in \mathbb{R}_+^L \) and a personalized price vector of public good \( \mathbf{q}_i^* \in \mathbb{R}_+^K, \forall i \) constitute a Lindahl equilibrium with transfer payments if there exist a sequence of wealth levels \( (I_1, \ldots, I_n) \) satisfying \( \sum_i I_i = \mathbf{p} \cdot \sum_i w_i \) and

(i) \( \mathbf{p}^*x_i^* + \mathbf{q}_i^*y^* \leq I_i \);

(ii) if \( (x_i, y) \succ_i (x_i^*, y^*) \), then \( \mathbf{p}^*x_i + \mathbf{q}_i^*y > I_i \);

(iii) for all \( (y, -v) \in Y \), we have \( \mathbf{q}^*y^* - \mathbf{p}^*v^* \geq \mathbf{q}^*y - \mathbf{p}^*v \);

(iv) \( (y^*, -v^*) \in Y \),

where, \( v^* = \sum_{t=1}^n w_i - \sum_{t=1}^n x_i^*, \sum_{t=1}^n q_i^* = \hat{q}^* \).

**Theorem 15.4.3 (The First Theorem of LE with Transfer payments)** Under price system \( (\mathbf{q}_1^*, \ldots, \mathbf{q}_n^*, \mathbf{p}^*) \), every Lindahl equilibrium with transfer payments \( (x^*, y^*) \) is weakly Pareto efficient. If the preference is local non-satiated, then it is Pareto efficient.

**Proof.** The proof is analogous to the proof of the first fundamental theorem and is thus omitted. ■

### 15.4.4 Core and Lindahl Equilibrium

Analogous to the private good economy, we can define the economic core as follows:

**Definition 15.4.4 (Blocking Coalition)** agents \( S \subseteq N \) form a block to allocation \( x \), if there exists \((x', y)\) such that

1. \((x', y')\) is feasible with respect to \( S \), that is \((y, \sum_{i \in S} (x'_i - w_i)) \in Y \),

2. for \( \forall i \in S \), we have \((x'_i, y') \succeq_i (x_i, y)\), and \( \exists j \in S \) such that \((x'_j, y') \succ_j (x_j, y)\).

**Definition 15.4.5 (Economic Core)** Feasible allocation \( x \) is an allocation in core, if there exists no blocking coalition with respect to \((x, y)\). The set of all allocations in core is called economic core or core.
Remark 15.4.2 Every allocation in core is Pareto efficient and individually rational, that is, \((x_i, y) \succeq_i (w_i, y), \forall i = 1, 2, \ldots, n.\)

Therefore, it can be proved similarly, every Lindahl equilibrium satisfies core property under local non-satiation condition.

Theorem 15.4.4 When local non-satiation condition holds, if \((x, y, p)\) is a Lindahl equilibrium, then \((x, y)\) is in the core.

Every Lindahl equilibrium is in the core under local non-satiation condition, the core convergence theorem does not hold necessarily. See Milleron (1972) for an counterexample.

15.4.5 The Second Fundamental Theorem

Similarly, we have the second fundamental theorem for public good economy.

Theorem 15.4.5 (The Second Fundamental Theorem with Public Goods)

For economy \(e = (e_1, \ldots, e_n, \{Y_j\})\), suppose that \(\succeq_i\) is continuous, convex, and strictly monotone , \(Y\) is closed and convex, and \(0 \in Y\). Then for any Pareto efficient allocation with interior private consumption \((x^*, y^*)\) (that is \(x^* \in R^{nL}_{++}\)), there exists a nonzero price vector \((q_1, \ldots, q_n, p) \in R^{L+nK}_+,\) such that \(((x, y), (q_1, \ldots, q_n), p)\) is a Lindahl equilibrium with transfer payments, that is, there exist a sequence of wealth levels \((I_1, \ldots, I_n)\) satisfying \(\sum_i I_i = p \cdot \sum_i w_i\) and

\[
(1) \text{ if } (x_i, y) \succ_i (x^*_i, y^*), \text{ then } px_i + q_i y > I_i \equiv px^*_i, \text{ } i = 1, \ldots, n; \\
(2) \text{ for all } (y, -v) \in Y, \text{ we have } qy^* - pv^* \geq qy - pv.
\]

where, \(v^* = \sum_{i=1}^{n} w_i - \sum_{i=1}^{n} x^*_i, \sum_{i=1}^{n} q_i = \hat{q}.\)

Proof. Define two subsets \(W\) and \(P(x^*, y^*)\) of \(R^{nK+L}\) as follows:

\[
W = \{(y, \ldots, y; -v) : (y, -v) \in Y\}.
\]

551
It follow from the closeness and convexity of $Y$ and $0 \in Y$, $W$ is non-empty, closed and convex.

$$P(x^*, y^*) = \{(y_1', \ldots, y_n'; -v') : v = \sum_{i=1}^n (x_i' - w_i) \& (x_i', y_i') \succ_i (x_i^*, y^*)\}.$$ 

Since $\succ_i$ is convex, $P(x^*, y^*)$ is convex.

Since $(x^*, y^*)$ is PO allocation, we must have $W \cap P(x^*) = \emptyset$, otherwise, there may exist a Pareto improvement. Therefore, applying the hyperplane separating theorem introduced in Chapter 2, we obtain that there exists $(q_1, \ldots, q_n, p) \neq 0$, such that for $\forall (y, \ldots, y; -v) \in W$ and $\forall (y_1', \ldots, y_n'; -v') \in P(x^*, y^*)$, we have

$$\sum_{i=1}^n q_i y_i' - pv' \geq \sum_{i=1}^n q_i y_i - pv. \tag{15.30}$$

We proceed in five steps.

1. Profit-maximization. For $\forall (y, -v) \in Y$, we have $\hat{q} y^* - pv^* \geq \hat{q} \cdot y - pv$.

   Let $\hat{z} = (y^*, \ldots, y^*; -v')$, where $v' = \sum_{i=1}^n (w_i - x_i')$ and $x_i' \geq x_i^*$. Then, from strong monotonicity, we have $(x_i', y^*) \succ_i (x_i^*, y^*)$, therefore $(y^*, \ldots, y^*; -v') \in P(x^*, y^*)$. Hence, it follows from (refseparating1) for any $(y, -v) \in Y$, we have

   $$\hat{q} y^* - pv' \geq \hat{q} y - pv.$$ 

   Let $v' \rightarrow v^*$, then we have $\hat{q} y^* - pv^* \geq \hat{q} \cdot y - pv, \forall (y, -v) \in Y$.

2. $(q_1, \ldots, q_n, p) \geq 0$, and $p \neq 0$.

   Firstly, we prove $q_i \geq 0$, $i = 1, \ldots, n$. Let

   $$\hat{z} = (y, \ldots, y; -v) + e_{yi}^k,$$

   where, $(y, \ldots, y; -v)$ is element in $W$, $e_{yi}^k = (0, \ldots, 1, 0, \ldots, 0)$ is a vector in $\mathbb{R}^{nK+L}$ such that the element associate with $q_i^k$ is 1 and the other elements are all zero. Then from the strong monotonicity of preference, and the fact
that $e^k_y$ is equally distributed among all agents, we have $\hat{z} \in P(x^*, y^*)$. Therefore, from (15.30), we have

$$\hat{y} - pv + q^k_i \geq \hat{q}_i y_i - pv. \quad (15.31)$$

Consequently,

$$q^k_i \geq 0, \quad k = 1, 2, \ldots, K; i = 1, 2, \ldots, n. \quad (15.32)$$

Now we prove that $p \geq 0$. Let $e^l_x = (0, \ldots, 1, 0, \ldots, 0)$ is a vector in $\mathbb{R}^{nK+L}$, with the element associate with good $l$ being 1 and other elements being zero. Repeat the above procedure, we have

$$p^l_i \geq 0, \quad l = 1, 2, \ldots, L. \quad (15.33)$$

Lastly, we prove $p \neq 0$. By contradiction, if $p = 0$, since $(q_1, \ldots, q_n, p) \neq 0$, then for some $k$, we must have $q^k = \sum_{i=1}^{n} q^k_i > 0$. Since the production technology of public good exhibits constant return to scale, when $p = 0$, all cost for all private good as inputs are zero. Then the profit may be infinitely large, then contradicts with the fact that $y$ is profit-maximizing plan.

3. For all $i$, if $(x_i, y) \succ (x^*_i, y^*)$, then $\sum_i px_i + \hat{q}y \geq \sum_i px^*_i + \hat{q}y^*$.

For every $i$ and $(x_i, y) \succ (x^*_i, y^*)$, by strong monotonicity of preferences, there exists $(x'_i, y')$ that is sufficiently close to $(x_i, y)$, such that $(x'_i, y') \succ_i (x_i, y) \succ_i (x^*_i, y^*)$, and thus $(y', \ldots, y'; \sum_i (x'_i - w_i)) \in P(x^*, y^*)$. Also, note that $(y^*, \ldots, y^*; \sum_i (x^*_i - w_i)) \in W$. Thus, by (15.30), we have

$$\sum_i px'_i + \hat{q}y' \geq \sum_i px^*_i + \hat{q}y^*. $$

Let $x'_i \rightarrow x_i$. We have $\sum_i px_i + \hat{q}y \geq \sum_i px^*_i + \hat{q}y^*$.

4. For all $i$, if $(x_i, y) \succ (x^*_i, y^*)$, then $px_i + q_i y \geq px^*_i + q_i y^*$. Let

$$(x'_i, y') = (x_i, y),$$

$$(x'_m, y') = (x^*_m, y^*), \quad m \neq i.$$
Then, it follows from step 3 that
\[ px_i + q_i y + \sum_{m \neq i} (px^*_m + q_m y^*) \geq \sum_j px^*_j + \sum_i q_i y^*, \]
therefore
\[ px_i + q_i y \geq px^*_i + q_i y^*. \]

5. for all \( i \), if \( (x_i, y) \succ (x^*_i, y^*) \), then \( px_i + q_i y > px^*_i + q_i y^* \equiv I_i \).

If the conclusion does not hold, then
\[ px_i + q_i y = px^*_i + q_i y^*. \tag{15.34} \]

Since \( (x_i, y) \succ (x^*_i, y^*) \), when \( 0 < \lambda < 1 \) is close sufficiently to 1, it follows from the continuity of preference that \( (\lambda x_i, \lambda y) \succ (x^*_i, y^*) \). From the conclusion of step 4, we have \( \lambda (px_i + q_i y) \geq px^*_i + q_i y^* = px_i + q_i y \).

Since \( x^* \in \mathcal{R}^{nL}_{++} \), from step 2, we have \( (q_1, \ldots, q_n, p) \geq 0 \) and \( p \neq 0 \), we already know that \( px_i + q_i y = px^*_i + q_i y^* > 0 \), therefore, \( \lambda \geq 1 \), this contradicts the fact that \( \lambda < 1 \).

The proof is thus finished. ■

15.4.6 Free-Rider Problem

When the MRS is known, a Pareto efficient allocation \( (x, y) \) can be determined from the Lindahl-Samuelson condition or the Lindahl solution. After that, the contribution of each consumer is given by \( g_i = w_i - x_i \). However, the society is hard to know the information about MRS. Of course, a naive method is that we could ask each individual to reveal his preferences, and thus determine the willingness-to-pay. However, since each consumer is self-interested, each person wants to be a free-rider and thus is not willing to tell the true MRS. If consumers realize that shares of the contribution for producing public goods (or the personalized prices) depend on their answers, they have “incentives to cheat.” That is, when the consumers are asked to report their utility functions or MRSs, they
will have incentives to report a smaller \( \text{MRS} \) so that they can pay less, and consume the public good (free riders). This causes the major difficulty in the public economies.

To see this, notice that the social goal is to reach Pareto efficient allocations for the public goods economy, but from the personal interest, each person solves the following problem:

\[
\max u_i(x_i, y) \tag{15.35}
\]

subject to

\[
g_i \in [0, w_i] \\
x_i + g_i = w_i \\
y = f(g_i + \sum_{j \neq i} g_j).
\]

That is, each consumer \( i \) takes others’ strategies \( g_{-i} \) as given, and maximizes his payoffs. From this problem, we can form a non-cooperative game:

\[
\Gamma = (G_i, \phi_i)_{i=1}^n
\]

where \( G_i = [0, w_i] \) is the strategy space of consumer \( i \) and \( \phi_i : G_1 \times G_2 \times \ldots \times G_n \to R \) is the payoff function of \( i \) which is defined by

\[
\phi_i(g_i, g_{-i}) = u_i[(w_i - g_i), f(g_i + \sum_{j \neq i} g_j)] \tag{15.36}
\]

**Definition 15.4.6** For the game, \( \Gamma = (G_i, \phi_i)_{i=1}^n \), the strategy \( g^* = (g_1^*, ..., g_n^*) \) is a *Nash Equilibrium* if

\[
\phi_i(g_i^*, g_{-i}^*) \succeq \phi_i(g_i, g_{-i}^*) \quad \text{for all } g_i \in G_i \text{ and all } i = 1, 2, ..., n,
\]

\( g^* \) is a *dominant strategy equilibrium* if

\[
\phi_i(g_i^*, g_{-i}) \succeq \phi_i(g_i, g_{-i}) \quad \text{for all } g \in G \text{ and all } i = 1, 2, ...
\]

**Remark 15.4.3** Note that the difference between Nash equilibrium (NE) and dominant strategy is that at NE, given best strategy of others, each consumer
chooses his best strategy while dominant strategy means that the strategy chosen by each consumer is best regardless of others’ strategies. Thus, a dominant strategy equilibrium is clearly a Nash equilibrium, but the converse may not be true. Only for a very special payoff functions, there is a dominant strategy while a Nash equilibrium exists for a continuous and quasi-concave payoff functions that are defined on a compact set.

For Nash equilibrium, if $u_i$ and $f$ are differentiable, then the first order condition is:

$$\frac{\partial \phi_i(g^*)}{\partial g_i} \leq 0 \quad \text{with equality if } g_i > 0 \quad \text{for all } i = 1, \ldots, n.$$  

(15.37)

Thus, we have

$$\frac{\partial \phi_i}{\partial g_i} = \frac{\partial u_i}{\partial x_i} (-1) + \frac{\partial u_i}{\partial y} f'(g_i^* + \sum_{j \neq i} g_j) \leq 0 \quad \text{with equality if } g_i > 0.$$

So, at an interior solution $g^*$, we have

$$\frac{\partial u_i}{\partial y} = \frac{1}{f'(g_i^* + \sum_{j \neq i} g_j)},$$

and thus

$$MRS_{yx_i}^i = MRTS_{yv},$$

which does not satisfy the Lindahl-Samuelson condition. Thus, the Nash equilibrium in general does not result in Pareto efficient allocations. The above equation implies that the low level of public good is produced rather than the Pareto efficient level of the public good when utility functions are quasi-concave (see Figure 15.1). Therefore, Nash equilibrium allocations are in general not consistent with Pareto efficient allocations. How can one solve this free-ride problem? We will answer this question in the mechanism design theory.

### 15.5 Reference

**Books and Monographs:**

556
Figure 15.1: Free-rider results in a lower provision of public goods than the level of Pareto efficient provision of public goods.


论文：


Part VI

Information, Incentives, and Mechanism Design
Information economics, incentive theory, mechanism design theory, principal-agent theory, contract theory, and auction theory have been very important and active research areas and had wide applications in various fields in economics, finance, management, and corporate law, political sciences in last five decades. Because of this, more than twenty economists of founding contributors of mechanism design and the associated files of game theory so far have been rewarded with the Nobel prize in economics, including Friedrich A. Hayek, Kenneth Arrow, George J. Stigler, Gerand Debreu, Ronald Coase, Herbert Simon, John Nash, Reinhard Selten, William Vickrey, James Mirrlees, George Akerlof, Joseph Stiglitz, Michael Spence, Robert Auman, Leo Hurwicz, Eric Maskin, Roger Myerson, Peter Diamond, Oliver Williamson, Alvin E. Roth, Lloyd S. Shapley, Jean Tirole, Oliver Hart, and Bengt Holmstrom.

The notion of incentives is a basic and key concept in modern economics. To many economists, economics is to a large extent a matter of incentives: incentives to work hard, to produce good quality products, to study, to invest, to save, etc.

Until about 40 year ago, economics was mostly concerned with understanding the theory of value in large economies. A central question asked in general equilibrium theory was whether a certain mechanism (especially the competitive mechanism) generated Pareto-efficient allocations, and if so – for what categories of economic environments. In a perfectly competitive market, the pressure of competitive markets solves the problem of incentives for consumers and producers. The major project of understanding how prices are formed in competitive markets can proceed without worrying about incentives.

The question was then reversed in the economics literature: instead of regarding mechanisms as given and seeking the class of environments for which they work, one seeks mechanisms which will implement some desirable outcomes (especially those which result in Pareto-efficient and individually rational allocations) for a given class of environments without destroying participants’ incentives, and have a low cost of operation and other desirable properties. In a sense, the theorists went back to basics.
The reverse question was stimulated by two major lines in the history of economics. Within the capitalist/private-ownership economics literature, a stimulus arose from studies focusing upon the failure of the competitive market to function as a mechanism for implementing efficient allocations in many nonclassical economic environments such as the presence of externalities, public goods, incomplete information, imperfect competition, increasing return to scale, etc. At the beginning of the seventies, works by Akerlof (1970), Hurwicz (1972), Spence (1974), and Rothschild and Stiglitz (1976) showed in various ways that asymmetric information was posing a much greater challenge and could not be satisfactorily imbedded in a proper generalization of the Arrow-Debreu theory.

A second stimulus arose from the socialist/state-ownership economics literature, as evidenced in the “socialist controversy” — the debate between Mises-Hayek and Lange-Lerner in twenties and thirties of the last century. The controversy was provoked by von Mises’s skepticism as to even a theoretical feasibility of rational allocation under socialism.

The incentives structure and information structure are thus two basic features of any economic system. The study of these two features is attributed to these two major lines, culminating in the theory of mechanism design. The theory of economic mechanism design which was originated by Hurwicz is very general. All economic mechanisms and systems (including those known and unknown, private-ownership, state-ownership, and mixed-ownership systems) can be studied with this theory.

At the micro level, the development of the theory of incentives has also been a major advance in economics in the last forty years. Before, by treating the firm as a black box the theory remains silent on how the owners of firms succeed in aligning the objectives of its various members, such as workers, supervisors, and managers, with profit maximization.

When economists began to look more carefully at the firm, either in agricultural or managerial economics, incentives became the central focus of their analysis. Indeed, delegation of a task to an agent who has different objectives
than the principal who delegates this task is problematic when information about
the agent is imperfect. This problem is the essence of incentive questions. Thus,
conflicting objectives and decentralized information are the two basic
ingredients of incentive theory.

We will discover that, in general, these informational problems prevent society
from achieving the first-best allocation of resources that could be possible in a
world where all information would be common knowledge.\(^1\) The additional costs
that must be incurred because of the strategic behavior of privately informed e-
nomic agents can be viewed as one category of the transaction costs. Although
they do not exhaust all possible transaction costs, economists have been rather
successful during the last forty years in modeling and analyzing these types of
costs and providing a good understanding of the limits set by these on the allo-
cation of resources. This line of research also provides a whole set of insights on
how to begin to take into account agents’ responses to the incentives provided by
institutions.

The three words — contracts, mechanisms and institutions are to a large
extent synonymous. They all mean “rules of the game,” which describe what
actions the parties can undertake, and what outcomes these actions would be
obtained. In most cases the rules of the game are given by designer: in chess,
basketball, etc. The rules are designed to achieve better outcomes. But there
is one difference. While mechanism design theory may be able answer “big”
questions, such as “socialism vs. capitalism,” contract theory is developed and
useful for more manageable smaller questions, concerning specific contracting

---

\(^1\)The term of first-best is relatively to the second-best. The theory of the second best
concerns what happens when one or more optimality conditions (typically the case of incomplete
information) cannot be satisfied. Canadian economist Richard Lipsey and Australian economist
Kelvin Lancaster showed in a 1956 paper that if one optimality condition in an economic model
cannot be satisfied, it is possible that the next-best solution involves changing other variables
away from the ones that are usually assumed to be optimal, see: Lipsey and Lancaster (1956),
2296233.
practices and mechanisms.

Thus, mechanism design is normative economics, in contrast to game theory, which is positive economics. Game theory is important because it predicts how a given game will be played by agents. Mechanism design goes one step further: given the physical environment and the constraints faced by the designer, what goal can be realized or implemented? What mechanisms are optimal among those that are feasible? In designing mechanisms one must take into account incentive constraints (e.g., consumers may not report truthfully how many pairs of shoes they need or how productive they are).

This part considers the design of economic mechanism in which one or many parties have private characteristics or hidden actions. The party who designs the mechanism will be called the designer or principal, while the other parties will be called agents, individuals, or participants. For the most part we will focus on the situation where the designer has no private information and the agents do. This framework is called screening, because the principal will in general try to screen different types of agents by inducing them to choose different bundles. The opposite situation, in which the designer has private information and agents do not, is called signaling, since the designer could signal his type with the design of his contract or mechanism.

We will briefly present the mechanism/contract theory in four chapters. Chapters 16 and 17 consider the principal-agent model where the principal delegates an action to a single agent with private information. This private information can be of two types: either the agent can take an action unobserved by the principal, the case of moral hazard or hidden action; or the agent has some private knowledge about his cost or valuation that is ignored by the principal, the case of adverse selection or hidden knowledge. The theory of optimal contract design considers when this private information is a problem for the principal, and what is the optimal way for the principal to cope with it. The design of the principal’s optimal contract can be regarded as a simple optimization problem. This simple focus will turn out to be enough to highlight the various
trade-offs between allocative efficiency and distribution of information
rents arising under incomplete information. The mere existence of informational constraints may generally prevent the principal from achieving allocative efficiency. We will characterize the allocative distortions that the principal finds desirable to implement in order to mitigate the impact of informational constraints. We should acknowledge that the materials in Chapters 16 and 17 notes are mainly drawn from Laffont and Martimort (2002).

Chapter 18 will consider situations with one principal and many agents. Moreover, maintaining the hypothesis that agents adopt an individualistic behavior, those organizational contexts require a solution concept of equilibrium, which describes the strategic interaction between agents under complete information.

Chapter 16 will discuss the case of incomplete information in which asymmetric information may not only affect the relationship between the principal and each of his agents, but it may also plague the relationships between agents. As such, agents do not know each other’s characteristics, and we need to consider Bayesian incentive compatible mechanism.

Chapter 17 will briefly study dynamic contract theory. We will discuss long-term incentive contracting in a dynamic principal-agent setting with one-agent and adverse selection. We will first consider the case where the principal (designer) can commit to a contract forever, and then consider what happens when she cannot commit against modifying the contract as new information arrives.
Chapter 16

Optimal Mechanism Design: Contracts with One-Agent and Hidden Information

16.1 Introduction

The optimal contract theory with two parties in large extent is also called principal-agent theory. One party is a principal and the other is an agent.

Incentive problems arise when a principal wants to delegate a task to an agent with private information. The exact opportunity cost of this task, the precise technology used, and how good the matching is between the agent’s intrinsic ability and this technology are all examples of pieces of information that may become private knowledge of the agent. In such cases, we will say that there is adverse selection.

Examples

1. The landlord delegates the cultivation of his land to a tenant, who will be the only one to observe the exact local weather conditions.

2. A client delegates his defense to an attorney who will be the only one to know the difficulty of the case.
3. An investor delegates the management of his portfolio to a broker, who will privately know the prospects of the possible investments.

4. A stockholder delegates the firm’s day-to-day decisions to a manager, who will be the only one to know the business conditions.

5. An insurance company provides insurance to agents who privately know how good a driver they are.

6. The Department of Defense procures a good from the military industry without knowing its exact cost structure.

7. A regulatory agency contracts for service with a public utility company without having complete information about its technology.

The common aspect of all those contracting settings is that the information gap between the principal and the agent has some fundamental implications for the design of the contract they sign. In order to reach an efficient use of economic resources, some information rent must be given up to the privately informed agent. At the optimal second-best contract, the principal trades off his desire to reach allocative efficiency against the costly information rent given up to the agent to induce information revelation. Implicit here is the idea that there exists a legal framework for this contractual relationship. The contract can be enforced by a benevolent court of law, the agent is bounded by the terms of the contract.

The main objective of this chapter is to characterize the optimal rent extraction-efficiency trade-off faced by the principal when designing his contractual offer to the agent under the set of incentive feasible constraints: incentive and participation constraints. In general, incentive constraints are binding at the optimum, showing that adverse selection clearly impedes the efficiency of trade. The main lessons of this optimization is that the optimal second-best contract calls for a distortion in the volume of trade away from the first-best and for giving up some strictly positive information rents to the most efficient agents.

\footnote{The term of second-best is relatively to the first-best. For the detailed explanation, see Footnote 1 in the introduction of Part VI.}
16.2 Basic Settings of Principal-Agent Model with Adverse Selection

16.2.1 Economic Environment (Technology, Preferences, and Information)

Consider a consumer or a firm (the principal) who wants to delegate to an agent the production of \( q \) units of a good. The value for the principal of these \( q \) units is \( S(q) \) where \( S' > 0 \), \( S'' < 0 \) and \( S(0) = 0 \).

The production cost of the agent is unobservable to the principal, but it is common knowledge that the fixed cost is \( F \) and the marginal cost belongs to the set \( \Phi = \{ \theta, \bar{\theta} \} \). The agent can be either efficient (\( \theta \)) or inefficient (\( \bar{\theta} \)) with respective probabilities \( \nu \) and \( 1 - \nu \). That is, he has the cost function

\[
C(q, \theta) = \theta q + F \quad \text{with probability } \nu \quad (16.1)
\]

or

\[
C(q, \bar{\theta}) = \bar{\theta} q + F \quad \text{with probability } 1 - \nu \quad (16.2)
\]

Denote by \( \Delta \theta = \bar{\theta} - \theta > 0 \) the spread of uncertainty on the agent’s marginal cost. This information structure is exogenously given to the players.

16.2.2 Contracting Variables: Outcomes

The contracting variables are the quantity produced \( q \) and the transfer \( t \) received by the agent. Let \( \mathcal{A} \) be the set of feasible allocations that is given by

\[
\mathcal{A} = \{(q, t) : q \in \mathbb{R}_+, t \in \mathbb{R}\} \quad (16.3)
\]

These variables are both observable and verifiable by a third party such as a benevolent court of law.
16.2.3 Timing

Unless explicitly stated, we will maintain the timing defined in the figure below, where \( A \) denotes the agent and \( P \) the principal.

![Diagram showing timing of contracting under hidden information]

Note that contracts are offered at the interim stage; there is already asymmetric information between the contracting parties when the principal makes his offer. We will briefly discuss the optimal contract at the ex ante stage and the ex post stage.

16.3 Complete Information Optimal Contract (Benchmark)

16.3.1 First-Best Production Levels

To get a reference system for comparison, let us first suppose that there is no asymmetry of information between the principal and the agent. The efficient (first-best) production levels are obtained by equating the principal’s marginal value and the agent’s marginal cost. Hence, we have the following first-order conditions

\[
S'(q^*) = \theta
\]

and

\[
S'(-q^*) = \bar{\theta}.
\]
The complete information efficient production levels $q^*$ and $\bar{q}^*$ should be both carried out if their social values, respectively $W^* = S(q^*) - \theta q^* - F$, and $\bar{W}^* = S(\bar{q}^*) - \bar{\theta} \bar{q}^* - F$, are non-negative.

Since

$$S(q^*) - \theta q^* \geq S(\bar{q}^*) - \bar{\theta} \bar{q}^* \geq S(\bar{q}^*) - \bar{\theta} \bar{q}^*$$

by definition of $\theta$ and $\bar{\theta} > \theta$, the social value of production when the agent is efficient, $W^*$, is greater than when he is inefficient, namely $\bar{W}^*$.

For trade to be always carried out, it is thus enough that production be socially valuable for the least efficient type, i.e., the following condition must be satisfied

$$W^* = S(q^*) - \theta q^* - F \geq 0.$$ (16.6)

As the fixed cost $F$ plays no role other than justifying the existence of a single agent, it is set to zero from now on in order to simplify notations.

Note that, since the principal’s marginal value of output is decreasing, the optimal production of an efficient agent is greater than that of an inefficient agent, i.e., $q^* > \bar{q}^*$.

### 16.3.2 Implementation of the First-Best

For a successful delegation of the task, the principal must offer the agent a utility level that is at least as high as the utility level that the agent obtains outside the relationship. We refer to these constraints as the agent’s participation constraints. If we normalize to zero the agent’s outside opportunity utility level (sometimes called his quo utility level), these participation constraints are written as

$$\ell - \theta q \geq 0,$$ (16.7)

$$\bar{\ell} - \bar{\theta} \bar{q} \geq 0.$$ (16.8)

To implement the first-best production levels, the principal can make the following take-it-or-leave-it offers to the agent: If $\theta = \bar{\theta}$ (resp. $\theta$), the principal
offers the transfer $\bar{t}^*$ (resp. $t^*$) for the production level $\bar{q}^*$ (resp. $q^*$) with $\bar{t}^* = \bar{\theta}\bar{q}^*$ (resp. $t^* = \theta q^*$). Thus, whatever his type, the agent accepts the offer and makes zero profit. The complete information optimal contracts are thus $(\bar{t}^*, \bar{q}^*)$ if $\theta = \bar{\theta}$ and $(\bar{t}^*, q^*)$ if $\theta = \bar{\theta}$. Importantly, under complete information delegation is costless for the principal, who achieves the same utility level that he would get if he was carrying out the task himself (with the same cost function as the agent).

**16.3.3 A Graphical Representation of the Complete Information Optimal Contract**

Since $\bar{\theta} > \theta$, the iso-utility curves for different types cross only once as shown in the above figure. This important property is called the single-crossing or Spence-
Mirrlees property.

The complete information optimal contract is finally represented Figure 16.3 by the pair of points \((A^*, B^*)\). Note that since the iso-utility curves of the principal correspond to increasing levels of utility when one moves in the southeast direction, the principal reaches a higher profit when dealing with the efficient type. We denote by \(\bar{V}^*\) (resp. \(V^*\)) the principal’s level of utility when he faces the \(\bar{\theta}–\) (resp. \(\theta–\)) type. Because the principal’s has all the bargaining power in designing the contract, we have \(\bar{V}^* = \bar{W}^*\) (resp. \(V^* = W^*\)) under complete information.
16.4 Incentive Feasible Contracts

16.4.1 Incentive Compatibility and Participation

Suppose now that the marginal cost $\theta$ is the agent’s private information and let us consider the case where the principal offers the menu of contracts $\{(t^*, q^*); (\bar{t}^*, \bar{q}^*)\}$ hoping that an agent with type $\theta$ will select $(t^*, q^*)$ and an agent with $\bar{\theta}$ will select instead $(\bar{t}^*, \bar{q}^*)$.

From Figure 16.3 above, we see that $B^*$ is preferred to $A^*$ by both types of agents. Offering the menu $(A^*, B^*)$ fails to have the agents self-selecting properly within this menu. The efficient type have incentives to mimic the inefficient one and selects also contract $B^*$. The complete information optimal contracts can no longer be implemented under asymmetric information. We will thus say that the menu of contracts $\{(t^*, q^*); (\bar{t}^*, \bar{q}^*)\}$ is not incentive compatible.

**Definition 16.4.1** A menu of contracts $\{(t, q); (\bar{t}, \bar{q})\}$ is incentive compatible when $(t, q)$ is weakly preferred to $(\bar{t}, \bar{q})$ by agent $\theta$ and $(\bar{t}, \bar{q})$ is weakly preferred to $(t, q)$ by agent $\bar{\theta}$.

Mathematically, these requirements amount to the fact that the allocations must satisfy the following incentive compatibility constraints:

\[
t - \theta q \geq \bar{t} - \bar{\theta} \bar{q} \tag{16.9}
\]

and

\[
\bar{t} - \bar{\theta} \bar{q} \geq t - \theta q \tag{16.10}
\]

Furthermore, for a menu to be accepted, it must satisfy the following two participation constraints:

\[
t - \theta q \geq 0, \tag{16.11}
\]

\[
\bar{t} - \bar{\theta} \bar{q} \geq 0. \tag{16.12}
\]

**Definition 16.4.2** A menu of contracts is incentive feasible if it satisfies both incentive and participation constraints (16.9) through (16.12).
The inequalities (16.9) through (16.12) fully characterize the set of incentive feasible menus of contracts. The restrictions embodied in this set express additional constraints imposed on the allocation of resources by asymmetric information between the principal and the agent.

### 16.4.2 Special Cases

**Bunching or Pooling Contracts:** A first special case of incentive feasible menu of contracts is obtained when the contracts targeted for each type coincide, i.e., when \( t = \bar{t} = t^p, q = \bar{q} = q^p \) and both types of agent accept this contract.

**Shutdown of the Least Efficient Type:** Another particular case occurs when one of the contracts is the null contract \((0,0)\) and the nonzero contract \((t^*, q^*)\) is only accepted by the efficient type. Then, (16.9) and (16.11) both reduce to

\[
\begin{align*}
    t^* - \theta q^* & \geq 0, \quad (16.13) \\
    0 & \geq t^* - \bar{\theta} q^*. \quad (16.14)
\end{align*}
\]

The incentive constraint of the bad type reduces to

\[
0 \geq t^* - \bar{\theta} q^*.
\]

As with the pooling contract, the benefit of the \((0,0)\) option is that it somewhat reduces the number of constraints since the incentive and participation constraints take the same form. The cost of such a contract may be an excessive screening of types. Here, the screening of types takes the rather extreme form of the least efficient type.

### 16.4.3 Monotonicity Constraints

Incentive compatibility constraints reduce the set of feasible allocations. Moreover, these quantities must generally satisfy a monotonicity constraint which does not exist under complete information. Adding (16.9) and (16.10), we immediately have

\[
q \geq \bar{q}. \quad (16.15)
\]
We will call condition (16.15) an implementability condition that is necessary and sufficient for implementability.

Indeed, suppose that (16.15) holds; it is clear that there exists transfers $\bar{t}$ and $\bar{t}$ such that the incentive constraints (16.9) and (16.10) both hold. It is enough to take those transfers such that

$$\theta(q - \bar{q}) \leq \bar{t} - \bar{t} \leq \bar{\theta}(q - \bar{q}).$$

(16.16)

**Remark 16.4.1** In this two-type model, the conditions for implementability take a simple form. With more than two types (or with a continuum), the characterization of these conditions might get harder. The conditions for implementability are also more difficult to characterize when the agent performs several tasks on behalf of the principal.

### 16.5 Information Rents

To understand the structure of the optimal contract it is useful to introduce the concept of information rent.

We know from previous discussion, under complete information, the principal is able to maintain all types of agents at their zero status quo utility level. Their respective utility levels $U^*$ and $\bar{U}^*$ at the first-best satisfy

$$U^* = t^* - \theta q^* = 0$$

(16.17)

and

$$\bar{U}^* = \bar{t}^* - \bar{\theta} \bar{q}^* = 0.$$  

(16.18)

Generally this will not be possible anymore under incomplete information, at least when the principal wants both types of agents to be active.

We use the notations $\underline{U} = t - \theta q$ and $\bar{U} = \bar{t} - \bar{\theta} \bar{q}$ to denote the respective information rent of each type.
Take any menu \( \{(\bar{t}, \bar{q}); (t, q)\} \) of incentive feasible contracts. How much would a \( \theta \)-agent get by mimicking a \( \bar{\theta} \)-agent? The high-efficient agent would get

\[
\bar{t} - \theta \bar{q} = \bar{t} - \bar{\theta} \bar{q} + \Delta \theta \bar{q} = \bar{U} + \Delta \theta \bar{q}.
\] (16.19)

Thus, as long as the principal insists on a positive output for the inefficient type, \( \bar{q} > 0 \), the principal must give up a positive rent to a \( \theta \)-agent. This information rent is generated by the informational advantage of the agent over the principal.

### 16.6 The Optimization Program of the Principal

According to the timing of the contractual game, the principal must offer a menu of contracts before knowing which type of agent he is facing. Then, the principal’s problem writes as

\[
\max_{\{(t, q); (\bar{t}, \bar{q})\}} \nu(S(q) - \bar{t}) + (1 - \nu)(S(\bar{q}) - \bar{t})
\]

subject to (16.9) to (16.12).

Using the definition of the information rents \( U = t - \theta q \) and \( \bar{U} = \bar{t} - \bar{\theta} \bar{q} \), we can replace transfers in the principal’s objective function as functions of information rents and outputs so that the new optimization variables are now \( \{(U, q); (\bar{U}, \bar{q})\} \).

The focus on information rents enables us to assess the distributive impact of asymmetric information, and the focus on outputs allows us to analyze its impact on allocative efficiency and the overall gains from trade. Thus an allocation corresponds to a volume of trade and a distribution of the gains from trade between the principal and the agent.

With this change of variables, the principal’s objective function can then be rewritten as

\[
\nu(S(q) - \theta q) + (1 - \nu)(S(\bar{q}) - \bar{\theta} \bar{q}) - (\nu U + (1 - \nu) \bar{U}).
\] (16.20)
The first term denotes expected allocative efficiency, and the second term denotes expected information rent which implies that the principal is ready to accept some distortions away from efficiency in order to decrease the agent’s information rent.

The incentive constraints (16.9) and (16.10), written in terms of information rents and outputs, becomes respectively

\[ U \geq \bar{U} + \Delta \theta \bar{q}, \]  
\[ \bar{U} \geq U - \Delta \theta q. \]  

(16.21) \hspace{1cm} (16.22)

The participation constraints (16.11) and (16.12) become respectively

\[ U \geq 0, \]  
\[ \bar{U} \geq 0. \]  

(16.23) \hspace{1cm} (16.24)

The principal wishes to solve problem \((P)\) below:

\[
\max \{ (U, q); (\bar{U}, \bar{q}) \}
\nu (S(q) - \theta q) + (1 - \nu) (S(\bar{q}) - \bar{\theta} \bar{q}) - (\nu U + (1 - \nu) \bar{U})
\]

subject to (16.21) to (16.24).

We index the solution to this problem with a superscript \(SB\), meaning second-best.

### 16.7 The Rent Extraction-Efficiency Trade-Off

#### 16.7.1 The Optimal Contract Under Asymmetric Information

The major technical difficulty of problem \((P)\) is to determine which of the many constraints imposed by incentive compatibility and participation are the relevant ones, i.e., the binding ones at the optimum or the principal’s problem.

Let us first consider contracts without shutdown, i.e., such that \(\bar{q} > 0\). This is true when the so-called Inada condition \(S'(0) = +\infty\) is satisfied and \(\lim_{q \to 0} S'(q)q = 0\).
Note that the $\theta$-agent’s participation constraint (16.23) is always strictly-satisfied. Indeed, (16.24) and (16.21) immediately imply (16.23). (16.22) also seems irrelevant because the difficulty comes from a $\theta$-agent willing to claim that he is inefficient rather than the reverse.

This simplification in the number of relevant constraints leaves us with only two remaining constraints, the $\theta$-agent’s incentive compatible constraint (16.21) and the $\tilde{\theta}$-agent’s participation constraint (16.24), and both constraints must be binding at the optimum of the principal’s problem ($P$):

\[ U = \Delta \theta \bar{q} \quad (16.25) \]

and

\[ \bar{U} = 0. \quad (16.26) \]

Substituting (16.25) and (16.26) into the principal’s objective function, we obtain a reduced program ($P'$) with outputs as the only choice variables:

\[
\max_{\{q, \bar{q}\}} \nu(S(q) - \theta q) + (1 - \nu)(S(\bar{q}) - \bar{\theta} \bar{q}) - (\nu \Delta \theta \bar{q}).
\]

Compared with the full information setting, asymmetric information alters the principal’s optimization simply by the subtraction of the expected rent that has to be given up to the efficient type. The inefficient type gets no rent, but the efficient type $\theta$ gets information rent that he could obtain by mimicking the inefficient type $\tilde{\theta}$. This rent depends only on the level of production requested from this inefficient type.

The first order conditions are then given by

\[ S'(\bar{q}^{SB}) = \theta \quad \text{or} \quad \bar{q}^{SB} = \bar{q}^*. \quad (16.27) \]

and

\[ (1 - \nu)(S'(\bar{q}^{SB}) - \bar{\theta}) = \nu \Delta \theta. \quad (16.28) \]

(16.28) expresses the important trade-off between efficiency and rent extraction which arises under asymmetric information.
To validate our approach based on the sole consideration of the efficient type’s incentive compatible constraint, it is necessary to check that the omitted incentive compatible constraint of an inefficient agent is satisfied. i.e., \(0 \geq \Delta \theta \bar{q}^{SB} - \Delta \bar{q}^{SB}\).

This latter inequality follows from the monotonicity of the second-best schedule of outputs since we have \(q^{SB} = q^* > \bar{q}^* > \bar{q}^{SB}\).

In summary, we have the following proposition.

**Proposition 16.7.1** Under asymmetric information, the optimal contracts entail:

1. **No output distortion for the efficient type in respect to the first-best**, \(q^{SB} = q^*\). A downward output distortion for the inefficient type, \(\bar{q}^{SB} < \bar{q}^*\) with

   \[S'(\bar{q}^{SB}) = \bar{\theta} + \frac{\nu}{1 - \nu} \Delta \theta.\]  

   (16.29)

2. **Only the efficient type gets a positive information rent given by**

   \[U^{SB} = \Delta \theta \bar{q}^{SB}.\]  

   (16.30)

3. **The second-best transfers are respectively given by**

   \(t^{SB} = \theta q^* + \Delta \theta \bar{q}^{SB}\) and \(\bar{t}^{SB} = \bar{\theta} \bar{q}^{SB}\).

**Remark 16.7.1** The basic idea and insides of principal-agent theory was somewhat revealed in the book of “the Art of War” written by Sun Tzu in the 6-th century BC. It was the ancient Chines military treatise. The book was China’s earliest and most outstanding and complete work on warcraft. Dubbed “the Bible of Military Science”, it is also the earliest work on military strategies in the world. He considered the critical importance of information by the saying that: “If you know the enemy and know yourself, you need not fear the result of a hundred battles. If you know yourself but not the enemy, for every victory gained you will also suffer a defeat. If you know neither the enemy nor yourself, you will succumb in every battle.” That is, the best you can do is the first best when information is complete, the best you can do is the second best when information is asymmetric. It would be worse you know thing about the others and yourself.
16.7.2 A Graphical Representation of the Second-Best Outcome

Starting from the complete information optimal contract \((A^*, B^*)\) that is not incentive compatible, we can construct an incentive compatible contract \((B^*, C)\) with the same production levels by giving a higher transfer to the agent producing \(q^*\) as shown in the figure above. The contract \(C\) is on the \(\theta\)-agent’s indifference curve passing through \(B^*\). Hence, the \(\theta\)-agent is now indifferent between \(B^*\) and \(C\). \((B^*, C)\) becomes an incentive-compatible menu of contracts. The rent that is given up to the \(\theta\)-firm is now \(\Delta \theta q^*\). This contract is not optimal by the
first order conditions (16.27) and (16.28). The optimal trade-off finally occurs at 
\((A^{SB}, B^{SB})\) as shown in the figure below.

Figure 16.5: Optimal second-best contract \(S^{SB}\) and \(B^{SB}\).

16.7.3 Shutdown Policy

If the first-order condition in (16.29) has no positive solution, \(\bar{q}^{SB}\) should be set at zero. We are in the special case of a contract with shutdown. \(B^{SB}\) coincides with 0 and \(A^{SB}\) with \(A^\ast\) in the figure above. No rent is given up to the \(\theta\)-firm by the unique non-null contract \((t^\ast, q^\ast)\) offered and selected only by agent \(\theta^\ast\). The benefit of such a policy is that no rent is given up to the efficient type.
Remark 16.7.2 The shutdown policy is dependent on the status quo utility levels. Suppose that, for both types, the status quo utility level is $U_0 > 0$. Then, from the principal’s objective function, we have

$$\frac{\nu}{1 - \nu} \Delta \theta \bar{q}^{SB} + U_0 \geq S(\bar{q}^{SB}) - \bar{q}^{SB}. \quad (16.31)$$

Thus, for $\nu$ large enough, shutdown occurs even if the Inada condition $S'(0) = +\infty$ is satisfied. Note that this case also occurs when the agent has a strictly positive fixed cost $F > 0$ (to see that, just set $U_0 = F$).

The occurrence of shutdown can also be interpreted as saying that the principal has another choice variable to solve the screening problem. This extra variable is the subset of types, which are induced to produce a positive amount. Reducing the subset of producing agents obviously reduces the rent of the most efficient type.

16.8 The Theory of the Firm Under Asymmetric Information

When the delegation of task occurs within the firm, a major conclusion of the above analysis is that, because of asymmetric information, the firm does not maximize the social value of trade, or more precisely its profit, a maintained assumption of most economic theory. This lack of allocative efficiency should not be considered as a failure in the rational use of resources within the firm. Indeed, the point is that allocative efficiency is only one part of the principal’s objective. The allocation of resources within the firm remains constrained optimal once informational constraints are fully taken into account.

Williamson (1975) has advanced the view that various transaction costs may impede the achievement of economic transactions. Among the many origins of these costs, Williamson stresses informational impact as an important source of inefficiency. Even in a world with a costless enforcement of contracts, a ma-
A source of allocative inefficiency is the existence of asymmetric information between trading partners.

Even though asymmetric information generates allocative inefficiencies, those efficiencies do not call for any public policy motivated by reasons of pure efficiency. Indeed, any benevolent policymaker in charge of correcting these inefficiencies would face the same informational constraints as the principal. The allocation obtained above is Pareto optimal in the set of incentive feasible allocations or incentive Pareto optimal.

### 16.9 Asymmetric Information and Marginal Cost Pricing

Under complete information, the first-best rules can be interpreted as price equal to marginal cost since consumers on the market will equate their marginal utility of consumption to price.

Under asymmetric information, price equates marginal cost only when the producing firm is efficient \((θ = \bar{θ})\). Using (16.29), we get the expression of the price \(p(\bar{θ})\) for the inefficient types output

\[
p(\bar{θ}) = \bar{θ} + \frac{ν}{1 - ν} \Delta θ.
\]

(16.32)

Price is higher than marginal cost in order to decrease the quantity \(\bar{q}\) produced by the inefficient firm and reduce the efficient firm’s information rent. Alternatively, we can say that price is equal to a generalized (or virtual) marginal cost that includes, in addition to the traditional marginal cost of the inefficient type \(\bar{θ}\), an information cost that is worth \(\frac{ν}{1 - ν} \Delta θ\).

### 16.10 The Revelation Principle

In the above analysis, we have restricted the principal to offer a menu of contracts, one for each possible type. One may wonder if a better outcome could be achieved
with a more complex contract allowing the agent possibly to choose among more options. The revelation principle ensures that there is no loss of generality in restricting the principal to offer simple menus having at most as many options as the cardinality of the type space. Those simple menus are actually examples of direct revelation mechanisms.

Definition 16.10.1 A direct revelation mechanism is a mapping $g(\cdot)$ from $\Theta$ to $A$ which writes as $g(\theta) = (q(\theta), t(\theta))$ for all belonging to $\Theta$. The principal commits to offer the transfer $t(\tilde{\theta})$ and the production level $q(\tilde{\theta})$ if the agent announces the value $\tilde{\theta}$ for any $\tilde{\theta}$ belonging to $\Theta$.

Definition 16.10.2 A direct revelation mechanism $g(\cdot)$ is truthful if it is incentive compatible for the agent to announce his true type for any type, i.e., if the direct revelation mechanism satisfies the following incentive compatibility constraints:

$$t(\theta) - \theta q(\theta) \geq t(\bar{\theta}) - \bar{\theta} q(\bar{\theta}),$$  \hspace{1cm} (16.33)  

$$t(\bar{\theta}) - \bar{\theta} q(\bar{\theta}) \geq t(\theta) - \theta q(\theta).$$  \hspace{1cm} (16.34)

Denoting transfer and output for each possible report respectively as $t(\theta) = t$, $q(\theta) = q$, $t(\bar{\theta}) = \bar{t}$ and $q(\bar{\theta}) = \bar{q}$, we get back to the notations of the previous sections.

A more general mechanism can be obtained when communication between the principal and the agent is more complex than simply having the agent report his type to the principal.

Let $M$ be the message space offered to the agent by a more general mechanism.

Definition 16.10.3 A mechanism is a message space $M$ and a mapping $\tilde{g}(\cdot)$ from $M$ to $A$ which writes as $\tilde{g}(m) = (\tilde{q}(m), \tilde{t}(m))$ for all $m$ belonging to $M$.

When facing such a mechanism, the agent with type $\theta$ chooses a best message $m^*(\theta)$ that is implicitly defined as

$$\tilde{t}(m^*(\theta)) - \theta \tilde{q}(m^*(\theta)) \geq \tilde{t}(\tilde{m}) - \theta \tilde{q}(\tilde{m}) \text{ for all } \tilde{m} \in M.$$  \hspace{1cm} (16.35)
The mechanism \((M, \tilde{g}(\cdot))\) induces therefore an allocation rule \(a(\theta) = (\tilde{q}(m^*(\theta)), \tilde{t}(m^*(\theta)))\) mapping the set of types \(\Theta\) into the set of allocations \(\mathcal{A}\).

Then we have the following revelation principle in the one agent case.

**Proposition 16.10.1** Any allocation rule \(a(\theta)\) obtained with a mechanism \((M, \tilde{g}(\cdot))\) can also be implemented with a truthful direct revelation mechanism.

**Proof.** The indirect mechanism \((M, \tilde{g}(\cdot))\) induces an allocation rule \(a(\theta) = (\tilde{q}(m^*(\theta)), \tilde{t}(m^*(\theta)))\) from \(M\) into \(\mathcal{A}\). By composition of \(\tilde{q}(\cdot)\) and \(m^*(\cdot)\), we can construct a direct revelation mechanism \(g(\cdot)\) mapping \(\Theta\) into \(\mathcal{A}\), namely \(g = \tilde{g} \circ m^*,\) or more precisely \(g(\theta) = (q(\theta), t(\theta)) = \tilde{g}(m^*(\theta)) = (\tilde{q}(m^*(\theta)), \tilde{t}(m^*(\theta)))\) for all \(\theta \in \Theta\).

We check now that the direct revelation mechanism \(g(\cdot)\) is truthful. Indeed, since \((16.35)\) is true for all \(\tilde{m}\), it holds in particular for \(\tilde{m} = m^*(\theta')\) for all \(\theta' \in \Theta\). Thus we have

\[
\tilde{t}(m^*(\theta)) - \theta \tilde{q}(m^*(\theta)) \geq \tilde{t}(m^*(\theta')) - \theta \tilde{q}(m^*(\theta')) \quad \text{for all } (\theta, \theta') \in \Theta^2. \tag{16.36}
\]

Finally, using the definition of \(g(\cdot)\), we get

\[
t(\theta) - \theta q(\theta) \geq t(\theta') - \theta q(\theta') \quad \text{for all } (\theta, \theta') \in \Theta^2. \tag{16.37}
\]

Hence, the direct revelation mechanism \(g(\cdot)\) is truthful.

Importantly, the revelation principle provides a considerable simplification of contract theory. It enables us to restrict the analysis to a simple aid well-defined family of functions, the truthful direct revelation mechanism.

The revelation principle can be illustrated by the following figure.

### 16.11 A More General Utility Function for the Agent

Still keeping quasi-linear utility functions, let \(U = t - C(q, \theta)\) now be the agent’s objective function in the assumptions: \(C_q > 0, C_\theta > 0, C_{qq} > 0\) and \(C_{q\theta} > 0\).
The generalization of the Spence-Mirrlees property is now $C_{\theta\theta} > 0$. This latter condition still ensures that the different types of the agent have indifference curves which cross each other at most once. This Spence-Mirrlees property is quite clear: a more efficient type is also more efficient at the margin.

Incentive feasible allocations satisfy the following incentive and participation constraints:

$$U = t - C(q, \theta) \geq \bar{t} - C(\bar{q}, \bar{\theta}), \quad (16.38)$$

$$U = \bar{t} - C(\bar{q}, \bar{\theta}) \geq t - C(q, \bar{\theta}), \quad (16.39)$$

$$\bar{U} = t - C(q, \bar{\theta}) \geq 0, \quad (16.40)$$

$$\bar{\bar{U}} = \bar{t} - C(\bar{q}, \bar{\theta}) \geq 0. \quad (16.41)$$

### 16.11.1 The Optimal Contract

Just as before, the incentive constraint of an efficient type in (16.38) and the participation constraint of an inefficient type in (16.41) are the two relevant constraints for optimization. These constraints rewrite respectively as

$$U \geq \bar{U} + \Phi(\bar{q}) \quad (16.42)$$

where $\Phi(\bar{q}) = C(\bar{q}, \bar{\theta}) - C(\bar{q}, \bar{\theta})$ (with $\Phi' > 0$ and $\Phi'' > 0$), and

$$\bar{U} \geq 0. \quad (16.43)$$

Those constraints are both binding at the second-best optimum, which leads to the following expression of the efficient type’s rent.
\[ U = \Phi(q). \]  

(16.44)

Since \( \Phi' > 0 \), reducing the inefficient agent’s output also reduces the efficient agent’s information rent.

With the assumptions made on \( C(\cdot) \), one can also check that the principal’s objective function is strictly concave with respect to outputs.

The solution of the principal’s program can be summarized as follows:

**Proposition 16.11.1** With general preferences satisfying the Spence-Mirrlees property, \( C_{q\theta} > 0 \), the optimal menu of contracts entails:

1. No output distortion with respect to the first-best outcome for the efficient type, \( \underline{q}^{SB} = \bar{q}^* \) with
   \[ S'(\bar{q}^*) = C_q(\bar{q}^*, \theta). \]  
   (16.45)

   A downward output distortion for the inefficient type, \( \bar{q}^{SB} < \bar{q}^* \) with
   \[ S'(\bar{q}^*) = C_q(\bar{q}^*, \bar{\theta}) \]  
   (16.46)

   and
   \[ S'(\bar{q}^{SB}) = C_q(\bar{q}^{SB}, \bar{\theta}) + \frac{\nu}{1 - \nu} \Phi'(\bar{q}^{SB}). \]  
   (16.47)

2. Only the efficient type gets a positive information rent given by
   \[ U^{SB} = \Phi(\bar{q}^{SB}). \]

3. The second-best transfers are respectively given by \( \bar{t}^{SB} = C(\bar{q}^*, \theta) + \Phi(\bar{q}^{SB}) \) and \( \bar{t}^{SB} = C(\bar{q}^{SB}, \bar{\theta}) \).

The first-order conditions (16.45) and (16.47) characterize the optimal solution if the neglected incentive constraint (16.39) is satisfied. For this to be true, we need to have

\[ \bar{t}^{SB} - C(\bar{q}^{SB}, \bar{\theta}) \geq t^{SB} - C(\bar{q}^{SB}, \bar{\theta}), \]

\[ = \bar{t}^{SB} - C(\bar{q}^{SB}, \bar{\theta}) + C(\bar{q}^{SB}, \bar{\theta}) - C(\bar{q}^{SB}, \bar{\theta}) \]  

(16.48)
by noting that (16.38) holds with equality at the optimal output such that \( t^{SB} = \bar{t}^{SB} - C(q^{SB}, \bar{\theta}) + C(q^{SB}, \bar{\theta}) \). Thus, we need to have

\[
0 \geq \Phi(q^{SB}) - \Phi(\bar{q}^{SB}).
\]  

(16.49)

Since \( \Phi' > 0 \) from the Spence-Mirrlees property, then (16.49) is equivalent to \( \bar{q}^{SB} \leq q^{SB} \). But from our assumptions we easily derive that \( q^{SB} = q^* \geq \bar{q}^{SB} \). So the Spence-Mirrlees property guarantees that only the efficient type’s incentive compatible constraint has to be taken into account.

**Remark 16.11.1** When there are the joint presence of asymmetric information and network externalities or nonzero type-dependent reservation utility, “no distortion on the top” rule may not be true so that the most efficient agent may not have the first-best outcome. Meng and Tian (2008) showed that the optimal contract for the principal may exhibit two-way distortion: the output of any agent is oversupplied (relative to the first-best) when his marginal cost of effort is low, and undersupplied when his marginal cost of effort is high. We will introduce these results in Subsection 13.6 of this chapter.

### 16.11.2 More than One Good

Let us now assume that the agent is producing a whole vector of goods \( q = (q_1, \ldots, q_n) \) for the principal. The agents’ cost function becomes \( C(q, \theta) \) with \( C(\cdot) \) being strictly convex in \( q \). The value for the principal of consuming this whole bundle is now \( S(q) \) with \( S(\cdot) \) being strictly concave in \( q \).

In this multi-output incentive problem, the principal is interested in a whole set of activities carried out simultaneously by the agent. It is straightforward to check that the efficient agent’s information rent is now written as \( U = \Phi(q) \) with \( \Phi(q) = C(q, \bar{\theta}) - C(q, \bar{\theta}) \). This leads to second-best optimal outputs. The efficient type produces the first-best vector of outputs \( q^{SB} = q^* \) with

\[
S_{qi}(q^*) = C_{qi}(q^*, \bar{\theta}) \quad \text{for all } i \in \{1, \ldots, n\}.
\]

(16.50)
The inefficient types vector of outputs $\tilde{q}^{SB}$ is instead characterized by the first-order conditions

$$S_{q_i}(\tilde{q}^{SB}) = C_{q_i}(\tilde{q}^{SB}, \bar{\theta}) + \frac{\nu}{1-\nu}\Phi_{q_i}(\tilde{q}^{SB})$$

for all $i \in \{1, \ldots, n\}$, \hspace{1cm} (16.51)

which generalizes the distortion of models with a single good.

Without further specifying the value and cost functions, the second-best outputs define a vector of outputs with some components $\tilde{q}_i^{SB}$ above $\bar{q}_i^*$ for a subset of indices $i$.

Turning to incentive compatibility, summing the incentive constraints $\bar{U} \geq \bar{U} + \Phi(\bar{q})$ and $\bar{U} \geq U - \Phi(q)$ for any incentive feasible contract yields

$$\Phi(q) = C(q, \bar{\theta}) - C(q, \bar{\theta}) \geq C(q, \bar{\theta}) - C(\bar{q}, \bar{\theta}) = \Phi(\bar{q})$$

for all implementable pairs $(\bar{q}, q)$. \hspace{1cm} (16.52)

Obviously, this condition is satisfied if the Spence-Mirrlees property $C_{q_i, \theta} > 0$ holds for each output $i$ and if the monotonicity conditions $\bar{q}_i < q_i$ for all $i$ are satisfied.

### 16.12 Ex Ante versus Ex Post Participation Constraints

The case of contracts we consider so far is offered at the interim stage, i.e., the agent already knows his type. However, sometimes the principal and the agent can contract at the ex ante stage, i.e., before the agent discovers his type. For instance, the contracts of the firm may be designed before the agent receives any piece of information on his productivity. In this section, we characterize the optimal contract for this alternative timing under various assumptions about the risk aversion of the two players.
16.12.1 Risk Neutrality

Suppose that the principal and the agent meet and contract ex ante. If the agent is risk neutral, his ex ante participation constraint is now written as

\[ \nu \bar{U} + (1 - \nu) \bar{\bar{U}} \geq 0. \]  
\[(16.54)\]

This ex ante participation constraint replaces the two interim participation constraints.

Since the principal’s objective function is decreasing in the agent’s expected information rent, the principal wants to impose a zero expected rent to the agent and have (16.54) be binding. Moreover, the principal must structure the rents \( U \) and \( \bar{U} \) to ensure that the two incentive constraints remain satisfied. An example of such a rent distribution that is both incentive compatible and satisfies the ex ante participation constraint with an equality is

\[ U^* = (1 - \nu) \Delta \theta q^* > 0 \quad \text{and} \quad \bar{U}^* = -\nu \Delta \theta \bar{q}^* < 0. \]  
\[(16.55)\]

With such a rent distribution, the optimal contract implements the first-best outputs without cost from the principal’s point of view as long as the first-best is monotonic as requested by the implementability condition. In the contract defined by (16.55), the agent is rewarded when he is efficient and punished when he turns out to be inefficient. In summary, we have

**Proposition 16.12.1** When the agent is risk neutral and contracting takes place ex ante, the optimal incentive contract implements the first-best outcome.

**Remark 16.12.1** The principal has in fact more options in structuring the rents \( U \) and \( \bar{U} \) in such a way that the incentive compatible constraints hold and the ex ante participation constraint (16.54) holds with an equality. Consider the following contracts \{\( (\bar{t}^*, q^*); (\bar{\bar{t}}^*, \bar{q}^*) \)\} where \( \bar{t}^* = S(q^*) - T^* \) and \( \bar{\bar{t}}^* = S(\bar{q}^*) - T^* \), with \( T^* \) being a lump-sum payment to be defined below. This contract is incentive compatible since

\[ \bar{t}^* - \theta q^* = S(q^*) - \theta q^* - T^* > S(\bar{q}^*) - \theta \bar{q}^* - T^* = \bar{\bar{t}}^* - \theta \bar{q}^* \]  
\[(16.56)\]
by definition of $q^*$, and

$$t^* - \theta q^* = S(q^*) - \theta q^* - T^* > S(q^*) - \theta q^* - T^* = \tilde{t}^* - \tilde{\theta} q^*$$  \hspace{1cm} (16.57)$$

by definition of $\tilde{q}^*$.

The rent distribution is not unique, and in fact, there are infinitely many as shown in Figure 16.7.

![Diagram](image)

Figure 16.7: information rent with a risk neutral agent

Note that the incentive compatibility constraints are now strict inequalities. Moreover, the fixed-fee $T^*$ can be used to satisfy the agent’s ex ante participation constraint with an equality by choosing $T^* = \nu(S(q^*) - \theta q^*) + (1 - \nu)(S(\tilde{q}^*) - \tilde{\theta} q^*)$. This implementation of the first-best outcome amounts to having the principal selling the benefit of the relationship to the risk-neutral agent for a fixed up-front
payment $T^*$. The agent benefits from the full value of the good and trades off the
value of any production against its cost just as if he was an efficiency maximizer.
We will say that the agent is residual claimant for the firm’s profits.

16.12.2 Risk Aversion

A Risk-Averse Agent

The previous section has shown us that the implementation of the first-best is
feasible with risk neutrality. What happens if the agent is risk-averse?

Consider now a risk-averse agent with a Von Neumann-Morgenstern utility
function $u(\cdot)$ defined on his monetary gains $t - \theta q$, such that $u' > 0$, $u'' < 0$
and $u(0) = 0$. Again, the contract between the principal and the agent is signed
before the agent discovers his type. The incentive-compatibility constraints are
unchanged but the agent’s ex ante participation constraint is now written as

$$
\nu u(U) + (1 - \nu) u(\hat{U}) \geq 0. \tag{16.58}
$$

As usual, one can check incentive-compatibility constraint (16.22) for the in-
efficient agent is slack (not binding) at the optimum, and thus the principal’s
program reduces now to

$$
\max \left\{ (\overline{U}, \overline{q}); (U, q) \right\} \nu \left( S(q) - \theta q - U \right) + (1 - \nu) \left( S(\overline{q}) - \overline{\theta q} - \overline{U} \right),
$$

subject to (16.21) and (16.58).

We have the following proposition.

Proposition 16.12.2 When the agent is risk-averse and contracting takes place
ex ante, the optimal menu of contracts entails:

1. No output distortion for the efficient $q^{SB} = q^*$. A downward
   output distortion for the inefficient type $q^{SB} < \overline{q}^*$, with

$$
S'(q^{SB}) = \hat{\theta} + \frac{\nu u'(U^{SB}) - u'(U^{SB})}{\nu u'(U^{SB}) + (1 - \nu) u'(U^{SB})} \Delta \theta. \tag{16.59}
$$
(2) Both (16.21) and (16.58) are the only binding constraints. The efficient (resp. inefficient) type gets a strictly positive (resp. negative) ex post information rent, $\bar{U}^{SB} > 0 > \bar{U}^{SB}$.

Proof: Define the following Lagrangian for the principals problem

$$L(q, \bar{q}, U, \bar{U}, \lambda, \mu) = \nu(S(q) - \theta q - \bar{U}) + (1 - \nu)(S(\bar{q}) - \bar{\theta} \bar{q} - \bar{U})$$
$$+ \lambda(U - \bar{U} - \Delta \theta \bar{q}) + \mu(\nu u(U) + (1 - \nu)u(\bar{U})) \tag{16.60}$$

Optimizing w.r.t. $U$ and $\bar{U}$ yields respectively

$$-\nu + \lambda + \mu \nu u'(\bar{U}^{SB}) = 0 \tag{16.61}$$
$$-(1 - \nu) - \lambda + \mu (1 - \nu) u'(\bar{U}^{SB}) = 0. \tag{16.62}$$

Summing the above two equations, we obtain

$$\mu(\nu u'(\bar{U}^{SB}) + (1 - \nu) u'(\bar{U}^{SB})) = 1. \tag{16.63}$$

and thus $\mu > 0$. Using (16.63) and inserting it into (16.61) yields

$$\lambda = \frac{\nu(1 - \nu)(u'(\bar{U}^{SB}) - u'(U^{SB}))}{\nu u'(U^{SB}) + (1 - \nu) u'(U^{SB})}. \tag{16.64}$$

Moreover, (16.21) implies that $U^{SB} = \bar{U}^{SB}$ and thus $\lambda \geq 0$, with $\lambda > 0$ for a positive output $y$.

Optimizing with respect to outputs yields respectively

$$S'(\bar{q}^{SB}) = \theta \tag{16.65}$$

and

$$S'(\bar{q}^{SB}) = \bar{\theta} + \frac{\lambda}{1 - \nu} \Delta \theta. \tag{16.66}$$

Simplifying by using (16.64) yields (16.59).

The rent distribution is shown in Figure 16.8.

Note that when the agent is risk neutral, the second term in the right of 16.59 is zero, and thus we get the same conclusion as in Proposition 16.12.1: the optimal incentive contract implements the first-best outcome.
Thus, with risk aversion, the principal can no longer costlessly structure the agent’s information rents to ensure the efficient type’s incentive compatibility constraint. Creating a wedge between $U$ and $\bar{U}$ to satisfy (16.21) makes the risk-averse agent bear some risk. To guarantee the participation of the risk-averse agent, the principal must now pay a risk premium. Reducing this premium calls for a downward reduction in the inefficient type’s output so that the risk borne by the agent is lower. As expected, the agent’s risk aversion leads the principal to weaken the incentives.

When the agent becomes infinitely risk averse, everything happens as if he
had an ex post individual rationality constraint for the worst state of the world given by (16.24) such that $\bar{U}_{SB} = 0$. In the limit, the inefficient agent’s output $\bar{q}^{SB}$ and the utility levels $U^{SB}$ and $\bar{U}^{SB}$ all converge toward the same solution. So, the previous model at the interim stage can also be interpreted as a model with an ex ante infinitely risk-agent at the zero utility level.

A Risk-Averse Principal

Consider now a risk-averse principal with a Von Neumann-Morgenstern utility function $\psi(\cdot)$ defined on his monetary gains from trade $S(q) - t$ such that $\psi' > 0$, $\psi'' < 0$ and $\psi(0) = 0$. Again, the contract between the principal and the risk-neutral agent is signed before the agent knows his type.

In this context, the first-best contract obviously calls for the first-best output $q^*$ and $\bar{q}^*$ being produced. It also calls for the principal to be fully insured between both states of nature and for the agent’s ex ante participation constraint to be binding. This leads us to the following two conditions that must be satisfied by the agent’s rents $U^*$ and $\bar{U}^*$:

$$S(q^*) - \theta q^* - U^* = S(\bar{q}^*) - \theta \bar{q}^* - \bar{U}^*$$  \hspace{1cm} (16.67)

and

$$\nu U^* + (1 - \nu) \bar{U}^* = 0.$$  \hspace{1cm} (16.68)

Solving this system of two equations with two unknowns $(U^*, \bar{U}^*)$ yields

$$U^* = (1 - \nu)(S(q^*) - \theta q^* - (S(\bar{q}^*) - \theta \bar{q}^*))$$  \hspace{1cm} (16.69)

and

$$\bar{U}^* = -\nu(S(q^*) - \theta q^* - (S(\bar{q}^*) - \theta \bar{q}^*)).$$  \hspace{1cm} (16.70)

Note that the first-best profile of information rents satisfies both types’ incentive compatibility constraints since

$$U^* - \bar{U}^* = S(q^*) - \theta q^* - (S(\bar{q}^*) - \theta \bar{q}^*) > \Delta \theta q^*$$  \hspace{1cm} (16.71)
(from the definition of $q^*$) and

$$\bar{U}^* - U^* = S(q^*) - \bar{\theta}q^* - (S(q^*) - \theta \bar{q}^*) > -\Delta \theta q^*,$$  \hspace{1cm} (16.72)

(from the definition of $\bar{q}^*$). Hence, the profile of rents $(U^*, \bar{U}^*)$ is incentive compatible and the first-best allocation is easily implemented in this framework. We can thus generalize the proposition for the case of risk neutral as follows:

**Proposition 16.12.3** When the principal is risk-averse over the monetary gains $S(q) - t$, the agent is risk-neutral, and contracting takes place ex ante, the optimal incentive contract implements the first-best outcome.

**Remark 16.12.2** It is interesting to note that $U^*$ and $\bar{U}^*$ obtained in (16.69) and (16.70) are also the levels of rent obtained in (16.56) and (16.57). Indeed, the lump-sum payment $T^* = \nu(S(q^*) - \theta q^*) + (1 - \nu)(S(\bar{q}^*) - \bar{\theta} \bar{q}^*)$, which allows the principal to make the risk-neutral agent residual claimant for the hierarchy’s profit, also provides full insurance to the principal. By making the risk-neutral agent the residual claimant for the value of trade, ex ante contracting allows the risk-averse principal to get full insurance and implement the first-best outcome despite the informational problem.

Of course this result does not hold anymore if the agent’s interim participation constraints must be satisfied. In this case, we still guess a solution such that (16.23) is slack at the optimum. The principal’s program now reduces to:

$$\max_{\{U, \bar{U}\} \in \mathcal{L}_2} \nu \nu(S(q) - \theta q - U) + (1 - \nu)\nu(S(\bar{q}) - \bar{\theta} \bar{q} - \bar{U})$$

subject to (16.21) to (16.24).

Inserting the values of $U$ and $\bar{U}$ that were obtained from the binding constraints in (16.21) and (16.24) into the principal’s objective function and optimizing with respect to outputs leads to $q^{SB} = q^*$, i.e., no distortion for the efficient type, just as in the case of risk neutrality and a downward distortion of the inefficient type’s output $\bar{q}^{SB} < \bar{q}^*$ given by

$$S'(\bar{q}^{SB}) = \theta + \frac{\nu \nu'(V^{SB})}{(1 - \nu)\nu'(V^{SB})} \Delta \theta. \hspace{1cm} (16.73)$$
where $V_{SB} = S(q^*) - \theta q^* - \Delta \theta q_{SB}$ and $\bar{V}_{SB} = S(\bar{q}_{SB}) - \bar{\theta} \bar{q}_{SB}$ are the principal’s payoffs in both states of nature. We can check that $\bar{V}_{SB} < V_{SB}$ since $S(\bar{q}_{SB}) - \bar{\theta} \bar{q}_{SB} < S(q^*) - \theta q^*$ from the definition of $q^*$. In particular, we observe that the distortion in the right-hand side of (16.73) is always lower than $\frac{\nu}{1-\nu} \Delta \theta$, its value with a risk-neutral principal. The intuition is straightforward. By increasing $\bar{q}$ above its value with risk neutrality, the risk-averse principal reduces the difference between $V_{SB}$ and $\bar{V}_{SB}$. This gives the principal some insurance and increases his ex ante payoff.

For example, if $\nu(x) = 1 - e^{-rx}$, (16.73) becomes $S'(\bar{q}_{SB}) = \bar{\theta} + \frac{\nu}{1-\nu} e^{r(\bar{V}_{SB} - V_{SB})} \Delta \theta$. If $r = 0$, we get back the distortion obtained before for the case of with a risk-neutral principal and interim participation constraints for the agent. Since $\bar{V}_{SB} < V_{SB}$, we observe that the first-best is implemented when $r$ goes to infinity.

In the limit, the infinitely risk-averse principal is only interested in the inefficient state of nature for which he wants to maximize the surplus, since there is no rent for the inefficient agent. Moreover, giving a rent to the efficient agent is now without cost for the principal.

Risk aversion on the side of the principal is quite natural in some contexts. A local regulator with a limited budget or a specialized bank dealing with relatively correlated projects may be insufficiently diversified to become completely risk neutral. See Lewis and Sappington (Rand J. Econ, 1995) for an application to the regulation of public utilities.

### 16.13 Commitment

To solve the incentive problem, we have implicitly assumed that the principal has a strong ability to commit himself not only to a distribution of rents that will induce information revelation but also to some allocative inefficiency designed to reduce the cost of this revelation. Alternatively, this assumption also means that the court of law can perfectly enforce the contract and that neither renegotiating nor reneging on the contract is a feasible alternative for the agent and (or) the
principal. What can happen when either of those two assumptions is relaxed?

16.13.1 Renegotiating a Contract

Renegotiation is a voluntary act that should benefit both the principal and the agent. It should be contrasted with a breach of contract, which can hurt one of the contracting parties. One should view a renegotiation procedure as the ability of the contracting partners to achieve a Pareto improving trade if any becomes incentive feasible along the course of actions.

Once the different types have revealed themselves to the principal by selecting the contracts \((t^{SB}, q^{SB})\) for the efficient type and \((\bar{t}^{SB}, \bar{q}^{SB})\) for the inefficient type, the principal may propose a renegotiation to get around the allocative inefficiency he has imposed on the inefficient agent’s output. The gain from this renegotiation comes from raising allocative efficiency for the inefficient type and moving output from \(\bar{q}^{SB}\) to \(\bar{q}^{*}\). To share these new gains from trade with the inefficient agent, the principal must at least offer him the same utility level as before renegotiation. The participation constraint of the inefficient agent can still be kept at zero when the transfer of this type is raised from \(\bar{t}^{SB} = \bar{\theta} \bar{q}^{SB}\) to \(\bar{t}^{*} = \bar{\theta} \bar{q}^{*}\). However, raising this transfer also hardens the ex ante incentive compatibility constraint of the efficient type. Indeed, it becomes more valuable for an efficient type to hide his type so that he can obtain this larger transfer, and truthful revelation by the efficient type is no longer obtained in equilibrium. There is a fundamental trade-off between raising efficiency ex post and hardening ex ante incentives when renegotiation is an issue.

16.13.2 Reneging on a Contract

A second source of imperfection arises when either the principal or the agent reneges on their previous contractual obligation. Let us take the case of the principal reneging on the contract. Indeed, once the agent has revealed himself to the principal by selecting the contract within the menu offered by the principal,
the latter, having learned the agent’s type, might propose the complete information contract which extracts all rents without inducing inefficiency. On the other hand, the agent may want to renege on a contract which gives him a negative ex post utility level as we discussed before. In this case, the threat of the agent reneging a contract signed at the ex ante stage forces the agent’s participation constraints to be written in interim terms. Such a setting justifies the focus on the case of interim contracting.

16.14 Informative Signals to Improve Contracting

In this section, we investigate the impacts of various improvements of the principal’s information system on the optimal contract. The idea here is to see how signals that are exogenous to the relationship can be used by the principal to better design the contract with the agent.

16.14.1 Ex Post Verifiable Signal

Suppose that the principal, the agent and the court of law observe ex post a viable signal $\sigma$ which is correlated with $\theta$. This signal is observed after the agent’s choice of production. The contract can then be conditioned on both the agent’s report and the observed signal that provides useful information on the underlying state of nature.

For simplicity, assume that this signal may take only two values, $\sigma_1$ and $\sigma_2$. Let the conditional probabilities of these respective realizations of the signal be $\mu_1 = \Pr(\sigma = \sigma_1/\theta = \bar{\theta}) \geq 1/2$ and $\mu_2 = \Pr(\sigma = \sigma_2/\theta = \bar{\theta}) \geq 1/2$. Note that, if $\mu_1 = \mu_2 = 1/2$, the signal $\sigma$ is uninformative. Otherwise, $\sigma_1$ brings good news the fact that the agent is efficient and $\sigma_2$ brings bad news, since it is more likely that the agent is inefficient in this case.

Let us adopt the following notations for the ex post information rents: $u_{11} =$
\( t(\theta, \sigma_1) - \theta q(\theta, \sigma_1), \) \( u_{12} = t(\bar{\theta}, \sigma_2) - \theta q(\bar{\theta}, \sigma_2), \) \( u_{21} = t(\bar{\theta}, \sigma_1) - \bar{\theta} q(\bar{\theta}, \sigma_1), \) and \( u_{22} = t(\bar{\theta}, \sigma_2) - \bar{\theta} q(\bar{\theta}, \sigma_2). \) Similar notations are used for the outputs \( q_{jj}. \) The agent discovers his type and plays the mechanism before the signal \( \sigma \) realizes. Then the incentive and participation constraints must be written in expectation over the realization of \( \sigma. \) Incentive constraints for both types write respectively as

\[
\mu_1 u_{11} + (1 - \mu_1) u_{12} \geq \mu_1 (u_{21} + \Delta \theta q_{21}) + (1 - \mu_1) (u_{22} + \Delta \theta q_{22}) \quad (16.74)
\]

\[
(1 - \mu_2) u_{21} + \mu_2 u_{22} \geq (1 - \mu_2) (u_{11} - \Delta \theta q_{11}) + \mu_2 (u_{12} - \Delta \theta q_{12}). \quad (16.75)
\]

Participation constraints for both types are written as

\[
\mu_1 u_{11} + (1 - \mu_1) u_{12} \geq 0, \quad (16.76)
\]

\[
(1 - \mu_2) u_{21} + \mu_2 u_{22} \geq 0. \quad (16.77)
\]

Note that, for a given schedule of output \( q_{ij}, \) the system (16.74) through (16.77) has as many equations as unknowns \( u_{ij}. \) When the determinant of the coefficient matrix of the system (16.74) to (16.77) is nonzero, one can find ex post rents \( u_{ij} \) (or equivalent transfers) such that all these constraints are binding. In this case, the agent receives no rent whatever his type. Moreover, any choice of production levels, in particular the complete information optimal ones, can be implemented this way. Note that the determinant of the system is nonzero when

\[
1 - \mu_1 - \mu_2 \neq 0 \quad (16.78)
\]

that fails only if \( \mu_1 = \mu_2 = \frac{1}{2}, \) which corresponds to the case of an uninformative and useless signal.

**16.14.2 Ex Ante Nonverifiable Signal**

Now suppose that a nonverifiable binary signal \( \sigma \), \( \sigma \) about \( \theta \) is available to the principal at the ex ante stage. Before offering an incentive contract, the principal
computes, using the Bayes law, his posterior belief that the agent is efficient for each value of this signal, namely

\begin{equation}
\hat{\nu}_1 = Pr(\theta = \theta/\sigma = \sigma_1) = \frac{\nu \mu_1}{\nu \mu_1 + (1 - \nu)(1 - \mu_2)},
\end{equation}

(16.79)

\begin{equation}
\hat{\nu}_2 = Pr(\theta = \theta/\sigma = \sigma_2) = \frac{\nu(1 - \mu_1)}{\nu(1 - \mu_1) + (1 - \nu)\mu_2}.
\end{equation}

(16.80)

Then the optimal contract entails a downward distortion of the inefficient agents production \(\bar{q}_{SB}(\sigma_i)\) which is for signals \(\sigma_1\), and \(\sigma_2\) respectively:

\begin{equation}
S'(\bar{q}_{SB}(\sigma_1)) = \bar{\theta} + \frac{\hat{\nu}_1}{1 - \hat{\nu}_1} \Delta \theta = \bar{\theta} + \frac{\nu \mu_1}{(1 - \nu)(1 - \mu_2)} \Delta \theta
\end{equation}

(16.81)

\begin{equation}
S'(\bar{q}_{SB}(\sigma_2)) = \bar{\theta} + \frac{\hat{\nu}_2}{1 - \hat{\nu}_2} \Delta \theta = \bar{\theta} + \frac{\nu(1 - \mu_1)}{(1 - \nu)\mu_2} \Delta \theta.
\end{equation}

(16.82)

In the case where \(\mu_1 = \mu_2 = \mu > \frac{1}{2}\), we can interpret \(\mu\) as an index of the informativeness of the signal. Observing \(\sigma_1\), the principal thinks that it is more likely that the agent is efficient. A stronger reduction in \(\bar{q}_{SB}\) and thus in the efficient type’s information rent is called for after \(\sigma_1\). (16.81) shows that incentives decrease with respect to the case without informative signal since \(\left(\frac{\mu}{1 - \mu} > 1\right)\). In particular, if \(\mu\) is large enough, the principal shuts down the inefficient firm after having observed \(\sigma_1\). The principal offers a high-powered incentive contract only to the efficient agent, which leaves him with no rent. On the contrary, because he is less likely to face an efficient type after having observed \(\sigma_2\), the principal reduces less of the information rent than in the case without an informative signal since \(\left(\frac{1 - \mu}{\mu} < 1\right)\). Incentives are stronger.

### 16.15 Contract Theory at Work

This section proposes several classical settings where the basic model of this chapter is useful. Introducing adverse selection in each of these contexts has proved to be a significative improvement of standard microeconomic analysis.
16.15.1 Regulation

In the Baron and Myerson (Econometrica, 1982) regulation model, the principal is a regulator who maximizes a weighted average of the agents’ surplus $S(q) - t$ and of a regulated monopoly’s profit $U = t - \theta q$, with a weight $\alpha < 1$ for the firms profit. The principal’s objective function is written now as $V = S(q) - \theta q - (1 - \alpha) U$.

Maximizing expected social welfare under incentive and participation constraints leads to $q^{SB} = q^*$ for the efficient type and a downward distortion for the inefficient type, $\bar{q}^{SB} < q^*$ which is given by

$$S'(q^{SB}) = \bar{\theta} + \frac{\nu}{1 - \nu} (1 - \alpha) \Delta \theta.$$  (16.83)

Note that a higher value of $\alpha$ reduces the output distortion, because the regulator is less concerned by the distribution of rents within society as $\alpha$ increases. If $\alpha = 1$, the firm’s rent is no longer costly and the regulator behaves as a pure efficiency maximizer implementing the first-best output in all states of nature.

The regulation literature of the last thirty years has greatly improved our understanding of government intervention under asymmetric information. We refer to the book of Laffont and Tirole (1993) for a comprehensive view of this theory and its various implications for the design of real world regulatory institutions.

16.15.2 Nonlinear Pricing by a Monopoly

In Maskin and Riley (Rand J. of Economics, 1984), the principal is the seller of a private good with production cost $cq$ who faces a continuum of buyers. The principal has thus a utility function $V = t - cq$. The tastes of a buyer for the private good are such that his utility function is $U = \theta u(q) - t$, where $q$ is the quantity consumed and $t$ his payment to the principal. Suppose that the parameter $\theta$ of each buyer is drawn independently from the same distribution on $\Theta = \{\theta, \bar{\theta}\}$ with respective probabilities $1 - \nu$ and $\nu$.

We are now in a setting with a continuum of agents. However, it is mathematically equivalent to the framework with a single agent. Now $\nu$ is the frequency of
type $\theta$ by the Law of Large Numbers.

Incentive and participation constraints can as usual be written directly in terms of the information rents $U = \theta u(q) - t$ and $\bar{U} = \bar{\theta} u(\bar{q}) - \bar{t}$ as

$$U \geq \bar{U} - \Delta \theta u(\bar{q}), \quad (16.84)$$

$$\bar{U} \geq U + \Delta \theta u(q), \quad (16.85)$$

$$U \geq 0, \quad (16.86)$$

$$\bar{U} \geq 0. \quad (16.87)$$

The principal’s program now takes the following form:

$$\max_{\left\{ (U, q) : (U, q) \right\}} \left\{ v(\bar{\theta} u(\bar{q}) + (1 - v)(\theta u(q) - cq) - (\nu \bar{U} + (1 - \nu) U) \right\}$$

subject to (16.84) to (16.87).

The analysis is the mirror image of that of the standard model discussed before, where now the efficient type is the one with the highest valuation for the good $\bar{\theta}$. Hence, (16.85) and (16.86) are the two binding constraints. As a result, there is no output distortion with respect to the first-best outcome for the high valuation type and $\bar{q}^{SB} = \bar{q}^*$, where $\bar{\theta} u'(\bar{q}^*) = c$. However, there exists a downward distortion of the low valuation agent’s output with respect to the first-best outcome. We have $q^{SB} < q^*$, where

$$\left( \bar{\theta} - \frac{\nu}{1 - \nu} \Delta \theta \right) u'(q^{SB}) = c \quad \text{and} \quad \theta u'(q^*) = c. \quad (16.88)$$

So the unit price is not the same if the buyers demand $q^*$ or $q^{SB}$, hence the expression of nonlinear prices.

### 16.15.3 Quality and Price Discrimination

Mussa and Rosen (JET, 1978) studied a very similar problem to the nonlinear pricing, where agents buy one unit of a commodity with quality $q$ but are vertically differentiated with respect to their preferences for the good. The marginal cost
(and average cost) of producing one unit of quality $q$ is $C(q)$ and the principal has the payoff function $V = t - C(q)$. The payoff function of an agent is now $U = \theta q - t$ with $\theta$ in $\Theta = \{\underline{\theta}, \bar{\theta}\}$, with respective probabilities $1 - \nu$ and $\nu$.

Incentive and participation constraints can still be written directly in terms of the information rents $U = \theta q - t$ and $\bar{U} = \bar{\theta} \bar{q} - \bar{t}$ as

$$U \geq \bar{U} - \Delta \theta \bar{q}, \quad (16.89)$$

$$\bar{U} \geq U + \Delta \theta q, \quad (16.90)$$

$$U \geq 0, \quad (16.91)$$

$$\bar{U} \geq 0. \quad (16.92)$$

The principal solves now:

$$\max_{\{U, q\} : (\bar{U}, \bar{q})} \{v(\bar{\theta} \bar{q} - C(\bar{q})) + (1 - \nu)(\theta q - C(q)) - (\nu \bar{U} + (1 - \nu)U)\}$$

subject to (16.89) to (16.92).

Following procedures similar to what we have done so far, only (16.90) and (16.91) are binding constraints. Finally, we find that the high valuation agent receives the first-best quality $\bar{q}^{SB} = \bar{q}^{*}$ where $\bar{\theta} = C'(\bar{q}^{*})$. However, quality is now reduced below the first-best for the low valuation agent. We have $q^{SB} < q^{*}$, where

$$\theta = C'(q^{SB}) + \frac{\nu}{1 - \nu} \Delta \theta \quad \text{and} \quad \bar{\theta} = C'(q^{*}) \quad (16.93)$$

Interestingly, the spectrum of qualities is larger under asymmetric information than under complete information. This incentive of the seller to put a low quality good on the market is a well-documented phenomenon in the industrial organization literature. Some authors have even argued that damaging its own goods may be part of the firm’s optimal selling strategy when screening the consumers’ willingness to pay for quality is an important issue.

16.15.4 Financial Contracts

Asymmetric information significantly affects the financial markets. For instance, in a paper by Freixas and Laffont (1990), the principal is a lender who provides
a loan of size $k$ to a borrower. Capital costs $Rk$ to the lender since it could be invested elsewhere in the economy to earn the risk-free interest rate $R$. The lender has thus a payoff function $V = t - Rk$. The borrower makes a profit $U = \theta f(k) - t$ where $\theta f(k)$ is the production with $k$ units of capital and $t$ is the borrower’s repayment to the lender. We assume that $f’ > 0$ and $f’’ < 0$. The parameter $\theta$ is a productivity shock drawn from $\Theta = \{\bar{\theta}, \theta\}$ with respective probabilities $1 - \nu$ and $\nu$.

Incentive and participation constraints can again be written directly in terms of the borrower’s information rents $\bar{U} = \theta f(k) - \bar{t}$ and $\bar{U} = \bar{\theta} f(\bar{k}) - \bar{t}$ as

$$U \geq \bar{U} - \Delta \theta f(\bar{k}),$$

(16.94)

$$\bar{U} \geq U + \Delta \theta f(k),$$

(16.95)

$$U \geq 0,$$

(16.96)

$$\bar{U} \geq 0.$$  

(16.97)

The principal’s program takes now the following form:

$$\max \{U(k), k; \bar{U}, \bar{k}\} = v(\bar{\theta} f(\bar{k}) - R\bar{k}) + (1 - \nu)(\theta f(k)) - Rk - (\nu \bar{U} + (1 - \nu)U)$$

subject to (16.94) to (16.97).

One can check that (16.95) and (16.96) are now the two binding constraints. As a result, there is no capital distortion with respect to the first-best outcome for the high productivity type and $\bar{k}_{SB} = k^*$ where $\bar{\theta} f’(\bar{k}^*) = R$. In this case, the return on capital is equal to the risk-free interest rate. However, there also exists a downward distortion in the size of the loan given to a low productivity borrower with respect to the first-best outcome. We have $k_{SB} < k^*$ where

$$\left(\theta - \frac{\nu}{1 - \nu} \Delta \theta\right) f'(k_{SB}) = R \quad \text{and} \quad \bar{\theta} f’(k^*) = R,$$

(16.98)

16.15.5 Labor Contracts

Asymmetric information also undermines the relationship between a worker and the firm for which he works. In Green and Kahn (QJE, 1983) and Hart (RES,
1983), the principal is a union (or a set of workers) providing its labor force \( l \) to a firm.

The firm makes a profit \( \theta f(l) - t \), where \( f(l) \) is the return on labor and \( t \) is the worker’s payment. We assume that \( f' > 0 \) and \( f'' < 0 \). The parameter \( \theta \) is a productivity shock drawn from \( \Theta = \{ \theta, \bar{\theta} \} \) with respective probabilities \( 1 - \nu \) and \( \nu \). The firm’s objective is to maximize its profit \( U = \theta f(l) - t \). Workers have a utility function defined on consumption and labor. If their disutility of labor is counted in monetary terms and all revenues from the firm are consumed, they get \( V = v(t - l) \) where \( l \) is their disutility of providing \( l \) units of labor and \( v(\cdot) \) is increasing and concave \( (v' > 0, v'' < 0) \).

In this context, the firm’s boundaries are determined before the realization of the shock and contracting takes place ex ante. It should be clear that the model is similar to the one with a risk-averse principal and a risk-neutral agent. So, we know that the risk-averse union will propose a contract to the risk-neutral firm which provides full insurance and implements the first-best levels of employments \( \bar{l} \) and \( l^* \) defined respectively by \( \bar{\theta} f'(\bar{l}) = 1 \) and \( \theta f'(l^*) = 1 \).

When workers have a utility function exhibiting an income effect, the analysis will become much harder even in two-type models. For details, see Laffont and Martimort (2002).

### 16.16 Challenges to “No Distortion at the Top”

**Rule**

As shown in the preceding sections, the canonical principal-agent model makes several simplifying assumptions. First, there is no externality among agents, that is, an agent’s behavior is assumed to have no impact on the welfare of others; second, an agent is assumed to obtain a minimal type-independent level of utility if he rejects the offer made by the principal. Under these assumptions, the optimal contract exhibits no-distortion for the “best” type agent and downward distortions for all other types.

607
However, as Meng and Tian (2009) showed, challenges to this “no distortion at the top” convention may arise if we relax the above two assumptions. To introduce these results, in this section we will discuss these two “challenges” in an integrated nonlinear pricing model.

16.16.1 Network Externalities

An externality is present whenever the well-being of a consumer or the production possibilities of a firm are directly affected by the actions of others in the economy.

A special case of externality is the so-called the “network externalities” that might arise for any of the following reasons: because the usefulness of the commodity depends directly on the size of the network (e.g., telephones, fax machines); or because of bandwagon effect, which means the desire to be in style: to have a good because almost everyone else has it; or indirectly through the availability of complementary goods and services (often known as the “hardware-software paradigm”) or of postpurchase services (e.g., for automobiles). Although “network externalities” are often regarded as positive impact on each others’ consumption, it may display negative properties in some cases. For example, people has the desire to own exclusive or unique goods, which is called “Snob effect”. The quantity demanded of a “snob” good is higher the fewer the people who own it.

Consider a principal-agent model in which the principal is a monopolist of a network good with marginal production cost $c$ and total output $q$ who faces a continuum of consumers. The principal’s payoff function is given by $V = t - cq$, where $t$ is the payment received from consumers. Consumer’s preference of the good is characterized by $\theta \in \Theta = \{\underline{\theta}, \overline{\theta}\}$, $Pr(\theta = \underline{\theta}) = v$, $Pr(\theta = \overline{\theta}) = 1 - v$, and then $v$ can be regarded as the frequency of type $\underline{\theta}$ by the Law of Large Numbers.

A consumer of $\theta$ type is assumed to have a utility function of $U(\theta) = \theta V(q(\theta)) + \Psi(Q) - t(\theta)$, where $q(\theta)$ is the amount of the good he consumes, $Q = vq + (1 - v)\overline{q} = E[q(\theta)]$ is the total amount of consumption (network size) and $t(\theta)$ is the tariff charged by the principal. $\theta V(q(\theta))$ is the intrinsic value of consuming, while $\Psi(Q)$
is the network value. Note that we assume the network effect is homogeneous among all the consumers, namely, the network value is independent of individual preference $\theta$ and individual consumption $q(\theta)$. It is assumed that $V'(q) > 0$ and $V''(q) < 0$.

**Definition 16.16.1** The network is decreasing, constant, or increasing if and only if $\Psi''(Q) < 0$, $\Psi''(Q) = 0$, or $\Psi''(Q) > 0$, respectively.

**Remark 16.16.1** When the network capacity is large and the maintaining technology is advanced enough, an increase in one consumer’s consumption will increase the marginal utilities of others, so $\Psi''(Q) > 0$. When network capacity and maintaining technology are limited, consumers are rivals to one another in the sense that an increase in one consumer’s consumption will decrease the marginal utilities of others, thus $\Psi''(Q) < 0$. When network expansion benefits all the consumers with constant margin, the network value term is a linear function: $\Psi''(Q) = 0$.

The objective of the monopolist is to design a menu of incentive-compatible and self-selecting quantity-price pairs $\{q(\hat{\theta}), t(\hat{\theta})\}$ to maximize his own revenue, where $\hat{\theta} \in \Theta$ is the consumer’s announcement. The network magnitude is $Q = vq + (1 - v)\bar{q}$. Under complete information, the monopolist’s problem is:

$$(P2) \left\{ \begin{array}{l} \max \left\{ v\left[ \theta V(q) - cq \right] + (1 - v)\left[ \theta V(\bar{q}) - c\bar{q} \right] + \Psi(Q) - \left[ vU + (1 - v)\bar{U} \right] \right\} \\
\text{s.t. } IR(\hat{\theta}) : U \geq 0 \\
IR(\bar{\theta}) : \bar{U} \geq 0 \end{array} \right.$$  

The first-best consumption is thus:

$$\begin{cases} \theta V'(q^{FB}) + \Psi'(vq^{FB} + (1 - v)\bar{q}^{FB}) = c, \\
\theta V'(\bar{q}^{FB}) + \Psi'(vq^{FB} + (1 - v)\bar{q}^{FB}) = c. \end{cases}$$  

(16.99)

Under asymmetric information, two incentive compatible constraints should be added to the above program, then we get:
The incentive compatible constraint $IC(\theta)$ of high demand type and the participation constraint of the low-demand type $IR(\theta)$ are binding, then the consumptions in the second-best contract are characterized by the following first-order conditions:

\[
\begin{align*}
\max \{&v[\theta V(q) - cq] + (1 - v)[\overline{\theta} V(q) - c\overline{q}] + \Psi(Q) - [vU + (1 - v)\overline{U}] \\
\text{s.t.} \quad &IR(\theta) : U \geq 0 \\
&IR(\overline{\theta}) : \overline{U} \geq 0 \\
&IC(\theta) : U \geq \overline{U} - \Delta \theta V(\overline{q}) \\
&IC(\overline{\theta}) : \overline{U} \geq U + \Delta \theta V(q)
\end{align*}
\]

We synthesize the first-best and second-best solution by considering them as solution to the following parameterized form:

\[
\max_{\{\theta, \overline{\theta}, q, \overline{q}, \alpha\}} \Pi(q, \overline{q}, \alpha) = v[\alpha V(q) - c_q] + (1 - v)[\overline{\theta} V(\overline{q}) - c\overline{q}] + \Psi(Q).
\]

We then have the following proposition.

**Proposition 16.16.1** In the presence of network externalities and asymmetric information, the direction of distortion in consumptions depends on the sign of $\Psi''(Q)$.
1. If the network is mildly increasing, i.e., $\Psi'(Q) > 0$ but is not too large such that $\Pi_{qq}$ is negative definite for all $\alpha \in [\theta - \frac{1-v}{v} \Delta \theta, \theta]$, then the consumptions of all types exhibit one-way distortion: $q^{SB} < q^{FB}$ and $\bar{q}^{SB} < \bar{q}^{FB}$.

2. If the network is decreasing, i.e., $\Psi'(Q) < 0$, then the consumptions exhibit two-way distortion: $q^{SB} < q^{FB}$ and $q^{SB} > q^{FB}$.

3. If the network is constant, i.e., $\Psi'(Q) = 0$, then the rules “no distortion on the top” and “one-way distortion” in canonical settings are still available: $q^{SB} < q^{FB}$, $q^{SB} = q^{FB}$.

For all these cases, the network magnitude is downsized: $Q^{SB} < Q^{FB}$.

**Proof.** The first order condition to (16.101) is:

$$\Pi_q(q, \alpha) = 0,$$  
(16.102)

this implies,

$$\begin{cases} 
\alpha V'(q) + \Psi'(vq + (1 - v)\bar{q}) = c, \\
\bar{V}'(q) + \Psi'(vq + (1 - v)\bar{q}) = c.
\end{cases}$$  
(16.103)

Differentiating (16.102) with respect to parameter $\alpha$, we get:

$$\Pi_{qq} \frac{dq}{d\alpha} + \Pi_{qa} = 0$$  
(16.104)

that is,

$$\begin{pmatrix} 
\alpha vV''(q) + v^2\Psi''(Q) & v(1-v)\Psi''(Q) \\
v(1-v)\Psi''(Q) & (1-v)\bar{V}''(\bar{q}) + (1-v)^2\Psi''(Q)
\end{pmatrix} \begin{pmatrix} 
\frac{dq}{d\alpha} \\
\frac{d\bar{q}}{d\alpha}
\end{pmatrix} + \begin{pmatrix} 
vV'(q) \\
0
\end{pmatrix} = \begin{pmatrix} 
0 \\
0
\end{pmatrix}$$

Solving the above equations, we have

$$\begin{cases} 
\frac{dq}{d\alpha} = \frac{-V'(q) [\bar{V}''(\bar{q}) + (1-v)\Psi''(Q)]}{\alpha V''(q) [\bar{V}''(\bar{q}) + (1-v)\Psi''(Q)] + v\bar{V}''(\bar{q})\Psi''(Q)} \\
\frac{d\bar{q}}{d\alpha} = \frac{vV'(q)\Psi''(Q)}{\alpha V''(q) [\bar{V}''(\bar{q}) + (1-v)\Psi''(Q)] + v\bar{V}''(\bar{q})\Psi''(Q)} \\
\frac{dQ}{d\alpha} = v \frac{dq}{d\alpha} + (1-v) \frac{d\bar{q}}{d\alpha} = \frac{-v\bar{V}'(q)V''(\bar{q})}{\alpha V''(q) [\bar{V}''(\bar{q}) + (1-v)\Psi''(Q)] + v\bar{V}''(\bar{q})\Psi''(Q)}
\end{cases}$$  
(16.105)
From the assumption that the Hessian matrix $\Pi_{qq}$ is negative definite, it can be verified that the $2^{th}$ diagonal element of $\Pi_{qq}$ is negative, and thus
\[ \bar{\theta}V''(\bar{q}) + (1 - v)\Psi''(Q) < 0 \quad (16.106) \]
and the determinant of $\Pi_{qq}$ is positive,
\[ \det(\Pi_{qq}) = v(1 - v)\left\{ \alpha V''(q) [\bar{\theta}V''(\bar{q}) + (1 - v)\Psi''(Q)] + v\bar{\theta}V''(\bar{q})\Psi''(Q) \right\} > 0. \quad (16.107) \]
The signs of derivatives in (16.105) can be determined, which are $\frac{dq}{da} > 0$ and $\frac{dQ}{da} > 0$. That means $q^{SB} < q^{FB}$ and $Q^{SB} < Q^{FB}$. The sign of $\frac{d\theta}{d\alpha}$ and thus the distortion direction of $q$ depends on the sign of $\Psi''(Q)$: if $\Psi''(Q) > 0$, $\frac{d\theta}{d\alpha} > 0$, then $q^{SB} < q^{FB}$; if $\Psi''(Q) < 0$, $\frac{d\theta}{d\alpha} < 0$, then $q^{SB} > q^{FB}$; if $\Psi''(Q) = 0$, $\frac{d\theta}{d\alpha} = 0$, then $q^{SB} = q^{FB}$.\hfill \blacksquare

**Remark 16.16.2** $\Psi''(\cdot) > 0$ implies that the marginal value from an increase in individual consumption is higher at a higher level of others’ consumption: $\frac{\partial^2U}{\partial q_i \partial q_j} > 0$ for $i \neq j$. Its interpretation is that the externalities in bigger network is larger than in small network so that an agent is more eager to consume more when other agent consume more. It is consistent with the critical “strategic complementarity” assumption in Segal (1999, 2003) and Csorba (2008). This condition allows them to characterize the optimal contracts by applying monotone comparative static tools, pioneered by Topkis (1978) and Milgrom and Shannon (1994).\footnote{By Topkis (1978) and Milgrom and Shannon (1994), a twice continuously differentiable function $\Pi = \Pi(q_1, q_2, \ldots, q_n; \epsilon)$ defined on a lattice $Q$ is supermodular if and only if for all $i \neq j$, $\frac{\partial^2\Pi}{\partial q_i \partial q_j} > 0$; furthermore, if $\frac{\partial^2\Pi}{\partial q_i \partial \epsilon} > 0, \forall i$, then function $\Pi$ has strictly increasing differences in $(q, \epsilon)$. Let $q(\epsilon) = \max_{q \in Q} \Pi(q, \epsilon)$. Then for a supermodular function with increasing differences in $(q, \epsilon)$, $q_i(\epsilon)$ is a strictly increasing function of $\epsilon$ for all $i$.}

### 16.16.2 Countervailing Incentives

In this section we discuss another cause of the failure of “no distortion on the top” and “one-way distortion” rules: the countervailing incentive problem. We
assume that the consumers can bypass the network offered by the incumbent firm and enter a competitive market including many homogenous firms. All these firms are the potential entrants of the market. Let $\omega$ denote the marginal production cost of the entrants. We assume that the goods or services offered by the entrants are incompatible with that of the incumbent monopolist,$^{3}$ and they have not yet formed their own consumers network. In the competitive outside market, each firm’s unit charge equals its marginal cost $\omega$, so the representative consumer’s utility derived from consuming the entrants’ goods is $\theta V(q) - \omega q$. Let $G(\theta) = \max_q [\theta V(q) - \omega q]$, $\underline{G} = G(\underline{\theta})$, $\overline{G} = G(\overline{\theta})$, and $\Delta G = \overline{G} - \underline{G}$. Throughout this section, we assume that the network is congestible.

The entry threat gives the consumers non-zero type-dependent reservation utilities, and thus the problem of the incumbent network supplier can be represented as

\[
(P5) \begin{cases}
\max_{\{ (U, q) : (\overline{U}, \overline{q}) \} \mid IR(\bar{\eta})} v [\theta V(q) - cq] + (1 - v) [\theta V(\overline{q}) - c\overline{q}] + \Psi(Q) - [vU + (1 - v)\overline{U}] \\
\text{s.t.} \quad IR(\bar{\eta}) : U \geq \underline{G} \\
\quad IR(\overline{\eta}) : U \geq \overline{G} \\
\quad IC(\theta) : U \geq U - \Delta \theta V(\overline{q}) \\
\quad IC(\overline{\theta}) : U \geq U + \Delta \theta V(q).
\end{cases}
\]

Note that (P5) is the same as (P3) except for the non-zero type-dependent reservation utilities $\underline{G}$ and $\overline{G}$. The following proposition characterizes the optimal entry-deterring contract.

In (P5), we have as many regimes as combinations of binding constraints among $IR(\bar{\eta})$, $IR(\overline{\eta})$, $IC(\theta)$ and $IC(\overline{\theta})$. To reduce the number of possible cases, we first give the following lemmas.

**Lemma 16.16.1** A pooling contract with $q = \overline{q}$ and $t = \overline{t}$ can never be optimal.

**Proof.** Suppose that the optimal contract is pooling with $q = \overline{q} = q$ and $t = \overline{t} = t$. There are two cases to be considered.

$^{3}$Otherwise, the entrants can share the present network with the incumbent monopolist.
(i) $\overline{\theta}V'(q) > c$. Then, increase $\overline{q}$ by $\varepsilon$ and the transfer by $\overline{\theta}V'(q)\varepsilon$, the $\overline{\theta}$–type consumers can remain indifferent. Since at $(q, t)$ the marginal rate of substitution between $q$ and $t$ is higher for $\overline{\theta}$–type consumers, this new allocation is incentive compatible. This raises the firm’s revenue by $(1 - v)[\overline{\theta}V'(q) - c]\varepsilon$.

(ii) $\overline{\theta}V'(q) \leq c$ and $\overline{\theta}V'(q) < c$. Then, decrease $\overline{q}$ by $\varepsilon$ and adjust $\overline{t}$ so that $\overline{\theta}$–type consumers can remain on the same indifference curve. Then the firm’s total charge will be increased by $[c - \overline{\theta}V'(q)]\varepsilon$.

Thus, in both cases, it contradicts with the fact $(q, t)$ is optimal contract. 

Lemma 16.16.2 If the two types of consumers are offered two different contracts, the two incentive constraints cannot be simultaneously bindings.

Proof. Suppose by way of contradiction that both ICs are binding. From $\underline{\theta}V(q) - \underline{t} + \Psi(Q) = \overline{\theta}V(\overline{q}) - \overline{t} + \Psi(Q)$ and $\overline{\theta}V(\overline{q}) - \overline{t} + \Psi(Q) = \overline{\theta}V(q) - \underline{t} + \Psi(Q)$, we have $q = \overline{q}$ and $t = \overline{t}$. But this is impossible by Lemma 1.

Lemma 16.16.3 The IC and IR constraints of the same type cannot be simultaneously slack.

Proof. If IR($\theta$) and IC($\theta$) are both slack, increase $t(\theta)$ by a tiny increment will not violate all the constraints, but the firm’s charge will be increased.

Applying the above three lemmas, only five possible regimes are needed to be considered, which are summarized in the following table.

<table>
<thead>
<tr>
<th>Constraints</th>
<th>Regime1</th>
<th>Regime2</th>
<th>Regime3</th>
<th>Regime4</th>
<th>Regime5</th>
</tr>
</thead>
<tbody>
<tr>
<td>IR($\theta$)</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>S</td>
</tr>
<tr>
<td>IR($\overline{\theta}$)</td>
<td>S</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>IC($\theta$)</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>IC($\overline{\theta}$)</td>
<td>B</td>
<td>B</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
</tbody>
</table>

Table1 – The five possible regimes
Where “B” denotes “binding” and “S” denotes “slack”.

The regimes are ordered from 1 to 5 when \( \Delta G \) increases, and \( \Delta G \) itself is determined by the entrants’ marginal cost \( \omega \). To find how \( \omega \) affects the nonlinear pricing contract of the incumbent firm, we give the following two lemmas. Lemma 16.16.4 states the change of \( \Delta G \) in different regimes, and Lemma 16.16.5 shows how \( \Delta G \) is affected by \( \omega \).

**Lemma 16.16.4** The optimal pricing contracts and the utility difference \( \Delta G \) in different regimes are:

1. In regime 1, the optimal solution to (P5) is \( \underline{q} = q^{SB} \), \( \overline{q} = \overline{q}^{SB} \), \( \underline{U} = G \), and \( \overline{U} = G + \Delta \theta V(q^{SB}) \). The value of utility difference satisfies \( \Delta G < \Delta \theta V(q^{SB}) \).

2. In regime 2, the optimal consumption level \( \underline{q} \) and \( \overline{q} \) are determined by
   \[
   \begin{align*}
   \underline{q} &= V^{-1} \left( \frac{\Delta G}{\Delta \theta} \right) \\
   \theta V'(\overline{q}) + \Psi'(v\underline{q} + (1-v)\overline{q}) &= c,
   \end{align*}
   \]
   where \( \underline{q} \in [q^{SB}, q^{FB}] \) and \( \overline{q} \in [\overline{q}^{FB}, q^{SB}] \). The consumers’ information rents are \( \underline{U} = G \) and \( \overline{U} = G \). The utility difference satisfies \( \Delta \theta V(q^{SB}) \leq \Delta G \leq \Delta \theta V(q^{FB}) \).

3. In regime 3, the optimal solution to (P5) is \( \underline{q} = q^{FB} \), \( \overline{q} = q^{FB} \), \( \underline{U} = G \), \( \overline{U} = G \), and \( \Delta \theta V(q^{FB}) < \Delta G < \theta V(q^{FB}) \).

4. In regime 4, the optimal consumption level \( \underline{q} \) and \( \overline{q} \) are determined by
   \[
   \begin{align*}
   \overline{q} &= V^{-1} \left( \frac{\Delta G}{\Delta \theta} \right) \\
   \theta V'(\underline{q}) + \Psi'(v\overline{q} + (1-v)\underline{q}) &= c,
   \end{align*}
   \]
   where \( \underline{q} \in [q^{CI}, q^{FB}] \) and \( \overline{q} \in [q^{FB}, q^{CI}] \). The consumers’ information rents are \( \underline{U} = G \) and \( \overline{U} = G \). The utility difference satisfies \( \Delta \theta V(q^{FB}) \leq \Delta G \leq \Delta \theta V(q^{CI}) \).
5. In regime 5, the optimal contract is \( q = q^{CI}, \overline{q} = \overline{q}^{CI}, U = G - \Delta \theta V(q^{CI}), \) and \( U = \overline{G}. \) The utility difference \( \Delta G > \theta V(q^{CI}). \) \( q^{CI} \) and \( \overline{q}^{CI} \) are determined by:

\[
\begin{align*}
\theta V'(q^{CI}) + \Psi' \left( v q^{CI} + (1 - v) \overline{q}^{CI} \right) &= c \\
\left( \bar{\theta} + \frac{v}{1 - v} \Delta \bar{\theta} \right) V'(q^{CI}) + \Psi' \left( v q^{CI} + (1 - v) \overline{q}^{CI} \right) &= c.
\end{align*}
\]

(16.110)

Proof.

- In regime 1, the constraints \( IR(\bar{\theta}) \) and \( IC(\bar{\theta}) \) are binding. Solving (P5) in the same way as (P3), we get the second best solution

\[
\{ q = q^{SB}, \overline{q} = \overline{q}^{SB}; U = \overline{G}, \overline{U} = \overline{G} + \Delta \theta V(q^{SB}) \},
\]

with \( \Delta G < \Delta U = \Delta \theta V(q^{SB}). \)

- In regime 2, the constraints \( IR(\theta), IR(\overline{\theta}) \) and \( IC(\overline{\theta}) \) are binding. The optimal contract set is thus given by

\[
\{ (q, \overline{q}, U, \overline{U}) : \Delta \theta V(q) = \Delta G, \bar{\theta} V(q) + \Psi'(q) = c; U = G, \overline{U} = \overline{G} \}.
\]

Substituting \( IR(\bar{\theta}) \) and \( IR(\overline{\theta}) \) into the objective function, the Lagrange function of is constructed as

\[
L(q, \overline{q}) = v \left[ \theta V(q) - cq \right] + (1 - v) \left[ \bar{\theta} V(q) - c\overline{q} \right] + \Psi(Q) - \left[ vG + (1 - v)\overline{G} \right] + \lambda \left[ \Delta G - \Delta \theta V(q) \right],
\]

where \( \lambda > 0 \) is the Lagrange multiplier of the binding constraint \( IC(\bar{\theta}). \)

Then \( q \) and \( \overline{q} \) are determined by:

\[
\begin{align*}
\left( \bar{\theta} + \frac{v}{1 - v} \Delta \bar{\theta} \right) V'(q) + \Psi' \left( v q + (1 - v) \overline{q} \right) &= c, \\
\bar{\theta} V'(\overline{q}) + \Psi' \left( v \overline{q} + (1 - v) q \right) &= c.
\end{align*}
\]

(16.111)
Because $\theta - \frac{1}{v} \Delta \theta < \bar{\theta}$, from formula (16.105) it is easy to verify that $q < q^{FB}$ and $\bar{q} > \bar{q}^{FB}$. Substituting $IR(\bar{\theta})$ and $IC(\bar{\theta})$ into the objective function of (P5) and letting $\delta > 0$ be the Lagrange multiplier associate with the binding constraint $IR(\bar{\theta})$, we obtain the following Lagrange function:

$$L(q, \bar{q}) = v \left[ \theta V(q) - cq \right] + (1 - v) \left[ \bar{\theta} V(\bar{q}) - c\bar{q} \right] + \Psi(Q) - \left[ vG + (1 - v)(\bar{G} + \Delta \theta V(q)) \right] + \delta \left[ \Delta \theta V(q) - \Delta \bar{G} \right].$$

The optimal consumptions $q$ and $\bar{q}$ are determined by:

$$\begin{cases}
(\theta - \frac{1 - v - \delta}{v} \Delta \theta) V'(q) + \Psi'(vq + (1 - v)\bar{q}) = c, \\
\bar{\theta} V'(\bar{q}) + \Psi'(vq + (1 - v)\bar{q}) = c.
\end{cases}$$

(16.112)

Because $\theta - \frac{1 - v - \delta}{v} \Delta \theta > \bar{\theta} - \frac{1 - v}{v} \Delta \theta$, from expression (16.105) we have $q > q^{SB}$ and $\bar{q} < \bar{q}^{SB}$. It then suffice to show that $\Delta \theta V(q^{SB}) \leq \Delta G \leq \Delta \theta V(q^{FB})$.

- In regime 3, $IR(\bar{\theta})$ and $IR(\theta)$ are binding. Then the optimal contract is given by

$$\left\{ q = q^{FB}, \bar{q} = \bar{q}^{FB}; U = G, \bar{U} = \bar{G} \right\}.$$  

From the slack ICs, we can verify that $\Delta G$ satisfies $\Delta \theta V(q^{FB}) < \Delta G < \Delta \theta V(q^{FB})$.

- In regime 4, $IR(\theta)$ $IR(\bar{\theta})$ and $IC(\theta)$ are binding. The optimal contract is:

$$\left\{ (q, \bar{q}, U, \bar{U}) : \Delta \theta V(\bar{q}) = \Delta G, \theta V(q) + \Psi(Q) = c; U = G, \bar{U} = \bar{G} \right\}.$$  

Substituting $IR(\theta)$ and $IR(\bar{\theta})$ into the objective function, and letting $\mu > 0$ the multiplier associate with the binding constraint $IC(\theta)$, we have the following Lagrange function:

$$L(q, \bar{q}) = v \left[ \theta V(q) - cq \right] + (1 - v) \left[ \bar{\theta} V(\bar{q}) - c\bar{q} \right] + \Psi(Q) - \left[ vG + (1 - v)\bar{G} \right] + \mu \left[ \Delta \theta V(\bar{q}) - \Delta G \right].$$
Thus, \( q \) and \( \bar{q} \) are determined by

\[
\begin{align*}
\theta V'(q) + \Psi'(vq + (1 - v)\bar{q}) &= c, \\
\left( \bar{\theta} + \frac{\mu}{1 - v} \Delta \bar{\theta} \right) V'(\bar{q}) + \Psi'(vq + (1 - v)\bar{q}) &= c.
\end{align*}
\]  

(16.113)

Substituting \( IR(\bar{\theta}) \) and \( IC(\bar{\theta}) \) into the objective function, and letting \( \eta > 0 \) be the multiplier associate with the binding constraint \( IR(\bar{\theta}) \), we have the Lagrange function:

\[
L(q, \bar{q}) = v(\theta V'(q) - cq) + (1 - v)\left( \theta V'(\bar{q}) - c\bar{q} \right) + \Psi(Q) - \left[ v(\Delta G - \Delta \theta V(\bar{q})) + (1 - v)\bar{G} \right] + \eta \left[ \Delta G - \Delta \theta V(\bar{q}) \right].
\]

Then the \( q \) and \( \bar{q} \) are determined by:

\[
\begin{align*}
\theta V'(q) + \Psi'(vq + (1 - v)\bar{q}) &= c, \\
\left( \bar{\theta} + \frac{v - \eta}{1 - v} \Delta \theta \right) V'(\bar{q}) + \Psi'(vq + (1 - v)\bar{q}) &= c.
\end{align*}
\]  

(16.114)

To compare the different consumption levels, we make some comparative static analysis. To do so, let

\[
\begin{align*}
\theta V'(q) + \Psi'(vq + (1 - v)\bar{q}) &= c, \\
\beta V'(q) + \Psi'(vq + (1 - v)\bar{q}) &= c.
\end{align*}
\]  

(16.115)

If \( \beta = \bar{\theta} \), it’s the expression of \( q^{FB} \) and \( \bar{q}^{FB} \); if \( \beta = \bar{\theta} + \frac{v}{1 - v} \Delta \theta \), it coincides with the countervailing incentives solution \( q^{CI} \) and \( \bar{q}^{CI} \).

Differentiating these two equations with respect to \( \beta \) leads to:

\[
\begin{align*}
\left[ \theta V''(q) + v\Psi''(Q) \right] \frac{dq}{d\beta} + (1 - v)\Psi''(Q) \frac{d\bar{q}}{d\beta} &= 0, \\
v\Psi''(Q) \frac{d\bar{q}}{d\beta} + [\beta V''(q) + (1 - v)\Psi''(Q)] \frac{d\bar{q}}{d\beta} &= -V'(q).
\end{align*}
\]  

(16.116)

Thus, when \( \Psi''(Q) < 0 \) we have

\[
\begin{align*}
\frac{dq}{d\beta} &= \frac{(1 - v)V'(q)\Psi''(Q)}{\beta V''(q)[\theta V''(q) + v\Psi''(Q)] + (1 - v)\theta V''(q)\Psi''(Q)} < 0, \\
\frac{d\bar{q}}{d\beta} &= \frac{-V'(q)\theta V''(q) + v\Psi''(Q)}{\beta V''(q)[\theta V''(q) + v\Psi''(Q)] + (1 - v)\theta V''(q)\Psi''(Q)} > 0.
\end{align*}
\]  

(16.117)
Because \( \bar{\theta} + \frac{\nu}{1-\nu} \Delta \theta > \bar{\theta} \) and \( \bar{\theta} + \frac{\nu}{1-\nu} \Delta \theta < \bar{\theta} + \frac{\nu}{1-\nu} \Delta \theta \), from formula (16.117), it can be verified that \( q > q^{FB} \), \( q < q^{FB} \), \( q < q^{CI} \), and \( q > q^{CI} \). Thus, \( \Delta G = \Delta \theta V(q) \in [\Delta \theta V(q^{FB}), \Delta \theta V(q^{CI})] \).

- In regime 5, \( IR(\bar{\theta}) \) and \( IC(\bar{\theta}) \) are binding constraints. Substituting \( \bar{U} = \bar{G} \) and \( \bar{U} = \bar{G} - \Delta \theta V(\bar{q}) \) into the objective function, we obtain the following first order conditions:

\[
\begin{align*}
\frac{\partial V'(q)}{\partial \bar{q}} + \Psi' \left(v \bar{q} + (1-v)\bar{q}\right) &= c, \\
\left(\bar{\theta} + \frac{\nu}{1-\nu} \Delta \theta\right) V'(q) + \Psi' \left(v \bar{q} + (1-v)\bar{q}\right) &= c. 
\end{align*}
\] (16.118)

It is the countervailing incentives consumption level. Note that \( \bar{\theta} + \frac{\nu}{1-\nu} \Delta \theta > \bar{\theta} \), and thus, from (16.117) \( q^{FB} > q^{CI} \), \( q^{FB} < q^{CI} \). The difference of reservation utility satisfies \( \Delta G > \Delta U = \Delta \theta V(\bar{q}^{CI}) \).

The following figures 16.9 to 16.13 give depictions of the above regimes. ■

![Figure 16.9: regime 1 \( \Delta G < \Delta \theta V(q^{SB}) \)](image)

**Lemma 16.16.5** Suppose \( V(0) = 0, V'(\cdot) > 0, V''(\cdot) < 0, \) and \( V(\cdot) \) satisfies the standard Inada conditions: \( \lim_{q \to +\infty} V'(q) = 0 \), and \( \lim_{q \to 0} V'(q) = +\infty \). Then the utility difference across different states \( \Delta G = \bar{G} - \bar{G} \) is a decreasing function of the marginal cost \( \omega \), \( \lim_{\omega \to 0} \Delta G = +\infty \), and \( \lim_{\omega \to +\infty} \Delta G = 0 \).

619
Figure 16.10: regime 2 $\Delta \theta V(q^{SB}) \leq \Delta G \leq \Delta \theta V(q^{FB})$

Figure 16.11: regime 3 $\Delta \theta V(q^{FB}) < \Delta G < \theta V(\bar{q}^{FB})$. 
Figure 16.12: regime 4 $\Delta \theta V(q^B) \leq \Delta G \leq \Delta \theta V(q^{CI})$

Figure 16.13: regime 5 $\Delta G > \theta V(q^{CI})$. 
Figure 16.14: The change of $\Delta G$ with $\omega$.

**Proof.** The first order condition of $G(\theta) = \max_q [\theta V(q) - \omega q]$ is given by $\theta V'(q^*) = \omega$. So the maximized utility derived from network bypassing is: $G(\theta) = \theta[V(q^*(\theta)) - q^*(\theta)V'(q^*(\theta))]$. Let $\Phi(q) = V(q) - qV'(q)$. Then $\Delta G = G(\bar{\theta}) - G(\bar{\theta}) = \bar{\theta}\Phi(q^*) - \theta\Phi(q^*)$, and its derivative with respect to marginal cost $\omega$ is:

$$
\frac{d\Delta G}{d\omega} = \bar{\theta}\Phi'(q^*) \frac{dq^*}{d\omega} - \theta\Phi'(q^*) \frac{dq^*}{d\omega}
$$

$$
= -\bar{\theta}\bar{q} V''(\bar{q}) \frac{dq^*}{d\omega} + \theta q^* V''(q^*) \frac{dq^*}{d\omega}
$$

$$
= -\bar{\theta}\bar{q} V''(\bar{q}) \frac{1}{\bar{V}''(\bar{q})} + \theta q^* V''(q^*) \frac{1}{q''(q^*)}
$$

$$
= -\bar{\theta}\bar{q} + q^* < 0.
$$

It is easy to verify that when the conditions $V(0) = 0$, $V'(\cdot) > 0$, $V''(\cdot) < 0$, $\lim_{q\to+\infty} V'(q) = 0$, and $\lim_{q\to0} V'(q) = +\infty$ are satisfied, $\lim_{\omega\to0} \Delta G = +\infty$ and $\lim_{\omega\to+\infty} \Delta G = 0$. Figure 6 depicts the relationship of $\Delta G$ and $\omega$. Figure 16.14 depicts the relationship between $\Delta G$ and $\omega$.

From the above lemmas, one can see that if the potential entrants’ competitiveness increases, the utility differences will increase from zero to infinity. Thus, there exist positive values $\omega_i$, $i = 1, 2, 3, 4$, such that $\omega_1 < \omega_2 < \omega_3 < \omega_4$ corresponding to $\Delta \theta V(\bar{q}_C)$, $\Delta \theta V(\bar{q}_B)$, $\Delta \theta V(\bar{q}_{FB})$ and $\Delta \theta V(\bar{q}_{SB})$, respectively, where
$q^{CI}$, $q^{FB}$, $q^{FB}$ and $q^{SB}$ are given in expressions (16.121), (16.99) and (16.100). Thus, combining Lemmas 16.16.4 and 16.16.5, we can easily get the following proposition.

**Proposition 16.16.2** The optimal entry-deterring contract depends on the marginal cost of potential entrant. Specifically, there exist positive values $\omega_1 < \omega_2 < \omega_3 < \omega_4$ such that:

1. If $\omega > \omega_4$, then $\Delta G < \Delta \theta V(q^{SB})$, and consequently the pricing contract is: $q = q^{SB}$, $q = q^{SB}$, $U = G$, and $\bar{U} = G + \Delta \theta V(q^{SB})$.

2. If $\omega_3 \leq \omega \leq \omega_4$, then $\Delta \theta V(q^{SB}) \leq \Delta G \leq \Delta \theta V(q^{FB})$, and consequently the optimal consumption level $q$ and $\bar{q}$ are determined by

$$q = V^{-1}\left(\frac{\Delta G}{\Delta \theta}\right),$$

(16.119)

$$\theta V'(q) + \Psi'(vq + (1 - v)q) = c,$$

where $q \in [q^{SB}, q^{FB}]$ and $\bar{q} \in [q^{FB}, q^{SB}]$. The consumers’ information rents are $U = G$ and $\bar{U} = G$.

3. If $\omega_2 < \omega < \omega_3$, then $\Delta \theta V(q^{FB}) < \Delta G < \Delta \theta V(q^{FB})$, and consequently the pricing contract is $q = q^{FB}$, $\bar{q} = q^{FB}$, $U = G$, and $\bar{U} = G$.

4. If $\omega_1 \leq \omega \leq \omega_2$, then $\Delta \theta V(q^{FB}) \leq \Delta G \leq \Delta \theta V(q^{CI})$, and consequently the optimal consumption level $q$ and $\bar{q}$ are given by

$$q = V^{-1}\left(\frac{\Delta G}{\Delta \theta}\right),$$

(16.120)

$$\theta V'(q) + \Psi'(vq + (1 - v)q) = c,$$

where $q \in [q^{CI}, q^{FB}]$ and $\bar{q} \in [q^{FB}, q^{CI}]$. “CI” denotes “countervailing incentives”. The consumers’ information rents are $U = G$ and $\bar{U} = G$.

5. If $0 < \omega < \omega_1$, then $\Delta G > \Delta \theta V(q^{CI})$, and consequently the optimal contract is $q = q^{CI}$, $\bar{q} = q^{CI}$, $U = G - \Delta \theta V(q^{CI})$, $\bar{U} = G$. 623
\[ q^{CI} \] and \[ \bar{q}^{CI} \] are given by:
\[
\begin{cases}
\frac{\theta V'(q^{CI}) + \Psi'(vq^{CI} + (1 - v)\bar{q}^{CI})}{\bar{\theta} + \frac{v}{1 - v}\Delta \theta} V'(q^{CI}) + \Psi'(vq^{CI} + (1 - v)\bar{q}^{CI}) = c \\
(16.121)
\end{cases}
\]

Remark 16.16.3 When \( \omega > \omega_4 \), the second best contract is also entry deterring. It means that when the outside competitors are not efficient enough to give high demand consumers enough utility exceeding their information rents acquired from the present network, the outside market is only attractive to low demand consumers. The incumbent firm need not to change its pricing contract when facing the entry threat of a firm with low competitiveness.

When \( \omega_3 \leq \omega \leq \omega_4 \), we have \( q \in [q^{SB}, q^{FB}] \) and \( \bar{q} \in [q^{FB}, \bar{q}^{SB}] \). That means when the marginal cost \( \omega \) decreases to the extent that the utility difference \( \Delta G \) is large enough to attract high demand consumers bypassing the present network, the monopolist must give up more information rent to him by increasing the consumption level of low demand consumers. The consumption level of high demand consumers themselves should also be lowered accordingly because of network effects. In this case, the sharper competitiveness of outside competitors makes the allocations less distorted.

When \( \omega_2 < \omega < \omega_3 \), asymmetric information imposes no distortion on both types’ allocation. As \( \omega \) decreases and \( \Delta G \) increases further, \( q \) will reach the first best level, it is suboptimal for the monopolist to increase the high type consumers’ information rent at the cost of distorting the consumption level of the low demand consumers upward. In this case, the main task for the firm toward the high type consumers is to prevent them from bypassing the incumbent market instead of preventing them from misreporting, the participation constraints are more difficult to be satisfied than the incentive compatible constraints. Thus only the IRs are binding, and the first best allocation is attained.

When \( \omega_1 \leq \omega \leq \omega_2 \), we have \( q \in [q^{CI}, q^{FB}] \) and \( \bar{q} \in [q^{FB}, \bar{q}^{CI}] \). The high difference of utilities induces the low type consumers to pretend to be a high
type, from which the countervailing incentives problem arises. The $IC(\theta)$, $IR(\theta)$ and $IR(\theta)$ in (P5) are binding. Again, the allocations of the two types will be distorted in opposite directions. But it is different from the distortions in cases 1 and 2. The monopolist then distorts $q$ downward to curb the rent of high demand consumers. In this case, however, information rent has to be given to low demand consumers to elicit them reporting their types truthfully. The information rent is a decreasing function of high demand consumers’ consumption $\overline{q}$, and so $\overline{q}$ has to be distorted upward to reduce the information rent gained by low demand consumers, while a certain amount of the low demand type’s consumption is “crowded” out of the network so that $\overline{q}$ is distorted downward.

When $0 < \omega < \omega_1$, the allocations remain at the countervailing incentives level: $q = q^{CI}$ and $\overline{q} = \overline{q}^{CI}$. The decrease in marginal cost $\omega$ demand further upward distortion on the consumption of high demand consumers (the consumption of low demand consumers will be distorted downward accordingly). Thus, the participation constraint of the low-type has to be slackened, which means certain amount of information rent should be given to the consumers with low willingness to pay. In this case, only $IC(\theta)$ and $IR(\theta)$ are binding constraints, the low-type consumers get information rent $\overline{G} - \Delta V (q^{CI})$. The incumbent firm keep reducing the tariffs ($\ell$ and $\overline{t}$ keep decreasing in this case) instead of distorting allocations to prevent high demand consumers from bypassing and low demand consumers from misreporting.

Changes in consumption with marginal cost $\omega$ are summarized in the following figure 16.15.

16.17 The Optimal Contract with a Continuum of Types

In this section, we give a brief account of the continuum type case. Most of the principal-agent literature is written within this framework.
Reconsider the standard model with $\theta$ in $\Theta = [\underline{\theta}, \bar{\theta}]$, with a cumulative distribution function $F(\theta)$ and a density function $f(\theta) > 0$ on $[\underline{\theta}, \bar{\theta}]$. Since the revelation principle is still valid with a continuum of types, and we can restrict our analysis to direct revelation mechanisms $\{(q(\tilde{\theta}), t(\tilde{\theta}))\}$, which are truthful, i.e., such that

$$t(\theta) - \theta q(\theta) \geq t(\tilde{\theta}) - \theta q(\tilde{\theta}) \quad \text{for any } (\theta, \tilde{\theta}) \in \Theta^2. \quad (16.122)$$

In particular, (16.122) implies

$$t(\theta) - \theta q(\theta) \geq t(\theta') - \theta q(\theta'), \quad (16.123)$$

$$t(\theta') - \theta' q(\theta') \geq t(\theta) - \theta' q(\theta) \quad \text{for all pairs } (\theta, \theta') \in \Theta^2. \quad (16.124)$$

Adding (16.123) and (16.124) we obtain

$$(\theta - \theta')(q(\theta') - q(\theta)) \geq 0. \quad (16.125)$$

Thus, incentive compatibility alone requires that the schedule of output $q(\cdot)$ has to be nonincreasing. This implies that $q(\cdot)$ is differentiable almost everywhere. So we will restrict the analysis to differentiable functions.

(16.122) implies that the following first-order condition for the optimal response $\tilde{\theta}$ chosen by type $\theta$ is satisfied

$$\dot{t}(\tilde{\theta}) - \theta \dot{q}(\tilde{\theta}) = 0. \quad (16.126)$$
For the truth to be an optimal response for all $\theta$, it must be the case that

$$\dot{t}(\theta) - \theta \dot{q}(\theta) = 0,$$

(16.127)

and (16.127) must hold for all $\theta$ in $\Theta$ since $\theta$ is unknown to the principal.

It is also necessary to satisfy the local second-order condition,

$$\ddot{i}(\tilde{\theta})|_{\tilde{\theta}=\theta} - \theta \ddot{q}(\tilde{\theta})|_{\tilde{\theta}=\theta} \leq 0$$

(16.128)

or

$$\ddot{i}(\theta) - \theta \ddot{q}(\theta) \leq 0.$$  

(16.129)

But differentiating (16.127), (16.129) can be written more simply as

$$-\dot{q}(\theta) \geq 0.$$  

(16.130)

(16.127) and (16.130) constitute the local incentive constraints, which ensure that the agent does not want to lie locally. Now we need to check that he does not want to lie globally either, therefore the following constraints must be satisfied

$$t(\theta) - \theta q(\theta) \geq t(\tilde{\theta}) - \theta q(\tilde{\theta}) \text{ for any } (\theta, \tilde{\theta}) \in \Theta^2.$$  

(16.131)

From (16.127) we have

$$t(\theta) - t(\tilde{\theta}) = \int_{\tilde{\theta}}^{\theta} \tau \dot{q}(\tau) d\tau = \theta q(\theta) - \tilde{\theta} q(\tilde{\theta}) - \int_{\tilde{\theta}}^{\theta} q(\tau) d\tau$$

(16.132)

or

$$t(\theta) - \theta q(\theta) = t(\tilde{\theta}) - \theta q(\tilde{\theta}) + (\theta - \tilde{\theta}) q(\tilde{\theta}) - \int_{\tilde{\theta}}^{\theta} q(\tau) d\tau.$$  

(16.133)

Indeed, since $q(\cdot)$ is nonincreasing, $(\theta - \tilde{\theta}) q(\tilde{\theta}) - \int_{\tilde{\theta}}^{\theta} q(\tau) d\tau \geq 0$. Thus, it turns out that the local incentive constraints (16.127) also imply the global incentive constraints.

In such circumstances, the infinity of incentive constraints (16.131) reduces to a differential equation and to a monotonicity constraint. Local analysis of
incentives is enough. Truthful revelation mechanisms are then characterized by
the two conditions (16.127) and (16.130).

Let us use the rent variable $U(\theta) = t(\theta) - \theta q(\theta)$. The local incentive constraint
is now written as (by using (16.127))

\[ \dot{U}(\theta) = -q(\theta). \]  
(16.134)

The optimization program of the principal becomes

\[ \max_{\{U(\cdot),q(\cdot)\}} \int_{\theta}^{\bar{\theta}} \left( S(q(\theta)) - \theta q(\theta) - U(\theta) \right) f(\theta) d\theta \]  
(16.135)

subject to

\[ \dot{U}(\theta) = -q(\theta), \]  
(16.136)
\[ \dot{q}(\theta) \leq 0, \]  
(16.137)
\[ U(\theta) \geq 0. \]  
(16.138)

Using (16.136), the participation constraint (16.138) simplifies to $U(\bar{\theta}) \geq 0$. As in the discrete case, incentive compatibility implies that only the participation constraint of the most inefficient type can be binding. Furthermore, it is clear from the above program that it will be binding. i.e., $U(\bar{\theta}) = 0$.

Momentarily ignoring (16.137), we can solve (16.136)

\[ U(\bar{\theta}) - U(\theta) = -\int_{\theta}^{\bar{\theta}} q(\tau) d\tau \]  
(16.139)

or, since $U(\bar{\theta}) = 0$,

\[ U(\theta) = \int_{\theta}^{\bar{\theta}} q(\tau) d\tau \]  
(16.140)

The principal’s objective function becomes

\[ \int_{\theta}^{\bar{\theta}} \left( S(q(\theta)) - \theta q(\theta) - \int_{\theta}^{\bar{\theta}} q(\tau) d\tau \right) f(\theta) d\theta, \]  
(16.141)

which, by an integration of parts, gives

\[ \int_{\theta}^{\bar{\theta}} \left( S(q(\theta)) - \left( \theta + \frac{F(\theta)}{f(\theta)} \right) q(\theta) \right) f(\theta) d\theta. \]  
(16.142)
Maximizing pointwise (16.142), we get the second-best optimal outputs

$$S'(q^{SB}(\theta)) = \theta + \frac{F(\theta)}{f(\theta)},$$

which is the first order condition for the case of a continuum of types.

If the monotone hazard rate property $\frac{d}{d\theta} \left( \frac{F(\theta)}{f(\theta)} \right) \geq 0$ holds, the solution $q^{SB}(\theta)$ of (16.143) is clearly decreasing, and the neglected constraint (16.137) is satisfied. All types choose therefore different allocations and there is no bunching in the optimal contract.

From (16.143), we note that there is no distortion for the most efficient type (since $F(\theta) = 0$ and a downward distortion for all the other types.

All types, except the least efficient one, obtain a positive information rent at the optimal contract

$$U^{SB}(\theta) = \int_{\theta}^{\bar{\theta}} q^{SB}(\tau) d\tau.$$  

(16.144)

Finally, one could also allow for some shutdown of types. The virtual surplus $S(q) - (\theta + \frac{F(\theta)}{f(\theta)}) q$ decreases with $\theta$ when the monotone hazard rate property holds, and shutdown (if any) occurs on an interval $[\theta^*, \bar{\theta}]$. $\theta^*$ is obtained as a solution to

$$\max_{\{\theta^*\}} \int_{\theta}^{\theta^*} \left( S(q^{SB}(\theta)) - \left( \theta + \frac{F(\theta)}{f(\theta)} \right) q^{SB}(\theta) \right) f(\theta) d\theta.$$ 

For an interior optimum, we find that

$$S(q^{SB}(\theta^*)) = \left( \theta^* + \frac{F(\theta^*)}{f(\theta^*)} \right) q^{SB}(\theta^*).$$

As in the discrete case, one can check that the Inada condition $S'(0) = +\infty$ and the condition $\lim_{q \to 0} S'(q) q = 0$ ensure the corner solution $\theta^* = \bar{\theta}$. 

**Remark 16.17.1** The optimal solution above can also be derived by using the Pontryagin principle. The Hamiltonian is then

$$H(q, U, \mu, \theta) = (S(q) - \theta q - U) f(\theta) - \mu q,$$

(16.145)

where $\mu$ is the co-state variable, $U$ the state variable and $q$ the control variable.
From the Pontryagin principle,

\[ \dot{\mu}(\theta) = -\frac{\partial H}{\partial U} = f(\theta). \] (16.146)

From the transversality condition (since there is no constraint on \( U(\cdot) \) at \( \theta \)),

\[ \mu(\theta) = 0. \] (16.147)

Integrating (16.146) using (16.147), we get

\[ \mu(\theta) = F(\theta). \] (16.148)

Optimizing with respect to \( q(\cdot) \) also yields

\[ S'(q^{SB}(\theta)) = \theta + \frac{\mu(\theta)}{f(\theta)}, \] (16.149)

and inserting the value of \( \mu(\theta) \) obtained from (16.148) again yields (16.143).

We have derived the optimal truthful direct revelation mechanism \( \{(q^{SB}(\theta), U^{SB}(\theta))\} \) or \( \{(q^{SB}(\theta), t^{SB}(\theta))\} \). It remains to be investigated if there is a simple implementation of this mechanism. Since \( q^{SB}(\cdot) \) is decreasing, we can invert this function and obtain \( \theta^{SB}(q) \). Then,

\[ t^{SB}(\theta) = U^{SB}(\theta) + \theta q^{SB}(\theta) \] (16.150)

becomes

\[ T(q) = t^{SB}(\theta^{SB}(q)) = \int_{\theta(q)}^{\bar{\theta}} q^{SB}(\tau)d\tau + \theta(q)q. \] (16.151)

To the optimal truthful direct revelation mechanism we have associated a nonlinear transfer \( T(q) \). We can check that the agent confronted with this nonlinear transfer chooses the same allocation as when he is faced with the optimal revelation mechanism. Indeed, we have \( \frac{d}{dq}(T(q) - \theta q) = T'(q) - \theta = \frac{dt^{SB}}{d\theta} \cdot \frac{dq^{SB}}{dq} - \theta = 0, \) since \( \frac{dt^{SB}}{d\theta} - \theta \frac{dq^{SB}}{dq} = 0. \)

To conclude, the economic insights obtained in the continuum case are not different from those obtained in the two-state case.
16.18 Further Extensions

The main theme of this chapter was to determine how the fundamental conflict between rent extraction and efficiency could be solved in a principal-agent relationship with adverse selection. In the models discussed, this conflict was relatively easy to understand because it resulted from the simple interaction of a single incentive constraint with a single participation constraint. Here we would mention some possible extensions.

One can consider a straightforward three-type extension of the standard model. One can also deal with a bidimensional adverse selection model, a two-type model with type-dependent reservation utilities, random participation constraints, the limited liability constraints, and the audit models. For detailed discussion about these topics and their applications, see Laffont and Martimort (2002).

16.19 Reference

Books and Monographs:


**Papers:**

632


Chapter 17

Optimal Mechanism Design:
Contracts with One-Agent and Hidden Action

17.1 Introduction

In the previous chapter, we stressed that the delegation of tasks creates an information gap between the principal and his agent when the latter learns some piece of information relevant to determining the efficient volume of trade. Adverse selection is not the only informational problem one can imagine. Agents may also choose actions that affect the value of trade or, more generally, the agent’s performance. The principal often loses any ability to control those actions that are no longer observable, either by the principal who offers the contract or by the court of law that enforces it. In such cases we will say that there is moral hazard.

The leading candidates for such moral hazard actions are effort variables, which positively influence the agent’s level of production but also create a disutility for the agent. For instance the yield of a field depends on the amount of time that the tenant has spent selecting the best crops, or the quality of their
harvesting. Similarly, the probability that a driver has a car crash depends on how safely he drives, which also affects his demand for insurance. Also, a regulated firm may have to perform a costly and nonobservable investment to reduce its cost of producing a socially valuable good.

As in the case of adverse selection, asymmetric information also plays a crucial role in the design of the optimal incentive contract under moral hazard. However, instead of being an exogenous uncertainty for the principal, uncertainty is now endogenous. The probabilities of the different states of nature, and thus the expected volume of trade, now depend explicitly on the agent’s effort. In other words, the realized production level is only a noisy signal of the agent’s action. This uncertainty is key to understanding the contractual problem under moral hazard. If the mapping between effort and performance were completely deterministic, the principal and the court of law would have no difficulty in inferring the agent’s effort from the observed output. Even if the agent’s effort was not observable directly, it could be indirectly contracted upon, since output would itself be observable and verifiable.

We will study the properties of incentive schemes that induce a positive and costly effort. Such schemes must thus satisfy an incentive constraint and the agent’s participation constraint. Among such schemes, the principal prefers the one that implements the positive level of effort at minimal cost.

The main theme of this chapter is to trade off incentives against insurance in a principal-agent relationship with hidden action. At the optimal second-best contract, the principal trades off his desire to reach allocative efficiency against the costly information rent given up to the agent to induce information revelation. This cost minimization yields the characterization of the second-best cost of implementing this effort. In general, this second-best cost is greater than the first-best cost that would be obtained by assuming that effort is observable. An allocative inefficiency emerges as the result of the conflict of interests between the principal and the agent.
17.2 Basic Settings of Principal-Agent Model with Moral Hazard

17.2.1 Effort and Production

We consider an agent who can exert a costly effort \( e \). Two possible values can be taken by \( e \), which we normalize as a zero effort level and a positive effort of one: \( e \in \{0, 1\} \). Exerting effort \( e \) implies a disutility for the agent that is equal to \( \psi(e) \) with the normalization \( \psi(0) = \psi_0 = 0 \) and \( \psi_1 = \psi \).

The agent receives a transfer \( t \) from the principal. We assume that his utility function is separable between money and effort, \( U = u(t) - \psi(e) \), with \( u(\cdot) \) increasing and concave \( (u' > 0, u'' < 0) \). Sometimes we will use the function \( h = u^{-1} \), the inverse function of \( u(\cdot) \), which is increasing and convex \( (h' > 0, h'' > 0) \).

Production is stochastic, and effort affects the production level as follows: the stochastic production level \( \tilde{q} \) can only take two values \( \{\bar{q}, \tilde{q}\} \), with \( \bar{q} - \tilde{q} = \Delta q > 0 \), and the stochastic influence of effort on production is characterized by the probabilities \( \Pr(\tilde{q} = \bar{q}|e = 0) = \pi_0 \), and \( \Pr(\tilde{q} = \bar{q}|e = 1) = \pi_1 \), with \( \pi_1 > \pi_0 \). We will denote the difference between these two probabilities by \( \Delta \pi = \pi_1 - \pi_0 \).

Note that effort improves production in the sense of first-order stochastic dominance, i.e., \( \Pr(\tilde{q} \leq \tilde{q}^*|e) \) is decreasing with \( e \) for any given production \( q^* \).

Indeed, we have \( \Pr(\tilde{q} \leq \bar{q}|e = 1) = 1 - \pi_1 < 1 - \pi_0 = \Pr(\tilde{q} \leq \bar{q}|e = 0) \) and \( \Pr(\tilde{q} \leq \bar{q}|e = 1) = 1 = \Pr(\tilde{q} \leq \bar{q}|e = 0) \).

17.2.2 Incentive Feasible Contracts

Since the agent’s action is not directly observable by the principal, the principal can only offer a contract based on the observable and verifiable production level, i.e., a function \( \{t(\tilde{q})\} \) linking the agent’s compensation to the random output \( \tilde{q} \).

With two possible outcomes \( \tilde{q} \) and \( \bar{q} \), the contract can be defined equivalently by a pair of transfers \( \tilde{t} \) and \( \bar{t} \). Transfer \( \tilde{t} \) (resp. \( \bar{t} \)) is the payment received by the agent if the production \( \tilde{q} \) (resp. \( \bar{q} \)) is realized.
The risk-neutral principal’s expected utility is now written as

\[ V_1 = \pi_1 (S(\bar{q}) - \bar{t}) + (1 - \pi_1) (S(q) - t) \]  

(17.1)

if the agent makes a positive effort \((e = 1)\) and

\[ V_0 = \pi_0 (S(\bar{q}) - \bar{t}) + (1 - \pi_0) (S(q) - t) \]  

(17.2)

if the agent makes no effort \((e = 0)\). For notational simplicity, we will denote the principal’s benefits in each state of nature by \(S(\bar{q}) = \bar{S}\) and \(S(q) = S\).

Each level of effort that the principal wishes to induce corresponds to a set of contracts ensuring moral hazard incentive compatibility constraint and participation constraint are satisfied:

\[ \pi_1 u(\bar{t}) + (1 - \pi_1) u(t) - \psi \geq \pi_0 u(\bar{t}) + (1 - \pi_0) u(t) \]  

(17.3)

\[ \pi_1 u(\bar{t}) + (1 - \pi_1) u(t) - \psi \geq 0. \]  

(17.4)

Note that the participation constraint is ensured at the ex ante stage, i.e., before the realization of the production shock.

**Definition 17.2.1** An incentive feasible contract satisfies the incentive compatibility and participation constraints (17.3) and (17.4).

The timing of the contracting game under moral hazard is summarized in the figure below.

Figure 17.1: Timing of contracting under moral hazard.
17.2.3 The Complete Information Optimal Contract

As a benchmark, let us first assume that the principal and a benevolent court of law can both observe effort. Then, if he wants to induce effort, the principal’s problem becomes

\[
\max_{\{\bar{t}, \tilde{t}\}} \pi_1(\bar{S} - \bar{t}) + (1 - \pi_1)(\bar{S} - \tilde{t})
\]

subject to (17.4).

Indeed, only the agents participation constraint matters for the principal, because the agent can be forced to exert a positive level of effort. If the agent were not choosing this level of effort, the agent could be heavily punished, and the court of law could commit to enforce such a punishment.

Denoting the multiplier of this participation constraint by \(\lambda\) and optimizing with respect to \(\bar{t}\) and \(\tilde{t}\) yields, respectively, the following first-order conditions:

\[
\begin{align*}
-\pi_1 + \lambda \pi_1 u'(\tilde{t}^*) &= 0, \\
-(1 - \pi_1) + \lambda (1 - \pi_1) u'(t^*) &= 0,
\end{align*}
\]

where \(\tilde{t}^*\) and \(t^*\) are the first-best transfers.

From (17.6) and (17.7) we immediately derive that \(\lambda = \frac{1}{u'(\tilde{t}^*)} = \frac{1}{u'(t^*)} > 0\), and finally that \(t^* = \tilde{t}^* = t^*\).

Thus, with a verifiable effort, the agent obtains full insurance from the risk-neutral principal, and the transfer \(t^*\) he receives is the same whatever the state of nature. Because the participation constraint is binding we also obtain the value of this transfer, which is just enough to cover the disutility of effort, namely \(t^* = h(\psi)\). This is also the expected payment made by the principal to the agent, or the first-best cost \(C^{FB}\) of implementing the positive effort level.

For the principal, inducing effort yields an expected payoff equal to

\[
V_1 = \pi_1 \bar{S} + (1 - \pi_1)\bar{S} - h(\psi)
\]

Had the principal decided to let the agent exert no effort, \(e_0\), he would make a zero payment to the agent whatever the realization of output. In this scenario,
the principal would instead obtain a payoff equal to

$$V_0 = \pi_0 \bar{S} + (1 - \pi_0) \bar{S}. \quad (17.9)$$

Inducing effort is thus optimal from the principal’s point of view when $V_1 \geq V_0$, i.e., $\pi_1 \bar{S} + (1 - \pi_1) \bar{S} - h(\psi) \geq \pi_0 \bar{S} + (1 - \pi_0) \bar{S}$, or to put it differently, when the expected gain of effect is greater than first-best cost of inducing effect, i.e.,

$$\Delta \pi \Delta S \geq h(\psi) \quad (17.10)$$

where $\Delta S = \bar{S} - \bar{S} > 0$.

Denoting the benefit of inducing a strictly positive effort level by $B = \Delta \pi \Delta S$, the first-best outcome calls for $e^* = 1$ if and only if $B > h(\psi)$, as shown in the figure below.

Figure 17.2: First-best level of effort.

### 17.3 Risk Neutrality and First-Best Implementation

If the agent is risk-neutral, we have (up to an affine transformation) $u(t) = t$ for all $t$ and $h(u) = u$ for all $u$. The principal who wants to induce effort must thus
choose the contract that solves the following problem:

$$\max_{\{l, \bar{l}\}} \pi_1(\bar{S} - \bar{l}) + (1 - \pi_1)(\bar{S} - \bar{\ell})$$

s.t. 

$$\pi_1 \bar{l} + (1 - \pi_1) \ell - \psi \geq \pi_0 \bar{l} + (1 - \pi_0) \ell$$ \hspace{1cm} (17.11)

$$\pi_1 \bar{l} + (1 - \pi_1) \ell - \psi \geq 0.$$ \hspace{1cm} (17.12)

With risk neutrality the principal can, for instance, choose incentive compatible transfers $\bar{l}$ and $\ell$, which make the agent’s participation constraint binding and leave no rent to the agent. Indeed, solving (17.11) and (17.12) with equalities, we immediately obtain

$$l^* = -\frac{\pi_0}{\Delta \pi} \psi$$ \hspace{1cm} (17.13)

and

$$\bar{l}^* = \frac{1 - \pi_0}{\Delta \pi} \psi = l^* + \frac{1}{\Delta \pi} \psi.$$ \hspace{1cm} (17.14)

The agent is rewarded if production is high. His net utility in this state of nature $\bar{U}^* = \bar{l}^* - \psi = \frac{1 - \pi_1}{\Delta \pi} \psi > 0$. Conversely, the agent is punished if production is low. His corresponding net utility $U^* = l^* - \psi = -\frac{\pi_1}{\Delta \pi} \psi < 0$.

The principal makes an expected payment $\pi_1 \bar{l}^* + (1 - \pi_1) \ell^* = \psi$, which is equal to the disutility of effort he would incur if he could control the effort level perfectly. The principal can costlessly structure the agent’s payment so that the latter has the right incentives to exert effort. Using (17.13) and (17.14), his expected gain from exerting effort is thus $\Delta \pi(\bar{l}^* - \ell^*) = \psi$ when increasing his effort from $e = 0$ to $e = 1$.

**Proposition 17.3.1** Moral hazard is not an issue with a risk-neutral agent despite the nonobservability of effort. The first-best level of effort is still implemented.

**Remark 17.3.1** One may find the similarity of these results with those described last chapter. In both cases, when contracting takes place ex ante, the incentive constraint, under either adverse selection or moral hazard, does not conflict with the ex ante participation constraint with a risk-neutral agent, and the first-best outcome is still implemented.
Remark 17.3.2 Inefficiencies in effort provision due to moral hazard will arise when the agent is no longer risk-neutral. There are two alternative ways to model these transaction costs. One is to maintain risk neutrality for positive income levels but to impose a limited liability constraint, which requires transfers not to be too negative. The other is to let the agent be strictly risk-averse. In the following, we analyze these two contractual environments and the different trade-offs they imply.

17.4 The Trade-Off Between Limited Liability Rent Extraction and Efficiency

Let us consider a risk-neutral agent. As we have already seen, (17.3) and (17.4) now take the following forms:

\[
\pi_1 \tilde{t} + (1 - \pi_1)\bar{t} - \psi \geq \pi_0 \bar{t} + (1 - \pi_0) \bar{t}
\]

(17.15)

and

\[
\pi_1 \tilde{t} + (1 - \pi_1)\bar{t} - \psi \geq 0.
\]

(17.16)

Let us also assume that the agent’s transfer must always be greater than some exogenous level \(-l\), with \(l \geq 0\). Thus, limited liability constraints in both states of nature are written as

\[
\tilde{t} \geq -l
\]

(17.17)

and

\[
\bar{t} \geq -l.
\]

(17.18)

These constraints may prevent the principal from implementing the first-best level of effort even if the agent is risk-neutral. Indeed, when he wants to induce a high effort, the principal’s program is written as

\[
\max_{\{\tilde{t}, \bar{t}\}} \pi_1 (\bar{S} - \tilde{t}) + (1 - \pi_1)(\bar{S} - \bar{t})
\]

subject to (17.15) to (17.18).

Then, we have the following proposition.
Proposition 17.4.1  With limited liability, the optimal contract inducing effort from the agent entails:

(1) For \( l > \frac{\pi_0}{\Delta \pi} \psi \), only (17.15) and (17.16) are binding. Optimal transfers are given by (17.13) and (17.14). The agent has no expected limited liability rent; \( EU^{SB} = 0 \).

(2) For \( 0 \leq l \leq \frac{\pi_0}{\Delta \pi} \psi \), (17.15) and (17.18) are binding. Optimal transfers are then given by:

\[
t^{SB} = -l, \quad (17.20)
\]
\[
\bar{t}^{SB} = -l + \frac{\psi}{\Delta \pi}. \quad (17.21)
\]

(3) Moreover, the agent’s expected limited liability rent \( EU^{SB} \) is non-negative:

\[
EU^{SB} = \pi_1 t^{SB} + (1 - \pi_1) \bar{t}^{SB} - \psi = -l + \frac{\pi_0}{\Delta \pi} \psi \geq 0. \quad (17.22)
\]

Proof. First suppose that \( 0 \leq l \leq \frac{\pi_0}{\Delta \pi} \psi \). We conjecture that (17.15) and (17.18) are the only relevant constraints. Of course, since the principal is willing to minimize the payments made to the agent, both constraints must be binding. Hence, \( t^{SB} = -l \) and \( \bar{t}^{SB} = -l + \frac{\psi}{\Delta \pi} \). We check that (17.17) is satisfied since \(-l + \frac{\psi}{\Delta \pi} > -l\). We also check that (17.16) is satisfied since \( \pi_1 t^{SB} + (1 - \pi_1) \bar{t}^{SB} - \psi = -l + \frac{\pi_0}{\Delta \pi} \psi \geq 0 \).

For \( l > \frac{\pi_0}{\Delta \pi} \psi \), note that the transfers \( \underline{t}^* = -\frac{\pi_0}{\Delta \pi} \psi \), and \( \bar{t}^* = -\psi + \frac{(1 - \pi_1) \psi}{\Delta \pi} > \underline{t}^* \) are such that both limited liability constraints (17.17) and (17.18) are strictly satisfied, and (17.15) and (17.16) are both binding. In this case, it is costless to induce a positive effort by the agent, and the first-best outcome can be implemented. The proof is completed.

Note that only the limited liability constraint in the bad state of nature may be binding. When the limited liability constraint (17.18) is binding, the principal is limited in his punishments to induce effort. The risk-neutral agent does not have enough assets to cover the punishment if \( \underline{q} \) is realized in order to induce
effort provision. The principal uses rewards when a good state of nature \( \tilde{q} \) is realized. As a result, the agent receives a non-negative ex ante limited liability rent described by (17.22). Compared with the case without limited liability, this rent is actually the additional payment that the principal must incur because of the conjunction of moral hazard and limited liability.

As the agent becomes endowed with more assets, i.e., as \( l \) gets larger, the conflict between moral hazard and limited liability diminishes and then disappears whenever \( l \) is large enough.

### 17.5 The Trade-Off Between Insurance and Efficiency

Now suppose the agent is risk-averse. The principal’s program is written as:

\[
\max_{(\tilde{t}, \tilde{t})} \pi_1 (\tilde{S} - \tilde{t}) + (1 - \pi_1)(\bar{S} - \bar{t}) \tag{17.23}
\]

subject to (17.3) and (17.4).

Since the principal’s optimization problem may not be a concave program for which the first-order Kuhn and Tucker conditions are necessary and sufficient, we make the following change of variables. Define \( \bar{u} = u(\tilde{t}) \) and \( \bar{u} = u(t) \), or equivalently let \( \tilde{t} = h(\bar{u}) \) and \( t = h(\bar{u}) \). These new variables are the levels of ex post utility obtained by the agent in both states of nature. The set of incentive feasible contracts can now be described by two linear constraints:

\[
\begin{align*}
\pi_1 \bar{u} + (1 - \pi_1)\bar{u} - \psi & \geq \pi_0 \bar{u} + (1 - \pi_0)\bar{u}, \tag{17.24} \\
\pi_1 \bar{u} + (1 - \pi_1)\bar{u} - \psi & \geq 0, \tag{17.25}
\end{align*}
\]

which replaces (17.3) and (17.4), respectively.

Then, the principal’s program can be rewritten as

\[
\max_{(\bar{u}, \bar{u})} \pi_1 (\tilde{S} - h(\bar{u})) + (1 - \pi_1)(\bar{S} - h(\bar{u})) \tag{17.26}
\]
subject to (17.24) and (17.25).

Note that the principal’s objective function is now strictly concave in \((\bar{u}, u)\) because \(h(\cdot)\) is strictly convex. The constraints are now linear and the interior of the constrained set is obviously non-empty.

### 17.5.1 Optimal Transfers

Letting \(\lambda\) and \(\mu\) be the non-negative multipliers associated respectively with the constraints (17.24) and (17.25), the first-order conditions of this program can be expressed as

\[
-\pi_1 h'(\bar{u}^{SB}) + \lambda \Delta \pi + \mu \pi_1 = -\frac{\pi_1}{u'(t^{SB})} + \lambda \Delta \pi + \mu \pi_1 = 0, \quad (17.27)
\]

\[
-(1 - \pi_1) h'(u^{SB}) - \lambda \Delta \pi + \mu (1 - \pi_1) = -\frac{(1 - \pi_1)}{u'(t^{SB})} - \lambda \Delta \pi + \mu (1 - \pi_1) = 0. \quad (17.28)
\]

where \(\bar{t}^{SB}\) and \(t^{SB}\) are the second-best optimal transfers. Rearranging terms, we get

\[
\frac{1}{u'(t^{SB})} = \mu + \frac{\lambda \Delta \pi}{\pi_1}, \quad (17.29)
\]

\[
\frac{1}{u'(t^{SB})} = \mu - \lambda \frac{\Delta \pi}{1 - \pi_1}. \quad (17.30)
\]

The four variables \((t^{SB}, \bar{t}^{SB}, \lambda, \mu)\) are simultaneously obtained as the solutions to the system of four equations (17.24), (17.25), (17.29), and (17.30). Multiplying (17.29) by \(\pi_1\) and (17.30) by \(1 - \pi_1\), and then adding those two modified equations we obtain

\[
\mu = \frac{\pi_1}{u'(t^{SB})} + \frac{1 - \pi_1}{u'(t^{SB})} > 0. \quad (17.31)
\]

Hence, the participation constraint (17.16) is necessarily binding. Using (17.31) and (17.29), we also obtain

\[
\lambda = \frac{\pi_1 (1 - \pi_1)}{\Delta \pi} \left( \frac{1}{u'(t^{SB})} - \frac{1}{u'(t^{SB})} \right), \quad (17.32)
\]

where \(\lambda\) must also be strictly positive. Indeed, from (17.24) we have \(\bar{u}^{SB} - u^{SB} \geq \frac{\psi}{\Delta \pi} > 0\) and thus \(t^{SB} > l^{SB}\), implying that the right-hand side of (17.32) is strictly
positive since \( u'' < 0 \). Using that (17.24) and (17.25) are both binding, we can immediately obtain the values of \( u(t^{SB}) \) and \( u(t^{SB}_1) \) by solving a system of two equations with two unknowns.

Note that the risk-averse agent does not receive full insurance anymore. Indeed, with full insurance, the incentive compatibility constraint (17.3) can no longer be satisfied. Inducing effort requires the agent to bear some risk, the following proposition provides a summary.

**Proposition 17.5.1** When the agent is strictly risk-averse, the optimal contract that induces effort makes both the agent’s participation and incentive constraints binding. This contract does not provide full insurance. Moreover, second-best transfers are given by

\[
\bar{t}^{SB} = h \left( \psi + \left(1 - \pi_1 \right) \frac{\psi}{\Delta \pi} \right) = h \left( \frac{1 - \pi_0}{\Delta \pi} \psi \right) \tag{17.33}
\]

and

\[
t^{SB}_1 = h \left( \psi - \pi_1 \frac{\psi}{\Delta \pi} \right) = h \left( -\frac{\pi_0}{\Delta \pi} \psi \right). \tag{17.34}
\]

### 17.5.2 The Optimal Second-Best Effort

Let us now turn to the question of the second-best optimality of inducing a high effort, from the principal’s point of view. The second-best cost \( C^{SB} \) of inducing effort under moral hazard is the expected payment made to the agent \( C^{SB} = \pi_1 \bar{t}^{SB} + (1 - \pi_1) t^{SB}_1 \). Using (17.33) and (17.34), this cost is rewritten as

\[
C^{SB} = \pi_1 h \left( \psi + \left(1 - \pi_1 \right) \frac{\psi}{\Delta \pi} \right) + (1 - \pi_1) h \left( \psi - \pi_1 \frac{\psi}{\Delta \pi} \right) = \pi_1 h \left( \frac{1 - \pi_0}{\Delta \pi} \psi \right) + (1 - \pi_1) h \left( -\frac{\pi_0}{\Delta \pi} \psi \right). \tag{17.35}
\]

The benefit of inducing effort is still \( B = \Delta \pi \Delta S \), and a positive effort \( e^* = 1 \) is the optimal choice of the principal whenever

\[
\Delta \pi \Delta S \geq C^{SB} = \pi_1 h \left( \psi + \left(1 - \pi_1 \right) \frac{\psi}{\Delta \pi} \right) + (1 - \pi_1) h \left( \psi - \pi_1 \frac{\psi}{\Delta \pi} \right) = \pi_1 h \left( \frac{1 - \pi_0}{\Delta \pi} \psi \right) + (1 - \pi_1) h \left( -\frac{\pi_0}{\Delta \pi} \psi \right). \tag{17.36}
\]
With $h(\cdot)$ being strictly convex, Jensen’s inequality implies that the right-hand side of (17.36) is strictly greater than the first-best cost of implementing effort $C^{FB} = h(\psi)$. Therefore, inducing a higher effort occurs less often with moral hazard than when effort is observable. The above figure represents this phenomenon graphically.

For $B$ belonging to the interval $[C^{FB}, C^{SB}]$, the second-best level of effort is zero and is thus strictly below its first-best value. There is now an under-provision of effort because of moral hazard and risk aversion.

**Proposition 17.5.2** With moral hazard and risk aversion, there is a trade-off between inducing effort and providing insurance to the agent. In a model with two possible levels of effort, the principal induces a positive effort from the agent less often than when effort is observable.

### 17.6 More than Two Levels of Performance

We now extend our previous $2 \times 2$ model to allow for more than two levels of performance. We consider a production process where $n$ possible outcomes can be realized. Those performances can be ordered so that $q_1 < q_2 < \cdots < q_i < \cdots < q_n$. We denote the principal’s return in each of those states of nature by $S_i = S(q_i)$. In this context, a contract is a $n$-tuple of payments $\{(t_1, \ldots, t_n)\}$. Also, let $\pi_{ik}$ be the probability that production $q_i$ takes place when the effort
level is $e_k$. We assume that $\pi_{ik}$ for all pairs $(i,k)$ with $\sum_{i=1}^{n} \pi_{ik} = 1$. Finally, we keep the assumption that only two levels of effort are feasible, i.e., $e_k$ in $\{0,1\}$. We still denote $\Delta \pi_i = \pi_{i1} - \pi_{i0}$.

### 17.6.1 Limited Liability

Consider first the limited liability model. If the optimal contract induces a positive effort, it solves the following program:

$$\max_{(t_1,\ldots,t_n)} \sum_{i=1}^{n} \pi_{i1}(S_i - t_i)$$

subject to

$$\sum_{i=1}^{n} \pi_{i1} t_i - \psi \geq 0,$$

$$\sum_{i=1}^{n} (\pi_{i1} - \pi_{i0}) t_i \geq \psi;$$

$$t_i \geq 0, \quad \text{for all } i \in \{1,\ldots,n\}. \quad (17.40)$$

(17.38) is the agent’s participation constraint. (17.39) is his incentive constraint. (17.40) are all the limited liability constraints by assuming that the agent cannot be given a negative payment.

First, note that the participation constraint (17.38) is implied by the incentive (17.39) and the limited liability (17.40) constraints. Indeed, we have

$$\sum_{i=1}^{n} \pi_{i1} t_i - \psi \geq \sum_{i=1}^{n} (\pi_{i1} - \pi_{i0}) t_i - \psi + \sum_{i=1}^{n} \pi_{i0} t_i \geq 0.$$

Hence, we can neglect the participation constraint (17.38) in the optimization of the principal’s program.

Denoting the multiplier of (17.39) by $\lambda$ and the respective multipliers of (17.40) by $\xi_i$, the first-order conditions lead to

$$-\pi_{i1} + \lambda \Delta \pi_i + \xi_i = 0. \quad (17.41)$$

with the slackness conditions $\xi_i t_i = 0$ for each $i$ in $\{1,\ldots,n\}$. 649
For such that the second-best transfer $t_{i}^{SB}$ is strictly positive, $\xi_{i} = 0$, and we must have $\lambda = \frac{\pi_{i1}}{\pi_{i1} - \pi_{i0}}$ for any such $i$. If the ratios $\frac{\pi_{i1} - \pi_{i0}}{\pi_{i1}}$ all different, there exists a single index $j$ such that $\frac{\pi_{j1} - \pi_{j0}}{\pi_{j1}}$ is the highest possible ratio. The agent receives a strictly positive transfer only in this particular state of nature $j$, and this payment is such that the incentive constraint (17.39) is binding, i.e., $t_{j}^{SB} = \frac{\psi}{\pi_{j1} - \pi_{j0}}$. In all other states, the agent receives no transfer and $t_{i}^{SB} = 0$ for all $i \neq j$.

Finally, the agent gets a strictly positive ex ante limited liability rent that is worth $EU^{SB} = \frac{\pi_{j0}\psi}{\pi_{j1} - \pi_{j0}}$.

The important point here is that the agent is rewarded in the state of nature that is the most informative about the fact that he has exerted a positive effort. Indeed, $\frac{\pi_{i1} - \pi_{i0}}{\pi_{i1}}$ can be interpreted as a likelihood ratio. The principal therefore uses a maximum likelihood ratio criterion to reward the agent. The agent is only rewarded when this likelihood ratio is maximized. Like an econometrician, the principal tries to infer from the observed output what has been the parameter (effort) underlying this distribution. But here the parameter is endogenously affected by the incentive contract.

**Definition 17.6.1** The probabilities of success satisfy the monotone likelihood ratio property (MLRP) if $\frac{\pi_{i1} - \pi_{i0}}{\pi_{i1}}$ is nondecreasing in $i$.

**Proposition 17.6.1** If the probability of success satisfies MLRP, the second-best payment $t_{i}^{SB}$ received by the agent may be chosen to be nondecreasing with the level of production $q_{i}$.

### 17.6.2 Risk Aversion

Suppose now that the agent is strictly risk-averse. The optimal contract that induces effort must solve the program below:

$$\max_{(t_{1},...,t_{n})} \sum_{i=1}^{n} \pi_{i1}(S_{i} - t_{i})$$

subject to

$$\sum_{i=1}^{n} \pi_{i1}u(t_{i}) - \psi \geq \sum_{i=1}^{n} \pi_{i0}u(t_{i})$$

(17.43)
and
\[ \sum_{i=1}^{n} \pi_{i1} u(t_i) - \psi \geq 0, \]  
(17.44)

where the latter constraint is the agent’s participation constraint.

Using the same change of variables as before, it should be clear that the program is again a concave problem with respect to the new variables \( u_i = u(t_i) \). Using the same notations as before, the first-order conditions of the principal’s program are written as:
\[ \frac{1}{u'(t_{iSB}^{SB})} = \mu + \lambda \left( \frac{\pi_{i1} - \pi_{i0}}{\pi_{i1}} \right) \quad \text{for all} \ i \in \{1, \ldots, n\}. \]  
(17.45)

Multiplying each of these equations by \( \pi_{i1} \) and summing over \( i \) yields \( \mu = E_q \left( \frac{1}{u'(t_{iSB}^{SB})} \right) > 0 \), where \( E_q \) denotes the expectation operator with respect to the distribution of outputs induced by effort \( e = 1 \).

Multiplying (17.45) by \( \pi_{i1} u(t_{iSB}^{SB}) \), summing all these equations over \( i \), and taking into account the expression of \( \mu \) obtained above yields
\[ \lambda \left( \sum_{i=1}^{n} (\pi_{i1} - \pi_{i0}) u(t_{iSB}^{SB}) - \psi \right) = E_q \left( u(t_{iSB}^{SB}) \left( \frac{1}{u'(t_{iSB}^{SB})} - E \left( \frac{1}{u'(t_{iSB}^{SB})} \right) \right) \right). \]  
(17.46)

Using the slackness condition \( \lambda \left( \sum_{i=1}^{n} (\pi_{i1} - \pi_{i0}) u(t_{iSB}^{SB}) - \psi \right) = 0 \) to simplify the left-hand side of (17.46), we finally get
\[ \lambda \psi = \text{cov} \left( u(t_{iSB}^{SB}), \frac{1}{u'(t_{iSB}^{SB})} \right). \]  
(17.47)

By assumption, \( u(\cdot) \) and \( u'(\cdot) \) covary in opposite directions. Moreover, a constant wage \( t_{iSB}^{SB} = t^{SB} \) for all \( i \) does not satisfy the incentive constraint, and thus \( t_{iSB}^{SB} \) cannot be constant everywhere. Hence, the right-hand side of (17.47) is necessarily strictly positive. Thus we have \( \lambda > 0 \), and the incentive constraint is binding.

Coming back to (17.45), we observe that the left-hand side is increasing in \( t_{iSB}^{SB} \) since \( u(\cdot) \) is concave. For \( t_{iSB}^{SB} \) to be nondecreasing with \( i \), MLRP must again hold. Then higher outputs are also those that are the more informative ones about the realization of a high effort. Hence, the agent should be more rewarded as output increases.
17.7 Contract Theory at Work

This section elaborates on the moral hazard paradigm discussed so far in a number of settings that have been discussed extensively in the contracting literature.

17.7.1 Efficiency Wage

Let us consider a risk-neutral agent working for a firm, the principal. This is a basic model studied by Shapiro and Stiglitz (AER, 1984). By exerting effort \( e \) in \( \{0, 1\} \), the firm’s added value is \( \bar{V} \) (resp. \( V \)) with probability \( \pi(e) \) (resp. \( 1 - \pi(e) \)). The agent can only be rewarded for a good performance and cannot be punished for a bad outcome, since they are protected by limited liability.

To induce effort, the principal must find an optimal compensation scheme \( \{(t, \bar{t})\} \) that is the solution to the program below:

\[
\max_{\{(t, \bar{t})\}} \pi_1(\bar{V} - \bar{t}) + (1 - \pi_1)(V - t) \tag{17.48}
\]

subject to

\[
\pi_1 \bar{t} + (1 - \pi_1)t - \psi \geq \pi_0 \bar{t} + (1 - \pi_0)t, \tag{17.49}
\]

\[
\pi_1 \bar{t} + (1 - \pi_1)t - \psi \geq 0, \tag{17.50}
\]

\[
t \geq 0. \tag{17.51}
\]

The problem is completely isomorphic to the one analyzed earlier. The limited liability constraint is binding at the optimum, and the firm chooses to induce a high effort when \( \Delta \pi \Delta V \geq \frac{\pi_1 \psi}{\Delta \pi} \). At the optimum, \( \bar{t}^{SB} = 0 \) and \( \bar{t}^{SB} > 0 \). The positive wage \( \bar{t}^{SB} = \frac{\psi}{\Delta \pi} \) is often called an efficiency wage because it induces the agent to exert a high (efficient) level of effort. To induce production, the principal must give up a positive share of the firm’s profit to the agent.

17.7.2 Sharecropping

The moral hazard paradigm has been one of the leading tools used by development economists to analyze agrarian economies. In the sharecropping model given in
Stiglitz (RES, 1974), the principal is now a landlord and the agent is the landlord’s tenant. By exerting an effort \( e \) in \( \{0, 1\} \), the tenant increases (decreases) the probability \( \pi(e) \) (resp. \( 1 - \pi(e) \)) that a large \( \bar{q} \) (resp. small \( q \)) quantity of an agricultural product is produced. The price of this good is normalized to one so that the principal’s stochastic return on the activity is also \( \bar{q} \) or \( q \), depending on the state of nature.

It is often the case that peasants in developing countries are subject to strong financial constraints. To model such a setting we assume that the agent is risk neutral and protected by limited liability. When he wants to induce effort, the principal’s optimal contract must solve

\[
\max_{\{t, \bar{t}\}} \pi_1(\bar{q} - \bar{t}) + (1 - \pi_1)(q - t) \quad \text{(17.52)}
\]

subject to

\[
\pi_1 \bar{t} + (1 - \pi_1) t - \psi \geq \pi_0 \bar{t} + (1 - \pi_0) t, \quad \text{(17.53)}
\]

\[
\pi_1 \bar{t} + (1 - \pi_1) t - \psi \geq 0, \quad \text{(17.54)}
\]

\[
t \geq 0. \quad \text{(17.55)}
\]

The optimal contract therefore satisfies \( t^{SB} = 0 \) and \( \bar{t}^{SB} = \frac{\psi}{\Delta \pi} \). This is again akin to an efficiency wage. The expected utilities obtained respectively by the principal and the agent are given by

\[
EV^{SB} = \pi_1 \bar{q} + (1 - \pi_1)q - \frac{\pi_1 \psi}{\Delta \pi}. \quad \text{(17.56)}
\]

and

\[
EU^{SB} = \frac{\pi_0 \psi}{\Delta \pi}. \quad \text{(17.57)}
\]

The flexible second-best contract described above has sometimes been criticized as not corresponding to the contractual arrangements observed in most agrarian economies. Contracts often take the form of simple linear schedules linking the tenant’s production to his compensation. As an exercise, let us now analyze a simple linear sharing rule between the landlord and his tenant, with the
landlord offering the agent a fixed share $\alpha$ of the realized production. Such a sharing rule automatically satisfies the agent’s limited liability constraint, which can therefore be omitted in what follows. Formally, the optimal linear rule inducing effort must solve

$$\max_{\alpha}(1 - \alpha)(\pi_1q + (1 - \pi_1)\bar{q})$$

subject to

$$\alpha(\pi_1\bar{q} + (1 - \pi_1)q) - \psi \geq \alpha(\pi_0\bar{q} + (1 - \pi_0)q), \quad (17.59)$$

$$\alpha(\pi_1\bar{q} + (1 - \pi_1)q) - \psi \geq 0 \quad (17.60)$$

Obviously, only (17.59) is binding at the optimum. One finds the optimal linear sharing rule to be

$$\alpha^{SB} = \frac{\psi}{\Delta\pi\Delta q}. \quad (17.61)$$

Note that $\alpha^{SB} < 1$ because, for the agricultural activity to be a valuable venture in the first-best world, we must have $\Delta\pi\Delta q > \psi$. Hence, the return on the agricultural activity is shared between the principal and the agent, with high-powered incentives ($\alpha$ close to one) being provided when the disutility of effort $\psi$ is large or when the principal’s gain from an increase of effort $\Delta\pi\Delta q$ is small.

This sharing rule also yields the following expected utilities to the principal and the agent, respectively

$$EV_\alpha = \pi_1\bar{q} + (1 - \pi_1)q - \left(\frac{\pi_1\bar{q} + (1 - \pi_1)q}{\Delta q}\right)\frac{\psi}{\Delta\pi} \quad (17.62)$$

and

$$EU_\alpha = \left(\frac{\pi_1\bar{q} + (1 - \pi_1)q}{\Delta q}\right)\frac{\psi}{\Delta\pi}. \quad (17.63)$$

Comparing (17.56) and (17.62) on the one hand and (17.57) and (17.63) on the other hand, we observe that the constant sharing rule benefits the agent but not the principal. A linear contract is less powerful than the optimal second-best contract. The former contract is an inefficient way to extract rent from the agent even if it still provides sufficient incentives to exert effort. Indeed, with a linear
sharing rule, the agent always benefits from a positive return on his production, even in the worst state of nature. This positive return yields to the agent more than what is requested by the optimal second-best contract in the worst state of nature, namely zero. Punishing the agent for a bad performance is thus found to be rather difficult with a linear sharing rule.

A linear sharing rule allows the agent to keep some strictly positive rent $EU_\alpha$. If the space of available contracts is extended to allow for fixed fees $\beta$, the principal can nevertheless bring the agent down to the level of his outside opportunity by setting a fixed fee $\beta^{SB}$ equal to

$$\left(\frac{\pi_1\bar{q} + (1 - \pi_1)q}{\Delta q}\right) \frac{\psi}{\Delta \pi} - \frac{\pi_0\psi}{\Delta \pi}.\quad (17.64)$$

17.7.3 Wholesale Contracts

Let us now consider a manufacturer-retailer relationship studied in Laffont and Tirole (1993). The manufacturer supplies at constant marginal cost $c$ an intermediate good to the risk-averse retailer, who sells this good on a final market. Demand on this market is high (resp. low) $\bar{D}(p)$ (resp. $D(p)$) with probability $\pi(e)$ where, again, $e$ is in $\{0, 1\}$ and $p$ denotes the price for the final good. Effort $e$ is exerted by the retailer, who can increase the probability that demand is high if after-sales services are efficiently performed. The wholesale contract consists of a retail price maintenance agreement specifying the prices $\bar{p}$ and $\bar{p}$ on the final market with a sharing of the profits, namely $\{(\bar{t}, p); (\bar{t}, \bar{p})\}$. When he wants to induce effort, the optimal contract offered by the manufacturer solves the following problem:

$$\max_{\{(\bar{t}, p); (\bar{t}, \bar{p})\}} \pi_1((\bar{p} - c)\bar{D}(\bar{p}) - \bar{t}) + (1 - \pi_1)((\bar{p} - c)\bar{D}(\bar{p}) - \bar{t}) \quad (17.64)$$

subject to (17.3) and (17.4).

The solution to this problem is obtained by appending the following expressions of the retail prices to the transfers given in (17.33) and (17.34): $\bar{p}^* + \frac{D(p^*)}{D(p^*)} = c$, and $\bar{p}^* + \frac{D(p^*)}{D(p^*)} = c$. Note that these prices are the same as those that would be chosen under complete information. The pricing rule is not affected by the incentive problem.
17.7.4 Financial Contracts

Moral hazard is an important issue in financial markets. In Holmstrom and Tirole (AER, 1994), it is assumed that a risk-averse entrepreneur wants to start a project that requires an initial investment worth an amount $I$. The entrepreneur has no cash of his own and must raise money from a bank or any other financial intermediary. The return on the project is random and equal to $\bar{V}$ (resp. $\bar{V}$) with probability $\pi(e)$ (resp. $1 - \pi(e)$), where the effort exerted by the entrepreneur $e$ belongs to $\{0,1\}$. We denote the spread of profits by $\Delta V = \bar{V} - \bar{V} > 0$. The financial contract consists of repayments $\{(\bar{z},\bar{z})\}$, depending upon whether the project is successful or not.

To induce effort from the borrower, the risk-neutral lender’s program is written as

$$\max_{\{(\bar{z},\bar{z})\}} \left\{ \pi_1 \bar{z} + (1 - \pi_1)\bar{z} - I \right\}$$

subject to

$$\pi_1 u(\bar{V} - \bar{z}) + (1 - \pi_1)u(\bar{V} - \bar{z}) - \psi \geq \pi_0 u(\bar{V} - \bar{z}) + (1 - \pi_0)u(\bar{V} - \bar{z}),$$

$$\pi_1 u(\bar{V} - \bar{z}) + (1 - \pi_1)u(\bar{V} - \bar{z}) - \psi \geq 0.$$  \hspace{1cm} (17.67)

Note that the project is a valuable venture if it provides the bank with a positive expected profit.

With the change of variables, $\bar{t} = \bar{V} - \bar{z}$ and $\bar{t} = \bar{V} - \bar{z}$, the principal’s program takes its usual form. This change of variables also highlights the fact that everything happens as if the lender was benefitting directly from the returns of the project, and then paying the agent only a fraction of the returns in the different states of nature.

Let us define the second-best cost of implementing a positive effort $C^{SB}$, and let us assume that $\Delta \pi \Delta V \geq C^{SB}$, so that the lender wants to induce a positive effort level even in a second-best environment. The lender’s expected profit is worth

$$V_1 = \pi_1 \bar{V} + (1 - \pi_1)\bar{V} - C^{SB} - I.$$  \hspace{1cm} (17.68)
Let us now parameterize projects according to the size of the investment $I$. Only the projects with positive value $V_1 > 0$ will be financed. This requires the investment to be low enough, and typically we must have

$$I < I^{SB} = \pi_1 \bar{V} + (1 - \pi_1)\bar{V} - C^{SB}. \quad (17.69)$$

Under complete information and no moral hazard, the project would instead be financed as soon as

$$I < I^* = \pi_1 \bar{V} + (1 - \pi_1)\bar{V} \quad (17.70)$$

For intermediary values of the investment, i.e., for $I$ in $[I^{SB}, I^*]$, moral hazard implies that some projects are financed under complete information but no longer under moral hazard. This is akin to some form of credit rationing.

Finally, note that the optimal financial contract offered to the risk-averse and cashless entrepreneur does not satisfy the limited liability constraint $t \geq 0$. Indeed, we have $t^{SB} = h \left( \psi - \frac{\pi_1 \psi}{\Delta \pi} \right) < 0$. To be induced to make an effort, the agent must bear some risk, which implies a negative payoff in the bad state of nature. Adding the limited liability constraint, the optimal contract would instead entail $t^{LL} = 0$ and $\bar{t}^{LL} = h \left( \frac{\psi}{\Delta \pi} \right)$. Interestingly, this contract has sometimes been interpreted in the corporate finance literature as a debt contract, with no money being left to the borrower in the bad state of nature and the residual being pocketed by the lender in the good state of nature.

Finally, note that

$$\bar{t}^{LL} - t^{LL} = h \left( \frac{\psi}{\Delta \pi} \right) < \bar{t}^{SB} - t^{SB} = h \left( \psi + (1 - \pi_1) \frac{\psi}{\Delta \pi} \right) \quad (17.71)$$

$$- h \left( \psi - \frac{\pi_1 \psi}{\Delta \pi} \right),$$

since $h(\cdot)$ is strictly convex and $h(0) = 0$. This inequality shows that the debt contract has less incentive power than the optimal incentive contract. Indeed, it becomes harder to spread the agent’s payments between both states of nature to induce effort if the agent is protected by limited liability by the agent, who is interested only in his payoff in the high state of nature, only rewards are attractive.
17.8 A Continuum of Performances

Let us now assume that the level of performance \( \tilde{q} \) is drawn from a continuous distribution with a cumulative function \( F(\cdot|e) \) on the support \([\tilde{q}, \bar{q}]\). This distribution is conditional on the agent’s level of effort, which still takes two possible values \( e \) in \( \{0, 1\} \). We denote by \( f(\cdot|e) \) the density corresponding to the above distributions. A contract \( t(q) \) inducing a positive effort in this context must satisfy the incentive constraint

\[
\int_{\tilde{q}}^{\bar{q}} u(t(q)) f(q|1) dq - \psi \geq \int_{\tilde{q}}^{\bar{q}} u(t(q)) f(q|0) dq, \tag{17.72}
\]

and the participation constraint

\[
\int_{\tilde{q}}^{\bar{q}} u(t(q)) f(q|1) dq - \psi \geq 0. \tag{17.73}
\]

The risk-neutral principal problem is thus written as

\[
\max_{\{t(q)\}} \int_{\tilde{q}}^{\bar{q}} (S(q) - t(q)) f(q|1) dq, \tag{17.74}
\]

subject to (17.72) and (17.73).

Denoting the multipliers of (17.72) and (17.73) by \( \lambda \) and \( \mu \), respectively, the Lagrangian is written as

\[
L(q, t) = (S(q) - t)f(q|1) + \lambda(u(t)(f(q|1) - f(q|0)) - \psi) + \mu(u(t)f(q|1) - \psi).
\]

Optimizing pointwise with respect to \( t \) yields

\[
\frac{1}{u'(t^{SB}(q))} = \mu + \lambda \left( \frac{f(q|1) - f(q|0)}{f(q|1)} \right). \tag{17.75}
\]

Multiplying (17.75) by \( f_1(q) \) and taking expectations, we obtain, as in the main text,

\[
\mu = E_q \left( \frac{1}{u'(t^{SB}(q))} \right) > 0, \tag{17.76}
\]

where \( E_q(\cdot) \) is the expectation operator with respect to the probability distribution of output induced by an effort \( e^{SB} \). Finally, using this expression of \( \mu \),
inserting it into (17.75), and multiplying it by \( f(q|1)u(t^{SB}(q)) \), we obtain

\[
\lambda(f(q|1) - f(q|0))u(t^{SB}(q)) = f(q|1)u(t^{SB}(q)) \left( \frac{1}{u'(t^{SB}(q))} - E_q \left( \frac{1}{u'(t^{SB}(q))} \right) \right).
\] (17.77)

Integrating over \([q, \tilde{q}]\) and taking into account the slackness condition \( \lambda(\int_{\tilde{q}}^{q}(f(q|1) - f(q|0))u(t^{SB}(q))dq - \psi) = 0 \) yields \( \lambda \psi = \text{cov}(u(t^{SB}(\tilde{q})), \frac{1}{u'(t^{SB}(\tilde{q}))}) \geq 0 \).

Hence, \( \lambda \geq 0 \) because \( u(\cdot) \) and \( u'(\cdot) \) vary in opposite directions. Also, \( \lambda = 0 \) only if \( t^{SB}(q) \) is a constant, but in this case the incentive constraint is necessarily violated. As a result, we have \( \lambda > 0 \). Finally, \( t^{SB}(\pi) \) is monotonically increasing in \( \pi \) when the monotone likelihood property \( \frac{d}{dq} \left( \frac{f(q|1) - f^*(q|0)}{f(q|1)} \right) \geq 0 \) is satisfied.

### 17.9 A Mixed Model of Moral Hazard and Adverse Selection

So far, when we discuss asymmetric information between a principal and an agent in the previous and current chapters, we only allow either adverse selection or moral hazard presented, but not both. However, in many cases, a principle may know information neither about agent’s action (such as his efforts) nor his characteristic (such as his risk aversion or cost). In this section, we will consider the case where both action and characteristic of an agent is unknown to the principal. In doing so, we introduce a mixed model of moral hazard and adverse selection, which was recently studied in Meng and Tian (GEB, 2013). This model investigates the optimal wage contract design when both efforts and risk aversion of the agent are unobservable. We first consider the benchmark case where agent’s efforts are unobservable, and then consider the case where either risk aversion or cost is also unobservable.
17.9.1 Optimal Wage Contract with Unobservable Efforts Only

Consider a principal-agent relationship in which the agent controls \( n \) activities that influence the principal’s payoff. The principal is risk neutral and her gross payoff is a linear function of the agent’s effort vector \( e \):

\[
V(e) = \beta' e + \eta, \tag{17.78}
\]

where the \( n \)-dimensional vector \( \beta \) characterizes the marginal effect of the agent’s effort \( e \) on \( V(e) \), and \( \eta \) is a noise term with zero mean. The agent chooses a vector of efforts \( e = (e_1, \cdots, e_n)' \in \mathbb{R}^n \) at quadratic personal cost \( e'Ce \), where \( C \) is a symmetric positive definite matrix. The diagonal element \( C_{ii} \) reflects the agent’s task-specific productivities, while the sign of off-diagonal elements \( C_{ij} \) indicates the relationship between two tasks \( i \) and \( j \), which are substitute (resp. complementary, independent) if \( C_{ij} > 0 \) (resp. \( < 0, = 0 \)). The agent’s preferences are represented by the negative exponential utility function \( u(x) = -e^{-rx} \), where \( r \) is the agent’s absolute risk aversion and \( x \) is his compensation minus personal cost.

It is assumed that there is a linear relation between the agent’s efforts and the expected levels of the performance measures:

\[
P_i(e) = b_i'e + \varepsilon_i, i = 1, \cdots, m, \tag{17.79}
\]

where \( b_i \in \mathbb{R}^n \) captures the marginal effect of the agent’s effort \( e \) on the performance measure \( P_i(e) \); \( B = (b_1, \cdots, b_m)' \) is an \( m \times n \) matrix of performance parameters, and it is assumed that the matrix \( B \) has full row rank \( m \) so that every performance measure cannot be replaced by the other measures; and \( \varepsilon = (\varepsilon_1, \cdots, \varepsilon_m)' \) is an \( m \times 1 \) vector of normally distributed variables with mean zero and variance-covariance matrix \( \Sigma \).

**Definition 17.9.1** (Orthogonality) A performance system is said to be orthogonal if and only if \( b_i'C^{-1}b_j = 0 \) and \( \text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \), for \( i \neq j \), that is, \( B'C^{-1}B \) and \( \Sigma \) are both diagonal matrices.
Definition 17.9.2 (Cost-adjusted Correlation) The cost-adjusted correlation between two performance measures $i$ and $j$ is the ratio of the cost-adjusted inner product of their vectors of sensitivities divided by the covariance of the error terms:

$$
\rho_{ij} = \frac{b_i'C_i^{-1}b_j}{\sigma_{ij}}
$$  

which measure performance not only about the agent’s task-specific abilities and interaction among tasks, but also what degree two performance measures are aligned with each other. If $n$ tasks are technologically independent and identical, i.e., $C = cI$, then we use the concept correlation $\rho_{ij} \equiv b_i'b_j/\sigma_{ij}$ to measure the degree of alignment between two performance measures.

Definition 17.9.3 (Cost-adjusted Congruence) The cost-adjusted congruence of a performance measure $P_i = b_i'e + \varepsilon_i$ is defined as

$$
\Gamma_i = \frac{b_i'C_i^{-1}\beta}{\sqrt{b_i'C_i^{-1}b_i}\sqrt{\beta'C_i^{-1}\beta}}.
$$  

A performance measure with nonzero cost-adjusted congruence is called congruent; a performance measure with unit cost-adjusted congruence is said to be perfectly congruent. We assume that there exists at least one congruent measure, i.e., $BC^{-1}\beta \neq 0$.

The principal compensates the agent’s effort through a linear contract:

$$
W(e) = w_0 + w'P(e),
$$

where $P(e) = (P_1(e), \cdots, P_m(e))'$, $w_0$ denotes the base wage, and $w = (w_1, \cdots, w_m)'$ the performance wage. Under this linear compensation rule, the principal’s expected profit is $\Pi_p = \beta'e - w_0 - w'Be$, and the agent’s certainty equivalent is

$$
CE_a = w_0 + w'Be - \frac{1}{2}\varepsilon'eC\epsilon - \frac{r}{2}w'\Sigma w.
$$

The principal’s problem is to design a contract $(w_0, w)$ that maximizes her expected profit $\Pi_p$ while ensuring the agent’s participation and eliciting his optimal effort.
The optimization problem of the principal is thus formulated as:

\[
\begin{align*}
\max_{\{w_0, w, e\}} & \quad \beta' e - w_0 - w' B e \\
\text{s.t.:} & \quad IR : w_0 + w' B e - \frac{1}{2} e' C e - \frac{r}{2} w' \Sigma w \geq 0 \\
& \quad IC : e \in \arg\max_{\tilde{e}} \left[ w_0 + w' B \tilde{e} - \frac{1}{2} \tilde{e}' C \tilde{e} - \frac{r}{2} w' \Sigma w \right].
\end{align*}
\]

The IR constraint ensures that the principal cannot force the agent into accepting the contract, and here the agent’s reservation utility is normalized to zero; the IC constraint represents the rationality of the agent’s effort choice.

We now consider the effort choosing problem of the agent for a given incentive scheme \((w_0, w)\). Since the objective is concave by noting that the second-order derivative of \(CE_a\) with respect to \(e\) is a negative definite matrix \(-C\), the maximizer can be determined by the first-order condition: \(Ce = B'w\). After replacing \(e\) with \(e^* = C^{-1}B'w\) and substituting the IR constraint written with equality into the principal’s objective function, the principal’s optimization problem simplifies to:

\[
\max_{w \in \mathbb{R}^m} \left[ \beta' C^{-1} B' w - \frac{1}{2} w' \left( BC^{-1} B' + r \Sigma \right) w \right].
\]

The optimal wage contract and effort to be elicited are therefore:

\[
\begin{align*}
w^p &= \left[ BC^{-1} B' + r \Sigma \right]^{-1} BC^{-1} \beta \\
w_0^p &= \frac{rw^p \Sigma w^p - w^p' BC^{-1} B' w^p}{2} \tag{17.85} \\
e^p &= C^{-1} B' w^p. \tag{17.86}
\end{align*}
\]

The resulting surplus of the principal is\(^1\)

\[
\Pi^p = \frac{1}{2} \beta' C^{-1} B' \left[ BC^{-1} B' + r \Sigma \right]^{-1} BC^{-1} \beta. \tag{17.87}
\]

A higher incentive pay could induce the agent to implement a higher effort, but it will also expose the agent to a higher risk. It therefore requires a premium to compensate the risk-averse agent for the risk he bears. The optimal power of

\(^1\)Superscript “p” denotes “pure moral hazard".
incentive is therefore determined by the tradeoff between incentive and insurance. Moreover, the results above show that in multi-task agency relationships, the degree of congruity of available performance measures and the agent’s task-specific abilities also affects the power and distortion of incentive contract.

### 17.9.2 Optimal Wage Contract with Unobservable Efforts and Risk Aversion

The pure moral hazard incentive contract stated above relies crucially on the agent’s attitude towards risk. In the following, we assume that risk aversion \( r \) is also private information of the agent, and its distribution function \( F(r) \) and density function \( f(r) \) supported on \([r, \bar{r}]\) are common knowledge to all parties. The principal then has to offer a contract menu \( \{w_0(\hat{r}), w(\hat{r})\} \) contingent on the agent’s reported “type” \( \hat{r} \) to maximize her expected payoff.

A contract \( \{w_0(\hat{r}), w(\hat{r})\} \) is said to be implementable if the following incentive compatibility condition is satisfied:

\[
w_0(r) + \frac{1}{2} w'(r) \left[ BC^{-1} B' - r \Sigma \right] w(r) \geq w_0(\hat{r}) + \frac{1}{2} w'(\hat{r}) \left[ BC^{-1} B' - r \Sigma \right] w(\hat{r}).
\]

(17.88)

Let \( U(r, \hat{r}) \equiv w_0(\hat{r}) + \frac{1}{2} w'(\hat{r}) \left[ BC^{-1} B' - r \Sigma \right] w(\hat{r}), \) and \( U(r) \equiv U(r, r), \) then the implementability condition of \( \{U(r), w(r)\} \) is stated equivalently as:

\[
\exists w_0 : [r, \bar{r}] \to \mathbb{R}_+, \forall (r, \hat{r}) \in [r, \bar{r}]^2, U(r) = \max_{\hat{r}} \left\{ w_0(\hat{r}) + \frac{1}{2} w'(\hat{r}) \left[ BC^{-1} B' - r \Sigma \right] w(\hat{r}) \right\},
\]

(17.89)

which is in turn equivalent to the following very similar condition

\[
\exists w_0 : \mathbb{R}^m \to \mathbb{R}_+, \forall r \in [r, \bar{r}], U(r) = \max_w \left\{ w_0(w) + \frac{1}{2} w'[BC^{-1} B' - r \Sigma] w \right\}.
\]

(17.90)

It is possible to show that \( U(\cdot) \) is continuous, convex \(^3\) (thus almost everywhere differentiable), and satisfies the envelop condition:

\[
U'(r) = -\frac{1}{2} w' \Sigma w.
\]

(17.91)

\(^2\)Substituting \( c^* = C^{-1} B' w \) into expression (17.83) yields \( U = w_0 + \frac{1}{2} w'[BC^{-1} B' - r \Sigma] w. \)

\(^3\)One way to define the convex functions is through representing them as maximum of the
Conversely, if (17.91) holds and \( U(r) \) is convex, then
\[
U(r) \geq U(\hat{r}) + (r - \hat{r})U'(\hat{r}) = U(\hat{r}) - \frac{1}{2}(r - \hat{r})w'(\hat{r})\Sigma w(\hat{r}),
\]
which implies the incentive compatibility condition \( U(r) \geq U(r, \hat{r}) \). Formally, we have

**Lemma 17.9.1** The surplus function \( U(r) \) and performance wage function \( w(r) \) are implementable if and only if:

1. envelop condition (17.91) is satisfied;
2. \( U(r) \) is convex in \( r \).

Substituting \( U(r) \) into the principal’s expected payoff, we get
\[
\Pi = \int_{\xi} \left[ \beta e^* - w_0(r) - w(r)'Be^* \right] f(r) dr
= \int_{\xi} \left\{ \beta' C^{-1}B'w(r) - \frac{1}{2}w(r)' \left[ BC^{-1}B' + r\Sigma \right] w(r) - U(r) \right\} f(r) dr.
\]
The principal’s optimization problem is therefore:
\[
\max_{U(r), w(r)} \Pi, \quad \text{s.t.:} \quad U(r) \geq 0, U'(r) = -\frac{1}{2}w(r)'\Sigma w(r), U(r) \text{ is convex} \quad (17.92)
\]
The following proposition summarizes the solution of the principal’s problem.

**Proposition 17.9.1** If \( \Phi(r) \) is nondecreasing, then the optimal wage contract is given by
\[
w^h(r) = \left[ BC^{-1}B' + \Phi(r)\Sigma \right]^{-1} BC^{-1}\beta
w^h_0(r) = \frac{1}{2} \int_r \left[ w^h(\tilde{r})'\Sigma w^h(\tilde{r}) d\tilde{r} - \frac{1}{2}w^h(r)' \left[ BC^{-1}B' - r\Sigma \right] w^h(r) \right],
\]
where \( \Phi(r) \equiv r + \frac{F(r)}{f(r)} \).

affine functions, that is, \( s(x) \) is convex if and only if
\[
s(x) = \max_{a,b \in \Omega} (a \cdot x + b)
\]
for some \( a \in \mathbb{R}^n, b \in \mathbb{R} \) and some \( \Omega \subset \mathbb{R}^{n+1} \). In this example \( a = -\frac{1}{2}w'\Sigma w, b = w_0(w) + \frac{1}{2}w'BC^{-1}B'w \), and thus \( U(r) = \max_{(a,b) \in \mathbb{R} \times \mathbb{R}_+} (ar + b) \) is a convex function in \( r \).

\(^4\)This condition is weaker than and could be implied by the monotone hazard rate property:
\[
\frac{d}{dr} \left[ \frac{F(r)}{f(r)} \right] \geq 0.
\]
Superscript “h” denotes “hybrid model of moral hazard and adverse selection”.

664
Proof. Using the envelop condition $U'(r) = -\frac{1}{2}w'\Sigma w$, the participation constraint $U(r) \geq 0$ simplifies to $U(\bar{r}) \geq 0$. Incentive compatibility implies that only the participation constraint of the most risk averse type can be binding, i.e., $U(\bar{r}) = 0$. We therefore get

$$U(r) = \int_{\bar{r}}^{r} \frac{1}{2} w(\bar{r})' \Sigma w(\bar{r}) \, dr. \quad (17.95)$$

The principal’s objective function becomes

$$\Pi = \int_{r}^{\bar{r}} \left\{ \beta' C^{-1} B' w(r) - \frac{1}{2} w(r)' [BC^{-1} B' + r \Sigma ] w(r) - \int_{\bar{r}}^{r} \frac{1}{2} w(\bar{r})' \Sigma w(\bar{r}) \, d\bar{r} \right\} f(r) \, dr$$

which, by an integration of parts, gives

$$\int_{r}^{\bar{r}} \left\{ \beta' C^{-1} B' w(r) - \frac{1}{2} w(r)' \left[ BC^{-1} B' + \left( r + \frac{F(r)}{f(r)} \right) \Sigma \right] w(r) \right\} f(r) \, dr.$$

Maximizing pointwise the above expression, we get

$$w^h(r) = \left[ BC^{-1} B' + \Phi(r) \Sigma \right]^{-1} BC^{-1} \beta$$

and

$$w^h_0(r) = \frac{1}{2} \int_{\bar{r}}^{r} w^h(\bar{r}) \Sigma w^h(\bar{r}) \, d\bar{r} - \frac{1}{2} w^h(r)' [BC^{-1} B' - r \Sigma ] w^h(r).$$

The only work left is to verify the convexity of $U(r)$. Notice that

$$U''(r) = -(D_r w^h)' \Sigma w^h = \Phi'(r) w^h(r)' \Sigma \left[ BC^{-1} B' + \Phi(r) \Sigma \right]^{-1} \Sigma w^h(r).$$

The second equality comes from the fact that the derivative of $w^h$ with respect to $r$ is

$$D_r w^h = - \left[ BC^{-1} B' + \Phi(r) \Sigma \right]^{-1} \Phi'(r) \Sigma \left[ BC^{-1} B' + \Phi(r) \Sigma \right]^{-1} BC^{-1} \beta$$

$$= -\Phi'(r) \left[ BC^{-1} B' + \Phi(r) \Sigma \right]^{-1} \Sigma w^h.$$

---

665 Let $A$ be a nonsingular, $m \times m$ matrix whose elements are functions of the scalar parameter $\alpha$, then

$$\frac{\partial A^{-1}}{\partial \alpha} = -A^{-1} \frac{\partial A}{\partial \alpha} A^{-1}.$$
It is clear that \( U''(r) \geq 0 \) because \( \Phi'(r) \geq 0 \) and the matrix \( \Sigma^{-1}(BC^{-1}B' + \Phi(r))\Sigma^{-1} \) is positive definite. The proof is completed.

The following conditions prove to be sufficient for the emergence of low-powered incentives.

**Condition 17.9.1** \( \Sigma \) is diagonal.

**Condition 17.9.2** Matrix \( BC^{-1}B' \) is diagonal.

**Condition 17.9.3** Matrices \( BC^{-1}B' \) and \( \Sigma \) commute: \( BC^{-1}B'\Sigma = \Sigma BC^{-1}B' \).

**Condition 17.9.4** The following inequality holds:

\[
2r\lambda_m^2 + \rho > 0 \quad (17.96)
\]

where

\[
\rho = \max \left\{ \min_{i=1,m} \lambda_i \mu_m \frac{(\sqrt{k_{\lambda} + 1})^2 - k_\mu (\sqrt{k_{\lambda} - 1})^2}{2\sqrt{k_{\lambda}}}, \min_{i=1,m} \mu_i \lambda_m \frac{(\sqrt{k_\mu + 1})^2 - k_\lambda (\sqrt{k_\mu - 1})^2}{2\sqrt{k_\mu}} \right\}
\]

\[
= \begin{cases} 
    \lambda_m \mu_m \frac{(\sqrt{k_{\lambda} + 1})^2 - k_\mu (\sqrt{k_{\lambda} - 1})^2}{2\sqrt{k_{\lambda}}} & \text{if } \sqrt{k_\mu} \leq \frac{\sqrt{k_{\lambda} + 1}}{\sqrt{k_{\lambda} - 1}}, k_\mu \geq k_\lambda \\
    \lambda_m \mu_m \frac{(\sqrt{k_\mu + 1})^2 - k_\mu (\sqrt{k_{\mu} - 1})^2}{2\sqrt{k_\mu}} & \text{if } \sqrt{k_\mu} \leq \frac{\sqrt{k_{\lambda} + 1}}{\sqrt{k_{\lambda} - 1}}, k_\mu < k_\lambda \\
    \lambda_1 \mu_m \frac{(\sqrt{k_{\lambda} + 1})^2 - k_\mu (\sqrt{k_{\lambda} - 1})^2}{2\sqrt{k_{\lambda}}} & \text{if } \sqrt{k_\mu} > \frac{\sqrt{k_{\lambda} + 1}}{\sqrt{k_{\lambda} - 1}}, k_\mu \geq k_\lambda \\
    \lambda_m \mu_1 \frac{(\sqrt{k_\mu + 1})^2 - k_\mu (\sqrt{k_{\mu} - 1})^2}{2\sqrt{k_\mu}} & \text{if } \sqrt{k_\mu} > \frac{\sqrt{k_{\lambda} + 1}}{\sqrt{k_{\lambda} - 1}}, k_\mu < k_\lambda 
\end{cases}
\]

\( \lambda_i, \mu_i \) are the \( i \)-th eigenvalues of \( \Sigma, BC^{-1}B' \) respectively in a descending enumeration. \( k_{\lambda} = \frac{\lambda}{\lambda_m} \) and \( k_\mu = \frac{\mu}{\mu_m} \) denote the spectral condition number of \( \Sigma \) and \( BC^{-1}B' \) respectively.

**Condition 17.9.5** There exists a positive number \( \lambda \) such that \( BC^{-1}B' = \lambda \Sigma \).

Condition 17.9.1 requires that the error terms of performance measures are stochastically independent. It assumes off the possibility that different measures are affected by common stochastic factor. Condition (17.9.2) states that 

---

\( \text{That is, } BC^{-1}B'\Sigma \text{ is symmetric.} \)
$b_i'C^{-1}b_j = 0$ for all $i \neq j$. Intuitively, it requires that different performance measures respond in distinct ways to the agent’s effort when cost is incorporated. Condition (17.9.4) holds when agent is sufficiently risk averse or when either matrix $BC^{-1}B'$ or $\Sigma$ is well-conditioned\(^8\). Several important special cases are:

- the performance measures system is orthogonal. In this case conditions (17.9.1) to (17.9.3) are all satisfied;
- $\Sigma$ is a scalar matrix, in which case (17.9.1), (17.9.3) and (17.9.4) are satisfied;
- $BC^{-1}B'$ is a scalar matrix, in which case conditions (17.9.2), (17.9.3) and (17.9.4) are satisfied.

Condition (17.9.5) emphasizes that the covariance matrix $\Sigma$ is a transformation of the measure-cost efficiency matrix $BC^{-1}B'$. That is to say, correlation between any pair of performance measures $i$ and $j$ is constant: $\rho_{ij} = \lambda$.

By comparing the wage contract obtained in the hybrid model to the benchmark pure moral hazard model, we find that the principal will reduce the power of incentives offered to the agent.

**Theorem 17.9.1**

1. Given any one of the conditions 17.9.1 to 17.9.4, there exists an $i \in \{1 \cdots m\}$, such that $|w_h^i(r)| < |w_p^i(r)|$ for all $r \in [\underline{r}, \bar{r}]$;

2. If both condition 17.9.1 and condition 17.9.2 are satisfied, then $|w_h^i(r)| < |w_p^i(r)|$ for all $r \in [\underline{r}, \bar{r}]$ and all $i \in \{1 \cdots m\}$.

3. Let $\omega_i, i \in K \equiv \{1, 2, \cdots, k\}$ denote $k$ distinct generalized eigenvalues of $BC^{-1}B'$ relative to $\Sigma$, $\forall_i \equiv \mathcal{N}(BC^{-1}B' - \omega_i \Sigma)$ be the eigenspace corresponding to $\omega_i$, and $\forall_i^\perp$ be its orthogonal complement. Suppose that $BC^{-1}\beta \not\in \bigcup_{i \in K} \forall_i^\perp$, then there exists a positive number $k \in (0,1)$ such that $w_h = kw_p$ if and only if condition 17.9.5 is met.

\(^8\)Matrices with condition numbers near 1 are said to be well-conditioned, while matrices with high condition numbers are said to be ill-conditioned.
Proof. Since the proof is long and complicated, it is omitted here and can be found in Meng and Tian (2013).

When the risk aversion parameter is unobservable to the principal, the less risk-averse agent gains information rent by mimicking the more risk-averse one. The amount of information rent gained by an agent depends on the performance wage of agents with larger risk aversion, and therefore the basic tradeoff between efficiency and rent extraction leads to low-powered incentive for all but the least risk-averse types. Under conditions (17.9.1) to (17.9.4), wage vector $w$ is shortened in different quadratic-form norms compared with the pure moral hazard case. Under condition (17.9.5), the wage vector that minimizes the cost of effort $e'Ce = w'BC^{-1}B'w$ points in the same direction as the wage vector that minimizes the risk premium $rw'Sw$. Consequently, the efficiency-rent tradeoff alters only the overall intensity of wage vector, not its relative allocation among performance measures.

17.9.3 Optimal Wage Contract with Unobservable Efforts and Cost

In this subsection we assume that the cost parameter is private information instead of risk aversion. To avoid the complicated multidimensional mechanism design issue, we assume that $C = cI$, that is, the tasks are technologically identical and independent. $\delta = \frac{1}{c}$ is assumed to be distributed on the support $[\delta, \delta]$, according to a cumulative distribution function $G(\delta)$ and density $g(\delta)$. The timeline of this problem is analogous to that in Section 17.9.2 except that the agent is now required to report $\hat{\delta}$. A contract menu $\{w_0(\delta), w(\delta)\}$ is said to be implementable if the following incentive compatibility condition is satisfied:

$$w_0(\delta) + \frac{1}{2}w(\delta)'[\delta BB' - r\Sigma]w(\delta) \geq w_0(\hat{\delta}) + \frac{1}{2}w(\hat{\delta})'[\delta BB' - r\Sigma]w(\hat{\delta}), \forall(\delta, \hat{\delta}) \in [\delta, \delta]^2.$$  

(17.97)
Let \( U(\delta, \hat{\delta}) \equiv w_0(\hat{\delta}) + \frac{1}{2} w'(\hat{\delta}) [\delta BB' - r \Sigma] w(\hat{\delta}) \), and \( U(\delta) \equiv U(\delta, \delta) \). Then \( \{ U(\delta), w(\delta) \} \) is called implementable if

\[
\exists w_0 : [\delta, \bar{\delta}] \rightarrow \mathbb{R}_+, \forall (\delta, \hat{\delta}) \in [\delta, \bar{\delta}]^2, U(\delta) = \max_{\delta} \left\{ w_0(\delta) + \frac{1}{2} w(\delta)' [\delta BB' - r \Sigma] w(\delta) \right\}
\]

(17.98)

or equivalently,

\[
\exists w_0 : \mathbb{R} \rightarrow \mathbb{R}_+, \forall \delta \in [\delta, \bar{\delta}], U(\delta) = \max_{w \in \mathbb{R}^m} \left\{ w_0(w) + \frac{1}{2} w'(\delta BB' - r \Sigma) w \right\}.
\]

(17.99)

\( U(\delta) \) is necessarily continuous, increasing and convex in \( \delta \) and satisfies the envelop condition:

\[
U'(\delta) = \frac{1}{2} w' BB' w.
\]

(17.100)

Conversely, similar to the case with unobservable risk aversion, the convexity of \( U(\delta) \) and envelop condition (17.100) implies

\[
U(\delta) \geq U(\hat{\delta}) + (\delta - \hat{\delta}) U'(\hat{\delta}) = U(\hat{\delta}) + \frac{1}{2} (\delta - \hat{\delta}) w' BB' w = U(\delta, \hat{\delta}),
\]

which in turn implies the implementability of contract. We summarize the above discussion in the following lemma.

**Lemma 17.9.2** The surplus function \( U(\delta) \) and wage function \( w(\delta) \) are implementable if and only if

1. \( U'(\delta) = \frac{1}{2} w' BB' w; \)
2. \( U(\delta) \) is convex in \( \delta \).

The second-best \( \delta \)–contingent contract solves the following optimization problem:

\[
\begin{align*}
\max \quad & \int_{\delta} g(\delta) d\delta \\
\text{s.t.:} \quad & U(\delta) \geq 0, U'(\delta) = \frac{w' BB' w}{2}, U(\delta) \text{ is convex}
\end{align*}
\]

\( ^9 \)In this case, let \( a = \frac{1}{2} w' BB' w, b = w_0(w) - \frac{1}{2} w' \Sigma w, \) then \( U(\delta) = \max_{a,b}(a\delta + b) \) is convex in \( \delta \).
Proposition 17.9.2  With unobservable cost, if \( \delta H(\delta) \) is nonincreasing, then the optimal wage is given by

\[
wh(\delta) = \left( H(\delta)BB' + \frac{r\Sigma}{\delta} \right)^{-1}B\beta \tag{17.101}
\]

\[
w_0h(\delta) = \frac{1}{2} \int_\delta \delta w^h(\tilde{\delta})'BB'w^h(\tilde{\delta})d\tilde{\delta} - \frac{1}{2}w^h(\delta)'[\delta BB' - r\Sigma]w^h(\delta) \tag{17.102}
\]

where \( H(\delta) \equiv 1 + \frac{1-G(\delta)}{\delta g(\delta)} \).

Proof. Again, the proof is omitted and referred to Meng and Tian (2013).

The following conditions justify the adoption of low-powered incentives in the case with unobservable cost parameter.

Condition 17.9.6  \( BB' \) is a diagonal matrix.

Condition 17.9.7  Matrices \( BB' \) and \( \Sigma \) commute: \( BB'\Sigma = \Sigma BB' \).

Condition 17.9.8  The following inequality holds:

\[
2\nu_m^2 + \frac{r}{\delta}\eta > 0.
\]

\[
\eta = \max \left\{ \min_{i=1,m} \lambda_i\nu_m \frac{(\sqrt{k\lambda} + 1)^2 - k\nu(\sqrt{k\lambda} - 1)^2}{2\sqrt{k\lambda}}, \min_{i=1,m} \nu_i\lambda_m \frac{(\sqrt{k\nu} + 1)^2 - k\lambda(\sqrt{k\nu} - 1)^2}{2\sqrt{k\nu}} \right\}
\]

\[
= \begin{cases} 
\lambda_m\nu_m \frac{(\sqrt{k\lambda} + 1)^2 - k\nu(\sqrt{k\lambda} - 1)^2}{2\sqrt{k\lambda}} & \text{if } \sqrt{k\lambda} \leq \frac{\sqrt{k\nu} + 1}{\sqrt{k\lambda} - 1}, k\nu \geq k\lambda \\
\lambda_m\nu_m \frac{(\sqrt{k\nu} + 1)^2 - k\nu(\sqrt{k\nu} - 1)^2}{2\sqrt{k\nu}} & \text{if } \sqrt{k\nu} \leq \frac{\sqrt{k\lambda} + 1}{\sqrt{k\lambda} - 1}, k\nu < k\lambda \\
\lambda_1\nu_1 \frac{(\sqrt{k\nu} + 1)^2 - k\nu(\sqrt{k\nu} - 1)^2}{2\sqrt{k\nu}} & \text{if } \sqrt{k\nu} > \frac{\sqrt{k\lambda} + 1}{\sqrt{k\lambda} - 1}, k\nu \geq k\lambda \\
\lambda_m\nu_1 \frac{(\sqrt{k\nu} + 1)^2 - k\nu(\sqrt{k\nu} - 1)^2}{2\sqrt{k\nu}} & \text{if } \sqrt{k\nu} > \frac{\sqrt{k\lambda} + 1}{\sqrt{k\lambda} - 1}, k\nu < k\lambda 
\end{cases}
\]

represents the lower bound of eigenvalues of Jordan product \( BB'\Sigma + \Sigma BB' \); \( \lambda_i, \nu_i \) are the \( i \)–th eigenvalues of \( \Sigma \) and \( BB' \) respectively in a descending enumeration. \( k\lambda = \frac{\lambda_m}{\lambda_m} \) and \( k\nu = \frac{\nu_1}{\nu_m} \) denote the spectral condition number of \( \Sigma \) and \( BB' \) respectively.

\[10\]This assumption is a bit stronger than the usual monotone inverse hazard rate condition. It holds for any nondecreasing \( g(.) \).

\[11\]Again, it is true if \( BB'\Sigma \) is symmetric.
Condition 17.9.9 There exists a positive number \( k \) such that \( BB' = k \Sigma \).

In the special case where performance measures system is orthogonal, conditions (17.9.1), (17.9.6) and (17.9.7) are satisfied. If \( \Sigma \) (resp. \( BB' \)) is a scalar matrix, then conditions (17.9.1) (resp. (17.9.6)) and (17.9.8) are both satisfied. Besides, condition (17.9.8) could hold even for nondiagonal \( BB' \) and \( \Sigma \), provided either of them is well-conditioned or \( \xi \) is sufficiently small.

Theorem 17.9.2 1. Given any of conditions (17.9.1), (17.9.6), (17.9.7), (17.9.8), there exists at least one \( i \in \{1, \cdots, m\} \), such that \( |w^h_i(\delta)| < |w^p_i(\delta)| \) for all \( \delta \in [\bar{\delta}, \delta] \);

2. If both conditions (17.9.1) and (17.9.6) are satisfied, namely, the performance measure system is orthogonal, then \( |w^h_i(\delta)| < |w^p_i(\delta)| \) for all \( \delta \in [\bar{\delta}, \delta] \) and all \( i \);

3. Let \( \tau_i, i \in L \equiv \{1, 2, \cdots, l\} \) denote \( l \) distinct generalized eigenvalues of \( BB' \) relative to \( \Sigma \), \( U_i \equiv \mathcal{N}(BB' - \tau_i \Sigma) \) be the eigenspace corresponding to \( \tau_i \), \( U_i^\perp \) be its orthogonal complement. Suppose that \( B \beta \notin \bigcup_{i \in L} U_i^\perp \), then there exists a positive number \( s \in (0, 1) \) such that \( w^h = sw^p \) if and only if condition (17.9.9) is met.

Proof. Again, it is omitted here and referred to Meng and Tian (2013).

When the agent possesses private information on his own cost, a more efficient agent (the agent with higher \( \delta \)) would accrue information rent by mimicking his less efficient counterpart. To minimize agency costs, optimality requires a downward distortion of the power of inefficient types’ incentive wage. Theorem 17.9.2 gives various conditions ensuring low-powered incentives. If the performance measure sensitivities are orthogonal to each other (\( b'_i b_j = 0 \) for \( i \neq j \)), or error terms are uncorrelated (\( \sigma_{ij} = 0 \) for \( i \neq j \)), or either \( BB' \) or \( \Sigma \) is well-conditioned (\( k_\nu \) or \( k_\lambda \) is close to one), or the agent is nearly risk neutral (\( r \) is very small), or the agent is highly efficient (\( \delta \) is very large), then the power of incentives will be
lowered for at least one performance measure. For an orthogonal system with all its performance measures congruent ($b_i' \beta \neq 0$ for all $i$), the wage vector in hybrid model is shorter than but points in the same direction as its pure moral hazard counterpart if and only if all the measures share the same signal-to-noise ratio ($b_i' b_i / \sigma_i^2 \equiv k$ for all $i$).

17.10 Further Extension

We have stressed the various conflicts that may appear in a moral hazard environment. The analysis of these conflicts, under both limited liability and risk aversion, was made easy because of our focus on a simple $2 \times 2$ environment with a binary effort and two levels of performance. The simple interaction between a single incentive constraint with either a limited liability constraint or a participation constraint was quite straightforward.

When one moves away from the $2 \times 2$ model, the analysis becomes much harder, and characterizing the optimal incentive contract is a difficult task. Examples of such complex contracting environment are abound. Effort may no longer be binary but, instead, may be better characterized as a continuous variable. A manager may no longer choose between working or not working on a project but may be able to fine-tune the exact effort spent on this project. Even worse, the agent’s actions may no longer be summarized by a one-dimensional parameter but may be better described by a whole array of control variables that are technologically linked. For instance, the manager of a firm may have to choose how to allocate his effort between productive activities and monitoring his peers and other workers.

Nevertheless, one can extend the standard model to the cases where the agent can perform more than two and possibly a continuum of levels of effort, to the case with a multitask model, the case where the agent’s utility function is no longer separable between consumption and effort. One can also analyze the trade-off between efficiency and redistribution in a moral hazard context. For detailed
discussion, see Chapter 5 of Laffont and Martimort (2002).

17.11 Reference

Books and Monographs:


Papers:


Chapter 18

General Mechanism Design: Contracts with Multi-Agents

18.1 Introduction

In the previous chapters on the principal-agent theory, we have introduced basic models to explain the core of the principal-agent theory with complete contracts. It highlights the various trade-offs between allocative efficiency and the distribution of information rents. Since the model involves only one agent, the design of the principal’s optimal contract has reduced to a constrained optimization problem without having to appeal to sophisticated game theory concepts.

In this chapter, we will introduce some of basic results and insights of the mechanism design in general, and implementation theory in particular for situations where there is one principal (also called the designer) and several agents. In such a case, asymmetric information may not only affect the relationship between the principal and each of his agents, but it may also plague the relationships between agents. To describe the strategic interaction between agents and the principal, the game theoretic reasoning is thus used to model social institutions as varied voting systems, auctions, bargaining protocols, and methods for deciding on public projects.
Incentive problems arise when the social planner cannot distinguish between things that are indeed different so that free-ride problem may appear. A free rider can improve his welfare by not telling the truth about his own un-observable characteristic. Like the principal-agent model, a basic insight of the incentive mechanism with more than one agent is that incentive constraints should be considered coequally with resource constraints. One of the most fundamental contributions of the mechanism theory has been shown that the free-rider problem may or may not occur, depending on the kind of game (mechanism) that agents play and other game theoretical solution concepts. A theme that comes out of the literature is the difficulty of finding mechanisms compatible with individual incentives that simultaneously results in a desired social goal.

Examples of incentive mechanism design that takes strategic interactions among agents exist for a long time. An early example is the Biblical story of the famous judgement of Solomon for determining who is the real mother of a baby. Two women came before the King, disputing who was the mother of a child. The King’s solution used a method of threatening to cut the lively baby in two and give half to each. One women was willing to give up the child, but another women agreed to cut in two. The King then made his judgement and decision: The first woman is the mother, do not kill the child and give him to the first woman. Another example of incentive mechanism design is how to cut a pie and divide equally among all participants.

The first major development was in the work of Gibbard-Hurwicz-Satterthwaite in 1970s. When information is private, the appropriate equilibrium concept is dominant strategies. These incentives adopt the form of incentive compatibility constraints where for each agent to tell truth about their characteristics must be dominant. The fundamental conclusion of Gibbard-Hurwicz-Satterthwaite’s impossibility theorem is that we have to have a trade-off between the truth-telling and Pareto efficiency (or the first best outcomes in general). Of course, if one is willing to give up Pareto efficiency, we can have a truth-telling mechanism, such as Vickery-Clark-Groves mechanism. In many cases, one can ignore the first-best
or Pareto efficiency, and so one can expect the truth-telling behavior.

On the other hand, we could give up the truth-telling requirement, and want to reach Pareto efficient outcomes. When the information about the characteristics of the agents is shared by individuals but not by the designer, then the relevant equilibrium concept is the Nash equilibrium. In this situation, one can give up the truth-telling, and uses a general message space. One may design a mechanism that Nash implements Pareto efficient allocations.

We will introduce these results and such trade-offs. In next chapter, we will also discuss the case of incomplete information in which agents do not know each other’s characteristics, and we need to consider Bayesian incentive compatible mechanism.

18.2 Basic Settings

Theoretical framework of the incentive mechanism design consists of five components: (1) economic environments (fundamentals of economy); (2) social choice goal to be reached; (3) economic mechanism that specifies the rules of game; (4) description of solution concept on individuals’ self-interested behavior, and (5) implementation of a social choice goal (incentive-compatibility of personal interests and the social goal at equilibrium).

18.2.1 Economic Environments

\[ e_i = (Z_i, w_i, \succeq_i, Y_i) \]: economic characteristic of agent \( i \) which consists of outcome space, initial endowment if any, preference relation, and the production set if agent \( i \) is also a producer;

\[ e = (e_1, \ldots, e_n) \]: an economy;

\( E \): The set of all priori admissible economic environments.

\( U = U_1 \times \ldots \times U_n \): The set of all admissible utility functions.
\[ \Theta = \Theta_1 \times \Theta_2 \cdots \times \Theta_n: \text{ The set of all admissible parameters } \theta = (\theta_1, \cdots, \theta_I) \in \Theta \text{ that determine types of parametric utility functions } u_i(\cdot, \theta_i), \text{ and so it is called the space of types or called the state of the world.} \]

Remark 18.2.1 Here, \( E \) is a general expression of economic environments. However, depending on the situations facing the designer, the set of admissible economic environments under consideration sometimes may be just given by \( E = U \), \( E = \Theta \), or by the set of all possible initial endowments, or production sets.

The designer is assumed that he does not know individuals’ economic characteristics. The individuals may or may not know the characteristics of the others. If they know, it is called the complete information case, otherwise it is called the incomplete information case.

For simplicity (but without loss of generality), when individuals’ economic characteristics are private information (the case of incomplete information), we assume preferences are given by parametric utility functions. In this case, each agent \( i \) privately observes a type \( \theta_i \in \Theta_i \), which determines his preferences over outcomes. The state \( \theta \) is drawn randomly from a prior distribution with density \( \varphi(\cdot) \) that can also be probabilities for finite \( \Theta \). Each agent maximizes von Neumann-Morgenstern expected utility over outcomes, given by (Bernoulli) utility function \( u_i(y, \theta_i) \). Thus, Information structure is specified by

1. \( \theta_i \) is privately observed by agent \( i \);
2. \( \{u_i(\cdot, \cdot)\}_{i=1}^n \) is common knowledge;
3. \( \varphi(\cdot) \) is common knowledge.

In this chapter, we will first discuss the case of complete information and then the case of incomplete information.
18.2.2 Social Goal

Given economic environments, each agent participates economic activities, makes decisions, receives benefits and pays costs on economic activities. The designer wants to reach some desired goal that is considered to be socially optimal by some criterion.

Let

\[ Z = Z_1 \times \ldots \times Z_n: \text{the outcome space (For example, } Z = X \times Y). \]
\[ A \subseteq Z: \text{the feasible set.} \]
\[ F: E \rightarrow A: \text{the social goal or called social choice correspondence in which } F(e) \text{ is the set of socially desired outcomes at the economy under some criterion of social optimality.} \]

*Examples of Social Choice Correspondences:*

- \( P(e) \): the set of Pareto efficient allocations.
- \( I(e) \): the set of individual rational allocations.
- \( W(e) \): the set of Walrasian allocations.
- \( L(e) \): the set of Lindahl allocations.
- \( FA(e) \): the set of fare allocations.

When \( F \) becomes a single-valued function, denoted by \( f \), it is called a social choice function.

*Examples of Social Choice Functions:*

- Solomon’s goal.
- Majority voting rule.

18.2.3 Economic Mechanism

Since the designer lacks the information about individuals’ economic characteristics, he needs to design an appropriate incentive mechanism (contract or rules
of game) to coordinate the personal interests and the social goal, i.e., under the mechanism, all individuals have incentives to choose actions which result in socially optimal outcomes when they pursue their personal interests. To do so, the designer informs how the information he collected from individuals is used to determine outcomes, that is, he first tells the rules of games. He then uses the information or actions of agents and the rules of game to determine outcomes of individuals. Thus, a mechanism consists of a message space and an outcome function. Let

\[ M_i : \text{the message space of agent } i. \]

\[ M = M_1 \times \ldots \times M_n : \text{the message space in which communications take place.} \]

\[ m_i \in M_i : \text{a message reported by agent } i. \]

\[ m = (m_1, \ldots, m_n) \in M : \text{a profile of messages.} \]

\[ h : M \rightarrow Z : \text{outcome function that translates messages into outcomes.} \]

\[ \Gamma = < M, h > : \text{a mechanism} \]

That is, a mechanism consists of a message space and an outcome function.

**Remark 18.2.2** A mechanism is often also referred to as a game form. The terminology of game form distinguishes it from a game in game theory in number of ways. (1) Mechanism design is normative analysis in contrast to game theory, which is positive economics. Game theory is important because it predicts how a given game will be played by agents. Mechanism design goes one step further: given the physical environment and the constraints faced by the designer, what goal can be realized or implemented? What mechanisms are optimal among those that are feasible? (2) The consequence of a profile of message is an outcome in mechanism design rather than a vector of utility payoffs. Of course, once the preference of the individuals are specified, then a game form or mechanism induces a conventional game. (3) The preferences of individuals in the mechanism
design setting vary, while the preferences of a game takes as given. This distinc-
tion between mechanisms and games is critical. Because of this, an equilibrium
(dominant strategy equilibrium) in mechanism design is much easier to exist than
a game. (4) In designing mechanisms one must take into account incentive con-
straints in a way that personal interests are consistent to the goal that a designer
want to implement it.

**Remark 18.2.3** In the implementation (incentive mechanism design) literature,
one requires a mechanism be incentive compatible in the sense that personal inter-
estests are consistent with desired socially optimal outcomes even when individual
agents are self-interested in their personal goals without paying much attention
to the size of message. In the realization literature originated by Hurwicz (1972,
1986b), a sub-field of the mechanism literature, one also concerns the size of
message space of a mechanism, and tries to find economic system to have small
operation cost. The smaller a message space of a mechanism, the lower (trans-
action) cost of operating the mechanism. For the neoclassical economies, it has
been shown that competitive market economy system is the unique most efficient
system that results in Pareto efficient and individually rational allocations (cf,
(1982), Tian (2004, 2005)).

### 18.2.4 Solution Concept of Self-Interested Behavior

A basic assumption in economics is that individuals are self-interested in the
sense that they pursue their personal interests. Unless they can be better off,
they in general does not care about social interests. As a result, different eco-
nomic environments and different rules of game will lead to different reactions of
individuals, and thus each individual agent’s strategy on reaction will depend on
his self-interested behavior which in turn depends on the economic environments
and the mechanism.

Let $b(e, \Gamma)$ be the set of equilibrium strategies that describes the self-interested
behavior of individuals. Examples of such equilibrium solution concepts include Nash equilibrium, dominant strategy, Bayesian Nash equilibrium, etc.

Thus, given $E$, $M$, $h$, and $b$, the resulting equilibrium outcome is the composite function of the rules of game and the equilibrium strategy, i.e., $h(b(e, \Gamma))$.

### 18.2.5 Implementation and Incentive Compatibility

In which sense can we see individuals’ personal interests do not have conflicts with a social interest? We will call such problem as implementation problem. The purpose of an incentive mechanism design is to implement some desired socially optimal outcomes. Given a mechanism $\Gamma$ and equilibrium behavior assumption $b(e, \Gamma)$, the implementation problem of a social choice rule $F$ studies the relationship of the intersection state of $F(e)$ and $h(b(e, \Gamma))$, which can be illustrated by the following diagram.

![Diagrammatic Illustration of Mechanism design Problem.](image)

Figure 18.1: Diagrammatic Illustration of Mechanism design Problem.

We have the following various definitions on implementation and incentive
compatibility of $F$.

A Mechanism $< M, h >$ is said to

(i) fully implement a social choice correspondence $F$ in equilibrium strategy $b(e, \Gamma)$ on $E$ if for every $e \in E$

(a) $b(e, \Gamma) \neq \emptyset$ (equilibrium solution exists),

(b) $h(b(e, \Gamma)) = F(e)$ (personal interests are fully consistent with social goals);

(ii) strongly implement a social choice correspondence $F$ in equilibrium strategy $b(e, \Gamma)$ on $E$ if for every $e \in E$

(a) $b(e, \Gamma) \neq \emptyset$,

(b) $h(b(e, \Gamma)) \subseteq F(e)$;

(iii) implement a social choice correspondence $F$ in equilibrium strategy $b(e, \Gamma)$ on $E$ if for every $e \in E$

(a) $b(e, \Gamma) \neq \emptyset$,

(b) $h(b(e, \Gamma)) \cap F(e) \neq \emptyset$.

A Mechanism $< M, h >$ is said to be $b(e, \Gamma)$ incentive-compatible with a social choice correspondence $F$ in $b(e, \Gamma)$-equilibrium if it (fully or strongly) implements $F$ in $b(e, \Gamma)$-equilibrium.

Note that we did not give a specific solution concept yet when we define the implementability and incentive-compatibility. As shown in the following, whether or not a social choice correspondence is implementable will depend on the assumption on the solution concept of self-interested behavior. When information is complete, the solution concept can be dominant equilibrium, Nash equilibrium, strong Nash equilibrium, subgame perfect Nash equilibrium, undominated equilibrium, etc. For incomplete information, equilibrium strategy can be Bayesian Nash equilibrium, undominated Bayesian Nash equilibrium, etc.
18.3 Examples

Before we discuss some basic results in the mechanism theory, we first give some economic environments which show that one needs to design a mechanism to solve the incentive compatible problems.

Example 18.3.1 (A Public Project) A society is deciding on whether or not to build a public project at a cost $c$. The cost of the public project is shared by individuals. Let $s_i$ be the share of the cost by $i$ so that $\sum_{i \in N} s_i = 1$.

The outcome space is then $Y = \{0, 1\}$, where 0 represents not building the project and 1 represents building the project. Individual $i$’s value from use of this project is $r_i$. In this case, the net value of individual $i$ is 0 from not having the project built and $v_i = r_i - s_i c$ from having a project built. Thus agent $i$’s valuation function can be represented as

$$v_i(y, v_i) = yr_i - ys_i c = yv_i.$$

Example 18.3.2 (Continuous Public Goods Setting) In the above example, the public good could only take two values, and there is no scale problem. But, in many case, the level of public goods depends on the collection of the contribution or tax. Now let $y \in R_+$ denote the scale of the public project and $c(y)$ denote the cost of producing $y$. Thus, the outcome space is $Z = R_+ \times R^n$, and the feasible set is $A = \{ (y, z_1(y), \ldots, z_n(y)) \in R_+ \times R^n : \sum_{i \in N} z_i(y) = c(y) \}$, where $z_i(y)$ is the share of agent $i$ for producing the public goods $y$. The benefit of $i$ for building $y$ is $r_i(y)$ with $r_i(0) = 0$. Thus, the net benefit of not building the project is equal to 0, the net benefit of building the project is $r_i(y) - z_i(y)$. The valuation function of agent $i$ can be written as

$$v_i(y) = r_i(y) - z_i(y).$$

Example 18.3.3 (Allocating an Indivisible Private Good) An indivisible good is to be allocated to one member of society. For instance, the rights to an exclusive license are to be allocated or an enterprise is to be privatized. In this
case, the outcome space is $Z = \{y \in \{0,1\}^n : \sum_{i=1}^n y_i = 1\}$, where $y_i = 1$ means individual $i$ obtains the object, $y_i = 0$ represents the individual does not get the object. If individual $i$ gets the object, the net value benefitted from the object is $v_i$. If he does not get the object, his net value is 0. Thus, agent $i$’s valuation function is

$$ v_i(y) = v_iy_i. $$

Note that we can regard $y$ as $n$-dimensional vector of public goods since $v_i(y) = v_iy_i = v^iy$, where $v^i$ is a vector where the $i$-th component is $v_i$ and the others are zeros, i.e., $v^i = (0, \ldots, 0, v_i, 0, \ldots, 0)$.

From these examples, a socially optimal decision clearly depends on the individuals’ true valuation function $v_i(\cdot)$. For instance, we have shown previously that a public project is produced if and only if the total values of all individuals is greater than its total cost, i.e., if $\sum_{i\in N} r_i > c$, then $y = 1$, and if $\sum_{i\in N} r_i < c$, then $y = 0$.

Let $V_i$ be the set of all valuation functions $v_i$, let $V = \prod_{i\in N} V_i$, let $h : V \to Z$ is a decision rule. Then $h$ is said to be efficient if and only if:

$$ \sum_{i\in N} v_i(h(v)) \geq \sum_{i\in N} v_i(h'(v')) \quad \forall v' \in V. $$

### 18.4 Dominant Strategy and Truthful Revelation Mechanisms

The most robust and strongest solution concept of describing self-interested behavior is dominant strategy. The dominant strategy identifies situations in which the strategy chosen by each individual is the best, regardless of choices of the others. The beauty of this equilibrium concept is in the weak rationality it demands of the agents: an agent need not forecast what the others are doing. An axiom in game theory is that agents will use it as long as a dominant strategy exists.
For \( e \in E \), a mechanism \( \Gamma = \langle M, h \rangle \) is said to have a dominant strategy equilibrium \( m^* \) if for all \( i \)

\[
h_i(m^*_i, m_{-i}) \succ_i h_i(m_i, m_{-i}) \text{ for all } m \in M. \tag{18.1}
\]

Denote by \( D(e, \Gamma) \) the set of dominant strategy equilibria for \( \Gamma = \langle M, h \rangle \) and \( e \in E \).

Under the assumption of dominant strategy, since each agent’s optimal choice does not depend on the choices of the others and does not need to know characteristics of the others, the required information is least when an individual makes decisions. Thus, if it exists, it is an ideal situation. Another advantage is that bad equilibria are usually not a problem. If an agent has two dominant strategies they must be payoff-equivalent, which is not a generic property.

In a given game, dominant strategies are not likely to exist. However, since we are designing the game, we can try to ensure that agents do have dominant strategies.

When the solution concept is given by dominant strategy equilibrium, i.e., \( b(e, \Gamma) = D(e, \Gamma) \), a mechanism \( \Gamma = \langle M, h \rangle \) is said to

(i) fully implement a social choice correspondence \( F \) in dominant equilibrium strategy \( D(e, \Gamma) \) on \( E \) if for every \( e \in E \)

(a) \( D(e, \Gamma) \neq \emptyset \) (equilibrium solution exists),

(b) \( h(D(e, \Gamma)) = F(e) \) (personal interests are fully consistent with social goals);

(ii) strongly implement a social choice correspondence \( F \) in dominant equilibrium strategy \( D(e, \Gamma) \) on \( E \) if for every \( e \in E \)

(a) \( D(e, \Gamma) \neq \emptyset \),

(b) \( h(D(e, \Gamma)) \subseteq F(e) \);

(iii) implement a social choice correspondence \( F \) in dominant equilibrium strategy \( D(e, \Gamma) \) on \( E \) if for every \( e \in E \)
The above definitions have applied to general (indirect) mechanisms, there is, however, a particular class of game forms which have a natural appeal and have received much attention in the literature. These are called direct or revelation mechanisms, in which the message space $M_i$ for each agent $i$ is the set of possible characteristics $E_i$. In effect, each agent reports a possible characteristic but not necessarily his true one.

A mechanism $\Gamma = < M, h >$ is said to be a revelation or direct mechanism if $M = E$.

**Example 18.4.1** The optimal contracts we discussed in Chapter 13 are revelation mechanisms.

**Example 18.4.2** The Groves mechanism we will discuss below is a revelation mechanism.

The most appealing revelation mechanisms are those in which truthful reporting of characteristics always turns out to be an equilibrium. It is the absence of such a mechanism which has been called the “free-rider” problem in the theory of public goods.

A revelation mechanism $< E, h >$ is said to implements a social choice correspondence $F$ truthfully in $b(e, \Gamma)$ on $E$ if for every $e \in E$,

(a) $e \in b(e, \Gamma)$;

(b) $h(e) \subset F(e)$.

That is, $F(\cdot)$ is truthfully implementable in dominant strategies if truth-telling is a dominant strategy for each agent in the direct revelation mechanism.

Truthfully implementable in dominant strategies is also called dominant strategy incentive compatible, strategy proof or straightforward.
Although the message space of a mechanism can be arbitrary, the following Revelation Principle tells us that one only needs to use the so-called revelation mechanism in which the message space consists solely of the set of individuals’ characteristics, and it is unnecessary to seek more complicated mechanisms. Thus, it will significantly reduce the complicity of constructing a mechanism.

**Theorem 18.4.1 (Revelation Principle)** Suppose a mechanism \(< M, h >\) implements a social choice rule \(F\) in dominant strategy. Then there is a revelation mechanism \(< E, g >\) which implements \(F\) truthfully in dominant strategy.

Proof. Let \(d\) be a selection of dominant strategy correspondence of the mechanism \(< M, h >\), i.e., for every \(e \in E\), \(m^* = d(e) \in D(e, \Gamma)\) such that \(h(d(e)) \in F(e)\). Since \(\Gamma = \langle M, h \rangle\) implements social choice rule \(F\), such a selection exists by the implementation of \(F\). Since the strategy of each agent is independent of the strategies of the others, each agent \(i\)'s dominant strategy can be expressed as \(m^*_i = d_i(e_i)\).

Define the revelation mechanism \(< E, g >\) by \(g(e) \equiv h(d(e))\) for each \(e \in E\). We first show that the truth-telling is always a dominant strategy equilibrium of the revelation mechanism \(\langle E, g \rangle\). Suppose not. Then, there exists a message \(e'\) and an agent \(i\) such that

\[u_i[g(e'_i, e'_{-i})] > u_i[g(e_i, e'_{-i})].\]

However, since \(g = h \circ d\), we have

\[u_i[h(d(e'_i), d(e'_{-i})] > u_i[h(d(e_i), d(e'_{-i})],\]

which contradicts the fact that \(m^*_i = d_i(e_i)\) is a dominant strategy equilibrium. This is because, when the true economic environment is \((e_i, e'_{-i})\), agent \(i\) has an incentive not to report \(m^*_i = d_i(e_i)\) truthfully, but have an incentive to report \(m'_i = d_i(e'_i)\), a contradiction.

Finally, since \(m^* = d(e) \in D(e, \Gamma)\) and \(< M, h >\) strongly implements a social choice rule \(F\) in dominant strategy, we have \(g(e) = h(d(e)) = h(m^*) \in F(e)\).
Hence, the revelation mechanism implements $F$ truthfully in dominant strategy. The proof is completed.

Thus, by the Revelation Principle, we know that, if truthful implementation rather than full or strong implementation is all that we require, we need never consider general mechanisms. In the literature, if a revelation mechanism $<E, h>$ truthfully implements a social choice rule $F$ in dominant strategy, the mechanism $\Gamma$ is sometimes said to be strongly individually incentive-compatible with a social choice correspondence $F$. In particular, when $F$ becomes a single-valued function $f$, $<E, f>$ can be regarded as a revelation mechanism. Thus, if a mechanism $<M, h>$ implements $f$ in dominant strategy, then the revelation mechanism $<E, f>$ is incentive compatible in dominant strategy, or called strongly individually incentive compatible.

**Remark 18.4.1** Notice that the Revelation Principle may be valid only for implementation. It does not apply to full or strong implementation. The Revelation Principle specifies a correspondence between a dominant strategy equilibrium of the original mechanism $<M, h>$ and the true profile of characteristics as a dominant strategy equilibrium, and it does not require the revelation mechanism has a unique dominant equilibrium so that the revelation mechanism $<E, g>$ may also exist non-truthful strategy equilibrium that does not corresponds to any equilibrium. Thus, in moving from the general (indirect) dominant strategy mechanisms to direct ones, one may introduce undesirable dominant strategies which are not truthful. More troubling, these additional strategies may create a situation where the indirect mechanism is a full or strong implantation of a given $F$, while the direct revelation mechanism is not. Thus, even if a mechanism fully or strongly implements a social choice function, the corresponding revelation mechanism $<E, g>$ may only implement, but not fully or strongly implement $F$.

However, if agents have strict preferences, i.e., $\forall a, b \in \mathcal{A}$, $a = b$ if and only if $a \sim_i b$, $\forall i \in n$, any two different outcomes cannot be indifferent. Thus the Revelation Principle is valid for full implementation, but not just for implementation.
The following result shows that, although \((\Gamma, \theta)\) may have more than one dominant strategy equilibrium, under condition of strict preferences, the resulting equilibrium outcome will be unique, and thus the revelation mechanism \(<E, g>\) by \(g(e) \equiv h(d(e))\) fully implements the social choice correspondence in dominant strategy.

**Proposition 18.4.1** Suppose agents have strict preferences. Then any dominant strategy outcome of \(\Gamma = < M, h >\) that implements a social choice goal \(F\) in dominant strategy is unique, and thus the revelation mechanism \(<E, g>\) by \(g(e) \equiv h(d(e))\) strongly implements \(F\) in dominant strategy.

**Proof.** Suppose \(\Gamma = < M, h >\) implements \(F\) in dominant strategy, but the equilibrium is not unique. Then, there is \(m^*_i, m'^*_i \in D_i(\Gamma, \theta)\) such that \(m^*_i \neq m'^*_i\). By dominant strategy, for all \(m_{-i} \in M_{-i}\), we have

\[
\begin{align*}
    h(m^*_i, m_{-i}) &>^e_i h(m'^*_i, m_{-i}) \\
    h(m'^*_i, m_{-i}) &>^e_i h(m^*_i, m_{-i})
\end{align*}
\]

Thus, \(h(m^*_i, m_{-i}) \sim^e_i h(m'^*_i, m_{-i})\). By strict preferences, we have \(h(m^*_i, m_{-i}) = h(m'^*_i, m_{-i})\). Repeating the process for each \(i\), we have \(h(m^*_i) = h(m'^*_i), \forall (m^*_i, m'^*_i) \in D(\Gamma, \theta)\). Therefore, dominant strategy is unique.

Thus, for all \(m^* \in D(e, \Gamma)\), \(g(e) \equiv h(m^*(e))\) is a single-valued mapping. So the revelation mechanism \(<E, g>\) strongly implements \(F\) in dominant strategy.

As a direct corollary of Proposition 18.4.1, any social choice goal that can be fully implemented in dominant strategy must be a single-valued social choice function. Formally, we have

**Corollary 18.4.1** Suppose agents have strict preferences. Any social choice goal \(F\) that can be fully implemented in dominant strategy (i.e., \(h(D(e, \Gamma)) = F(e)\)) must be a single-valued social choice function.
Proof. Since $\Gamma = \langle M, h \rangle$ fully implements $F$ in dominant strategy, we have $h(D(e, \Gamma)) = F(e)$. Then, by Proposition 18.4.1, we have $h(m^*) = h(m^{**}), \forall (m^*, m^{**}) \in D(\Gamma, \theta).$ Thus, for $\forall \theta \in \Theta, f(\theta) = h(D(\Gamma, \theta))$ is a single-valued function.

18.5 Gibbard-Satterthwaite Impossibility Theorem

The Revelation Principle is very useful to find a dominant strategy mechanism. If one hopes a social choice goal $f$ can be (weakly) implemented in dominant strategy, one only needs to show the revelation mechanism $< E, f >$ is strongly incentive compatible.

However, the more is known a priori about agents’ characteristics, the fewer incentive-compatibility constraints an implementable choice function has to satisfy, and the more likely it is to be implementable. Thus, the worst possible case is when nothing is known about the agents’ preferences over $X$. The Gibbard-Satterthwaite impossibility theorem in Chapter 12 tells us that, if the domain of economic environments is unrestricted, such a mechanism does not exist unless it is a dictatorial mechanism. From the angle of the mechanism design, we restate this theorem here.

Definition 18.5.1 A social choice function is dictatorial if there exists an agent whose optimal choice is the social optimal.

Now we state the Gibbard-Satterthwaite Theorem without the proof that is very complicated. A proof can be found, say, in Salanié’s book (2000): Microeconomics of Market Failures.

Theorem 18.5.1 (Gibbard-Satterthwaite Theorem) If $X$ has at least 3 alternatives, a social choice function which is strongly individually incentive compatible and defined on a unrestricted domain is dictatorial.
Thus, to have positive results, we must relax assumptions either on the unrestricted domain or on \( \geq 3 \) alternatives.

When there are only two alternatives, Gibbard-Satterthwaite Theorem may not be true.

**Example 18.5.1** [Voting Two Candidates by the Majority Rule.](#) Suppose \( n \) people vote for two candidates \( a \) and \( b \) by the majority rule. Then the social choice rule is given by

\[
F(\theta) = \begin{cases} 
  a & \text{if } \#\{i \in I | a \succ^\theta_i b\} > \#\{i \in I | b \succ^\theta_i a\} \\
  b & \text{if } \#\{i \in I | a \succ^\theta_i b\} \leq \#\{i \in I | b \succ^\theta_i a\} 
\end{cases}
\]

Then, one can see that participant has incentive to truthfully reveal his/her preferences on candidates regardless reports of others. Thus, \( F(\cdot) \) can be implementable truthfully in dominant strategy.

### 18.6 Hurwicz Impossibility Theorem

The Gibbard-Satterthwaite impossibility theorem is a very negative result. This result is essentially equivalent to Arrow’s impossibility result. However, as we will show, when the admissible set of economic environments is restricted, the result may be positive as the Groves mechanism defined on quasi-linear utility functions. Unfortunately, the following Hurwicz’s impossibility theorem shows the Pareto efficiency and the truthful revelation is fundamentally inconsistent even for the class of neoclassical economic environments.

**Theorem 18.6.1** (Hurwicz Impossibility Theorem, 1972) For the neoclassical private goods economies, there is no mechanism \( < M, h > \) that implements Pareto efficient and individually rational allocations in dominant strategy. Consequently, any revelation mechanism \( < M, h > \) that yields Pareto efficient and individually rational allocations is not strongly individually incentive compatible. (Truth-telling about their preferences is not Nash Equilibrium).
Proof: By the Revelation Principle, we only need to show that any revelation mechanism cannot implement Pareto efficient and individually rational allocations truthfully in dominant equilibrium for a particular pure exchange economy. In turn, it is enough to show that truth-telling is not a Nash equilibrium for any revelation mechanism that yields Pareto efficient and individually rational allocations for a particular pure exchange economy.

Consider a private goods economy with two agents \((n = 2)\) and two goods \((L = 2)\),

\[
\begin{align*}
  w_1 &= (0, 2), \quad w_2 = (2, 0) \\
  u_i(x, y) &= \begin{cases} 
  3x_i + y_i & \text{if } x_i \leq y_i \\
  x_i + 3y_i & \text{if } x_i > y_i.
\end{cases}
\end{align*}
\]

Figure 18.2: An illustration of the proof of Hurwicz’s impossibility theorem.
Thus, feasible allocations are given by

\[
A = \{ [(x_1, y_1), (x_2, y_2)] \in \mathbb{R}_+^4 : \\
\quad x_1 + x_2 = 2 \\
\quad y_1 + y_2 = 2 \}.
\]

Let \( U_i \) be the set of all neoclassical utility functions, i.e. they are continuous and quasi-concave, which agent \( i \) can report to the designer. Thus, the true utility function \( \hat{u}_i \in U_i \). Then,

\[
U = U_1 \times U_2 \\
\quad h : U \to A.
\]

Note that, if the true utility function profile \( \hat{u}_i \) were a Nash Equilibrium, it would satisfy

\[
\hat{u}_i(h_i(\hat{u}_i, \hat{u}_{-i})) \geq \hat{u}_i(h_i(u_i, \hat{u}_{-i})) \tag{18.2}
\]

We want to show that \( \hat{u}_i \) is not a Nash equilibrium. Note that,

1. \( P(e) = O_1O_2 \) (contract curve);
2. \( IR(e) \cap P(e) = \overline{ab} \);
3. \( h(\hat{u}_1, \hat{u}_2) = d \in \overline{ab} \).

Now, suppose agent 2 reports his utility function by cheating:

\[
u_2(x_2, y_2) = 2x + y \tag{18.3}
\]

Then, with \( u_2 \), the new set of individually rational and Pareto efficient allocations is given by

\[
IR(e) \cap P(e) = \overline{ae} \tag{18.4}
\]

Note that any point between \( a \) and \( e \) is strictly preferred to \( d \) by agent 2. Thus, an allocation determined by any mechanism which is IR and Pareto efficient
allocation under \((\hat{\mathbf{u}}_1, \mathbf{u}_2)\) is some point, say, the point \(c\) in the figure, between the segment of the line determined by \(a\) and \(e\). Hence, we have

\[
\hat{u}_2(h_2(\hat{u}_1, u_2)) > \hat{u}_2(h_2(\hat{u}_1, \hat{u}_2)) \tag{18.5}
\]

since \(h_2(\hat{u}_1, u_2) = c \in \overline{ae}\). Similarly, if \(d\) is between \(ae\), then agent 1 has incentive to cheat. Thus, no mechanism that yields Pareto efficient and individually rational allocations is incentive compatible. The proof is completed.

Thus, the Hurwicz’s impossibility theorem implies that Pareto efficiency and the truthful revelation about individuals’ characteristics are fundamentally incompatible. However, if one is willing to give up Pareto efficiency, say, one only requires the efficient provision of public goods, is it possible to find an incentive compatible mechanism which results in the efficient provision of a public good and can truthfully reveal individuals’ characteristics? The answer is positive. For the class of quasi-linear utility functions, the so-called Vickrey-Clark-Groves Mechanism can be such a mechanism.

### 18.7 Vickrey-Clark-Groves Mechanisms

From Chapter 15 on public goods, we have known that public goods economies may present problems by a decentralized resource allocation mechanism because of the free-rider problem. Private provision of a public good generally results in less than an efficient amount of the public good. Voting may result in too much or too little of a public good. Are there any mechanisms that result in the “right” amount of the public good? This is a question of the incentive compatible mechanism design.

Again, we will focus on the quasi-linear environment, where all agents are known to care for money. In this environment, the results are more positive, and we can even implement the efficient decision rule (the efficient provision of public goods). For simplicity, let us first return to the model of discrete public good.
18.7.1 Vickrey-Clark-Groves Mechanisms for Discrete Public Good

Consider a provision problem of a discrete public good. Suppose that the economy has \( n \) agents. Let

\[
\begin{align*}
  c & : \text{the cost of producing the public project.} \\
  r_i & : \text{the maximum willingness to pay of } i. \\
  s_i & : \text{the share of the cost by } i. \\
  v_i & = r_i - s_i c : \text{the net value of } i.
\end{align*}
\]

The public project is determined according to

\[
y = \begin{cases} 
  1 & \text{if } \sum_{i=1}^{n} v_i \geq 0 \\
  0 & \text{otherwise.}
\end{cases}
\]

From the discussion in Chapter 15, it is efficient to produce the public good, \( y = 1 \), if and only if

\[
\sum_{i=1}^{n} v_i = \sum_{i=1}^{n} (r_i - s_i g_i) \geq 0.
\]

Since the maximum willingness to pay for each agent, \( r_i \), is private information and so is the net value \( v_i \), what mechanism one should use to determine if the project is built? One mechanism that we might use is simply to ask each agent to report his or her net value and provide the public good if and only if the sum of the reported value is positive. The problem with such a scheme is that it does not provide right incentives for individual agents to reveal their true willingness-to-pay. Individuals may have incentives to underreport their willingness-to-pay.

Thus, a question is how we can induce each agent to truthfully reveal his true value for the public good. The so-called Vickrey-Clark-Groves (VCG) mechanism gives such a mechanism.
Suppose the utility functions are quasi-linear in net increment in private good, \( x_i - w_i \), which have the form:

\[
\bar{u}_i(x_i - w_i, y) = x_i - w_i + r_i y
\]

s.t. \( x_i + g_i y = w_i + t_i \),

where \( t_i \) is the transfer to agent \( i \). Then, we have

\[
u_i(t_i, y) = t_i + r_i y - g_i y = t_i + (r_i - g_i)y = t_i + v_i y.
\]

- Groves Mechanism:

In a Groves mechanism, agents are required to report their net values. Thus the message space of each agent \( i \) is \( M_i = \mathbb{R} \). The Groves mechanism is defined as follows:

\[ \Gamma = (M_1, \ldots, M_n, t_1(\cdot), t_2(\cdot), \ldots, t_n(\cdot), y(\cdot)) \equiv (M, t(\cdot), y(\cdot)), \]

where

(1) \( b_i \in M_i = \mathbb{R} \): each agent \( i \) reports a “bid” for the public good, i.e., report the net value of agent \( i \) which may or may not be his true net value \( v_i \).

(2) The level of the public good is determined by

\[
y(b) = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} b_i \geq 0 \\ 0 & \text{otherwise.} \end{cases}
\]

(3) Each agent \( i \) receives a side payment (transfer)

\[
t_i(b) = \begin{cases} \sum_{j \neq i} b_j & \text{if } \sum_{i=1}^{n} b_i \geq 0 \\ 0 & \text{otherwise.} \end{cases}
\]

Then, the payoff of agent \( i \) is given by

\[
\phi_i(b) = \begin{cases} v_i + t_i(b) = v_i + \sum_{j \neq i} b_j & \text{if } \sum_{i=1}^{n} b_i \geq 0 \\ 0 & \text{otherwise.} \end{cases}
\]
We want to show that it is optimal for each agent to report the true net value, \( b_i = v_i \), regardless of what the other agents report. That is, truth-telling is a dominant strategy equilibrium.

There are two cases to be considered.

Case 1: \( v_i + \sum_{j \neq i} b_j > 0 \). Then agent \( i \) can ensure the public good is provided by reporting \( b_i = v_i \). Indeed, if \( b_i = v_i \), then \( \sum_{j \neq i} b_j + v_i = \sum_{i=1}^{n} b_j > 0 \) and thus \( y = 1 \). In this case, \( \phi_i(v_i, b_{-i}) = v_i + \sum_{j \neq i} b_j > 0 \).

Case 2: \( v_i + \sum_{j \neq i} b_j \leq 0 \). Agent \( i \) can ensure that the public good is not provided by reporting \( b_i = v_i \) so that \( \sum_{i=1}^{n} b_i \leq 0 \). In this case, \( \phi_i(v_i, b_{-i}) = 0 \geq v_i + \sum_{j \neq i} b_j \).

Thus, for either cases, agent \( i \) has incentives to tell the true value of \( v_i \). Hence, it is optimal for agent \( i \) to tell the truth. There is no incentive for agent \( i \) to misrepresent his true net value regardless of what other agents do.

The above preference revelation mechanism has a major fault: the total side-payment may be very large. Thus, it is very costly to induce the agents to tell the truth.

Ideally, we would like to have a mechanism where the sum of the side-payment is equal to zero so that the feasibility condition holds, and consequently it results in Pareto efficient allocations, but in general it impossible by Hurwicz’s impossibility theorem. However, we could modify the above mechanism by asking each agent to pay a “tax”, but not receive payment. Because of this “waster” tax, the allocation of public goods is still not Pareto efficient.

The basic idea of paying a tax is to add an extra amount to agent \( i \)’s side-payment, \( d_i(b_{-i}) \) that depends only on what the other agents do.

**General Groves Mechanism (Vickrey-Clark-Groves Mechanism):** Ask each agent to pay additional tax, \( d_i(b_{-i}) \).

In this case, the transfer is given by

\[
t_i(b) = \begin{cases} 
\sum_{j \neq i} b_j + d_i(b_{-i}) & \text{if } \sum_{i=1}^{n} b_i \geq 0 \\
d_i(b_{-i}) & \text{if } \sum_{i=1}^{n} b_i < 0.
\end{cases}
\]
The payoff to agent $i$ now takes the form:

$$
\phi_i(b) = \begin{cases} 
  v_i + t_i(b) = v_i + \sum_{j \neq i} b_j + d_i(b_{-i}) & \text{if } \sum_{i=1}^n b_i \geq 0 \\
  d_i(b_{-i}) & \text{otherwise.}
\end{cases} \quad (18.8)
$$

For exactly the same reason as for the mechanism above, one can prove that it is optimal for each agent $i$ to report his true net value. In conclusion, we have the following proposition.

**Proposition 18.7.1** For discrete public good economies under consideration, the truth-telling is a dominant strategy under the Vickrey-Clark-Groves mechanism that implements truthfully the efficient decision rule (the efficient provision of public goods) in dominant strategy.

If the function $d_i(b_{-i})$ is suitably chosen, the size of the side-payment can be significantly reduced. A nice choice was given by Clarke (1971). He suggested a particular Groves mechanism known as the Clarke mechanism (also called Pivot mechanism):

The Pivotal Mechanism is a special case of the general Groves Mechanism in which $d_i(b_{-i})$ is given by

$$
d_i(b_{-i}) = \begin{cases} 
  -\sum_{j \neq i} b_j & \text{if } \sum_{i=1}^n b_i \geq 0 \\
  0 & \text{if } \sum_{j \neq i} b_j < 0.
\end{cases}
$$

In this case, it gives

$$
t_i(b) = \begin{cases} 
  0 & \text{if } \sum_{i=1}^n b_i \geq 0 \text{ and } \sum_{j \neq i} b_j \geq 0, \\
  \sum_{j \neq i} b_j & \text{if } \sum_{i=1}^n b_i \geq 0 \text{ and } \sum_{j \neq i} b_j < 0 \\
  -\sum_{j \neq i} b_j & \text{if } \sum_{i=1}^n b_i < 0 \text{ and } \sum_{j \neq i} b_j \geq 0 \\
  0 & \text{if } \sum_{i=1}^n b_i < 0 \text{ and } \sum_{j \neq i} b_j < 0
\end{cases} \quad (18.9)
$$

i.e.,

$$
t_i(b) = \begin{cases} 
  -|\sum_{j \neq i} b_j| & \text{if } (\sum_{i=1}^n b_i)(\sum_{j \neq i} b_i) < 0 \\
  -|\sum_{j \neq i} b_j| & \text{if } \sum_{i=1}^n b_i = 0 \text{ and } \sum_{j \neq i} b_j < 0 \\
  0 & \text{otherwise.}
\end{cases} \quad (18.10)
$$
Therefore, the payoff of agent $i$

$$\phi_i(b) = \begin{cases} 
    v_i & \text{if } \sum_{i=1}^{n} b_i \geq 0 \text{ and } \sum_{j\neq i} b_j \geq 0 \\
    v_i + \sum_{j\neq i} b_j & \text{if } \sum_{i=1}^{n} b_i \geq 0 \text{ and } \sum_{j\neq i} b_j < 0 \\
    -\sum_{j\neq i} b_j & \text{if } \sum_{i=1}^{n} b_i < 0 \text{ and } \sum_{j\neq i} b_j \geq 0 \\
    0 & \text{if } \sum_{i=1}^{n} b_i < 0 \text{ and } \sum_{j\neq i} b_j < 0. 
\end{cases}$$

(18.11)

**Remark 18.7.1** Thus, from the transfer given in (18.10), adding in the side-payment has the effect of taxing agent $i$ only if he changes the social decision. Such an agent is called the pivotal person. The amount of the tax agent $i$ must pay is the amount by which agent $i$’s bid damages the other agents. The price that agent $i$ must pay to change the amount of public good is equal to the harm that he imposes on the other agents.

### 18.7.2 Vickrey-Clark-Groves Mechanisms with Continuous Public Goods

Now we are concerned with the provision of continuous public goods. Consider a public goods economy with $n$ agents, one private good, and $K$ public goods. Denote

- $x_i$: the consumption of the private good by $i$;
- $y$: the consumption of the public goods by all individuals;
- $t_i$: transfer payment to $i$;
- $g_i(y)$: the contribution made by $i$;
- $c(y)$: the cost function of producing public goods $y$ that satisfies the condition:

$$\sum g_i(y) = c(y).$$

Then, agent $i$’s budget constraint should satisfy

$$x_i + g_i(y) = w_i + t_i$$

(18.12)
and his utility functions are given by

\[ \bar{u}_i(x_i - w, y) = x_i - w + u_i(y). \] (18.13)

By substitution,

\[ u_i(t_i, y) = t_i + (u_i(y) - g_i(y)) \equiv t_i + v_i(y), \]

where \( v_i(y) \) is called the valuation function of agent \( i \). From the budget constraint,

\[ \sum_{i=1}^{n} \{ x_i + g_i(y) \} = \sum_{i=1}^{n} w_i + \sum_{i=1}^{n} t_i, \] (18.14)

we have

\[ \sum_{i=1}^{n} x_i + c(y) = \sum_{i=1}^{n} w_i + \sum_{i=1}^{n} t_i \] (18.15)

The feasibility (or balanced) condition then becomes

\[ \sum_{i=1}^{n} t_i = 0. \] (18.16)

Recall that Pareto efficient allocations are completely characterized by

\[ \max \sum a_i \bar{u}_i(x_i, y) \]

s.t. \[ \sum_{i=1}^{n} x_i + c(y) = \sum_{i=1}^{n} w_i. \]

For quasi-linear utility functions it is easily seen that the weights \( a_i \) must be the same for all agents since \( a_i = \lambda \) for interior Pareto efficient allocations (no income effect), the Lagrangian multiplier, and \( y \) is thus uniquely determined for the special case of quasi-linear utility functions \( u_i(t_i, y) = t_i + v_i(y) \). Then, the above characterization problem becomes

\[ \max_{t_i, y} \left[ \sum_{i=1}^{n} (t_i + v_i(y)) \right], \] (18.17)

or equivalently
\[ \max_{y} \sum_{i=1}^{n} v_i(y); \]

(2) (feasibility condition): \( \sum_{i=1}^{n} t_i = 0. \)

Then, the Lindahl-Samuelson condition is given by:
\[ \sum_{i=1}^{n} \frac{\partial v_i(y)}{\partial y_k} = 0, \]
that is,
\[ \sum_{i=1}^{n} \frac{\partial u_i(y)}{\partial y_k} = \frac{\partial c(y)}{\partial y_k}. \]

Thus, Pareto efficient allocations for quasi-linear utility functions are completely characterized by the Lindahl-Samuelson condition \( \sum_{i=1}^{n} v'_i(y) = 0 \) and feasibility condition \( \sum_{i=1}^{n} t_i = 0. \)

In a Groves mechanism, it is supposed that each agent is asked to report the valuation function \( v_i(y) \). Denote his reported valuation function by \( b_i(y) \).

To get the efficient level of public goods, the government may announce that it will provide a level of public goods \( y^* \) that maximizes
\[ \max_{y} \sum_{i=1}^{n} b_i(y). \]

The Groves mechanism has the form:
\[ \Gamma = (V,h), \quad (18.18) \]

where \( V = V_1 \times \ldots \times V_n \) : is the message space that consists of the set of all possible valuation functions with element \( b_i(y) \in V_i, \ h = (t_1(b), t_2(b), \ldots, t_n(b), y(b)) \) are outcome functions. It is determined by:

(1) Ask each agent \( i \) to report his/her valuation function \( b_i(y) \) which may or may not be the true valuation \( v_i(y) \);

(2) Determine \( y^* \): the level of the public goods, \( y^* = y(b) \), is determined by
\[ \max_{y} \sum_{i=1}^{n} b_i(y); \quad (18.19) \]
(3) Determine \( t_i \): transfer of agent \( i \), \( t_i \) is determined by

\[
t_i(b) = \sum_{j \neq i} b_j(y^*). \tag{18.20}
\]

The payoff of agent \( i \) is then given by

\[
\phi_i(b(y^*)) = v_i(y^*) + t_i(b) = v_i(y^*) + \sum_{j \neq i} b_j(y^*). \tag{18.21}
\]

The social planner’s goal is to have the optimal level \( y^* \) that solves the problem:

\[
\max_y \sum_{i=1}^{n} b_i(y).
\]

In which case is the individual’s interest consistent with social planner’s interest? Under the rule of this mechanism, it is optimal for each agent \( i \) to truthfully report his true valuation function \( b_i(y) = v_i(y) \) since agent \( i \) wants to maximize

\[
v_i(y) + \sum_{j \neq i} b_j(y).
\]

By reporting \( b_i(y) = v_i(y) \), agent \( i \) ensures that the government will choose \( y^* \) which also maximizes his payoff while the government maximizes the social welfare. That is, individual’s interest is consistent with the social interest that is determined by the Lindahl-Samuelson condition. Thus, truth-telling, \( b_i(y) = v_i(y) \), is a dominant strategy equilibrium.

In general, \( \sum_{i=1}^{n} t_i(b(y)) \neq 0 \), which means that a Groves mechanism in general does not result in Pareto efficient outcomes even if it satisfies the Lindahl-Samuelson condition, i.e., it is Pareto efficient to provide the public goods.

As in the discrete case, the total transfer can be very large, just as before, they can be reduced by an appropriate side-payment. The Groves mechanism can be modified to

\[
t_i(b) = \sum_{j \neq i} b_j(y) + d_i(b_{-i}).
\]

The general form of the Groves Mechanism (Vickrey-Clark-Groves Mechanism) is then \( \Gamma = \langle V, t, y(b) \rangle \) such that
\[ \sum_{i=1}^{n} b_i(y(b)) \geq \sum_{i=1}^{n} b_i(y) \text{ for } y \in Y; \]

\[ t_i(b) = \sum_{j \neq i} b_j(y) + d_i(b \setminus i). \]

In summary, we have the following proposition.

**Proposition 18.7.2** For continuous public good economies under consideration, the truth-telling is a dominant strategy under the Vickrey-Clark-Groves mechanism that implements truthfully the efficient decision rule \( y^* (\cdot) \) (the efficient provision of public goods) in dominant strategy.

A special case of the Vickrey-Clark-Groves mechanism is independently described by Clark and is called the Clark mechanism (also called the pivotal mechanism) in which \( d_i(b \setminus i) \) is given by

\[ d_i(b \setminus i) = \max_y \sum_{j \neq i} b_j(y). \quad (18.22) \]

That is, the pivotal mechanism, \( \Gamma = < V, t, y(b) > \), is to choose \( (y^*, t^*_i) \) such that

1. \( \sum_{i=1}^{n} b_i(y^*) \geq \sum_{i=1}^{n} b_i(y) \text{ for } y \in Y; \)
2. \( t_i(b) = \sum_{j \neq i} b_j(y^*) - \max_y \sum_{j \neq i} b_j(y). \)

It is interesting to point out that the Clark mechanism contains the well-known Vickery auction mechanism (the second-price auction mechanism) as a special case. Under the Vickery mechanism, the highest bidding person obtains the object, and he pays the second highest bidding price. To see this, let us explore this relationship in the case of a single good auction (Example 9.3.3 in the beginning of this chapter). In this case, the outcome space is

\[ Z = \{ y \in \{0, 1\}^n : \sum_{i=1}^{n} y_i = 1 \} \]

where \( y_i = 1 \) implies that agent \( i \) gets the object, and \( y_i = 0 \) means the person does not get the object. Agent \( i \)'s valuation function is then given by

\[ v_i(y) = v_i y_i. \]
Since we can regard $y$ as a $n$-dimensional vector of public goods, by the Clark mechanism above, we know that

$$y^* = g(b) = \{y \in Z : \max_{i=1}^n b_i y_i\} = \{y \in Z : \max_{i \in N} b_i\},$$

and the truth-telling is a dominate strategy. Thus, if $g_i(v) = 1$, then $t_i(v) = \sum_{j \neq i} v_j y_j^* - \max_y \sum_{j \neq i} v_j y_j = -\max_{j \neq i} v_j$. If $g_i(b) = 0$, then $t_i(v) = 0$. This means that the object is allocated to the individual with the highest valuation and he pays an amount equal to the second highest valuation. No other payments are made. This is exactly the outcome predicted by the Vickery mechanism.

### 18.7.3 Uniqueness of VCG for Efficient Decision

Do there exist other mechanisms implementing the efficient decision rule $y^*$? The answer is no if valuation functions $v(y, \cdot)$ is sufficiently “rich” in a sense. To see this, consider the class of parametric valuation functions $v_i(y, \cdot)$ with continuum type space $\Theta_i = [\theta, \bar{\theta}]$.

**Proposition 18.7.3 (Laffont and Maskin (1980))** Suppose $Y = \mathbb{R}$, $\Theta = [\theta, \bar{\theta}]$ and $v_i : Y \times \Theta_i \to \mathbb{R}$, $\forall i \in N$, is differentiable. Then any mechanism implementing an efficient decision rule $y^*(\cdot)$ in dominant strategy is a Vickrey-Clark-Groves mechanism. That is, if a social choice rule $f(\theta) = \{y(\theta), t_1(\theta), \ldots, t_n(\theta)\}$ is implemented truthfully in dominant strategy, and

$$y(\theta) = \arg\max_{y \in Y} \sum_{i=1}^n v_i(y, \theta_i),$$

then we must have $t_i(\theta) = \sum_{j \neq i} v_j(y(\theta), \theta_j) + d_i(\theta_{-i})$, where $d_i(\theta_{-i})$ is independent of $\theta_i$.

Proof. Since

$$y(\theta) = \arg\max_{y \in Y} \sum_{i=1}^n v_i(y, \theta_i),$$

we have

$$\sum_{i=1}^n \frac{\partial v_i}{\partial y}(y(\theta), \theta_i) = 0. \quad (18.23)$$
By the requirement of truth implementation in dominant strategy,

\[ v_i(y(\theta_i, \theta_{-i}), \theta_i) + t_i(\theta_i, \theta_{-i}) \geq v_i(y(\hat{\theta}_i, \theta_{-i}), \theta_i) + t_i(\hat{\theta}_i, \theta_{-i}), \forall \theta_i, \hat{\theta}_i, \theta_{-i}. \]

Then

\[ \frac{\partial v_i}{\partial y}(y(\theta), \theta_i) \frac{\partial y}{\partial \theta_i}(\theta) + \frac{\partial t_i}{\partial \theta_i}(\theta_i, \theta_{-i}) = 0, \forall (\theta_i, \theta_{-i}) \in \Theta. \] (18.24)

Let \( d_i(\theta) \triangleq t_i(\theta) - \sum_{j \neq i} v_j(y(\theta), \theta_j) \). We need to show \( d_i(\theta_i, \theta_{-i}) \) is independent of \( \theta_i \). Indeed, since

\[
\frac{\partial d_i}{\partial \theta_i} = \frac{\partial t_i}{\partial \theta_i}(\theta_i, \theta_{-i}) - \sum_{j \neq i} \frac{\partial v_i}{\partial y}(y(\theta), \theta_j) \frac{\partial y}{\partial \theta_i}(\theta_i, \theta_{-i}) \\
= \frac{\partial t_i}{\partial \theta_i}(\theta_i, \theta_{-i}) + \frac{\partial v_i}{\partial y}(y(\theta), \theta_i) \frac{\partial y}{\partial \theta_i}(\theta_i, \theta_{-i}) - \sum_{j=1}^{n} \frac{\partial v_j}{\partial y}(y(\theta), \theta_j) \frac{\partial y}{\partial \theta_i}(\theta_i, \theta_{-i}) \\
= 0,
\] (18.25)

we have \( d_i(\theta) = d_i(\theta_{-i}) \).

The intuition behind the proof is that each agent \( i \) only has the incentive to tell the truth when he is made the residual claimant of total surplus, i.e., \( U_i(\theta) = S(\theta) + d_i(\theta_{-i}) \), which requires a VCG mechanism. This result was also obtained by Green and Laffont (1979), but under much more restrictive assumptions. A particular case of the proposition was also obtained by Green and Laffon (1979). Namely, they allow agents to have all possible valuations over a finite decision set \( X \), which can be described by the Euclidean type space \( \Theta_i = \mathbb{R}^{|K|} \) (the valuation of agent \( i \) for decision \( k \) being represented by \( \theta_{ik} \)). The proof can be seen in Mas-Colell, Whinston, Green, 1995) without imposing differentiability on \( v_i \).

### 18.7.4 Balanced VCG Mechanisms

Implementation of a social surplus-maximizing decision rule \( y^*(\cdot) \), is a necessary condition for (ex-post) Pareto efficiency, but it is not sufficient. To have Pareto efficiency, we must also guarantee that there is an ex post Balanced Budget (no
waste of numeraire):
\[ \sum_{i=1}^{n} t_i(\theta) = 0, \forall \theta \in \Theta. \] (18.26)

A trivial case for which we can achieve ex post efficiency is given in the following example.

**Example 18.7.1** If there exists some agent \( i \) such that \( \Theta_i = \{ \bar{\theta}_i \} \) a singleton, then we have no incentive problem for agent \( i \); and we can set
\[ t_i(\theta) = -\sum_{j \neq i} t_j(\theta), \forall \theta \in \Theta. \]

This trivially guarantees a balanced budget.

However, in general, there is no such a positive result. When the set of \( v(\cdot, \cdot) \) functions is “rich”, then there may be no choice rule \( f(\cdot) = (y^*(\cdot), t_1(\cdot), \cdots, t_n(\cdot)) \) where \( y^*(\cdot) \) is ex post optimal, which is truthfully implementable in dominant and satisfies (18.26), so that it does not result in Pareto efficient outcome.

To have an example where budget balance cannot be achieved, consider a setting with two agents. Under the assumptions of Proposition 18.7.3, any mechanism implementing an efficient decision rule \( y^*(\cdot) \) in dominant strategy is a VCG mechanism, hence
\[ t_i(\theta) = v_{-i}(y^*(\theta), \theta_{-i}) + d_i(\theta_{-i}). \]

Budget balance requires that
\[ 0 = t_1(\theta) + t_2(\theta) = v_1(y^*(\theta), \theta_1) + v_2(y^*(\theta), \theta_2) + d_1(\theta_2) + d_2(\theta_1). \]

Letting \( S(\theta) = v_1(y^*(\theta), \theta_1) + v_2(y^*(\theta), \theta_2) \) denote the maximum social surplus in state \( \theta \), we must therefore have
\[ S(\theta) = -d_1(\theta_2) - d_2(\theta_1). \]

Thus, efficiency can only be achieved when maximal total surplus is additively separable in the agents’ types, which is unlikely.
To have a robust example where additive separability does not hold, consider the "public good setting" in which $Y = [y, \bar{y}] \subset \mathbb{R}$ with $y < \bar{y}$, and for each agent $i$, $\Theta_i = [\underline{\theta}_i, \bar{\theta}_i] \subset \mathbb{R}, \underline{\theta}_i < \bar{\theta}_i$ (to rule out the situation described in the above Example), $v_i(y, \theta_i)$ is differentiable in $\theta_i$, $\frac{\partial v_i(y, \theta_i)}{\partial \theta_i}$ is bounded, $v_i(y, \theta_i)$ has the single-crossing property (SCP) in $(y, \theta_i)$ and $y^*(\theta_1, \theta_2)$ is in the interior of $Y$.

Note that by the Envelope Theorem, we have (almost everywhere)

$$\frac{\partial S(\theta)}{\partial \theta_1} = \frac{\partial v_1(y^*(\theta_1, \theta_2), \theta_1)}{\partial \theta_1}.$$

It is clear that if $\frac{\partial v_1(y, \theta_1)}{\partial \theta_1}$ depends on $y$, and $y^*(\theta_1, \theta_2)$ depends on $\theta_2$, in general $\frac{\partial S(y, \theta)}{\partial \theta_1}$ will depend on $\theta_2$, and therefore $S(\theta)$ will not be additively separable. In fact,\

- SCP of $v_1(\cdot)$ implies that $\frac{\partial v_1(y, \theta_1)}{\partial y}$ is strictly increasing in $\theta_1$.

- SCP of $v_2(\cdot)$ implies that $y^*(\theta_1, \theta_2)$ is strictly increasing in $\theta_2$.

Consequently, $\frac{\partial S(\theta)}{\partial \theta_1}$ is strictly increasing in $\theta_2$, thus $S(\theta)$ is not additively separable. In this case, no ex post efficient choice rule is truthfully implementable in dominant strategies.

As such, if the wasted money is subtracted from the social surplus, it is no longer socially efficient to implement the decision rule $y^*(\cdot)$. Instead, we may be interested, e.g., in maximizing the expectation

$$E \left[ \sum_i v_i(x(\theta), \theta_i) + \sum_i t_i(\theta) \right]$$

in the class of choice rules $(y(\cdot), t_1(\cdot), \cdots, t_I(\cdot))$ satisfying dominant incentive compatibility, given a probability distribution $\varphi$ over states $\theta$. In this case, we may have interim efficiency as we will see in the section on Bayesian implementation.

Nevertheless, when economic environments are "thin", we can get some positive results. Groves and Loeb (1975), Tian (1996a), and Liu and Tian (1999) provide such a class of economic environments. An example of such utility functions are: $u_i(t_i, y) = t_i + v_i(y)$ with $v_i(y) = -1/2y^2 + \theta_iy$ for $i = 1, 2, \ldots, n$ with $n \geq 3$. 

709
18.8 Nash Implementation

18.8.1 Nash Equilibrium and General Mechanism Design

From Hurwicz’s impossibility theorem, we know that, if one wants to have a mechanism that results in Pareto efficient and individually rational allocations, one must give up the dominant strategy implementation, and then, by Revelation Principle, we must look at more general mechanisms \(< M, h >\) instead of using a revelation mechanism.

We know the dominant strategy equilibrium is a very strong solution concept. Now, if we adopt the Nash equilibrium as a solution concept to describe individuals’ self-interested behavior, can we design an incentive compatible mechanism which implements Pareto efficient allocations?

For \(e \in E\), a mechanism \(< M, h >\) is said to have a Nash equilibrium \(m^* \in M\) if

\[
h_i(m^*) >_i h_i(m_i, m^*_{-i})
\]

for all \(m_i \in M_i\) and all \(i\). Denote by \(NE(e, \Gamma)\) the set of all Nash equilibria of the mechanism \(\Gamma\) for \(e \in E\).

It is clear every dominant strategy equilibrium is a Nash equilibrium, but the converse may not be true.

A mechanism \(\Gamma = < M, h >\) is said to strongly Nash-implement a social choice correspondence \(F\) on \(E\) if for every \(e \in E\)

(a) \(NE(e, \Gamma) \neq \emptyset\);

(b) \(h(NE(e, \Gamma)) \subseteq F(e)\).

It fully Nash implements a social choice correspondence \(F\) on \(E\) if for every \(e \in E\)

(a) \(NE(e, \Gamma) \neq \emptyset\)

(b) \(h(NE(e, \Gamma)) = F(e)\).
The following proposition shows that, if a truth-telling about their characteristics is a Nash equilibrium of the revelation mechanism, it must be a dominant strategy equilibrium of the mechanism.

**Proposition 18.8.1** For a revelation mechanism \( \Gamma = \langle E, h \rangle \), a truth-telling \( e^* \) is a Nash equilibrium if and only if it is a dominant equilibrium

\[
h(e^*_i, e_{-i}) \succ_i h(e_i, e_{-i}) \quad \forall (e_i, e_{-i}) \in E \land i \in N.
\] (18.28)

Proof. Since for every \( e \in E \) and \( i \), by Nash equilibrium, we have

\[
h(e_i, e_{-i}) \succ_i h(e'_i, e_{-i}) \quad \text{for all } e'_i \in E_i.
\]

Since this is true for any \( (e'_i, e_{-i}) \), it is a dominant strategy equilibrium. The proof is completed.

Thus, we cannot get any new results if one insists on the choice of revelation mechanisms. To get more satisfactory results, one must give up the revelation mechanism, and look for a more general mechanism with general message spaces.

Notice that, when one adopts the Nash equilibrium as a solution concept, implementation, rather than full or strong implement, in Nash solution may not be a useful requirement. To see this, consider any social choice correspondence \( F \) and the following mechanism: each individual’s message space consists of the set of economic environments, i.e., it is given by \( M_i = E \). The outcome function is defined as \( h(m) = a \in F(e) \) when all agents report the same economic environment \( m_i = e \), and otherwise it is seriously punished by giving a worse outcome. Then, it is clear the truth-telling is a Nash equilibrium. However, it has a lot of Nash equilibria, in fact infinity number of Nash equilibria. Any false reporting about the economic environment \( m_i = e' \) is also a Nash equilibrium. So, when we use Nash equilibrium as a solution concept, we need a social choice rule to be strongly implemented or full implemented in Nash equilibrium.
18.8.2 Characterization of Nash Implementation

Now we discuss what kind of social choice rules can be fully or strongly implemented through Nash incentive compatible mechanism. Maskin in 1977 gave necessary and sufficient conditions for a social choice rule to be fully Nash implementable (This paper was not published till 1999 due to the incorrectness of the original proof. It then appeared in Review of Economic Studies, 1999). Maskin’s result is fundamental since it not only helps us to understand what kind of social choice correspondence can be fully Nash implemented, but also gives basic techniques and methods in studying implementability of a social choice rule under other solution concepts.

Maskin’s monotonicity condition can be stated in two different ways although they are equivalent.

\[ F : E \rightarrow A \]

Figure 18.3: An illustration of Maskin’s monotonicity.

**Definition 18.8.1 (Maskin’s Monotonicity)** A social choice correspondence \( F : E \rightarrow A \) is said to be Maskin’s monotonic if for any \( e, \bar{e} \in E, x \in F(e) \) such
that for all $i$ and all $y \in A$, $x \succ_i y$ implies that $x \succ_i y$, then $x \in F(\bar{e})$.

In words, Maskin’s monotonicity requires that if an outcome $x$ is socially optimal with respect to economy $e$, changing economy $e$ to $\bar{e}$ makes all individuals even more like $x$, then $x$ remains socially optimal with respect to $\bar{e}$.

**Definition 18.8.2 (Another Version of Maskin’s Monotonicity)** A equivalent condition for a social choice correspondence $F : E \to A$ to be Maskin’s monotonic is that, if for any two economic environments $e, \bar{e} \in E$, $x \in F(e)$ such that $x \notin F(\bar{e})$, there is an agent $i$ and another $y \in A$ such that $x \succ_i y$ and $y \succ_i x$.

Maskin’s monotonicity is a reasonable condition, and many well known social choice rules satisfy this condition.

**Example 18.8.1 (Weak Pareto Efficiency)** The weak Pareto optimal correspondence $P_w : E \to A$ is Maskin’s monotonic.

Proof. If $x \in P_w(e)$, then for all $y \in A$, there exists $i \in N$ such that $x \succ_i y$. Now if for any $j \in N$ such that $x \succ_j y$ implies $x \succ_j y$, then we have $x \succ_i y$ for particular $i$. Thus, $x \in P_w(\bar{e})$.

**Example 18.8.2 (Majority Rule)** The majority rule or call the Condorcet correspondence $CON : E \to A$ for strict preference profile, which is defined by

$$CON(e) = \{x \in A : \#\{i | x \succ_i y\} \geq \#\{i | y \succ_i x\} \text{ for all } y \in A\}$$

is Maskin’s monotonic.

Proof. If $x \in CON(e)$, then for all $y \in A$,

$$\#\{i | x \succ_i y\} \geq \#\{i | y \succ_i x\}. \quad (18.29)$$

But if $\bar{e}$ is an economy such that, for all $i$, $x \succ_i y$ implies $x \succ_i y$, then the left-hand side of (18.29) cannot fall when we replace $e$ by $\bar{e}$. Furthermore, if the right-hand side rises, then we must have $x \succ_i y$ and $y \succ_i x$ for some $i$, a contradiction of the relation between $e$ and $\bar{e}$, given the strictness of preferences. So $x$ is still a majority winner with respect to $\bar{e}$, i.e., $x \in CON(\bar{e})$. 713
In addition to the above two examples, Walrasian correspondence and Lindahl correspondence with interior allocations are Maskin’s monotonic. The class of preferences that satisfy “single-crossing” property and individuals’ preferences over lotteries satisfy the von Neumann-Morgenstern axioms also automatically satisfy Maskin’ monotonicity.

The following theorem shows the Maskin’s monotonicity is a necessary condition for full Nash- implementability.

**Theorem 18.8.1** For a social choice correspondence \( F : E \rightarrow A \), if it is fully Nash implementable, then it must be Maskin’s monotonic.

**Proof.** For any two economies \( e, \bar{e} \in E, x \in F(e) \), then by full Nash implementability of \( F \), there is \( m \in M \) such that \( m \) is a Nash equilibrium and \( x = h(m) \). This means that \( h(m) \succ_i h(m_i', m_{-i}) \) for all \( i \) and \( m_i' \in M_i \). Given \( x \succ_i y \) implies \( x \succ_i y, h(m) \succ_i h(m_i', m_{-i}) \), which means that \( m \) is also a Nash equilibrium at \( \bar{e} \). Thus, by full Nash implementability again, we have \( x \in F(\bar{e}) \).

Maskin’s monotonicity itself cannot guarantee a social choice correspondence is fully Nash implementable. However, under the so-called no-veto power, it becomes sufficient.

**Definition 18.8.3 (No-Veto Power)** A social choice correspondence \( F : E \rightarrow A \) is said to satisfy no-veto power if whenever for any \( i \) and \( e \) such that \( x \succ_j y \) for all \( y \in A \) and all \( j \neq i \), then \( x \in F(e) \).

The no-veto power (NVP) condition implies that if \( n - 1 \) agents regard an outcome is the best to them, then it is social optimal. This is a rather weaker condition. NVP is satisfied by virtually all “standard” social choice rules, including weak Pareto efficient and Condorect correspondences. It is also often automatically satisfied by any social choice rules when the references are restricted. For example, for private goods economies with at least three agents, if each agent’s utility function is strong monotonic, then there is no other allocation such that \( n - 1 \) agents regard it best, so the no-veto power condition holds. The following
Theorem is given by Maskin (1977, 1999), but a complete proof of the theorem was due to Repullo (1987).

**Theorem 18.8.2** Under no-veto power, if Maskin’s monotonicity condition is satisfied, then $F$ is fully Nash implementable.

**Proof.** The proof is by construction. For each agent $i$, his message space is defined by

$$M_i = E \times A \times \mathcal{N}$$

where $\mathcal{N} = \{1, 2, \ldots, N\}$. Its elements are denoted by $m_i = (e^i, a^i, v^i)$, which means each agent $i$ announces an economic profile, an outcome, and a real number. Notice that we have used $e^i$ and $a^i$ to denote the economic profile of all individuals’ economic characteristics and the outcome announced by individual $i$, but not just agent $i$’s economic characteristic and outcome.

The outcome function is constructed in three rules:

Rule (1). If $m_1 = m_2 = \ldots = m_n = (e, a, v)$ and $a \in F(e)$, the outcome function is defined by

$$h(m) = a.$$  

In words, if players are unanimous in their strategy, and their proposed alternative $a$ is $F$-optimal, given their proposed profile $e$, the outcome is $a$.

Rule (2). For all $j \neq i$, $m_j = (e, a, v)$, $m_i = (e^i, a^i, v^i) \neq (e, a, v)$, and $a \in F(e)$, define:

$$h(m) = \begin{cases} 
  a^i & \text{if } a^i \in L(a, e^i) \\
  a & \text{if } a^i \notin L(a, e^i).
\end{cases}$$

where $L(a, e^i) = \{b \in A : a R_i b\}$ which is the lower contour set of $R_i$ at $a$. That is, suppose that all players but one play the same strategy and, given their proposed profile, their proposed alternative $a$ is $F$-optimal. Then, the outcome is selected from the worse one of $a$ and $a^i$ according to agent $i$’s preference $\succeq_i$. As such, no one can gain by deviating from a unanimous strategy.

Rule (3). If neither Rule (1) nor Rule (2) applies, then define

$$h(m) = a^{i^*}$$

715
where \( i^* = \max \{ i \in \mathcal{N} : v^i = \max_j v^j \} \). In other words, when neither Rule (1) nor Rule (2) applies, the outcome is the alternative proposed by player with the highest index among those whose proposed number is maximal.

Now we show that the mechanism \( \langle M, h \rangle \) defined above fully Nash implements social choice correspondence \( F \), i.e., \( h(N(e)) = F(e) \) for all \( e \in E \). We first show that \( F(e) \subset h(N(e)) \) for all \( e \in E \), i.e., we need to show that for all \( e \in E \) and \( a \in F(e) \), there exists a \( m \in M \) such that \( a = h(m) \) is a Nash equilibrium outcome. To do so, we only need to show that any \( m \) which is given by Rule (1) is a Nash equilibrium. Note that \( h(m) = a \) and for any given \( m'_i = (e'^i,a'^i,v'^i) \neq m_i \), by Rule (2), we have

\[
h(m'_i, m_{-i}) = \begin{cases} a^i & \text{if } a^i \in L(a, e_i) \\ a & \text{if } a^i \notin L(a, e_i). \end{cases}
\]

and thus

\[
h(m) R_i h(m'_i, m_{-i}) \forall m'_i \in M_i.
\]

Hence, \( m \) is a Nash equilibrium.

We now show that for each economic environment \( e \in E \), if \( m \) is a Nash equilibrium, then \( h(m) \in F(e) \). First, consider the Nash equilibrium \( m \) is given by Rule (1) so that \( a \in F(e) \), but the true economic profile is \( e' \), i.e., \( m \in NE(e', \Gamma) \) (if it is a Nash equilibrium at \( e \), then it is proved). We need to show \( a \in F(e') \). By Rule (1), \( h(m) = a \). Let \( b \in L(a, e_i) \), so the new message for \( i \) is \( m'_i = (e, b, v^i) \). Then \( h(m'_i, m_{-i}) = b \) by Rule (2). Now, since \( a \) is a Nash equilibrium outcome with respect to \( e' \), we have \( a = h(m) R'_i h(m'_i, m_{-i}) = b \). Thus, we have shown that for all \( i \in \mathcal{N} \) and \( b \in A \), \( a R_i b \) implies \( a R'_i b \). Thus, by Maskin’s monotonicity condition, we have \( a \in F(e') \).

Next, suppose Nash equilibrium \( m \) for \( e' \) is in the Case (2), i.e., for all \( j \neq i \), \( m_j = (e, a, v) \), \( m_i \neq (e, a, v) \). Let \( a' = h(m) \). By Case (3), each \( j \neq i \) can induce any outcome \( b \in A \) by choosing \( (R^j, b, v^j) \) with sufficiently a large \( v^j \) (which is greater than \( \max_{k \neq j} v_k \)), as the outcome at \( (m'_j, m_{-j}) \), i.e., \( b = h(m'_j, m_{-j}) \). Hence, \( m \) is a Nash equilibrium with respect to \( e' \) implies that for all \( j \neq i \), we
have

\[ a' \succ_j b. \]

Thus, by no-veto power assumption, we have \( a' \in F(e') \).

The same argument as the above, if \( m \) is a Nash equilibrium for \( e' \) is given by Case (3), we have \( a' \in F(e') \). The proof is completed.

Although Maskin’s monotonicity is very weak, it is violated by some social choice rules such as Solomon’s famous judgement. Solomon’s solution falls under full Nash implementation, since each woman knows who is the real mother. His solution, which consisted in threatening to cut the baby in two, is not entirely foolproof. What would be have done if the impostor has had the presence of mind to scream like a real mother? Solomon’s problem can be formerly described by the language of mechanism design as follows.

Two women: Anne and Bets

Two economies (states): \( E = \{\alpha, \beta\} \), where

\[ \begin{align*}
\alpha & : \text{Anne is the real mother} \\
\beta & : \text{Bets is the real mother}
\end{align*} \]

Solomon has three alternatives so that the feasible set is given by \( A = \{a, b, c\} \), where

\[ \begin{align*}
a & : \text{give the baby to Anne} \\
b & : \text{give the baby to Bets} \\
c & : \text{cut the baby in half.}
\end{align*} \]

Solomon’s desirability (social goal) is to give the baby to the real mother,

\[ \begin{align*}
f(\alpha) = a & \text{ if } \alpha \text{ happens} \\
f(\beta) = b & \text{ if } \beta \text{ happens}
\end{align*} \]

Preferences of Anne and Bets:
For Anne,

\begin{align*}
\text{at state } \alpha: \ &a \succ^A \alpha b \succ^A \alpha c \\
\text{at state } \beta: \ &a \succ^A \beta c \succ^A \beta b
\end{align*}

For Bets,

\begin{align*}
\text{at state } \alpha: \ &b \succ^B \alpha c \succ^B \alpha a \\
\text{at state } \beta: \ &b \succ^B \beta a \succ^B \beta c
\end{align*}

To see Solomon’s solution does not work, we only need to show his social choice goal is not Maskin’s monotonic. Notice that for Anne, since

\begin{align*}
a \succ^A \alpha b, c, \\
a \succ^A \beta b, c,
\end{align*}

and \( f(\alpha) = a \), by Maskin’s monotonicity, we should have \( f(\beta) = a \), but we actually have \( f(\beta) = b \). So Solomon’s social choice goal is not fully Nash implementable. As such, one needs to adopt a solution concept that is a refinement of Nash equilibrium so that the set of its equilibria is smaller.

### 18.9 Better Mechanism Design

Maskin’s theorem gives necessary and sufficient conditions for a social choice correspondence to be fully Nash implementable. However, due to the general nature of the social choice rules under consideration, the implementing mechanisms in proving characterization theorems turn out to be quite complex. Characterization results show what is possible for the full or strong implementation of a social choice rule, but not what is realistic. Thus, like most characterization results in the literature, Maskin’s mechanism is not natural in the sense that it is not continuous; small variations in an agent’s strategy choice may lead to large jumps in the resulting allocations, and further it has a message space of infinite dimension since it includes preference profiles as a component. In this section, we give some mechanisms that have some desired properties.
18.9.1 Groves-Ledyard Mechanism

Groves-Ledyard Mechanism (1977, Econometrica) was the first to give a specific mechanism that strongly Nash implements Pareto efficient allocations for public goods economies.

To show the basic structure of the Groves-Ledyard mechanism, consider a simplified Groves-Ledyard mechanism. Public goods economies under consideration have one private good \(x_i\), one public good \(y\), and three agents \((n = 3)\). The production function is given by \(y = v\).

The mechanism is defined as follows:

\[
M_i = R_i, \quad i = 1, 2, 3. \quad \text{Its elements, } m_i, \text{ can be interpreted as the proposed contribution (or tax) that agent } i \text{ is willing to make.}
\]

\[
t_i(m) = m_i^2 + 2m_jm_k: \text{ the actual contribution } t_i \text{ determined by the mechanism with the reported } m_i.
\]

\[
y(m) = (m_1 + m_2 + m_3)^2: \text{ the level of public good } y.
\]

\[
x_i(m) = w_i - t_i(m): \text{ the consumption of the private good.}
\]

Then the mechanism is balanced since

\[
\sum_{i=1}^{3} x_i + \sum_{i=1}^{3} t_i(m) = \sum_{i=1}^{3} x_i + (m_1 + m_2 + m_3)^2 = \sum_{i=3}^{n} x_i + y = \sum_{i=1}^{3} w_i
\]

The payoff function is given by

\[
v_i(m) = u_i(x_i(m), y(m)) = u_i(w_i - t_i(m), y(m)).
\]

To find a Nash equilibrium, we set

\[
\frac{\partial v_i(m)}{\partial m_i} = 0
\]

(18.30)
Then, \[ \frac{\partial v_i(m)}{\partial m_i} = \frac{\partial u_i}{\partial x_i}(-2m_i) + \frac{\partial u_i}{\partial y}2(m_1 + m_2 + m_3) = 0 \quad (18.31) \]

and thus \[ \frac{\partial u_i}{\partial y} = \frac{m_i}{m_1 + m_2 + m_3}. \quad (18.32) \]

When \( u_i \) are quasi-concave, the first order condition will be a sufficient condition for Nash equilibrium.

Making summation, we have at Nash equilibrium

\[ \sum_{i=1}^{3} \frac{\partial u_i}{\partial y} = \sum_{i=1}^{3} \frac{m_i}{m_1 + m_2 + m_3} = 1 = \frac{1}{f'(v)} \quad (18.33) \]

that is,

\[ \sum_{i=1}^{3} MRS_{yx_i} = MRTS_{yv}. \]

Thus, the Lindahl-Samuelson and balanced conditions hold which means every Nash equilibrium allocation is Pareto efficient.

They claimed that they have solved the free-rider problem in the presence of public goods. However, economic issues are usually complicated. Some agreed that they indeed solved the problem, some did not. There are two weakness of Groves-Ledyard Mechanism: (1) it is not individually rational: the payoff at a Nash equilibrium may be lower than at the initial endowment, and (2) it is not individually feasible: \( x_i(m) = w_i - t_i(m) \) may be negative.

How can we design the incentive mechanism to pursue Pareto efficient and individually rational allocations?

18.9.2 Walker’s Mechanism

Walker (1981, Econometrica) gave such a mechanism. Again, consider public goods economies with \( n \) agents, one private good, and one public good, and the production function is given by \( y = f(v) = v \). We assume \( n \geq 3 \).

The mechanism is defined by:
$M_i = R$

$g(m) = \sum_{i=1}^{n} m_i$: the level of public good.

$q_i(m) = \frac{1}{n} + m_{i+1} - m_{i+2}$: personalized price of public good.

$t_i(m) = q_i(m)y(m)$: the contribution (tax) made by agent $i$.

$x_i(m) = w_i - t_i(m) = w_i - q_i(m)y(m)$: the private good consumption.

Then, the budget constraint holds:

$$x_i(m) + q_i(m)y(m) = w_i \quad \forall m_i \in M_i$$  \hspace{1cm} (18.34)

Making summation, we have

$$\sum_{i=1}^{n} x_i(m) + \sum_{i=1}^{n} q_i(m)y(m) = \sum_{i=1}^{n} w_i$$

and thus

$$\sum_{i=1}^{n} x_i(m) + y(m) = \sum_{i=1}^{n} w_i$$

which means the mechanism is balanced.

The payoff function is

$$v_i(m) = u_i(x_i, y) = u_i(w_i - q_i(m)y(m), y(m))$$

The first order condition for interior allocations is given by

$$\frac{\partial v_i}{\partial m_i} = - \frac{\partial u_i}{\partial x_i} \left[ \frac{\partial q_i}{\partial m_i} y(m) + q_i(m) \frac{\partial y(m)}{\partial m_i} \right] + \frac{\partial u_i}{\partial y} \frac{\partial y(m)}{\partial m_i} = 0$$

$$\Rightarrow \frac{\partial u_i}{\partial y} = q_i(m)$$ \hspace{1cm} (FOC) for the Lindahl Allocation

$$\Rightarrow N(e) \subseteq L(e)$$

Thus, if $u_i$ are quasi-concave, it is also a sufficient condition for Lindahl equilibrium. We can also show every Lindahl allocation is a Nash equilibrium allocation, i.e.,

$$L(e) \subseteq N(e)$$  \hspace{1cm} (18.35)
Indeed, suppose \([(x^*, y^*), q_1^*, \ldots, q_n^*]\) is a Lindahl equilibrium. Let \(m^*\) be the solution of the following equation

\[
q_i^* = \frac{1}{n} + m_{i+1} - m_{i+2}, \\
y^* = \sum_{i=1}^{n} m_i
\]

Then we have \(x_i(m^*) = x_i^*, y(m^*) = y^*\) and \(q_i(m^*) = q_i^*\) for all \(i \in N\). Thus from \((x(m_i^*, m_{-i}), y(m_i^*, m_{-i})) \in \mathbb{R}_+^2\) and \(x_i(m_i^*, m_{-i}) + q_i(m^*) y(m_i^*, m_{-i}) = w_i\) for all \(i \in N\) and \(m_i \in M_i\), we have \((x_i(m^*), y(m^*)) R_i (x_i(m_i^*, m_{-i}), y(m_i^*, m_{-i}))\), which means that \(m^*\) is a Nash equilibrium.

Thus, Walker’s mechanism fully implements Lindahl allocations which are Pareto efficient and individually rational.

Walker’s mechanism also has a disadvantage that it is not feasible although it does solve the individual rationality problem. If a person claims large amounts of \(t_i\), consumption of private good may be negative, i.e., \(x_i = w_i - t_i < 0\). Tian proposed a mechanism that overcomes Walker’s mechanism’s problem. Tian’s mechanism is individually feasible, balanced, and continuous.

### 18.9.3 Tian’s Mechanism

In Tian’s mechanism (JET, 1991), everything is the same as Walker’s, except that \(y(m)\) is given by

\[
y(m) = \begin{cases} 
  a(m) & \text{if} \quad \sum_{i=1}^{n} m_i > a(m) \\
  \sum_{i=1}^{n} m_i & \text{if} \quad 0 \leq \sum_{i=1}^{n} m_i \leq a(m) \\
  0 & \text{if} \quad \sum_{i=1}^{n} m_i < 0
\end{cases}
\]

where \(a(m) = \min_{i \in N'(m)} \frac{w_i}{q_i(m)}\) that can be regarded as the feasible upper bound for having a feasible allocation. Here \(N'(m) = \{i \in N : q_i(m) > 0\}\).

An interpretation of this formulation is that if the total contributions that the agents are willing to pay were between zero and the feasible upper bound, the level of public good to be produced would be exactly the total taxes; if the total
taxes were less than zero, no public good would be produced; if the total taxes exceeded the feasible upper bound, the level of the public good would be equal to the feasible upper bound.

Figure 18.4: The feasible public good outcome function $Y(m)$.

To show this mechanism has all the nice properties, we need to assume that preferences are strongly monotonically increasing and convex, and further assume that every interior allocation is preferred to any boundary allocations: For all $i \in N$, $(x_i, y) \in \mathbb{R}^2_{++}$ and $(x'_i, y') \in \partial \mathbb{R}^2_+$, where $\partial \mathbb{R}^2_+$ is the boundary of $\mathbb{R}^2_+$.

To show the equivalence between Nash allocations and Lindahl allocations. We first prove the following lemmas.

**Lemma 18.9.1** If $(x(m^*), y(m^*)) \in N_{M,h}(e)$, then $(x_i(m^*), y(m^*)) \in \mathbb{R}^2_{++}$ for all $i \in N$.

**Proof:** We argue by contradiction. Suppose $(x_i(m^*), y(m^*)) \in \partial \mathbb{R}^2_+$. Then $x_i(m^*) = 0$ for some $i \in N$ or $y(m^*) = 0$. Consider the quadratic equation

$$y = \frac{w^*}{2(y + c)}, \quad (18.37)$$

723
where \( w^* = \min_{i \in N} w_i \); \( c = b + n \sum_{i=1}^{n} |m_i^*| \), where \( b = 1/n \). The larger root of (18.37) is positive and denoted by \( \tilde{y} \). Suppose that player \( i \) chooses his/her message \( m_i = \tilde{y} - \sum_{j \neq i} m_j^* \). Then \( \tilde{y} = m_i + \sum_{j \neq i} m_j^* > 0 \) and

\[
  w_j - q_j(m_i^*, m_{-i}) \tilde{y} \geq w_j - [b + (n \sum_{s=1}^{n} |m_s^*| + \tilde{y})] \tilde{y}
\]

\[
= w_j - (\tilde{y} + b + n \sum_{s=1}^{n} |m_s^*|) \tilde{y}
\]

\[
= w_j - w^*/2 \geq w_j/2 > 0
\]

(18.38)

for all \( j \in N \). Thus, \( y(m_i^*, m_{-i}) = \tilde{y} > 0 \) and \( x_j(m_i^*, m_{-i}) = w_j - q_j(m_i^*, m_{-i})y(m^*/m_i, i) = w_j - q_j(m_i^*, m_{-i})\tilde{y} > 0 \) for all \( j \in N \). Thus \( (x_i(m_i^*, m_{-i}), y(m_i^*, m_{-i})) \) is a Lindahl allocation with \( (x_i(m^*), y(m^*)) \) as the Lindahl price vector, which contradicts the hypothesis \( (x(m^*), y(m^*)) \) is an interior point of \( N_{M,h}(e) \). Q.E.D.

**Lemma 18.9.2** If \( (x(m^*), y(m^*)) \in N_{M,h}(e) \), then \( y(m^*) \) is an interior point of \( [0, a(m)] \) and thus \( y(m^*) = \sum_{i=1}^{n} m_i^* \).

**Proof:** By Lemma 18.9.1, \( y(m^*) > 0 \). So we only need to show \( y(m^*) < a(m^*) \). Suppose, by way of contradiction, that \( y(m^*) = a(m^*) \). Then \( x_j(m^*) = w_j - q_j(m^*)y(m^*) = w_j - q_j(m^*)a(m^*) = w_j - w_j = 0 \) for at least some \( j \in N \). But \( x(m^*) > 0 \) by Lemma 18.9.1, a contradiction. Q.E.D.

**Proposition 18.9.1** If the mechanism has a Nash equilibrium \( m^* \), then \( (x(m^*), y(m^*)) \) is a Lindahl allocation with \( (q_1(m^*), \ldots, q_n(m^*)) \) as the Lindahl price vector, i.e., \( N_{M,h}(e) \subseteq L(e) \).

**Proof:** Let \( m^* \) be a Nash equilibrium. Now we prove that \( (x(m^*), y(m^*)) \) is a Lindahl allocation with \( (q_1(m^*), \ldots, q_n(m^*)) \) as the Lindahl price vector. Since the mechanism is feasible and \( \sum_{i=1}^{n} q_i(m^*) = 1 \) as well as \( x_i(m^*) + q_i(m^*)y(m^*) = w_i \) for all \( i \in N \), we only need to show that each individual is maximizing his/her preference. Suppose, by way of contradiction, that there is some \( (x, y) \in \mathbb{R}^2_+ \) such that \( (x, y) \) is a Lindahl allocation with \( (X_i(m^*), Y(m^*)) \) and \( x_i + q_i(m^*)y \leq w_i \). Because of monotonicity of
preferences, it will be enough to confine ourselves to the case of \(x_i + q_i(m^*) y = w_i\).
Let \((x_{i\lambda}, y_{\lambda}) = (\lambda x_i + (1 - \lambda)x_i(m^*), \lambda y + (1 - \lambda)y(m^*))\). Then by convexity of preferences we have \((x_{i\lambda}, y_{\lambda}) P_i (x_i(m^*), y(m^*))\) for any \(0 < \lambda < 1\). Also \((x_{i\lambda}, y_{\lambda}) \in \mathbb{R}_+^2\) and \(x_{i\lambda} + q_i(m^*) y_{\lambda} = w_i\).

Suppose that player \(i\) chooses his/her message \(m_i = y_{\lambda} - \sum_{j \neq i}^n m_j^*\). Since \(y(m^*) = \sum_{j = 1}^n m_j^*\) by Lemma 18.9.2, \(m_i = y_{\lambda} - y(m^*) + m_j^*\). Thus as \(\lambda \to 0\), \(y_{\lambda} \to y(m^*)\), and therefore \(m_i \to m_i^*\). Since \(x_j(m^*) = w_j - q_j(m^*) y(m^*) > 0\) for all \(j \in N\) by Lemma 18.9.1, we have \(w_j - q_j(m_i^*, m_{-i}) y_{\lambda} > 0\) for all \(j \in N\) as \(\lambda\) is a sufficiently small positive number. Therefore, \(y(m_i^*, m_{-i}) = y_{\lambda}\) and \(x_i(m_i^*, m_{-i}) = w_i - q_i(m^*) y(m_i^*, m_{-i}) = w_i - q_i(m^*) y_{\lambda} = x_{i\lambda}\). From \((x_{i\lambda}, y_{\lambda}) P_i (x_i(m^*), y(m^*))\), we have \((x_i(m_i^*, m_{-i}), y(m_i^*, m_{-i})) P_i (x_i(m^*), y(m^*))\). This contradicts \((x(m^*), y(m^*)) \in N_{M,h}(e)\). Q.E.D.

**Proposition 18.9.2** If \((x^*, y^*)\) is a Lindahl allocation with the Lindahl price vector \(q^* = (q^*_1, \ldots, q^*_n)\), then there is a Nash equilibrium \(m^*\) of the mechanism such that \(x_i(m^*) = x_i^*, \quad q_i(m^*) = q_i^*\), for all \(i \in N\), \(y(m^*) = y^*\), i.e., \(L(e) \subseteq N_{M,h}(e)\).

Proof: We need to show that there is a message \(m^*\) such that \((x^*, y^*)\) is a Nash allocation. Let \(m^*\) be the solution of the following linear equations system:

\[
q^*_i = \frac{1}{n} + m_{i+1} - m_{i+2},
\]

\[
y^* = \sum_{i=1}^n m_i
\]

Then we have \(x_i(m^*) = x_i^*, \ y(m^*) = y^*\) and \(q_i(m^*) = q_i^*\) for all \(i \in N\). Thus from \((x(m_i^*, m_{-i}), y(m_i^*, m_{-i})) \in \mathbb{R}_+^2\) and \(x_i(m_i^*, m_{-i}) + q_i(m^*) y(m_i^*, m_{-i}) = w_i\) for all \(i \in N\) and \(m_i \in M_i\), we have \((x_i(m^*), y(m^*)) R_i (x_i(m_i^*, m_{-i}), y(m_i^*, m_{-i}))\). Q.E.D.

Thus, Tian’s mechanism fully Nash implements Lindahl allocations.
18.9.4 Informational Efficiency and Uniqueness of Competitive Market Mechanism

Consider production economies with $L$ private goods, $I$ consumers and $J$ firms so that total number of agents is $n := I + J$. Consumer $i$’s characteristic is given by $e_i = (X_i, w_i, R_i)$, where $X_i \subset \mathbb{R}^L$, $w_i \in \mathbb{R}^L_+$, and $R_i$ is convex$^1$, continuous on $X_i$, and strictly monotone on the set of interior points of $X_i$. Producer $j$’s characteristic is given by $e_j = (Y_j)$. We assume that, for $j = I + 1, \ldots, n$, $Y_j$ is nonempty, closed, convex, contains 0 (possibility of inaction), and $Y_j - \mathbb{R}^L_+ \subseteq Y_j$ (free-disposal). We also assume that the economies under consideration have no externalities or public goods.

An economy is the full vector $e = (e_1, \ldots, e_I, e_{I+1}, \ldots, e_N)$ and the set of all such production economies is denoted by $E$ and is called neoclassical production economies. $E$ is assumed to be endowed with the product topology.

Let $x_i$ denote the net increment in commodity holdings (net trade) by consumer $i$ and $y_j$ producer $j$’s (net) output vector. Denoted by $x = (x_1, \ldots, x_I)$ and $y = (y_{I+1}, \ldots, y_N)$.

An allocation of the economy $e$ is a vector $z := (x, y) \in \mathbb{R}^{NL}$. An allocation $z = (x, y)$ is said to be individually feasible if $x_i + w_i \in X_i$ for $i = 1, \ldots, I$, and $y_j \in Y_j$ for $j = I + 1, \ldots, n$. An allocation $z = (x, y)$ is said to be balanced if $\sum_{i=1}^I x_i = \sum_{j=I+1}^N y_j$. An allocation $z = (x, y)$ is said to be feasible if it is balanced and individually feasible for every individual.

An allocation $z = (x, y)$ is said to be Pareto efficient if it is feasible and there does not exist another feasible allocation $z' = (x', y')$ such that $(x'_i + w_i)R_i(x_i + w_i)$ for all $i = 1, \ldots, I$ and $(x'_i + w_i)R_i(x_i + w_i)$ for some $i = 1, \ldots, I$. Denote by $P(e)$ the set of all such allocations.

An important characterization of a Pareto optimal allocation is associated with the following concept. Let $\Delta^{L-1}_+ = \{ p \in \mathbb{R}^{L-1}_+ : \sum_{i=1}^L p^i = 1 \}$ be the $L - 1$ $^1R_i$ is convex if for bundles $a, b, c$ with $0 < \lambda \leq 1$ and $c = \lambda a + (1 - \lambda)b$, the relation $a P_i b$ implies $c P_i b$. Note that the term “convex” is defined as in Debreu (1959), not as in some recent textbooks.
dimensional unit simplex.

A nonzero vector \( p \in \Delta^{I-1} \) is called a vector of efficiency prices for a Pareto optimal allocation \((x, y)\) if

(a) \( p \cdot x_i \leq p \cdot x'_i \) for all \( i = 1, \ldots, I \) and all \( x'_i + w_i \in X_i \)
and \((x'_i + w_i) R_i (x_i + w_i)\);

(b) \( p \cdot y_j \geq p \cdot y'_j \) for all \( y'_j \in Y_j, j = I + 1, \ldots, N \).

In the terminology of Debreu (1959, p. 93), \((x, y)\) is an equilibrium relative to the price system \( p \). It is well known that under certain regularity conditions such as convexity, local non-satiation, etc, every Pareto optimal allocation \((x, y)\) has an efficiency price associated with it as shown in Second Theorem of Welfare Economics in Chapter 11.

We also want a mechanism is individually rational. However, as Hurwicz (1979b) pointed out, it is not quite obvious what the appropriate generalization of the individual rationality concept should be for an economy with production. The following definition of individual rationality of an allocation for an economy with production was introduced by Hurwicz (1979b).

An allocation \( z = (x, y) \) is said to be individually rational with respect to the fixed share guarantee structure \( \gamma_i(e; \theta) \) if \((x_i + w_i) R_i (\gamma_i(e) + w_i)\) for all \( i = 1, \ldots, I \). Here, \( \gamma_i(e; \theta) \) is given by

\[
\gamma_i(e; \theta) = \frac{p \cdot \sum_{j=I+1}^N \theta_{ij} y_j}{p \cdot w_i}, \quad i = 1, \ldots, I,
\]

where \( p \) is an efficiency price vector for \( e \) and the \( \theta_{ij} \) are non-negative fractions such that \( \sum_{i=1}^n \theta_{ij} = 1 \) for \( j = I + 1, \ldots, N \), which can be interpreted as the profit shares of consumer \( i \) from producer \( j \). Note that this definition on the individual rationality contains pure exchange as well as constant returns as special cases. Denote by \( I_\theta(e) \) the set of all such allocations.

Now we define the competitive equilibria of a private ownership economy in which the \( i \)-th consumer owns the share \( \theta_{ij} \) of the \( j \)-th producer, and is, consequently, entitled to the corresponding fraction of its profits. Thus, the ownership
structure can be denoted by the matrix $\theta = (\theta_{ij})$. Denoted by $\Theta$ the set of all such ownership structures.

An allocation $z = (x, y) = (x_1, x_2, \ldots, x_I, y_{I+1}, y_{I+2}, \ldots, y_N) \in \mathbb{R}^{IL} \times \mathcal{Y}$ is a $\theta$-Walrasian allocation for an economy $e$ if it is feasible and there is a price vector $p \in \Delta^{L-1}$ such that

1. $p \cdot x_i = \sum_{j=I+1}^{N} \theta_{ij} p \cdot y_j$ for all $i = 1, \ldots, I$;
2. for all $i = 1, \ldots, I$, $(x'_i + w_i) P_i (x_i + w_i)$ implies $p \cdot x'_i > \sum_{j=I+1}^{N} \theta_{ij} p \cdot y_j$; and
3. $p \cdot y_j \geq p \cdot y'_j$ for all $y'_j \in \mathcal{Y}_j$ and $j = I + 1, \ldots, N$.

Denoted by $W_\theta(e)$ the set of all such allocations, and by $W_\theta(e)$ the set of all such price-allocations pair $(p, z)$.

Let $E^c \subset E$ be the subset of production economies on which $W(e) \neq \emptyset$ for all $e \in E^c$ and call such a subset as the Walrasian production economies.

It may be remarked that, every $\theta$-Walrasian allocation is clearly individually rational with respect to the $\gamma_i(e; \theta)$, and also, by the strong monotonicity of preferences, it is Pareto efficient. Thus we have $W_\theta(e) \subset I_\theta(e) \cap P(e)$ for all $e \in E^c$.

We now define the Walrasian process that is a privacy-preserving process and realizes the Walrasian correspondence $W_\theta$, and in which messages consist of prices and trades of all agents. In defining the Walrasian process, it is assumed that the private ownership structure matrix $\theta$ is common knowledge for all the agents.

Define the excess demand correspondence of consumer $i$ ($i = 1, \ldots, I$) $D_i : \Delta^{L-1} \times \Theta \times \mathbb{R}^{I} \times E_i \rightarrow \mathbb{R}^L$ by

$$D_i(p, \theta, \pi_{I+1}, \ldots, \pi_N, e_i) = \{ x_i : x_i + w_i \in X_i, p \cdot x_i = \sum_{j=I+1}^{N} \theta_{ij} \pi_j \}$$

$$(x'_i + w_i) P_i (x_i + w_i) \text{ implies } p \cdot x'_i > \sum_{j=I+1}^{N} \theta_{ij} \pi_j$$

where $\pi_j$ is the profit of firm $j$ ($j = I + 1, \ldots, N$).
Define the supply correspondence of firm \( j \) \( (j = I + 1, \ldots, N) \) \( S_j : \Delta^{L-1} \times E_j \rightarrow \mathbb{R}^L \) by

\[
S_j(p, e_j) = \{ y_j : y_j \in \mathcal{Y}_j, p \cdot y_j \geq p \cdot y'_j \forall y'_j \in \mathcal{Y}_j \}. \tag{18.41}
\]

Note that \((p, x, y)\) is a \( \theta \)-Walrasian (competitive) equilibrium for economy \( e \) with the private ownership structure \( \theta \) if \( p \in \Delta^{L-1}, x_i \in D_i(p, \theta, p \cdot y_{I+1}, \ldots, p \cdot y_N) \) for \( i = 1, \ldots, I \), \( y_j \in S_j(p, e_j) \) for \( j = I = 1, \ldots, N \), and the allocation \((x, y)\) is balanced.

The Walrasian (competitive) process \( \langle M_c, \mu_c, h_c \rangle \) is defined as follows.

Define \( M_c = \Delta^{L-1} \times Z \).

Define \( \mu_c : E \rightarrow M_c \) by

\[
\mu_c(e) = \cap_{i=1}^N \mu_{ci}(e_i), \tag{18.42}
\]

where \( \mu_{ci} : E_i \rightarrow M_c \) is defined as follows:

(1) For \( i = 1, \ldots, I \), \( \mu_{ci}(e_i) = \{ (p, x, y) : p \in \Delta^{L-1}, x_i \in D_i(p, \theta, p \cdot y_{I+1}, \ldots, p \cdot y_N, e_i) \) and \( \sum_{i=1}^I x_i = \sum_{j=I+1}^N y_j \} \).

(2) For \( i = I + 1, \ldots, N \), \( \mu_{ci}(e_i) = \{ (p, x, y) : p \in \Delta^{L-1}, y_i \in S_i(p, e_i) \) and \( \sum_{i=1}^I x_i = \sum_{j=I+1}^N y_j \} \).

Thus, we have \( \mu_c(e) = \mathcal{W}_e(e) \) for all \( e \in E \).

Finally, the Walrasian outcome function \( h_c : M_c \rightarrow Z \) is defined by

\[
h_c(p, x, y) = (x, y), \tag{18.43}
\]

which is an element in \( \mathcal{W}_e(e) \).

The Walrasian process can be viewed as a formalization of resource allocation, simulating the competitive mechanism, which is non-wasteful and individually rational with respect to the fixed share guarantee structure \( \gamma_i(e; \theta) \). The competitive message process is privacy-preserving by the construction of the Walrasian process.
Remark 18.9.1 For a given private ownership structure matrix $\theta$, since an element, $m = (p, x_1, \ldots, x_I, y_{I+1}, \ldots, y_N) \in \mathbb{R}^L_++ \times \mathbb{R}^{(N)L}$, of the Walrasian message space $M_c$ satisfies the conditions $\sum_{l=1}^L p^l = 1$, $\sum_{i=1}^I x_i = \sum_{j=I+1}^N y_j$, and $p \cdot x_i = \sum_{j=I+1}^N \theta_{ij} p \cdot y_j$ ($i = 1, \ldots, I$) and one of these equations is not independent by Walras Law, any Walrasian message is contained within a Euclidean space of dimension $(L + IL + JL) - (1 + L + I) + 1 = (L - 1)I + LJ$ and thus, an upper bound on the Euclidean dimension of $M_c$ is $(L - 1)I + LJ$.

To establish the informational efficiency of the competitive mechanism, we consider a special class of economies, denoted by $E^{cf} = \prod_{i=1}^N E^{cf}_i$, where preference orderings are characterized by Cobb-Douglas utility functions, and efficient production technology are characterized by quadratic functions.

For $i = 1, \ldots, I$, consumer $i$’s admissible economic characteristics in $E^{cf}_i$ are given by the set of all $e_i = (x_i, w_i, R_i)$ such that $x_i = \mathbb{R}^L_+$, $w_i > 0$, and $R_i$ is represented by a Cobb-Douglas utility function $u(\cdot, a_i)$ with $a_i \in \Delta^{L-1}$ such that $u(x_i + w_i, a_i) = \prod_{l=1}^L (x_i^l + w_i^l)^{a_l}$.

For $i = I+1, \ldots, N$, producer $i$’s admissible economic characteristics are given by the set of all $e_i = Y_i = Y(b_i)$ such that

$$Y(b_i) = \{ y_i \in \mathbb{R}^L : b_1^i y_i^1 + \sum_{l=2}^L \left( y_l^l + \frac{b_l^i}{2} (y_l^l)^2 \right) \leq 0, \quad -\frac{1}{b_l^i} \leq y_l^i \leq 0 \text{ for all } l \neq 1 \},$$

where $b_i = (b_1^i, \ldots, b_L^i)$ with $b_l^i > \frac{1}{w_i^l}$. It is clear that any economy in $E^{cf}$ is fully specified by the parameters $a = (a_1, \ldots, a_I)$ and $b = (b_{I+1}, \ldots, b_N)$. Furthermore, production sets are nonempty, closed, and convex by noting that $0 \in Y(b_j)$ and their efficient points are represented by quadratic production functions in which $(y_2^i, \ldots, y_L^i)$ are inputs and $y_1^i$ is possibly an output.

Given an initial endowment $\bar{w} \in \mathbb{R}^{LI}_+$, define a subset $\bar{E}^{cf}$ of $E^{cf}$ by $\bar{E}^{cf} = \{ e \in E^{cd} : w_i = \bar{w}_i \forall i = 1, \ldots, I \}$. That is, endowments are constant over $\bar{E}^{cf}$. Let $E^c$ be the set of all production economies in which Walrasian equilibrium exists.
We then have the following theorem that shows that the Walrasian mechanism is informationally efficient among all smooth resource allocation mechanisms that are informationally decentralized and non-wasteful over the set $E^c$.

**Theorem 18.9.1** (Informational Efficiency Theorem) Then the Walrasian allocation mechanism $⟨M_c, µ_c, h_c⟩$ is informationally efficient among all allocation mechanisms $⟨M, µ, h⟩$ defined on $E^c$ that

(i) are informationally decentralized;

(ii) are non-wasteful with respect to $P$;

(iii) have Hausdorff topological message spaces;

(iv) satisfy the local threadedness property at some point $e ∈ E^cq$.

That is, $M_c = FR^{(L-1)I+LJ} \subseteq_F M$.

This theorem establishes the informational efficiency of the competitive mechanism within the class of all smooth resource allocation mechanisms which are informationally decentralized and non-wasteful over the class of Walrasian production economies $E^c$.

The following theorem further shows that the competitive allocation process is the only most informationally efficient decentralized mechanism for production economies among mechanisms that achieve Pareto optimal and individually rational allocations.

**Theorem 18.9.2** Suppose that $⟨M, µ, h⟩$ is an allocation mechanism on the class of production economies $E^cq$ such that:

(i) it is informationally decentralized;

(ii) it is non-wasteful with respect to $P$;

---

A stationary message correspondence $µ$ is said to be locally threaded at $e ∈ E$ if there is a neighborhood $N(e) ⊂ E$ and a continuous function $f : N(e) \rightarrow M$ such that $f(e') ∈ µ(e')$ for all $e' ∈ N(e)$. The stationary message correspondence $µ$ is said to be locally threaded on $E$ if it is locally threaded at every $e ∈ E$. 

731
(iii) it is individually rational with respect to the fixed share guarantee structure \( \gamma_i(e; \theta) \);

(iv) \( M \) is a \((L-1)I + LJ\) dimensional manifold;

(v) \( \mu \) is a continuous function on \( E^{eq} \).

Then, there is a homeomorphism \( \phi \) on \( \mu(E^{eq}) \) to \( M_c \) such that

(a) \( \mu_c = \phi \cdot \mu \);

(b) \( h_c \cdot \phi = h \).

The conclusion of the theorem is summarized in the following commutative homeomorphism diagram:

\[
\begin{array}{ccc}
E & \xrightarrow{\mu_c} & M_c \\
\downarrow{\mu} & \phi & \downarrow{h_c} \\
\mu(E) & \xrightarrow{h} & Z \\
\end{array}
\]

Thus, the above theorem shows that the competitive mechanism is the unique informationally efficient process that realizes Pareto optimal and individually rational allocations over the class of production economies \( E^c \). For a mechanism \( \langle M, \mu, h \rangle \), when there does not any homeomorphism \( \phi \) on \( \mu(E^{eq}) \) to \( M_c \) such that

(1) \( \mu_L = \phi \cdot \mu \) and

(2) \( h_L \cdot \phi = h \),

we may call such a mechanism \( \langle M, \mu, h \rangle \) a non-Walrasian allocation mechanism. The Uniqueness Theorem then implies that any non-Walrasian allocation mechanism defined on \( E^{eq} \) must use a larger message space.

For public goods economies, Tian (2000e) obtains the similar results, and shows that Lindahl mechanism is the unique informationally efficient process that realizes Pareto optimal and individually rational allocations over the class of production economies that guarantees the existence of Lindahl equilibrium.

### 18.10 Reference

Books and Monographs


**Papers**


Reichelstein, S., and S. Reiter (1988), ”Game forms with minimal message s-

Repullo, R. (1987). “A Simple Proof of Maskin’s Theorem on Nash Implemen-

Saijo, T. (1987), ”On constant Maskin monotonic social choice functions”, *Jour-

56: 693-700.

Samuelson, PA (1954),”The pure theory of public expenditure”, *The review of

Samuelson, PA (1955),”Diagrammatic Exposition of a Theory of Public Expen-
diture”, *The review of economics and statistics*,37, 350-356.

Satterthwaite, M.A. (1975), ”Strategy-proofness and Arrow’s conditions: exis-
tence and correspondence theorems for voting procedures and social welfare


*Journal of Public Economics*, 134, 35-41

*Games and Economic Behavior*, forthcoming.

Suh, S.C. (1997), ”Double implementation in Nash and strong Nash equilibria”,


739


Chapter 19

Incomplete Information and Bayesian-Nash Implementation

19.1 Introduction

Nash implementation has imposed a strong assumption on information requirement. Although the designer does not know information about agents’ characteristics, Nash equilibrium assume each agent knows characteristics of the others. This assumption is hardly satisfied in many cases in the real world. Can we remove this assumption? The answer is positive. One can use the Bayesian-Nash equilibrium, introduced by Harsanyi, as a solution concept to describe individuals’ self-interested behavior. Although each agent does not know economic characteristics of the others, he knows the probability distribution of economic characteristics of the others. In this case, we can still design an incentive compatible mechanism.
19.2 Basic Analytical Framework

19.2.1 Model

Throughout we follow the notation introduced in the previous chapter. Let $Z$ denote the set of outcomes, $A \subseteq Z$ the set of feasible outcomes, and $\Theta_i$ agent $i$’s space of types. Unless indicted explicitly, for simplicity, we assume each individual’s preferences are given by parametric utility functions with private values, denoted by $u_i(x, \theta_i)$, where $x \in Z$ and $\theta_i \in \Theta_i$. Assume that all agents and the designer know that the vector of types, $\theta = (\theta_1, \ldots, \theta_n)$ is distributed according to $\varphi(\theta)$ a priori on a set $\Theta$.

Each agent knows his own type $\theta_i$, and therefore computes the conditional distribution of the types of the other agents:

$$\varphi(\theta_{-i} | \theta_i) = \frac{\varphi(\theta_i, \theta_{-i})}{\int_{\Theta_{-i} \varphi(\theta_i, \theta_{-i})} d\theta_{-i}}.$$ (19.1)

As usual, a mechanism is a pair, $\Gamma = \langle M, h \rangle$. Given $m$ with $m_i : \Theta_i \to M_i$, agent $i$’s expected utility at $\theta_i$ is given by

$$\Pi_{\langle M, h \rangle}(m; \theta_i) = E_{\theta_{-i}}[u_i(h(m_i(\theta_i), m_{-i}(\theta_{-i})), \theta_i) | \theta_i] = \int_{\Theta_{-i}} u_i(h(m(\theta)), \theta_i) \varphi(\theta_{-i} | \theta_i) d\theta_{-i}. \quad (19.1)$$

A strategy $m^*(\cdot)$ is a Bayesian-Nash equilibrium (BNE) of $\langle M, h \rangle$ if for all $\theta_i \in \Theta_i$,

$$\Pi^i(m^*(\theta_i); \theta_i) \geq \Pi^i(m_i, m^*_{-i}(\theta_{-i}); \theta_i) \quad \forall m_i \in M_i.$$ That is, if player $i$ believes that other players are using strategies $m^*_{-i}(\cdot)$ then he maximizes his expected utility by using strategy $m_i^*(\cdot)$. Denote by $B(\theta, \Gamma)$ the set of all Bayesian-Nash equilibria of the mechanism.

Remark 19.2.1 In the present private value incomplete information setting, a message $m_i : \Theta_i \to M_i$ is a dominant strategy strategy for agent $i$ in mechanism $\Gamma = \langle M, h \rangle$ if for all $\theta_i \in \Theta_i$ and all possible strategies $m_{-i}(\theta_{-i})$,

$$E_{\theta_{-i}}[u_i(h(m_i(\theta_i), m_{-i}(\theta_{-i})), \theta_i) | \theta_i] \geq E_{\theta_{-i}}[u_i(h(m_i^*(\theta_i), m_{-i}(\theta_{-i})), \theta_i) | \theta_i]. \quad (19.2)$$
for all $m_i' \in M_i$.

Since condition (19.2) holding for all $\theta_i$ and all $m_{-i}(\theta_{-i})$ is equivalent to the condition that, for all $\theta_i \in \Theta_i$,

$$u_i(h(m_i(\theta_i), m_{-i}), \theta_i) \geq u_i(h(m_i', m_{-i}), \theta_i)$$  \hspace{1cm} (19.3)

for all $m_i' \in M_i$ and all $m_{-i} \in M_{-i}$, this leads the following definition of dominant equilibrium.

Thus, for such a private value model, a message $m_i^*$ is a dominant strategy equilibrium of a mechanism $\Gamma = \langle M, h \rangle$ if for all $i$ and all $\theta_i \in \Theta_i$,

$$u_i(h(m_i^*(\theta_i), m_{-i}), \theta_i) \geq u_i(h(m_i'(\theta_i), m_{-i}), \theta_i)$$

for all $m_i' \in M_i$ and all $m_{-i} \in M_{-i}$, which is the same as that in the case of complete information.

**Remark 19.2.2** It is clear every dominant strategy equilibrium is a Bayesian-Nash equilibrium, but the converse may not be true. Bayesian-Nash equilibrium requires more sophistication from the agents than dominant strategy equilibrium. Each agent, in order to find his optimal strategy, must have a correct prior $\varphi(\cdot)$ over states, and must correctly predict the equilibrium strategies used by other agents.

### 19.2.2 Bayesian Incentive-Compatibility and Bayesian Implementation

To see specifically what is implementable in Bayesian-Nash equilibrium (BNE), till the last two sections, the social choice goal is given by a single-valued social choice function rather than the set of social choice functions. We focuses on the issues of Bayesian Incentive-Compatibility and Bayesian Implementation of a
social choice function. Furthermore, by Revelation Principle below, without loss of generality, we can focus on direct revelation mechanisms.

Let \( f : \Theta \rightarrow A \) be a social choice function. Since at Bayesian-Nash equilibrium, every agent reaches his interim utility maximization, Bayesian implementation sometimes is called interim implementation. Like implementation in dominant strategy and Nash strategy, we similarly have the strong Bayesian-Nash implementation and Baysian-Nash implementation.

**Definition 19.2.1** A mechanism \( \Gamma = < M, h > \) is said to strongly Bayesian implement a social choice function \( f \) on \( \Theta \) if for all Bayesian-Nash equilibrium \( m^* \), we have \( h(m^*(\theta)) = f(\theta), \forall \theta \in \Theta \). If such a mechanism exists, we call \( f \) is strongly Bayesian implementable.

When there are multiple Bayesian-Nash equilibria, it may result in some undesirable equilibrium outcomes, i.e., it is not equal to \( f \) (we will provide such an example. Then, we have the following weaker concept of Bayesian implementability.

**Definition 19.2.2** A mechanism \( \Gamma = < M, h > \) is said to Bayesian implement a social choice function \( f \) on \( \Theta \) if there exists a Bayesian-Nash equilibrium \( m^* \), such that \( h(m^*(\theta)) = f(\theta), \forall \theta \in \Theta \). If such a mechanism exists, we call \( f \) is Bayesian implementable.

Like dominant implementation, we will see that a social choice function is Bayesian implementable if and only if it is truthfully Bayesian implementable.

**Definition 19.2.3** We call a social choice function \( f \) is truthfully Bayesian implementable or Bayesian incentive compatible, if truth-telling: \( m^*(\theta) = \theta, \forall \theta \in \Theta \) is a Bayesian-Nash equilibrium of revelation mechanism \( \Gamma = (\Theta, f) \), i.e.,

\[
E_{\theta_{-i}}[u_i(f(\theta_i, \theta_{-i})), \theta_i] \geq E_{\theta_{-i}}[u_i(f(\theta'_{i}, \theta_{-i})), \theta_i], \quad \forall i, \forall \theta_{i}, \theta'_{i} \in \Theta_{i}.
\]
**Bayesian incentive-compatibility** means that for social choice rule \( f \), every agent will report his type truthfully provided all other agents are employing their truthful strategies and thus every truthful strategy profile is a Bayesian equilibrium of the direct mechanism \((\Theta, f)\). Notice that Bayesian incentive compatibility does not say what is the best response of an agent when other agents are not using truthful strategies. When a mechanism has multiple equilibria, non-truth-telling may be a Bayesian-Nash equilibrium, resulting in an undesirable outcome that is not equal to \( f \). When Bayesian incentive compatibility is just a basic requirement, it in general only ensures a social choice rule is truthfully Bayesian implementable but not strongly Bayesian implementable, i.e., it may also contain some undesirable equilibrium outcomes when a mechanism has multiple equilibria.

Similarly, we have the following revelation principle.

**Proposition 19.2.1 (Revelation Principle)** A social choice rule \( f(\cdot) \) is implementable in Bayesian-Nash equilibrium (in short, BNE) if and only if it is truthfully implementable in BNE.

Proof. The proof is the same as before: Suppose that there exists a mechanism \( \Gamma = (M_1, \cdots, M_n, g(\cdot)) \) and an equilibrium strategy profile \( m^*(\cdot) = (m_1^*(\cdot), \cdots, m_n^*(\cdot)) \) such that \( g(m^*(\cdot)) = f(\cdot) \) and \( \forall i, \forall \theta_i \in \Theta_i \). We then have
\[
E_{\theta_i}[u_i(g(m_1^*(\theta), m_{-i}^*(\theta_{-i})), \theta_i) \vert \theta_i] \geq E_{\theta_i}[u_i(g(m'_1(\hat{\theta}), m_{-i}^*(\theta_{-i})), \theta_i) \vert \theta_i], \forall m'_1 \in M_1.
\]

One way to deviate for agent \( i \) is by pretending that his type is \( \hat{\theta}_i \) rather than \( \theta_i \), i.e., sending message \( m'_i = m_i^*(\hat{\theta}_i) \). This gives
\[
E_{\theta_i}[u_i(g(m_1^*(\theta), m_{-i}^*(\theta_{-i})), \theta_i) \vert \theta_i] \geq E_{\theta_i}[u_i(g(m'_1(\hat{\theta}), m_{-i}^*(\theta_{-i})), \theta_i) \vert \theta_i], \forall \hat{\theta}_i \in \Theta_i.
\]

But since \( g(m^*(\theta)) = f(\theta), \forall \theta \in \Theta \), we must have \( \forall i, \forall \theta_i \in \Theta_i \), there is
\[
E_{\theta_i}[u_i(f(\theta, \theta_{-i}), \theta_i) \vert \theta_i] \geq E_{\theta_i}[u_i(f(\hat{\theta}_i, \theta_{-i}), \theta_i) \vert \theta_i], \forall \hat{\theta}_i \in \Theta_i.
\]

Although Bayesian incentive-compatibility is a necessary and sufficient for \( f \) to be Bayesian implementable, it is not a sufficient condition for \( f \) to be strongly Bayesian implementable. Strong Bayesian implementability requires all BNE
outcomes are equal to $f(\theta)$. The goal for designing a mechanism is to reach a desirable equilibrium outcome, but it may also result in an undesirable equilibrium outcome. Thus, while the incentive compatibility requirement is central, it may not be sufficient for a mechanism to give all of desirable outcomes. The severity of this multiple equilibrium problem has been exemplified by Demski and Sappington (JET, 1984), Postlewaite and Schmeidler (JET, 1986), Repullo (Econometrica, 1988), and others. The implementation problem involves designing mechanisms to ensure that all equilibria result in desirable outcomes.

Under what conditions, does there exists a mechanism that has a unique Bayesian incentive-compatible equilibrium outcome? If such a mechanism exists, by definition, we know a social choice function $f$ is strongly Bayesian implementable. One such condition is the nonexistence of undominated Bayesian equilibrium.

**Definition 19.2.4** A message $m \in M$ is weakly dominated, if there is an agent $i$, $\theta_i$, and another Bayesian message $m_i'(\theta_i) \neq m_i$, such that for all $m_{-i} \in M_{-i}$,

$$\Pi^i(m_i(\theta_i), m_{-i}; \theta_i) \leq \Pi^i(m_i'(\theta_i), m_{-i}; \theta_i)$$

with strick inequality for some $m_{-i} \in M_{-i}$.

**Definition 19.2.5** $m \in M$ is a undominated Bayesian equilibrium of $\Gamma = (M, h)$, if it is a Bayesian-Nash equilibrium that is not weakly dominated.

Palfrey and Srivastava(1989, JPE) prove the following proposition.

**Proposition 19.2.2** For the class of private value economic environments, if there exists a mechanism, such that all agents do not have weakly dominated strategies, then any Bayesian incentive-compatible social choice function is strongly Bayesian implementable.

In the remainder of the chapter, we mainly consider truthful Bayesian implementability of a social choice rule. The last two sections of the chapter will
discuss the necessary and sufficient conditions for full Bayesian implementability of a general social choice rule—the set of social choice functions under more general interdependent value model.

19.3 Truthful Implementation of Pareto Efficient Outcomes

Once again, let us return to the quasilinear setting. From the section on dominant strategy implementation in the last chapter, we know that there is, in general, no ex-post Pareto efficient implementation if the dominant equilibrium solution behavior is assumed. However, in quasilinear environments, relaxation of the equilibrium concept from DS to BNE allows us to implement ex post Pareto efficient choice rule $f(\cdot) = (y^*(\cdot), t_1(\cdot), \cdots, t_i(\cdot))$, where $\forall \theta \in \Theta$,

$$y(\theta) \in \arg\max_{y \in Y} \sum_{i=1}^{n} v_i(x, \theta_i),$$

and there is a balanced budget:

$$\sum_{i=1}^{n} t_i(\theta) = 0.$$ (19.5)

The Expected Externality Mechanism described below was suggested independently by D’Aspermont and Gerard-Varet (1979) and Arrow (1979) and are also called AGV mechanisms in the literature. This mechanism enables us to have ex-post Pareto efficient implementation under the following additional assumption.

**Assumption 19.3.1** Types are distributed independently: $\phi(\theta) = \Pi_i \phi_i(\theta_i), \forall \theta \in \Theta$.

To see this, take the VCG transfer for agent $i$ and instead of using other agents’ announced types, take the expectation over their possible types

$$t_i(\hat{\theta}) = E_{\theta_{-i}} \left[ \sum_{j \neq i} v_j(y^*(\hat{\theta}_i, \theta^{-i}_j), \theta_j) \right] + d_i(\hat{\theta}_{-i})$$
(by the above assumption the expectation over $\theta_{-i}$ does not have to be taken conditionally on $\theta_i$). Note that unlike in VCG mechanisms, the first term only depends on agent $i$’s announcement $\hat{\theta}_i$, and not on other agents’ announcements. This is because it sums the expected utilities of agents $j \neq i$ assuming that they tell the truth and given that $i$ announced $\hat{\theta}_i$, and does not depend on the actual announcements of agents $j \neq i$. This means that $t_i(\cdot)$ is less “variable”, but on average it will cause $i$’s incentives to be lined up with the social welfare.

To see that $i$’s incentive compatibility is satisfied given that agent $j \neq i$ announce truthfully, observe that agent $i$ solves

$$
\max_{\hat{\theta}_i, \theta_i} E_{\theta_{-i}}[v_i(y(\hat{\theta}_i, \theta_i), \theta_i)] + E_{\theta_{-i}}[\sum_{j=1}^n v_j(y(\hat{\theta}_i, \theta_{-i}), \theta_j)] + E_{\theta_{-i}} d_i(\theta_{-i}).
$$

(19.6)

Again, agent $i$’s announcement only matters through the decision $y^*(\hat{\theta}_i, \theta_{-i})$. Furthermore, by the definition of the efficient decision rule $y^*(\cdot)$, the designer always chooses $y^*(\cdot)$ to max social surplus (19.4) for each realization of $\theta_{-i} \in \Theta_{-i}$. Then, when agent $i$ announces truthfully: $\hat{\theta}_i = \theta_i$, maximization of (19.6) is consist with the social goal (19.4). As such, when the social surplus (19.4) is maximized, agent $i$’s expected utility (19.6) is maximized by choosing the decision $y^*(\theta_i, \theta_{-i})$, which can be achieved by announcing truthfully: $\hat{\theta}_i = \theta_i$. Therefore, truthful announcement maximizes the agent $i$’s expected utility as well. Thus, BIC is satisfied.

**Remark 19.3.1** The argument relies on the assumption that other agents announce truthfully. Thus, it is in general not a dominant strategy for agent $i$ to announce the truth. Indeed, if agent $i$ expects the other agents to lie, i.e., announce $\hat{\theta}_{-i}(\theta_{-i}) \neq \theta_{-i}$, then agent $i$’s expected utility is

$$
E_{\theta_{-i}}[v_i(y^*(\hat{\theta}_i, \hat{\theta}_{-i}(\theta_{-i})), \theta_i)] + E_{\theta_{-i}}[\sum_{j \neq i} v_j(y^*(\hat{\theta}_i, \theta_{-i}), \theta_j)] + E_{\theta_{-i}} d_i(\theta_{-i}),
$$

which may not be maximized by truthfully announcement.
Furthermore, we can now choose functions $d_i(\cdot)$ so that the budget is balanced. To see this, let

$$\xi_i(\theta_i) = E_{\theta_{-i}}[\sum_{j \neq i} v_j(y^*(\theta_i, \theta_{-i}), \theta_j)],$$

so that the transfers in the Expected Externality mechanism are $t_i(\theta) = \xi_i(\theta_i) + d_i(\theta_{-i})$. We will show that we can use the $d(\cdot)$ functions to "finance" the $\xi(\cdot)$ functions in the following way:

Let

$$d_j(\theta_{-j}) = -\sum_{i \neq j} \frac{1}{n-1} \xi_i(\theta_i).$$

Then

$$\sum_{j=1}^{n} d_j(\theta_{-j}) = -\frac{1}{n-1} \sum_{j=1}^{n} \sum_{i \neq j} \xi_i(\theta_i) = -\frac{1}{n-1} \sum_{j \neq i} \sum_{i=1}^{n} \xi_i(\theta_i)$$

$$= -\frac{1}{n-1} \sum_{i=1}^{n} (n-1) \xi_i(\theta_i) = -\sum_{i=1}^{n} \xi_i(\theta_i),$$

and therefore

$$\sum_{i=1}^{n} t_i(\theta) = \sum_{i=1}^{n} \xi_i(\theta_i) + \sum_{i=1}^{n} d_i(\theta_{-i}) = 0.$$

Thus, we have shown that when agents’ Bernoulli utility functions are quasi-linear and agents’ types are statistically independent, there is an ex-post efficient social choice function that is implementable in Bayesian-Nash equilibrium.

### 19.4 Characterization of DIC and BIC under Linear Mode

Consider the quasi-linear environment with the decision set $Y \subset \mathbb{R}^n$, and the type spaces $\Theta_i = [\theta_i; \theta_i] \subset \mathbb{R}$ for all $i$. Each agent $i$’s utility takes the form

$$\theta_i y_i + t_i.$$

Note that these payoffs satisfy SCP.
The decision set \( Y \) depends on the application.

Two examples:

1. Allocating a private good: \( Z = \{ y \in \{0, 1\}^n : \sum_i y_i = 1 \} \).

2. Provision of a nonexcludable public good: \( Y = \{(q, \cdots, q) \in \mathbb{R}^n : q \in \{0, 1\}\} \).

We can fully characterize both dominant incentive compatibility (DIC) and Bayesian incentive compatibility (BIC) social choice rules in this environment.

Let \( U_i(\theta) = \theta_i y_i(\theta) + t_i(\theta) \). We first have the following proposition.

**Proposition 19.4.1 (DIC Characterization Theorem)** In the linear model, social choice rule \((y(\cdot), t_1(\cdot), \cdots, t_I(\cdot))\) is Dominant Incentive Compatible if and only if for all \( i \in N \),

\[
(1) \quad y_i(\theta_i, \theta_{-i}) \text{ is nondecreasing in } \theta_i, \quad (DM);
\]

\[
(2) \quad U_i(\theta_i, \theta_{-i}) = U_i(\theta_i, \theta_{-i}) + \int_{\theta_i}^{\theta_i'} y_i(\tau, \theta_{-i}) d\tau, \forall \theta_i \in \Theta. \quad (DICFOC)
\]

**Proof. Necessity:** Dominant strategy incentive compatibility implies that for each \( \theta_i' > \theta_i \), we have

\[
U_i(\theta_i, \theta_{-i}) \geq \theta_i y_i(\theta_i', \theta_{-i}) + t_i(\theta_i', \theta_{-i}) = U_i(\theta_i', \theta_{-i}) + (\theta_i - \theta_i') y_i(\theta_i', \theta_{-i})
\]

and

\[
U_i(\theta_i', \theta_{-i}) \geq \theta_i' y_i(\theta_i', \theta_{-i}) + t_i(\theta_i, \theta_{-i}) = U_i(\theta_i, \theta_{-i}) + (\theta_i' - \theta_i) y_i(\theta_i, \theta_{-i}).
\]

Thus

\[
y_i(\theta_i', \theta_{-i}) \geq \frac{U_i(\theta_i', \theta_{-i}) - U_i(\theta_i, \theta_{-i})}{\theta_i' - \theta_i} \geq y_i(\theta, \theta_{-i}). \quad (19.7)
\]

Expression (19.7) implies that \( y_i(\theta_i, \theta_{-i}) \) must be nondecreasing in \( \theta_i \) by noting that \( \theta_i' > \theta_i \). In addition, letting \( \theta_i' \rightarrow \theta_i \) in (19.7) implies that for all \( \theta_i \), we have

\[
\frac{\partial U_i(\theta)}{\partial \theta_i} = y_i(\theta),
\]

753
and thus
\[ U_i(\theta_i, \theta_{-i}) = U_i(\theta_i, \theta_{-i}) + \int_{\theta_i}^{\theta_{-i}} y_i(\tau, \theta_{-i})d\tau, \forall \theta \in \Theta. \]

**Sufficiency.** Consider any two types \( \theta'_i \) and \( \theta_i \). Without loss of generality, suppose \( \theta_i > \theta'_i \). If DM and DICFOC hold, then
\[
U_i(\theta_i, \theta_{-i}) - U_i(\theta'_i, \theta_{-i}) = \int_{\theta'_i}^{\theta_i} y_i(\tau', \theta_{-i})d\tau' \geq \int_{\theta'_i}^{\theta_i} y_i(\theta_i, \theta_{-i})d\tau' = (\theta_i - \theta'_i)y_i(\theta'_i, \theta_{-i}).
\]

Thus, we have
\[
U_i(\theta_i, \theta_{-i}) \geq U_i(\theta'_i, \theta_{-i}) + (\theta_i - \theta'_i)y_i(\theta'_i, \theta_{-i}) = \theta_iy_i(\theta'_i, \theta_{-i}) + t_i(\theta'_i, \theta_{-i}).
\]

Similarly, we have derive that
\[
U_i(\theta'_i, \theta_{-i}) \geq U_i(\theta_i, \theta_{-i}) + (\theta'_i - \theta_i)y_i(\theta_i, \theta_{-i}) = \theta'_iy_i(\theta_i, \theta_{-i}) + t_i(\theta_i, \theta_{-i}).
\]

Hence, the decision rule \((y(\cdot), t_1(\cdot), \ldots, t_n(\cdot))\) is dominant strategy incentive compatible.

When agent \( i \) announces \( \hat{\theta} \) and the others announce \( \theta_{-i} \) truthfully, we have the interim expected consumption \( \bar{y}_i(\hat{\theta}_i) \equiv E_{\theta_{-i}}y_i(\hat{\theta}_i, \theta_{-i}) \) and transfer \( \bar{t}_i(\hat{\theta}_i) \equiv E_{\theta_{-i}}t_i(\hat{\theta}_i, \theta_{-i}) \). Thus agent \( i \)'s interim expected utility \( \bar{U}_i(\hat{\theta}_i, \hat{\theta}_i) = \theta_i\bar{y}_i(\hat{\theta}_i) + \bar{t}_i(\hat{\theta}_i) \) depends only on the interim expected consumption \( E_{\theta_{-i}}y_i(\hat{\theta}_i, \theta_{-i}) \) and interim transfer \( E_{\theta_{-i}}t_i(\hat{\theta}_i, \theta_{-i}) \).

Define
\[
\bar{U}_i(\theta_i) \equiv \bar{U}_i(\theta_i, \theta_i) = E_{\theta_{-i}}U_i(\theta) = \theta_i\bar{y}_i(\theta_i) + \bar{t}_i(\theta_i),
\]
which is agent \( i \)'s interim expected utility when he announces \( \theta_{-i} \) truthfully. Then, BIC means that truth-telling is optimal at the interim stage. By the same method in the proof of the above proposition for \( \bar{U}(\theta_i) \), we can similarly prove the following result on BIC.
Proposition 19.4.2 (BIC Characterization Theorem) In the linear model, social choice rule \((y(\cdot), t_1(\cdot), \cdots, t_I(\cdot))\) is Bayesian Incentive Compatible if and only if for all \(i \in I\),

1. \(E_{\theta - i} y_i(\theta_i, \theta_{-i}) = \bar{y}_i(\theta_i)\) is nondecreasing in \(\theta_i\) (BM);

2. \(E_{\theta - i} U_i(\theta_i, \theta_{-i}) = E_{\theta - i} U_i(\theta_i, \theta_{-i}) + \int_{\theta_i}^{\theta_{i+1}} E_{\theta - i} y_i(\tau, \theta_{-i}) d\tau, \forall \theta_i \in \Theta_i\).

(BICFOC)

Note that (BICFOC) is a pooling of (DICFOC), and similarly (BM) is a pooling of (DM). The latter means that BIC allows to implement some decision rules that are not DS implementable.

The proposition implies that for any two BIC mechanisms that implement the same decision rule \(y(\cdot)\), the interim expected utilities \(E_{\theta - i} U_i(\theta)\) and thus the transfers \(E_{\theta - i} t_i(\theta)\) coincide up to a constant.

Corollary 19.4.1 In the linear model, for any BIC mechanism that implements the ex post Pareto efficient decision rule \(y(\cdot)\), there exists a VCG mechanism with the same interim expected transfers and utilities.

Proof. If \((y^*(\cdot), t(\cdot))\) is a BIC mechanism and \((y^*(\cdot), \tilde{t}(\cdot))\) is a VCG mechanism, then by BIC Characterization Theorem (Proposition 19.4.2), \(E_{\theta - i} t_i(\theta) = E_{\theta - i} \tilde{t}_i(\theta) + c_i, \forall \theta_i \in \Theta_i\). Then letting \(\tilde{t}_i(\theta) = \tilde{t}_i(\theta) + c_i, (y^*(\cdot), \tilde{t}(\cdot))\) is also a VCG mechanism, and \(E_{\theta - i} \tilde{t}_i(\theta) = E_{\theta - i} t_i(\theta)\).

For example, the expected externality mechanism is interim-equivalent to a VCG mechanism. Its only advantage is that it allows to balance the budget ex post, i.e., in each state of the world. More generally, if a decision rule \(y(\cdot)\) is DS implementable, the only reason to implement it in a BIC mechanism that is not DIC is we care about ex post transfers/utilities rather than just their interim or ex ante expectations.
19.5 Impossibility of BIC Pareto Efficient Optimal Contract

19.5.1 Participation Constraints

In general mechanism design considered in this chapter so far, we has not been yet thinking of the mechanism as a contract or principal-agent issue. Since if a mechanism is a contract, it must be voluntary, i.e. it should satisfy participation constraint. Thus, if we would put the principal-agent theory into the framework of general mechanism design, a mechanism should be individually rational. As such, when agents’ types are not observed, a social choice function that can be successfully implemented should satisfy not only the incentive compatibility either in a dominant strategy or Bayesian Nash dominant strategy, depending on the equilibrium concept used, but also participation constraints that are relevant in the environment under study.

When using dominant strategy as a solution concept, Hurwicz impossibility theorem tells us that, there is no any truth-telling mechanism that results in Pareto efficient allocations. As for Bayesian incentive compatibility, such as a mechanism exists such as the Expected Externality mechanism. Then, a question is, if imposing the individual rationality constraint, so that we can think of a mechanism as a contract, is such a mechanism still exist? The answer is negative.

If thinking of the mechanism as a contract, the following issues are raised:

- Will the agents accept it voluntarily, i.e., are their participation constraints satisfied?

- If one of the agents designs the contract, he will try to maximize his own payoff subject to the other agents $i$ participation constraints. What will the optimal contract look like?

To analyze these questions, we need to imposes additional restrictions on the social choice rule in the form of participation constraints. These constraints
depend on when agents can withdraw from the mechanism, and what they get when they do. Let \( \hat{u}_i(\theta_i) \) be the utility of agent \( i \) if he withdraws from the mechanism. (This assumes that when an agent withdraws from the mechanism he does not care what the mechanism does with other agents.)

**Definition 19.5.1** The social choice rule \( f(\cdot) \) is

1. **Ex Post Individually Rational** if for all \( i \),
   \[
   U_i(\theta) \equiv u_i(f(\theta), \theta) \geq \hat{u}_i(\theta_i), \forall \theta \in \Theta.
   \]

2. **Interim Individually Rational** if for all \( i \),
   \[
   E_{\theta-}[U_i(\theta_i, \theta-)] \geq \hat{u}_i(\theta_i), \forall \theta \in \Theta.
   \]

3. **Ex ante Individually Rational** if for all \( i \),
   \[
   E_\theta[U_i(\theta)] \geq E_\theta[\hat{u}_i(\theta_i)].
   \]

Note that ex post IRs imply interim IR, which in turn imply ex ante IRs, but the reverse may not be true. Then, the constraints imposed by voluntary participation are most sever when agents can withdraw at the ex post stage.

The ex post IRs arise when the agent can withdraw in any state after the announcement of the outcome. For example, they are satisfied by any decentralized bargaining procedure. These are the hardest constraints to satisfy.

The Interim IRs arise when the agent can withdraw after learning his type \( \theta_i \); but before learning anything about other agents’ types. Once the agent decides to participate, the outcome can be imposed on him. These constraints are easier to satisfy than ex post IR.

With the ex-ante participation constraint the agent can commit to participating even before his type is realized. These are the easiest constraints to satisfy. For example, in a quasi-linear environment, whenever the mechanism generates a positive expected total surplus, i.e.,

\[
E_\theta[\sum_i U_i(\theta)] \geq E_\theta[\sum_i \hat{u}_i(\theta_i)],
\]
all agents’ ex ante IR can be satisfied by reallocating expected total surplus among agents through lump-sum transfers, which will not disturb agents’ incentive constraints or budget balance.

For this reason, we will focus mainly on interim IR. In the following, we will illustrate further the limitations on the set of implementable social choice functions that may be caused by participation constraints by the important theorem given by Myerson-Satterthwaite (1983).

19.5.2 Myerson-Satterthwaite Impossibility Theorem

Even though Pareto efficient mechanisms do exist (e.g. the expected externality mechanism), it remains unclear whether such a mechanism may result from private contracting among the parties. We have already seen that private contracting need not yield efficiency. In the Principal-Agent model, the Principal offers an efficient contract to extract the agent’s information rent. However, this leaves open the question of whether there is some contracting/bargaining procedure that would yield efficiency. For example, in the P-A model, if the agent makes an offer to the principal, who has no private information on his own, the agent would extract all the surplus and implement efficient as a result. Therefore, we focus on a bilateral situation in which both parties have private information. In this situation, it turns out that generally there does not exist an efficient mechanism that satisfies both agents’ participation constraints.

Consider the setting of allocating an indivisible object with two agents - a seller and a buyer: \( I = \{S, B\} \). Each agent’s type is \( \theta_i \in \Theta_i = [\underline{\theta}_i, \overline{\theta}_i] \subset \mathbb{R} \), where \( \theta_i \sim \varphi_i(\cdot) \) are independent, and \( \varphi_i(\cdot) > 0 \) for all \( \theta_i \in \Theta_i \). Let \( y \in \{0, 1\} \) indicate whether \( B \) receives the good. A social choice rule is then \( f(\theta) = (y(\theta), t_1(\theta), t_2(\theta)) \). The agents’ utilities can then be written as

\[
\begin{align*}
    u_B(y, \theta_B) &= \theta_B y + t_B, \\
    u_S(y, \theta_S) &= -\theta_S y + t_S.
\end{align*}
\]
It is easy to see that an efficient decision rule \( y^*(\theta) \) must have\(^2\)

\[
y^*(\theta_B, \theta_S) = \begin{cases} 
1 & \text{if } \theta_B > \theta_S, \\
0 & \text{if } \theta_B < \theta_S.
\end{cases}
\]

We could use an expected externality mechanism to implement an efficient decision rule in BNE with ex post budget balance. However suppose that we have to satisfy Interim IR:

\[
E_{\theta_S}[\theta_By(\theta_B, \theta_S) + t_B(\theta_B, \theta_S)] \geq 0,
\]

\[
E_{\theta_B}[-\theta_Sy(\theta_B, \theta_S) + t_S(\theta_B, \theta_S)] \geq 0.
\]

As for budget balance, let us relax this requirement by requiring only ex ante Budget Balance:

\[
E[\theta_B(t_B(\theta_B, \theta_S) + t_S(\theta_B, \theta_S))] \leq 0.
\]

Unlike ex post budget balance considered before, ex ante budget balance allows us to borrow money as long as we break even on average. It also allows us to have a surplus of funds. The only constraint is that we cannot have an expected shortage of funds.

We can then formulate a negative result:

**Theorem 19.5.1 (Myerson-Satterthwaite Theorem, JET 1983)** *In the two-party trade setting above, suppose each agent’s type is \( \theta_i \in \Theta_i = [\theta_i, \bar{\theta}_i] \subset \mathbb{R} \), where \( \theta_i \sim \varphi_i(\cdot) \) are independent and \( (\theta_B, \bar{\theta}_B) \cap (\theta_S, \bar{\theta}_S) \neq \emptyset \) (gains from trade are possible but not certain). Then there is no BIC social choice rule that has the efficient decision rule and satisfies ex ante Budget Balance and interim IR.*

Proof. Consider first the case where \([\theta_B, \bar{\theta}_B] = [\theta_S, \bar{\theta}_S] = [\bar{\theta}, \bar{\theta}]\).

By Corollary 19.4.1 above, we know that any BIC mechanism in the linear environment that implements the ex post Pareto efficient decision rule \( y(\cdot) \), there exists a VCG mechanism with the same interim expected transfers and utilities.

\(^2\)Given our full support assumption on the distributions, ex ante efficiency dictates that the decision rule coincide with \( y^*(\cdot) \) almost everywhere.
As such, we can restrict attention to VCG mechanisms while preserving ex ante budget balance and interim IR. Such mechanisms take the form

\[ t_B(\theta_B, \theta_S) = -\theta_S y^*(\theta_B, \theta_S) + d_B(\theta_S), \]
\[ t_S(\theta_B, \theta_S) = \theta_B y^*(\theta_B, \theta_S) + d_S(\theta_B). \]

By interim IR of B’s type \( \theta \), \( E_{\theta_S}[\theta_B y(\theta_B, \theta_S) + t_B(\theta_B, \theta_S)] \geq 0 \), using the fact that \( y^*(\theta, \theta_S) = 0 \) with probability 1, we must have \( E_{\theta_S} d_B(\theta_S) \geq 0 \). Similarly, by interim IR of S’s type \( \bar{\theta} \), \( E_{\theta_B}[-\theta_S y(\theta_B, \bar{\theta}_S) + t_S(\theta_B, \bar{\theta}_S)] \geq 0 \), using the fact that \( y^*(\theta_B, \bar{\theta}) = 0 \) with probability 1, we must have \( E_{\theta_B} d_S(\theta_B) \geq 0 \). Thus, adding the transfers, we have

\[ E_{\theta}[t_B(\theta_B, \theta_S) + t_S(\theta_B, \theta_S)] = E_{\theta}[(\theta_B - \theta_S)y^*(\theta_B, \theta_S)] + E_{\theta_S}[d_B(\theta_S)] + E_{\theta_B}[d_S(\theta_B)] \geq E_{\theta}[\max\{\theta_B - \theta_S\}] > 0 \]

since \( \Pr(\theta_B > \theta_S) > 0 \). Therefore, ex ante budget balance cannot be satisfied.

Now, in the general case, let \((\theta, \bar{\theta}) = (\theta_B, \bar{\theta}_B) \cap (\theta_S, \bar{\theta}_S)\), and observe that any type of either agent above \( \bar{\theta} \) [or below \( \theta \)] has the same decision rule, and therefore must have the same transfer, as this agent’s type \( \bar{\theta} \) [resp. \( \theta \)]. Therefore, the payments are the same as if both agents having valuations distributed on \([\theta, \bar{\theta}]\); with possible atoms on \( \theta \) and \( \bar{\theta} \). The argument thus still applies.

The intuition for the proof is simple: In a VCG mechanism, in order to induce truthful revelation, each agent must become the residual claimant for the total surplus (since it is given by the sum of individuals’ values plus transfers). This means that in case trade is implemented, the buyer pays the seller’s cost for the object, and the seller receives the buyer’s valuation for the object. Any additional payments to the agents must be nonnegative, in order to satisfy interim IRs of the lowest-valuation buyer and the highest-cost seller. Thus, each agent’s utility must be at least equal to the total surplus. In BNE implementation, agents receive the same expected utilities as in the VGC mechanism, thus again each agent’s expected utility must equal at least to the total expected surplus. This cannot be done without having an expected infusion of funds equal to the expected surplus.
Myerson-Satterthwaite Impossibility Theorem is based on the assumption that agents’ types are independent. If agents’ types are dependent, Cremer-McLean Full Surplus Extraction Theorem to be discussed later tells us that the conclusion may not be true.

**Interpretation of the Theorem:** Who is designing the mechanism to maximize expected total surplus subject to agents’ interim IRs?

- Agents themselves at the ex ante stage? They would face ex ante rather than interim IRs, which would be easy to satisfy. For example, in a quasi-linear environment, whenever the mechanism (such as AVG mechanism) generates a positive expected total surplus $E_\theta[\sum_i U_i(\theta)] \geq E_\theta[\sum_i \hat{u}_i(\theta_i)]$, all agents’ ex ante IR can be satisfied by reallocating expected surplus among agents through lump-sum transfers, which will not disturb agents’ incentive constraints or budget balance.

- An agent at the interim stage? He would be interested not in efficiency but in maximizing his own payoff. As such, it cannot form incentive compatibility.

- A benevolent mediator at the interim stage? But where would such a mediator come from?

A better interpretation of the result is as an upper bound on the efficiency of decentralized bargaining procedures. Indeed, any such procedure can be thought of as a mechanism, and must satisfy interim IR (indeed, even ex post IR), and ex ante budget balance (indeed, even ex post budget balance). The Theorem says that decentralized bargaining in this case cannot be efficient. In the terminology of the Coase Theorem, private information creates a “transaction cost”.

Cramton-Gibbons-Klemperer (1987) show that when the object is divisible, and is jointly owned initially, efficiency can be attained if initial ownership is sufficiently evenly allocated. For example, suppose that the buyer initially owns $\hat{y} \in [0, 1]$ of the object. Efficiency can be achieved when $\hat{y}$ is sufficiently close to
1/2. Thus, when a partnership is formed, the partners can choose initial shares so as to eliminate inefficiencies in dissolution. While this result can be applied to study optimal initial allocations of property rights, it does not explain why property rights are good at all. That is, interim IRs stemming from property rights can only hurt the parties in the model. For example, the parties could achieve full efficiency by writing an ex ante contract specifying the AGV mechanism and not allowing withdrawal at the interim stage. One would have to appeal to a difficulty of specify in complicated mechanisms such as AGV to explain why the parties would look for optimal property rights that facilitate efficient renegotiation rather than specifying an efficient mechanism ex ante without allow in subsequent withdrawal or renegotiation.

19.6 The Revenue Equivalence Theorem in Auctions

Let us consider again the setting of allocating an indivisible object among \( n \) risk-neutral buyers: \( Y = \{ y \in \{0, 1\}^n : \sum_i y_i = 1 \} \), and the payoff of buyer \( i \) is

\[ \theta_i y_i + t_i. \]

The buyers’ valuations are independently drawn from \([\theta_i, \bar{\theta}_i]\) with \( \theta_i < \bar{\theta}_i \) according to a strictly positive density \( \varphi_i(\cdot) \), and c.d.f. denoted by \( \Phi_i(\cdot) \).

Suppose that the object initially belongs to a seller (auctioneer), who is the residual sink of payments. The auctioneer’s expected revenue in a social choice rule \((y(\cdot), t_1, \ldots, t_n)\) can be written as

\[
- \sum_i E_{\theta} t_i(\theta) = \sum_i E_{\theta}[\theta_i y_i(\theta) - U_i(\theta)] = \sum_i E_{\theta}[\theta_i y_i(\theta)] - \sum_i E_{\theta}[\theta_i y_i(\theta)] = \sum_i E_{\theta}[U_i(\theta, \theta_{-i})].
\]

The first term is the agents’ expected total surplus, while the second term is the sum of the agents’ expected utilities. By BIC Characterization Theorem (Proposition 19.4.2), the second term is fully determined by the decision (object allocation) rule \( y(\cdot) \) together with the lowest types’ interim expected utilities.
Since the total surplus is pinned down as well, we have the following result.

**Theorem 19.6.1 (The Revenue Equivalence Theorem)** Suppose that two different auction mechanisms have Bayesian-Nash Equilibria in which (i) the same decision (object allocation) rule \( y(\cdot) \) is implemented, and (ii) each buyer \( i \) has the same interim expected utility when his valuation is \( \theta_i \). Then the two equilibria of the two auction mechanisms generate the same expected revenue for the seller.

Note that even when the decision rule and lowest agents’ utilities are fixed, the seller still has significant freedom in designing the auction procedure, since there are many ways to achieve given interim expected transfers \( \bar{t}_i(\theta_i) \) with different ex post transfer \( t_i(\theta) \).

For example, suppose that the buyers are symmetric, and that the seller wants to implement an efficient decision rule \( y^*(\cdot) \), and make sure that the lowest-valuation buyers receive zero expected utility. We already know that this can be done in dominant strategies - using the Vickrey (second-price sealed-bid) auction. More generally, consider a \( k^{th} \) price sealed-bid auction, with \( 1 \leq k \leq n \). Here the winner is the highest bidder, but he pays the \( k^{th} \) highest bid. Suppose that buyers’ valuations are i.i.d. Then it can be shown that the auction has a unique equilibrium, which is symmetric, and in which an agent’s bid is an increasing function of his valuation, \( b(\theta_i) \). (See, e.g., Fudenberg- Tirole “Game Theory,” pp. 223-225.) Agent \( i \) receives the object when he submits the maximum bid, which happens when he has the maximum valuation. Therefore, the auction implements an efficient decision rule. Also, a buyer with valuation \( \theta \) wins with zero probability, hence he has zero expected utility. Hence, the Revenue Equivalence Theorem establishes that \( k^{th} \) price sealed-bid auctions generate the same expected revenue for the seller for all \( k \).

How can this be true, if for any given bid profile, the seller receives a higher revenue when \( k \) is lower? The answer is that the bidders will submit lower bids when \( k \) is lower, which exactly offsets the first effect. For example, we know
that in the second-price auction, buyer will bid their true valuation. In the first price auction, on the other hand, buyers will obviously bid less than their true valuation, since bidding their true valuation would give them zero expected utility. The revenue equivalence theorem establishes that the expected revenue ends up being the same. Remarkably, we know this without solving for the actual equilibrium of the auction. With \( k > 2 \), the revenue equivalence theorem implies that buyers will bid more than their true valuation.

Now we can also solve for the seller’s optimal auction. Suppose the seller also has the option of keeping the object to herself. The seller will be called “agent zero”, her valuation for the object denoted by \( \theta_0 \), and whether she keeps the object will be denoted by \( y_0 \in \{0, 1\} \). Thus, we must have \( \sum_{i=0}^{n} y_i = 1 \).

The seller’s expected payoff can be written as

\[
\theta_0 E_\theta y_0(\theta) + \text{Expected Revenue} = \theta_0 E_\theta y_0(\theta) + \sum_{i=1}^{n} E_\theta[\theta_i y_i(\theta)] - \sum_{i=1}^{n} E_\theta E_{\theta_{-i}}[U_i(\theta_i, \theta_{-i})].
\]

Thus, the seller’s expected payoff is the difference between total surplus and the agents’ expected information rents.

By BIC Characterization Theorem (Proposition 19.4.2), the buyer’s expected interim utilities must satisfy

\[
E_{\theta_{-i}}[U_i(\theta_i, \theta_{-i})] = E_{\theta_{-i}}[U_i(\theta_i, \theta_{-i})] + \int_{\theta_i}^{\theta_{-i}} E_{\theta_{-i}} y_i(\tau, \theta_{-i}) d\tau, \forall \theta_i \in \Theta_i. \quad (BICFOC)
\]

By the buyers’ interim IR, \( E_{\theta_{-i}}[U_i(\theta_i, \theta_{-i})] \geq 0, \forall i \), and for a given decision rule \( y \), the seller will optimally chose transfers to set \( E_{\theta_{-i}}[U_i(\theta_i, \theta_{-i})] = 0, \forall i \).

Furthermore, just as in the one-agent case, upon integration by parts, buyer \( i \)’s expected information rent can be written as

\[
E_\theta E_{\theta_{-i}} \left[ \frac{1}{h_i(\theta_i)} y_i(\theta) \right],
\]

where \( h_i(\theta_i) = \frac{\varphi_i(\theta_i)}{1 - \Phi_i(\theta_i)} \) is the hazard rate of agent \( i \). Substituting in the seller’s payoff, we can write it as the expected “virtual surplus”:

\[
E_\theta \left[ \theta_0 y_0(\theta) + \sum_{i=1}^{n} \left( \theta_i - \frac{1}{h_i(\theta_i)} \right) y_i(\theta) \right].
\]
Finally, for $i \geq 1$, let $\nu_i(\theta_i) = \theta_i - 1/h_i(\theta_i)$, which we will call the “virtual valuation” of agent $i$. Let also $\nu_0(\theta_0) = \theta_0$. Then the seller’s program can be written as

$$\max_{\tilde{x}(\cdot)} E_{\theta} \left[ \sum_{i=0}^{n} \nu_i(\theta_i) y_i(\theta) \right] \text{ s.t. } \sum_{i=0}^{n} y_i(\theta) = 1,$$

where $E_{\theta_i} y_i(\theta_i, \theta_{-i})$ is nondecreasing in $\theta_i$, $\forall i \geq 1$. (BM)

If we ignore Bayesian Monotonicity (BM) above, we can maximize the above expectation for each state $\theta$ independently. The solution then is to give the object to the agent who has the highest virtual valuation in the particular state: we have $y_i(\theta) = 1$ when $\nu_i(\theta_i) > \nu_j(\theta_j)$ for all agents $j \neq i$. The decision rule achieving this can be written as $y(\theta) = y^*(\nu_0(\theta_0), \nu_1(\theta_1), \cdots, \nu_n(\theta_n))$, where $y^*(\cdot)$ is the first-best efficient decision rule. Intuitively, the principal uses agents’ virtual valuation rather than their true valuations because she cannot extract all of the agents’ rents. An agent’s virtual valuation is lower than his true valuation because it accounts for the agent’s information rents which cannot be extracted by the principal. The profit-maximizing mechanism allocates the object to the agent with the highest virtual valuation.

Under what conditions can we ignore the monotonicity constraint? Note that when the virtual valuation function $\nu_i(\theta_i) = \theta_i - 1/h_i(\theta_i)$ is an increasing function\(^3\), then an increase in an agent’s valuation makes him more likely to receive the object in the solution to the relaxed problem. Therefore, $y_i(\theta_i, \theta_{-i})$ is nondecreasing in $\theta_i$ for all $\theta_{-i}$, and so by DIC Characterization Theorem (Proposition 19.4.1), the optimal allocation rule is implementable not just in BNE, but also in DS. The DIC transfers could be constructed by integration using DICFOC; the resulting transfers take a simple form:

$$t_i(\theta) = p_i(\theta_{-i}) y_i(\theta), \text{ where } p_i(\theta_{-i}) = \inf\{\hat{\theta}_i \in [\theta_i, \bar{\theta}_i] : y_i(\hat{\theta}_i, \theta_{-i}) = 1\}.$$

\(^3\)A sufficient condition for this is that the hazard rate $h_i(\theta_i)$ is a nondecreasing function. This is the same argument as in the one-agent case (note that $\nu_{y\theta \theta} = 0$ for the linear utility function $\nu(y, \theta) = \theta y$).
Thus, for each agent $i$ there is a “pricing rule” $p_i(\theta_{-i})$ that is a function of others’ announcements, so that the agent wins if his announced valuation exceeds the price, and he pays the price whenever he wins. This makes truth-telling a dominant strategy by the same logic as in the Vickrey auction: lying cannot affect your own price but only whether you win or lose, and you want to win exactly when your valuation is above the price you face.

19.7 Cremer-McLean Full Surplus Extraction Theorem

A central theme of mechanism design under incomplete information is about surplus extraction ability of agents. Whatever the realistic observation or economic intuition tells us that the existence of private information makes the principal has to give up some positive information rent. However, the Myerson-Satterthwaite Theorem is based not only on the assumption of private value but also on the assumption that agents’ types are independent. If agents’ types are dependent, can the Myerson-Satterthwaite Impossibility Theorem that there is no BIC social choice rule that has the efficient decision rule and satisfies ex ante Budget Balance and interim IR be still true?

The answer is no, and we will have a positive result. The above observation and economic intuition are no longer true. This is the basic ideal of the full surplus extraction theorem in Cremer and McLean (1988). This theorem proves that, when private types are correlated, the principal of the mechanism can extract all information rents of agents through verifying information reported, and thus obtains Pareto efficient outcome. Using the ideas of Cremer and McLean (1988), we can see that with correlated types, “almost anything is implementable”. Such a conclusion is called Cremer-McLean Theorem, which has a wide application. In auction theory, seller can design a Pareto efficient design mechanism. Cremer-McLean’s full surplus extraction theorem holds for any kind of correlated types.

Consider an economy with $n$ agents. Each agent’s type is discrete, $\theta_i \in \Theta_i \equiv$
\{\theta_1^i, \theta_2^i, \ldots, \theta_k^i\}, \forall i \in N = \{1, \ldots, n\}, and his utility function is quasi-linear with private values, i.e., \(U_i(y, t, \theta_i) = v_i(y, \theta_i) + t_i\). Assume agents are correlated and have density function \(\pi(\theta)\). When agent \(i\)'s type is \(\theta_i\), his beliefs about other agents' types are given by

\[
\pi_i(\theta_{-i}|\theta_i) = \frac{\pi(\theta)}{\sum_{\theta'_{-i} \in \Theta_{-i}} \pi(\theta'_{-i}, \theta_i)}.
\]

For each agent \(i\), form matrix \(\Pi_i\) with element \(\pi_i(\theta_{-i}|\theta_i)\). As such, matrix \(\Pi_i\) has \(k_i\) rows and \(\sum_{j \neq i} k_j\) columns. Each row represents agent \(i\)'s beliefs distribution about other agents' type when his type is \(\theta_i\). If agents' types are independent, all rows are the same, and thus the ranks of the matrix \(\Pi_i\) is 1. If types are correlated, then different rows represent different beliefs. We assume \(\Pi_i\) is row full rank that is \(k_i\), which implies an agent's beliefs distribution about others' beliefs is different.

The following well-known Cremer-McLean Full Surplus Extraction Theorem proves that, even under requirement of dominant incentive-compatibility, the principal can extract all surplus. The ideal of the proof is pretty simpler. We first take any efficient and dominant incentive-compatible mechanism such as VCG mechanism, and then form a corresponding efficient and dominant incentive-compatible mechanism with equality individual rational constraint so that information rent is zero. As a result, we obtain a Pareto efficient dominant incentive-compatible mechanism. Formally, we have the following Cremer-McLean Full Surplus Extraction Theorem.

**Theorem 19.7.1 (Cremer-McLean Full Surplus Extraction Theorem)** Suppose agents' types under private value model are correlated, and the information matrix \(\Pi_i\) has row full rank. Then, for any efficient and dominant incentive-compatible social choice rule \((y(\cdot), t_1(\cdot), t_2(\cdot))\), there exists another dominant incentive-compatible and efficient social choice rule \((\tilde{y}(\cdot), \tilde{t}_1(\cdot), \tilde{t}_2(\cdot))\) with the same decision rule \(y(\cdot)\) in which interim individual rationality constraints are binding, and thus the first-best is implementable.
Proof. For any dominant incentive-compatible social choice rule \((y(\cdot), t_1(\cdot), t_2(\cdot))\), since agents have private quasi-linear utility functions, agent \(i\)’s interim expected utility with \(\theta_i\) at equilibrium can be written as

\[
\bar{U}_i(\theta_i) = \sum_{\theta_{-i}} \pi_i(\theta_{-i}|\theta_i)[u_i(y(\theta), \theta_i)] + t_i(\theta).
\]

Let \(\bar{u}_i^* = [\bar{U}(\theta^1), \bar{U}(\theta^2), \ldots, \bar{U}(\theta^m)]'\). Since \(\Pi_i\) is row full rank, there is a vector \(c_i = (c_i(\theta_{-i}))_{\theta_{-i} \in \Theta_{-i}}\) with \(\sum_{j \neq i} m_j\) columns, such that

\[
\Pi_i c_i = -\bar{u}_i^*,
\]

that is, for \(\forall \theta_i\), we have

\[
\bar{U}_i(\theta_i) + \sum_{\theta_{-i}} \pi_i(\theta_{-i}|\theta_i)c_i(\theta_{-i}) = 0.
\]

To guarantee \(i\)’s interim rationality is binding, let

\[
\bar{t}_i(\theta) = t_i(\theta) + c_i(\theta_{-i}).
\]

Construct a new mechanism, called Cremer-McLean mechanism:

\[
(y(\theta), \bar{t}_i(\cdot)) = (y(\theta), t_i(\theta) + c_i(\theta_{-i}), \forall i \in N).
\]

Since \(c_i(\theta_{-i})\) is uncorrelated with \(\theta_i\), Cremer-McLean mechanism is still dominant incentive-compatible and efficient. But, under Cremer-McLean mechanism, agent \(i\)’s interim expected utility at equilibrium becomes

\[
\bar{V}_i(\theta_i) = \sum_{\theta_{-i}} \pi_i(\theta_{-i}|\theta_i)[u_i(y(\theta), \theta_i)] + \bar{t}_i(\theta)]
\]

\[
= \sum_{\theta_{-i}} \pi_i(\theta_{-i}|\theta_i)[u_i(y(\theta), \theta_i)] + t_i(\theta) + c_i(\theta_{-i})
\]

\[
= \bar{U}_i(\theta_i) + \sum_{\theta_{-i}} \pi_i(\theta_{-i}|\theta_i)c_i(\theta_{-i})
\]

\[
= 0.
\]

Thus, all agents’ interim individual rationality constraints are binding, and the designer extracts all the surplus, especially, the designer can extract agents’ surplus through VCG mechanism to reach the first best outcome.
The above Cremer-McLean Theorem is based on private value model. In fact, Cremer-McLean Theorem still holds for interdependent model. Although VCG mechanism with interdependent values is not dominant incentive-compatible, the generalized VCG mechanism (by modifying VCG mechanism) is dominant incentive-compatible so that Cremer-McLean Theorem still holds for interdependent model. Such a mechanism is discussed in auction theory with interdependent values.

It may be pointed out, besides Cremer-McLean Full Surplus Extraction Theorem rely on the assumptions of quasi-linear utility function and unlimited liability, it also relies on an implicit assumption that there is no collusion, or it does not holds. As such, a major criticism towards CM’s result comes from its vulnerability to collusion among agents. In the FSE mechanism, payments to and from agents depend on the reports of other agents. Therefore, it is highly susceptible to collusion among the agents, especially in nearly independent environments where these payments are very large.

The pioneering work that studies collusion in principal-multiagent setting is due to Laffont and Mortimort (1997, 2000). In procurement/public good settings with two agents, they show that if the types are correlated, preventing collusion entails strict cost to the principal (LM, 2000). Che and Kim(2006) shows that the results obtained by Laffont and Mortimort (1997, 2000) depends the two-agent assumption. For \( n > 2 \), that show that agents’ collusion, including both reporting manipulation and arbitrage, is harmless to the principal in a broad class of circumstances.

It is still unknown what outcome could be implemented for two-agent nonlinear pricing environment when types are correlated and arbitrage is allowed. Meng and Tian (2015) recently show that it makes a big difference if types are positively or negatively related. Under negative correlation, the principal can exploit the conflict of interest inside the coalition to prevent collusion at no cost, i.e., CM’s result is still true. However, under positive correlation, however, the threat collusion forces the principal to distort the allocation away from the first-best
level obtained without collusion.

19.8 Characterization of Bayesian Implementability

This section investigates the necessary and sufficient conditions for full Bayesian implementability of any set of social choice functions in general economic environments (say, allowing interdependent types).

Assume agent $i$’s parametric utility function depends on type $\theta$, denoted by $u_i(x, \theta)$, where $x \in Z$ and $\theta_i \in \Theta_i$. The designer and agents know $\theta = (\theta_1, \cdots, \theta_n)$ has density function $\varphi(\theta)$ on $\Theta = \prod_{i \in N} \Theta_i$.

Let $X = \{x : \Theta \rightarrow A\}$ be the set of all feasible outcomes, and $\hat{F} \subseteq X$ a social choice rule that is the set of social choice functions $\hat{F} = \{f_1, f_2, \cdots\}$. When a social choice rule contains only one social choice function $\hat{F} = \{f\}$, it is called social choice function, denoted by $f$. We will see below that a set of social choice functions $\hat{F}$ in general differs from social choice correspondence, unless every status $\theta \in \Theta$ is common knowledge event and the set satisfies closureness.

Given a mechanism $\langle M, h \rangle$, like Nash implementation, Bayesian implementation also involves the relationship between $\hat{F}$ and $B(\Gamma)$.

**Definition 19.8.1** A mechanism $\Gamma = \langle M, h \rangle$ is said to fully Bayesian implements $\hat{F}$, if

(i) for every $f \in \hat{F}$, there is a Bayesian-Nash equilibrium $m^*$, such that $h(m^*) = f(\cdot)$;

(ii) If $m^*$is Bayesian-Nash equilibrium, then $h(m^*) \in \hat{F}$.

If such a mechanism exists, we call $\hat{F}$ is fully Bayesian implementable. If only condition (ii) is satisfied, we call $\hat{F}$ is strongly Bayesian implementable. When $\hat{F}$ contains only one social choice function, the above definition reduces to the definition for a social choice function is strongly Bayesian implementable.
Like implementation issue under other solution concepts, full Bayesian implementability needs to deal with two issues: Condition (i) requires to seek a mechanism such that any outcome determined by social choice rule is a Bayesian-Nash equilibrium outcome of the mechanism, while Condition (ii) requires all Bayesian-Nash equilibrium outcomes of the mechanism are the outcomes under social choice rules. As for Bayesian-Nash implementation, while the incentive compatibility requirement is central, it may not be sufficient for a mechanism to give all of desirable outcomes because of multiple equilibria. The severity of this multiple equilibrium problem is a critical issue to be solved in full or strong Bayesian implementation. Then, we need find conditions such that there exists a mechanism in which all Bayesian-Nash equilibrium outcomes are the outcomes under social choice rule.

We first consider an example given by Palfrey and Srivastava(1989), which shows that even for social choice function, how the issue of multiple Bayesian-Nash equilibria affects full Bayesian implementability.

**Example 19.8.1** Consider an exchange economy with two agents and two goods. Agent 1 has two preferences (types) $\theta_1$ and $\theta_1'$, and agent 2 only has one preference $\theta_2$. Agent 1’s preference is private information. Under preference profile $\theta = (\theta_1, \theta_2)$, Pareto efficient allocation is $x(\theta)$, under $\theta' = (\theta_1', \theta_2)$, Pareto efficient allocation is $x(\theta')$, see Figure 19.1. In the above two allocations $x(\theta)$ and $x(\theta')$, agent 1 with type $\theta_1$ likes $x(\theta)$ more, while agent 1 with $\theta_1'$ feels two allocations are indifferent, but $\theta_2$ likes $x(\theta')$ more. Consider social choice rule $f(\theta) = x(\theta)$ and $f(\theta') = x(\theta')$, it is Pareto efficient.

In the above exchange economy, consider the following mechanism $\Gamma = (\Theta, h(\cdot), m(\cdot))$: Each agent (mainly agent 1) reports his types $M_1 = \Theta_1 = \{\theta_1, \theta_1'\}, M_2 = \{\theta_2\}$. If reported message is $m = (\theta_1, \theta_2)$, then $h(\theta_1, \theta_2) = x(\theta)$; if reported message is $m = (\theta_1', \theta_2)$, then $h(\theta_1', \theta_2) = x(\theta')$. This mechanism has Bayesian-Nash equilibria: One is $m_1(\tilde{\theta}_1) = \tilde{\theta}_1, \forall \tilde{\theta}_1 \in \Theta_1; m_2(\theta_2) = \theta_2$, another is $m_1(\tilde{\theta}_1) = \theta_1', \forall \tilde{\theta}_1 \in \Theta_1; m_2(\theta_2) = \theta_2$. However, these two Bayeisan-Nash equilibria are equally desirable from Pareto optimality criterion. When the report is $\theta' = (\theta_1', \theta_2)$, the second
Bayesian-Nash equilibrium outcome is not Pareto efficient. This example shows that we have multiple equilibria problem in Bayesian implementation.

Figure 19.1: Bayesian implementation under exchange economies with two agents and two goods.

To eliminate undesirable Bayesian-Nash equilibria, we consider the following mechanism, see Table 19.1.

<table>
<thead>
<tr>
<th>Agent 1</th>
<th>θ₁</th>
<th>θ₁'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent 2</td>
<td>θ₂</td>
<td>ρ</td>
</tr>
<tr>
<td>x(θ)</td>
<td>x(θ')</td>
<td></td>
</tr>
<tr>
<td>x(θ')</td>
<td>x(θ)</td>
<td></td>
</tr>
</tbody>
</table>

In this mechanism, agent 2 has two different reports θ₂ and ρ. If agent reports ρ, then \( h(θ₁, ρ) = x(θ') \) and \( h(θ₁', ρ) = x(θ) \). Under this mechanism, \( m₁(θ₁) = θ₁ \) is not agent 1’s Bayesian-Nash equilibrium, or agent 2 would report ρ, resulting
outcome $x(\theta')$. This is because $x(\theta')$ is strictly preferred to $x(\theta)$ to agent 2. When agent 2 chooses $N$, the optimal report of agent 1 with $\theta_1$ will be $m_1(\theta_1) = \theta'_1$.

One can easily see the mechanism has two Bayesian-Nash equilibria:

Equilibrium 1: $m(\theta_1) = \theta_1, m(\theta'_1) = \theta'_1, m_2(\theta_2) = \theta_2$, its equilibrium outcome is $h(\theta_1, \theta_2) = x(\theta)$ and $h(\theta'_1, \theta_2) = x(\theta')$;

Equilibrium 2: $m(\theta_1) = \theta'_1, m(\theta'_1) = \theta_1, m_2(\theta_2) = \rho$, resulting in equilibrium outcome $h(\theta_1, \rho) = x(\theta)$ and $h(\theta'_1, \rho) = x(\theta')$.

Thus, although this mechanism has two Bayesian-Nash equilibria, the outcome at these two equilibria is the same as outcome under social choice rule, i.e., this mechanism fully Bayesian implements $f$.

For a general social choice set $\hat{F}$, Pastlewaite–Schmeidler (JET, 1986), Palfrey-Srivastava (RES, 1989), Mookherjee-Reichelstein (RES, 1990), Jackson (Econometrica, 1991), Dutta-Sen (Econometrica, 1994), Tian (Social Choices and Welfare, 1996; Journal of Math. Econ., 1999) provided necessary and sufficient conditions for a social choice set $\hat{F}$ to be fully Bayesian implementable. Under some technical conditions, they showed that a social choice set $\hat{F}$ is fully Bayesian implementable if and only if $\hat{F}$ is Bayesian incentive compatible (it deals with the first problem that any outcome determined by social choice rule is a Bayesian-Nash equilibrium outcome of the mechanism), and it is Bayesian monotonic (it solves the second problem: all Bayesian-Nash equilibrium outcomes of the mechanism are the outcomes under social choice rules).

We first discuss these conditions. It will be seen that Bayesian monotonicity is similar to Maskin monotonicity, using expected utility function replaces Bernoulli utility function.

**Definition 19.8.2** A mechanism $\Gamma = \langle M, h \rangle$ Bayesian implements social choice set $\hat{F}$, if there exists a Bayesian-Nash equilibrium $m^*$, such that $h(m^*) \in \hat{F}$. If such a mechanism exists, we call that $\hat{F}$ is Bayesian Implementable.

The following important concept of Bayesian incentive-compatibility solves the first problem of full Bayesian implementability.
Definition 19.8.3 A social choice set \( \hat{F} \) is said to be Bayesian incentive-compatibility or truthfully Bayesian implementable, if for all \( f \in \hat{F} \), truthtelling, i.e., \( m^*(\theta) = \theta \), \( \forall \theta \in \Theta \), is Bayesian-Nash equilibrium of a revelation mechanism \( \Gamma = (\Theta, f) \), i.e., for any \( f \in \hat{F} \), we have

\[
E_{\theta_i}[u_i(f(\theta_i, \theta_{-i})), \theta_i] \geq E_{\theta_i}[u_i(f(\theta'_i, \theta_{-i})), \theta_i], \quad \forall i, \forall \theta_i, \theta'_i \in \Theta_i.
\]

We then have the first necessary condition for full Bayesian implementability of a social choice set \( \hat{F} \).

Proposition 19.8.1 If social choice set \( \hat{F} \) is Bayesian implementable, then it satisfies Bayesian incentive-compatibility

Proof. Suppose mechanism \( \Gamma = (M, h) \) Bayesian implements social choice set \( \hat{F} \). If some social choice function \( f \in \hat{F} \) is not Bayesian incentive-compatible, then there is \( i \) and \( \theta_i, \theta'_i \in \Theta_i \), such that

\[
E_{\theta_i}[u_i(f(\theta_i, \theta_{-i})), \theta_i] < E_{\theta_i}[u_i(f(\theta'_i, \theta_{-i})), \theta_i]. \tag{19.8}
\]

Let \( m \in B(\Gamma) \) such that \( h \circ m = f \). When agent \( i \)'s type is \( \theta_i \), if he chooses \( m_i(\theta_i) \), then his expected utility is given by

\[
E_{\theta_i}[u_i(h(m_i(\theta_i)), \theta_i), \theta_i] = E_{\theta_i}[u_i(f(\theta_i, \theta_{-i})), \theta_i].
\]

If he chooses a message \( m'_i = m_i(\theta'_i) \), his expected utility is

\[
E_{\theta_i}[u_i(h(m_i(\theta'_i)), m_i(\theta_{-i})), \theta_i] = E_{\theta_i}[u_i(f(\theta'_i, \theta_{-i})), \theta_i].
\]

By (19.8), we know the agent has incentive to deviate from \( m_i(\theta_i) \), contradicting to \( m \in B(\Gamma) \).

We now discuss the second necessary condition for a social choice set to fully Bayesian implementable, i.e., Bayesian monotonicity condition. Consider a revelation mechanism, an agent \( i \), and a strategy \( \alpha_i : \Theta_i \to \Theta_i \). If agent \( i \) tell the truth, it implies \( \alpha_i(\theta_i) = \theta_i, \forall \theta_i \in \Theta_i \), otherwise \( \alpha_i \) is called a deception strategy.
of agent $i$. We call $\alpha(\theta) = (\alpha_1(\theta_1), \cdots, \alpha_n(\theta_n))$ is a deception, if at least one agent has a deception strategy.

Let

$$\alpha_{-i}(\theta_{-i}) = (\alpha_1(\theta_1), \cdots, \alpha_{i-1}(\theta_{i-1}), \alpha_{i+1}(\theta_{i+1}), \cdots, \alpha_n(\theta_n)).$$

For a social choice function $f$ and deception $\alpha$, $f \circ \alpha$ denotes the social choice under deception. When social status is $\theta$, social choice outcome is $f \circ \alpha(\theta) = f(\alpha(\theta))$. For every $\theta' \in \Theta$, define $f_{\alpha_i(\theta_i)}(\theta') \equiv f(\alpha_i(\theta_i), \theta'_{-i})$, i.e. it is the outcome when only agent $i$ chooses deception strategy.

Similar to Maskin monotonicity, we have the following Bayesian monotonicity condition.

**Definition 19.8.4 (Bayesian monotonicity)** A social choice set $\hat{F}$ is said to satisfy **Bayesian monotonicity**, if for any $f \in \hat{F}$ and deception $\alpha$ that results in $f \circ \alpha \notin \hat{F}$, there exists a agent $i$ and a function $y : \Theta_{-i} \rightarrow A$, such that

$$E[u_i(f(\theta_i, \theta_{-i}), \theta)|\theta_i] \geq E[u_i(y(\theta_{-i}), \theta)|\theta_i]$$

for all $\theta_i \in \Theta_i$, and for some $\theta'_{i}$, we have

$$E[u_i(f(\theta'_{i}, \theta_{-i}), \theta')|\theta'_{i}] < E[u_i(y(\theta_{-i}), \theta')|\theta'_{i}].$$

Bayesian monotonicity is an variation of Maskin monoticity under incomplete information. Its role is to avoid undesirable outcomes become Bayesian-Nash equilibria. Consider a mechanism $\Gamma = (M, h)$ that fully Bayesian implements a social choice set $\tilde{F}$, and a social choice function $f \in \tilde{F}$ that can be Bayesian implemented by Bayesian-Nash equilibrium $m$, i.e., for any $\theta \in \Theta$, $h(m(\theta)) = f(\theta)$. Suppose agent $i$ adopts a deception strategy $\alpha$. Then strategy $m \circ \alpha$ results in an outcome given by $f \circ \alpha$. If $f \circ \alpha \notin \tilde{F}$, then $m \circ \alpha$ is not Bayesian-Nash equilibrium. The inequality in Bayesian monotonicity condition (19.10) avoids the possibility that $m \circ \alpha$ becomes a Bayesian-Nash equilibrium, while another inequality (19.9) ensures no one has incentive to cheat.

We then have another necessary condition for full Bayesian implementability, i.e., Bayesian monotonicity.
Proposition 19.8.2  If a social choice set \( \hat{F} \) is fully Bayesian implementable, then \( \hat{F} \) must satisfy Bayesian monotonicity.

Proof. Suppose \( \Gamma = (M, h) \) fully Bayesian implements a social choice set \( \hat{F} \). Then, for any social choice function \( f \in \hat{F} \), there is a Bayesian-Nash equilibrium \( m^* \) such that \( f(\theta) = h(m^*(\theta)), \forall \theta \in \Theta \). Let \( \alpha \) be a deception such that \( f \circ \alpha \notin \hat{F} \).

Consider strategy \( m^* \circ \alpha \circ \alpha \). Under this strategy, for any \( \theta \in \Theta \), agents’s strategy profile is \( m^* \circ \alpha \circ \alpha(\theta) = m^*(\alpha(\theta)) \). Since \( f \circ \alpha \notin \hat{F} \), Full Bayesian implementability implies that \( m^* \circ \alpha \circ \alpha \) is not a Bayesian-Nash equilibrium, which implies that there is \( \theta_i' \in \Theta_i \) such that the agent has incentive to choose some \( m_i' \neq m_i^*(\alpha_i(\theta_i')) \) such that

\[
E[u_i(h(m_i', m_{-i}^*(\theta_{-i})), m_{-i}^*(\theta_{-i})))|\theta_i'] < E[u_i(h(m_i', m_{-i}^*(\alpha_{-i}(\theta_{-i})), m_{-i}^*(\theta_{-i})))|\theta_i].
\]  

(19.11)

Define \( y : \Theta_{-i} \to A \): \( y(\theta_{-i}) = h(m_i', m_{-i}^*(\theta_{-i})) \). We have

\[
y(\alpha_{-i}(\theta_{-i})) = h(m_{-i}^*(m_i', \alpha_{-i}(\theta_{-i}))).
\]

Thus, from the above inequality (19.11), we can obtained (19.10), while (19.9) comes from the fact that \( f \) can be Bayesian implemented by \( h \circ m^* \). In this case, for any \( \theta_i \), all agents have incentive to tell the truth. \( \square \)

Addition to Bayesian incentive-compatibility and Bayesian monotonicity, to enable them also to be sufficient conditions, some technic conditions are needed. For full implementable social choice set \( \hat{F} \), first they need satisfy closureness condition. We call a subset of a type space \( \Theta' \subseteq \Theta \) a common knowledge event, if for any \( \theta' = (\theta_i', \theta_{-i}') \in \Theta', \theta = (\theta_i, \theta_{-i}) \notin \Theta' \), we have \( \varphi(\theta_{-i}'|\theta_i) = 0, \forall i \).

If an agent does not know the true status, for all possible statuses, the agent need predict what messages the other agents will report. All such possible states consist of common knowledge, which are bases for reporting messages.

Definition 19.8.5 (Closureness of Social Choice Set) Let \( \Theta_1 \) and \( \Theta_2 \) be a partition of \( \Theta \), i.e., \( \Theta_1 \cap \Theta_2 = \Theta \) and \( \Theta_1 \cup \Theta_2 = \Theta \). A social choice set \( \hat{F} \) is said to have closureness, if for any \( f_1, f_2 \in \hat{F} \), there is \( f \in \hat{F} \), satisfying \( f(\theta) = f_1(\theta), \forall \theta \in \Theta_1 \); \( f(\theta) = f_2(\theta), \forall \theta \in \Theta_2 \).

776
If every state $\theta \in \Theta$ is common knowledge, it reduces to economic environment with complete information discussed in last chapter, and thus a social choice set that satisfies closureness becomes social choice correspondence. If a social choice set $\hat{F}$ does not satisfy closureness, say, $\Theta = \{\theta, \theta'\}$, and every state is common knowledge, $\hat{F} = \{f_1, f_2\}$ satisfies $f_1(\theta) = f_2(\theta') = a; f_1(\theta') = f_2(\theta) = b, a \neq b$, then the social choice set $\hat{F}$ is not fully Bayesian implementable. This is because, if $\hat{F}$ is fully Bayesian implementable, we need they are Bayesian-Nash equilibrium under two states $a$ and $b$, and then by the definitions of $f_1$ and $f_2$, no way ensures implementable outcomes are different under two different states. We should notice that social choice set does not equal to social choice correspondence, i.e., $\hat{F} \neq F$, where $F(\theta) = F(\theta') = \{a, b\}$.

Jackson (1991) characterizes the sufficient condition for full Bayesian implementation of a social choice set in incomplete information environments. We first give the definition of economic environment in incomplete information. 

**Definition 19.8.6** An environment is called a *economic environment*, if for any state $\theta \in \Theta$ and any outcome $y \in Y$, there are two agents $i$ and $j$, and two outcomes $y_i$ and $y_j$, we have

$$u_i(y_i, \theta) > u_i(y, \theta)$$

and

$$u_j(y_j, \theta) > u_j(y, \theta).$$

Intuitively, for incomplete information economic environments, for any social choice function and state, there are at least two agents who hope to change social choice outcome under the state, which implies that a social choice outcome cannot make all agents reach their satiated points. If every utility functions are monotonic in private goods economies, this condition is satisfied. Such a condition is also satisfied for public good economies and economies with externalities.

We now state the following theorem given by Jackson (1991) without proof. 给出
Proposition 19.8.3 (Necessity and Sufficiency for Full Bayesian Implementability) For economic environments with \( N \geq 3 \) agents, suppose that a social choice set \( \hat{F} \) satisfies closureness condition. Then, \( \hat{F} \) is fully Bayesian implementable if and only if it satisfies Bayesian incentive-compatibility and Bayesian monotonicity.

Under more general environments, Jackson (1991) strengthens Bayesian monotonicity by introducing monotonicity-no-veto-power condition and shows that in economic environments with \( N \geq 3 \) agents, if a social choice function Bayesian incentive-compatibility and monotonicity-no-veto-power condition, it is fully Bayesian implementable. For environments with two agents, Dutta and Sen (1994) provide sufficient conditions for a social choice function to fully Bayesian implementable.

One can similarly investigate refinement of Bayesian implementation and virtual Bayesian implementation such as in Palfrey–Srivastava (JPE, 1989) as we discuss in the earlier section of this chapter.

19.9 Reference

Books and Monographs


Papers


Chapter 20

Dynamic Mechanism Design

20.1 Introduction

In Chapters 16-19, we discussed the optimal contract design and the general mechanism design, which have a common feature. Contracts, mechanisms and institutions are to a large extent synonymous. They all mean rules of the game, which describe what actions the parties can undertake, and what outcomes these actions would entail. Indeed, almost all games observed in daily life are not given by nature, but designed by someone or organization: in chess, basketball, football. Constitution is another typical example of mechanism design. The rules are designed to achieve desired outcomes: given an institutional environment and certain constraints faced by the designer, what rules are feasible? What mechanisms are optimal among those that are feasible?

While mechanisms design theory considers designing rules of more general games such as institutional design, contract theory proved useful for more manageable smaller questions, concerning specific contracting practices. In addition, a contract is characterized by the following features: (1) a contract is designed by one of the parties themselves; (ii) participation is voluntary; (iii) parties may be able to renegotiate the contract later on.

In practical environments, people often repeatedly interact each other, which
is different from one time interaction. In this chapter, we discuss the dynamics of incentive contracts and dynamic mechanism design in different backgrounds.

20.2 Dynamic Contracts under Full Commitment

In this section we discuss a simple dynamic contracting problem, where the principal has perfect commitment capacity. We derive second best dynamic contracts in three cases: (1) constant types, (2) independent types, and (3) correlated types over time.

20.2.1 Constant Types

For simplicity, consider a situation of monopoly selling in two periods, where the type of the agent (or consumer) is constant over time. We discuss the second best contracting of the principal (or monopolist). Let us first start with the simplest case where the agent’s type, $\theta \in \{\theta_H, \theta_L\}$, is constant over time. As in Chapter 16, we assume that the probability of type $\theta_L$ is $\beta$.

The timing of contracting is described in figure (20.1). At time 0, the agent learns his type $\theta$; at time 0.25, the principal offers a long term contract $(q_1, T_1(q_1); q_2, T(q_1, q_2))$ to the agent for both periods; at time 0.5, the agent decides whether to accept the contract or not, the interaction continues if accepting and is over otherwise; in period 1, the agent buys $q_1$ units, and pays for $T_1(q_1)$; in period 2, the agent buys $q_2$ units, and pays for $T(q_1, q_2)$ to the principal, where the transfer is dependent on the current and the past activity.

The agent’s utility function is:

$$u(\theta, q_1, q_2, T_1, T_2) = \theta v(q_1) - T_1 + \delta [\theta v(q_2) - T_2(q_1, q_2)],$$

where $\delta$ is the discount rate, and his status quo utility level is 0.
The principal’s utility function is:

\[ T_1 - cq_1 + \delta[T_2 - cq_2]. \]

As the previous discussion of static adverse selection, if there is only one period, we know the second best contract is described as \((q_{SB}^H, T_{SB}^H; q_{SB}^L, T_{SB}^L)\), where \(q_{SB}^H = q_{FB}^H\) and \(q_{SB}^L\) satisfy

\[ \theta_H v'(q_{SB}^H) = c, \]
\[ \theta_L v'(q_{SB}^L) = \frac{c}{1 - (\frac{1-\beta}{\beta^2} \theta_H - \theta_L)}. \]

\(T_{SB}^H\) and \(T_{SB}^L\) satisfy

\[ \theta_H v(q_{SB}^H) - T_{SB}^H = U_H = \Delta \theta v(q_{SB}^L), \]
\[ \theta_L v(q_{SB}^L) - T_{SB}^L = U_L = 0. \]

This is the result we discussed in Chapter 16 on nonlinear pricing that was first studied by Maskin and Riley (Rand J. of Economics, 1984).

If the principal has full commitment power, the second best contract for two periods is twice the repetition of the static second-best contract as shown below. Since the principal can commit intertemporally, she can omit the information learned in period one, and strictly enforce the static second-best contract. In this case, the revelation principle remains valid. In the following, we discuss this long-term second-best contract, denoted by \((q_1, q_2, T)\), where \(T\) is the two-period
discounted transfer, i.e., \( T = T_1 + \delta T_2 \). The complete form of the contract is \((q_{1H}, q_{2H}, T_H; q_{1L}, q_{2L}, T_L)\). If the contract is incentive feasible, then the below four conditions are satisfied:

\[
U_H \equiv \theta_H(v(q_{1H}) + \delta v(q_{2H})) - T_H \geq U_L + \Delta \theta(v(q_{1L}) + \delta v(q_{2L})), \quad (20.1)
\]

\[
U_L \equiv \theta_L(v(q_{1L}) + \delta v(q_{2L})) - T_L \geq U_H - \Delta \theta(v(q_{1H}) + \delta v(q_{2H})); \quad (20.2)
\]

\[
U_H \geq 0, \quad (20.3)
\]

\[
U_L \geq 0. \quad (20.4)
\]

The above can be described the following constrained optimal problem:

\[
\max_{(q_{1H}, q_{2H}, U_H; q_{1L}, q_{2L}, U_L)} (1 - \beta)\left[ \theta_H(v(q_{1H}) + \delta v(q_{2H})) - U_H - c(q_{1H} + \delta q_{2H}) \right]
\]

\[
+ \beta \left[ \theta_L(v(q_{1L}) + \delta v(q_{2L})) - U_L - c(q_{1L} + \delta q_{2L}) \right] \quad (20.5)
\]

and the same time satisfying the conditions of (20.1), (20.2), (20.3), and (20.4).

As the same precious logic, only the conditions of (20.1) and (20.4) are binding, so we have: \( U_L = 0, \ U_H = \Delta \theta(v(q_{1L}) + \delta v(q_{2L})) \). Substituting them into the equation (20.5), we get the first order conditions for \( q_{1H}, q_{2H}; q_{1L}, q_{2L} \) respectively:

\[
\theta_H v'(q_{1H}^{SB}) = \theta_H v'(q_{2H}^{SB}) = c,
\]

\[
\theta_L v'(q_{1L}^{SB}) = \theta_L v'(q_{2L}^{SB}) = \frac{c}{1 - \left(1 - \frac{\beta}{\beta} \frac{\theta_H - \theta_L}{\theta_L} \right)}.
\]

Hence, the second-best long-term contract with full commitment for two periods is twice the repetition of the static second-best contract.

### 20.2.2 Independent Dynamic Types

Now we discuss the other extreme case of dynamic contracting, i.e., independent type over time.

The timing of contracting with agent’s independent type is described in figure (20.2). Comparing the figure (20.1), at time 0, the agent only learns his type in period 1, and only at the time 1.5, he can learn his type in period 2. The
type distribution of the consumer is i.i.d, the probability of $\theta_L$ is $\beta$, and the other setting is the same to the case of constant type.

We start with the interaction in the period 2. When $\theta_1$ is the agent’s first-period announcement on his type, then the contract $(q_{2H}(\theta_1), T_{2H}(\theta_1); q_{2L}(\theta_1), T_{2H}(\theta_1))$ in the second period depends on $\theta_1$, and the following constraints have to be specified: contract in period 2 should satisfy incentive compatibility

$$U_{2H}(\theta_1) \equiv \theta_H(\theta_1)v(q_{2H}(\theta_1)) - T_{2H}(\theta_1) \geq U_{2L}(\theta_1) + \Delta \theta v(q_{2L}(\theta_1)), \quad (20.6)$$

$$U_{2L}(\theta_1) \equiv \theta_L v(q_{2L}(\theta_1)) - T_{2L}(\theta_1) \geq U_{2H}(\theta_1) - \Delta \theta v(q_{2H}(\theta_1)), \quad (20.7)$$

where $U_{2H}(\theta_1)$ and $U_{2L}(\theta_1)$ are the second-period equilibrium utility for $\theta_H$ and $\theta_L$ respectively.

The participation constraint at second period is not binding since in two periods interaction under full commitment, if the agent’s total discounted utility is not below the status quo utility, he should accept the two-period contracting. As such, $(q_{1H}, T_{1H}, q_{2H}(\theta_H), T_{2H}(\theta_H))$ and $(q_{1L}, T_{1L}, q_{2L}(\theta_L), T_{2L}(\theta_L))$ constitutes a two-period incentive compatible contract.

Let us also denote the first-period rents by $U_{1H} = \theta_H v(q_{1H}) - T_{1H}$ and $U_{1L} = \theta_L v(q_{1L}) - T_{1L}$. At date 1, beside the conditions of (20.6) and (20.7), the feasible
incentive constraints at the first-period should include the following conditions:

\[ U_{1H} + \delta[(1 - \beta)U_{2H}(\theta_H) + \beta U_{2L}(\theta_H)] \geq U_{1L} + \Delta \theta v_{q1L} + \delta[(1 - \beta)U_{2H}(\theta_L)] + \beta U_{2L}(\theta_L); \quad (20.8) \]

\[ U_{1L} + \delta[(1 - \beta)U_{2H}(\theta_L) + \beta U_{2L}(\theta_L)] \geq U_{1H} - \Delta \theta v_{q1H} + \delta[(1 - \beta)U_{2H}(\theta_H)] + \beta U_{2L}(\theta_H), \quad (20.9) \]

\[ U_{1H} + \delta[(1 - \beta)U_{2H}(\theta_H) + \beta U_{2L}(\theta_H)] \geq 0, \quad (20.10) \]

\[ U_{1L} + \delta[(1 - \beta)U_{2H}(\theta_L) + \beta U_{2L}(\theta_L)] \geq 0, \quad (20.11) \]

where the inequality functions (20.8) and (20.9) are the incentive compatibility constraints, and (20.10) and (20.11) the participation constraints.

Obviously, only the constraints of (20.8) and (20.11) are binding at the optimum of the principal’s problem, and thus we have:

\[ \delta[(1 - \beta)U_{2H}(\theta_L) + \beta U_{2L}(\theta_L)] = -U_{1L}, \quad (20.12) \]

\[ \delta[(1 - \beta)U_{2H}(\theta_H) + \beta U_{2L}(\theta_H)] = -U_{1H} + \Delta \theta v(q_{1L}). \quad (20.13) \]

Thus, the incentive feasibility of second-period contract means the ex ante participation constraints, i.e. (20.12) and (20.13) should be binding.

Then, following the same logic of adverse selection under ex ante participation constraints discussed in Chapter 16, there is no allocative distortion, i.e., \( q_{2H}(\theta_H) = q_{2H}(\theta_L) = q_{H}^* \) and \( q_{2L}(\theta_H) = q_{2L}(\theta_L) = q_{L}^* \). At the same time, the second period rents are given by

\[ U_{2H}(\theta_L) = \frac{-U_{1L}}{\delta} + \beta \Delta \theta v(q_{H}^*), \quad (20.14) \]

\[ U_{2L}(\theta_L) = \frac{-U_{1L}}{\delta} - (1 - \beta)\Delta \theta v(q_{H}^*), \quad (20.15) \]

\[ U_{2H}(\theta_H) = \frac{-U_{1H} + \Delta \theta v(q_{SB}^L)}{\delta} + \beta \Delta \theta v(q_{H}^*), \quad (20.16) \]

\[ U_{2L}(\theta_H) = \frac{-U_{1H} + \Delta \theta v(q_{SB}^L)}{\delta} - (1 - \beta)\Delta \theta v(q_{H}^*). \quad (20.17) \]

Hence, the optimal consumptions corresponding to the inefficient draws of types in both periods are such that \( q_{1L} = q_{SB}^L \) and \( q_{2L} = q_{L}^* \), respectively. The agents gets
a positive rent only when his type is $\theta_H$ at date 1, and his expected intertemporal informational rent over both periods is $\Delta \theta v(q^{SB}_L)$, which is positive.

We then have the following proposition.

**Proposition 20.2.1** With independent types and a risk-neutral agent, the optimal long-term contract for two periods with full commitment combines the optimal static contract written interim for period 1 and the optimal static contract written ex ante for periods 2. In particular, the expected rent of the agent only equals the expectation of the agent’s rent when he is efficient in period one and is worth $\Delta \theta v(q^{SB}_L)$, which is positive.

The same result would also be obtained if the risk-neutral agent can drop the contract at the period 2 when he does not get a positive rent in the second period. In this case, the participation constraint is needed, i.e.,

$$(1 - \beta)U_{2H}(\theta) + \beta U_{2L}(\theta) \geq 0, \forall \theta \in \{\theta_H, \theta_L\}.$$ 

Thus, it is enough to $U_{1L} = 0$ and $U_{1H} = \Delta \theta v(q^{SB}_L)$ so that the rights-hand sides of (20.12) and (20.13) equal to zero.

**20.2.3 Correlated Types**

Let us generalize the previous information structures and turn now to the more general case where the agent’s type are imperfectly correlated over time. In this case, there are new features of second-best dynamic contracting. The problem firstly were studied by Baron and Besanko (1984), in which they derived the second-best contracts with correlated over time and full commitment of the principal.

In this subsection, we remain the hypothesis of full commitment, and study the long-term contract in the background of intertemporal price discrimination. In this case, the agent learned his first-period type $\theta_1 \in \{\theta_H, \theta_L\}$, and the principal only know its distribution. The agent’s second-period type is imperfectly
correlated to the first-period one, we assume their correlation is:

\[ \beta_i = \text{prob}(\theta_2 = \theta_H|\theta_i) \text{ for } i = H, L. \]

We also suppose that \( \beta_H \geq \beta_L \), which means the intertemporal correlation is positive, i.e., when type at period 1 is high-type, the probability that the type is a high-type is also higher than the probability that the type is a low-type at period 2. When \( \beta_H = 1 > \beta_L = 0 \), it is equivalent to the case of constant types; when \( \beta_H = \beta_L = \beta \), it is equivalent to the case of independent types.

The timing of the contract is the same as in figure (20.2). In this framework, a direct revelation mechanism requires that the agent report the new information in each period he has learned on his current type. Typically, a direct revelation mechanism is a four-tuple \( \{(T_1(\tilde{\theta}_1), q_1(\tilde{\theta}_1)), (T_2(\tilde{\theta}_1, \tilde{\theta}_2), q_2(\tilde{\theta}_1, \tilde{\theta}_2))\} \) for all pair \( (\tilde{\theta}_1, \tilde{\theta}_2) \), where \( \tilde{\theta}_1 \) (resp. \( \tilde{\theta}_2 \)) is the date 1 (resp. data 2) announcement on his first-period (resp. second-period) type.

The important point to note here is that the first-period report can now be used by the principal to update his beliefs on the agent’s second period type. This report can be viewed as an informative signal that is useful for improving second-period contracting. The difference is that now the signal used by the principal to improve second-period contracting is not exogenously given by nature but comes from the first-period report \( \tilde{\theta}_1 \) of the agent on his type \( \theta_1 \). Hence, this signal can be strategically manipulated by the agent in the first period in order to improve his second-period rent.

If the agent’s type \( \theta_i \) in period \( t \) accepts the contract \( (T_t, q_t) \), his utility in period \( t \) is \( U_{it} = \theta_i v(q_t) - T_t \). We assume that \( v(q_t) \) is concave, and further the agent is infinitely averse. Assume that the common discount rate is \( \delta \). In deriving the second-best long-term contract, the principal’s objective is to maximize her total discount utility while the contract is subject to some intertemporal incentive feasible constraints.

We first discuss the second-period constraints, given the agent’s first-period report \( \tilde{\theta}_1 \). Because the agent is infinitely risk averse below zero wealth, his ex
Post participation constraint in period 2 is written as:

\[
\begin{align*}
U_{H2}(\hat{\theta}_1) &= \theta_H v(q_{H2}(\hat{\theta}_1)) - T_{H2}(\hat{\theta}_1) \geq 0, \quad (20.19) \\
U_{L2}(\hat{\theta}_1) &= \theta_L v(q_{L2}(\hat{\theta}_1)) - T_{L2}(\hat{\theta}_1) \geq 0. \quad (20.20)
\end{align*}
\]

Moreover, inducing information revelation by the agent in period 2 requires to satisfy the following incentive constraints:

\[
\begin{align*}
U_{H2}(\hat{\theta}_1) &= \theta_H v(q_{H2}(\hat{\theta}_1)) - T_{H2}(\hat{\theta}_1) \geq U_{L2}(\hat{\theta}_1) + \Delta \theta v(q_{L2}(\hat{\theta}_1)), \quad (20.21) \\
U_{L2}(\hat{\theta}_1) &= \theta_L v(q_{L2}(\hat{\theta}_1)) - T_{L2}(\hat{\theta}_1) \geq U_{H2}(\hat{\theta}_1) - \Delta \theta v(q_{H2}(\hat{\theta}_1)). \quad (20.22)
\end{align*}
\]

From the equations of (20.21) and (20.22), we get \(q_{H2}(\hat{\theta}_1) \geq q_{L2}(\hat{\theta}_1)\), and thus the monotonic condition remains valid.

Given the agent’s first-period announcement \(\hat{\theta}_1\), in deriving the second-best contract, the principal is to solve the following problem:

\[
\begin{align*}
\pi_2(\hat{\theta}_1, q_{H2}(\hat{\theta}_1), q_{L2}(\hat{\theta}_1)) &= \max \beta(\hat{\theta}_1)(T_{H2}(\hat{\theta}_1) - cq_{H2}(\hat{\theta}_1)) \\
&\quad + (1 - \beta(\hat{\theta}_1))(T_{L2}(\hat{\theta}_1) - cq_{H2}(\hat{\theta}_1)), \quad (20.23)
\end{align*}
\]

subject to the constraints of (20.21), (20.22), (20.19), and (20.20).

Again, only the constraints of (20.21) and (20.20) are binding, and then the second-period monopolist’s profit is:

\[
\begin{align*}
\pi_2(\hat{\theta}_1, q_{H2}(\hat{\theta}_1), q_{L2}(\hat{\theta}_1)) &= \beta(\hat{\theta}_1)(\theta_H v(q_{H2}(\hat{\theta}_1)) - cq_{H2}(\hat{\theta}_1)) + \\
&\quad (1 - \beta(\hat{\theta}_1))(\theta_L v(q_{L2}(\hat{\theta}_1)) - cq_{L2}(\hat{\theta}_1)) - \\
&\quad \beta(\hat{\theta}_1)\Delta \theta v(q_{L2}(\hat{\theta}_1)).
\end{align*}
\]

Let us now consider period 1. The type \(\theta_i\)-agent’s rent in period 1 is:

\[
\begin{align*}
U_{H1} &= \theta_H v(q_{H1}) - T_{H1}, \\
U_{L1} &= \theta_L v(q_{L1}) - T_{L1},
\end{align*}
\]

where \(q_{i1} = q_i(\theta_i), T_{i1} = T_1\theta_i, i \in \{H, L\}\).
The type $\theta_i$ agent’s second-period rent is:

\[ EU_2(\theta_i) = \beta_i \Delta \theta v(q_{L2}(\theta_i)), \]

\[ EU_2(\theta_i) = \beta_i \Delta \theta v(q_{L2}(\theta_i)). \]

Thus, the incentive compatibility constraints can be written as:

\[ U_H^1 + \delta \beta_H \Delta \theta v(q_{L2}(\theta_H)) \geq U_L^1 + \Delta \theta v(q_{L1}) + \delta \beta_H \Delta \theta v(q_{L2}(\theta_H)), \quad (20.24) \]

\[ U_L^1 + \delta \beta_L \Delta \theta v(q_{L2}(\theta_L)) \geq U_H^1 - \Delta \theta v(q_{H1}) + \delta \beta_L \Delta \theta v(q_{L2}(\theta_H)). \quad (20.25) \]

From the inequality functions of (20.24) and (20.25), we get:

\[ \delta \Delta \theta (v(q_{L2}) - v(q_{H2}))(\beta_H - \beta_L) + \Delta \theta (v(q_{H1}) - v(q_{L1})) \geq 0. \]

From the above, we have $q_{H2} \geq q_{L2}$, and thus: $q_{H1} \geq q_{L1}$. Hence, at the both periods, the incentive compatible consumption plans both satisfy the monotonic condition.

Due to the agent’s infinite risk aversion, his participation constraint in period one is written as

\[ U_H^1 \geq 0, \quad (20.26) \]

\[ U_L^1 \geq 0. \quad (20.27) \]

Then, in designing whole long-term second-best contract, the principle is to solve the following problem:

\[ \pi = \max \beta (T_H^1 - c q_{H1}) + (1 - \beta)(T_L^1 - c q_{L1}) + \delta[\beta \pi_2(\theta_H, q_{H2}(\theta_H), q_{L2}(\theta_H))] + (1 - \beta)\pi_2(\theta_L, q_{H2}(\theta_L), q_{L2}(\theta_L))] \]

subject to the constraints of (20.24) and (20.25), (20.26), and (20.27).

From the same logic, only the constraints of (20.24) and (20.27) are binding.
The above optimal problem can be rewritten as:

\[
\pi = \max_{\{(q_{1L}, q_{2L}(\theta_L)), (q_{1H}, q_{2H}(\theta_H))\}} \beta(\theta Hv(q_{1H}) - cq_{1H} - U_{H1}) + (1 - \beta)(\theta Lv(q_{1L}) - cq_{1L}) + \\
\delta[\beta\pi_2(\theta_H, q_{2H}(\theta_H), q_{2L}(\theta_H)) + (1 - \beta)\pi_2(\theta_L, q_{2H}(\theta_L), q_{2L}(\theta_L))],
\]

where

\[
U_{H1} = \Delta \theta v(q_{1L}) + \delta \beta_H \Delta \theta (v(q_{2L}(\theta_L)) - v(q_{2L}(\theta_H))),
\]

and \(\pi_2(\theta_H, q_{2H}(\theta_H), q_{2L}(\theta_H)), \pi_2(\theta_L, q_{2H}(\theta_L), q_{2L}(\theta_L))\) is derived from the inequality (20.24).

The first order condition is the following:

\[
\theta Hv'(q_{1H}) = c. \tag{20.28}
\]

Hence, \(q_{1H} = q^*_H\), which means that for the type \(\theta_H\)-agent, his first-period consumption is efficient (first best). From the following equation

\[
(1 - \frac{\Delta \theta}{\theta_L} \frac{\beta}{1 - \beta}) \theta Hv'(q_{1H}) = c, \tag{20.29}
\]

we have \(q_{1L} = q^*_{SP_H}\), and thus, for the type \(\theta_L\) agent, his first-period is not efficient, and downward distortion exists.

At the same time, because

\[
\theta Hv'(q_{2H}(\theta_H)) = c, \tag{20.30}
\]
\[
\theta Hv'(q_{2H}(\theta_L)) = c, \tag{20.31}
\]

we have \(q_{2H}(\theta_H) = q_{H2}(\theta_L) = q^*_H\), and then for the second-period type \(\theta_H\) agent, his consumption is also efficient.

Also, by

\[
\theta Lv'(q_{2L}(\theta_H)) = c, \tag{20.32}
\]

we get \(q_{2L}(\theta_H) = q^*_L\). Thus, for both the first-period type \(\theta_H\) and the second-period type \(\theta_L\), their consumption are efficient too.
Since
\[ \left(1 - \frac{\Delta \theta}{\theta_L} (1 - \beta)\theta_L + \beta \beta_H \right) \theta_L v'(q_{L2}(\theta_L)) = c, \tag{20.33} \]
at both periods, for the type $\theta_L$ agent, his consumption $q_{L2}(\theta_L)$ are equivalent to
the case of constant types, i.e., $\beta_H = 1, \beta_L = 0$, and then we have
\[ \left(1 - \frac{\Delta \theta}{\theta_L} (1 - \beta)\theta_L \frac{1 - \beta}{1 - \theta_L} \right) \theta_L v'(q_{L2}(\theta_L)) = c, \]
we get the second-best consumption, $q_{L2}(\theta_L) = q_{L1} = q_{S\beta}^H$.

As the case of independent types, i.e., $\beta_H = \beta_L = \beta$, we have:
\[ \left(1 - \frac{\Delta \theta}{\theta_L} \frac{\beta}{1 - \beta} \right) \theta_L v'(q_{L2}(\theta_L)) = c. \]

Hence, we have $q_{L2}(\theta_L) < q_{L1} = q_{S\beta}^H$.

### 20.3 Dynamic Contracts under Different Commitment Power

Now we discuss how the different degree of commitment affects the agent’s incentives. If the principal has no full commitment power, the agent may have no incentives to reveal his type in early time, since the principal can use her information to reduce the agent’s rent in later time. In the regulation background, the regulated price is dependent on the cost reported by a regulated monopolist. If the cost information is used in the future regulation, there is phenomena of rachet effect, which destroys the revelation incentives of true information by the regulated monopolist. This problem is deeply analyzed by Freixas, Guesnerie and Tirole (1985).

In this section, we first assume that the principal (designer) can commit to a contract forever, and then consider what happens when she cannot commit against modifying the contract as new information arrives. We discuss the dynamic contracting of monopoly selling. Hart and Tirole (1988) analyzed this problem in detail, and we adopt the simplified version from Segal (2010).
For this purpose, we consider a principal-agent relationship that is repeated over time, over a finite or infinite horizon. First, as a benchmark, we suppose that the agent’s type $\theta$ is realized ex ante and is then constant over time. For simplicity, we focus on a 2-period P-A model with a constant type and stationary payoffs. That is, suppose the payoffs are stationary. The common discount rate is $\delta$.

The agent’s type $\theta$ is realized before the relationship starts and is the same in both periods. Suppose the type space is $\{\theta_H, \theta_L\}$, $\theta_H > \theta_L > 0$, and the probability of $\theta_H$ is $\beta$. We will allow $\delta$ to be smaller or greater than one, the latter can be interpreted as capturing situations in which the second period is very long.

The outcome of contracting in this model depends on the degree of the principal’s ability to commit not to modify the contract after the rest period. If the principal cannot commit, she may modify the optimal commitment contract using the information revealed by the agent in the rest period. To what extent can the principal commit not to modify the contract and avoid the ratchet and renegotiation problems? The literature has considered three degrees of commitment:

1. **Full Commitment**: The principal can commit to any contract ex-ante. We have already considered this case. The Revelation Principle works, and we obtain a simple replication of the static model.

2. **Long Term Renegotiable Contracts**: The principal cannot modify the contract unilaterally, but the contract can be renegotiated if both the principal and the agent agree to do it. Thus, the principal cannot make any modifications that make the agent worse off, but can make modifications that make the agent better off. Thus, the ratchet problem does not arise in this setting (the high type would not accept a modification making him worse off), but the renegotiation problem does arise.

3. **Short-Term Contracts**: The principal can modify the contract unilaterally after the rest period. This means that the ex ante contract has no effect.
in the second period, and after the rest period the parties contract on the second-period outcome. This setting gives rise to both the ratchet problem and the renegotiation problem.

A key feature of the cases without commitment is that when $\delta$ is sufficiently high, the principal will no longer want the agent to reveal his type fully in the rest period, she prefers to commit herself against contract modification by having less information at the modification stage. For intermediate levels of $\delta$, the principal will prefer to have partial revelation of information in the rest period, which allows her to commit herself against modification. This partial revelation will be achieved by the agent using a mixed strategy, revealing his type with some probability and pooling with some probability.

For simplicity, we assume the consumer (agent) has unit demand of the good. Denote $x_{it} \in \{0, 1\}$ as the purchase decision of type $i$ agent. Suppose the production cost is normalized to 0, $p_t$ is the price in time $t$.

The total discounted utility of the type $i$ agent is:

$$U(x_{i1}, x_{i2}) = \sum_{t=1}^{2} \delta^{t-1}(\theta_{i} x_{it} - p_{t}).$$

The total discounted profit of the principal is

$$\pi(p_1, p_2) = \sum_{t=1}^{2} \delta^{t-1} p_{t} [\beta x_{H,t} + (1 - \beta)x_{L,t}],$$

where $p_t$ is the price of the good at time $t$.

### 20.3.1 Contracting with Full Commitment

As a benchmark, we first discuss the simplest case, i.e. the one-period optimal contract.

Note that if $p \leq \theta_{L}$, $x_{H} = x_{L} = 1$, and the profit is $p$; if $\theta_{L} < p \leq \theta_{H}$, $x_{H} = 1, x_{L} = 0$, and the profit is $\beta p$; if $p > \theta_{H}$, $x_{H} = x_{L} = 0$, and the profit is 0.
Let \( \bar{\beta} = \frac{\theta_L}{\theta_H} \). Then the principal’s optimal price then is:

\[
p^*(\beta) = \begin{cases} 
\theta_H & \text{if } \beta > \bar{\beta}, \\
\theta_L & \text{if } \beta < \bar{\beta}, \\
\theta_H \text{ or } \theta_L & \text{if } \beta = \bar{\beta}.
\end{cases}
\]

and the agent’s choice is:

\[
(x_H^*, x_L^*) = \begin{cases} 
(1, 0) & \text{if } \beta > \bar{\beta}, \\
(1, 1) & \text{if } \beta < \bar{\beta}, \\
(1, 0) \text{ or } (1, 1) & \text{if } \bar{\beta}.
\end{cases}
\]

The principal’s one-period profit is:

\[
\pi_1^*(\beta) = \max\{\beta \theta_H, \theta_L\}.
\]

If \( \beta < \bar{\beta} \), the utility of type \( \theta_H \) is \( U(\theta_H) = \theta_H - \theta_L \), which is also the rent of the type \( \theta_H \).

Suppose now that there are two periods. We already know that any two-period contract can be replaced with a repetition of the same static contract. Therefore, the optimal commitment contract sets the same price, i.e., \( p_t^* = p^*, t = 1, 2 \). The principal’s profit is \( \pi^*(\beta) = (1 + \delta) \pi_1^*(\beta) \), which is the most profit as she can earn.

### 20.3.2 Dynamic Contracting with No Commitment

If the principal lacks full commitment, will she want to modify this contract after the rest period? This depends on \( p^*(\beta) \). If \( \beta \leq \bar{\beta} \), and therefore \( p_1^*(\beta) = \theta_L \), then the two types pool in the first period, and the principal receives no information. As such, she has no incentives to modify the same optimal static contract in the second period.

However, when \( \beta > \bar{\beta} \), the monopolist may have incentives to modify the contract unilaterally. Indeed, if the consumer does not purchase the good, i.e., \( x_1 = 0 \), the monopolist know the consumer is \( \theta_L \), and then she will modify the contract so that \( p_2^*(\beta) = \theta_H \). Once the principal knows following a purchase that
she deals with a high type and following no purchase that she deals with a low type, she wants to restore efficiency for the low type by the price to $\theta_L$ in the second period. Expecting this price reduction following no purchase, the high type will not buy in the rest period, and then the contract is not implementable. Thus, when the principal deals optimally with the renegotiation problem, the ratcheting problem may arise and further hurt the principal. As result, there are only short-term contracts in this situation.

In the following, we suppose that $\beta > \bar{\beta}$. Let the parties play the following two-period game: (1) Principal offers price $p_1$; (2) Agent chooses $x_1 \in \{0, 1\}$; (3) Principal offers price $p_2$; (4) Agent chooses $x_2 \in \{0, 1\}$.

We solve for the principal’s preferred weak Perfect Bayesian Equilibrium (PBE) in this extensive-form game of incomplete information. We do this by considering possible continuation PBEs following different rest-period price choices $p_1$. Then the principal’s optimal PBE can be constructed by choosing the optimal continuation PBE following any price choice $p_1$, and finding the price $p_1^*$ that gives rise to the best continuation PBE for the principal.

The principal’s strategy in the continuation game is a function $p_2(x_1)$, which sets the second-period price following the agent’s first-period choice $x_1$. Let $\hat{\beta}(x_1) = prob(\theta) = prob(\theta_H|x_1)$ be the principal’s posterior belief that the agent is a high type after $x_1$ is chosen. In the following, we analyze three possible equilibrium types: “full revelation” equilibrium, “no revelation” equilibrium and “partial revelation” equilibria.

“Full Revelation” Equilibrium:

Under the first-period price $p_1$, there is only one possible “full revelation” (or separating) equilibrium: $x_H = 1$ and $x_L = 0$. The principal’s posterior belief is given by

$$\hat{\beta}(x_1) = \begin{cases} 1 & \text{if } x_1 = 1, \\ 0 & \text{if } x_1 = 0. \end{cases}$$
The principal’s second-period price is:

\[ p_2(x_1) = \begin{cases} 
\theta_H & \text{if } x_1 = 1, \\
\theta_L & \text{if } x_1 = 0. 
\end{cases} \]

Therefore, if such equilibrium exists, the incentive compatibility constraints for the types \( \{\theta_H, \theta_L\} \) should hold:

\[ \theta_H - p_1 \geq \delta(\theta_H - \theta_L), \quad \text{if type is } \theta_H, \quad (20.34) \]
\[ \theta_L - p_1 \leq 0, \quad \text{if type is } \theta_L. \quad (20.35) \]

In the incentive compatibility inequality function (20.34) for the type \( \theta_H \), the left hand is his utility of first-period consumption, and the right hand is his utility of second-period consumption (with second-period price \( p_2 = \theta_L \)), the inequality (20.34) can be rewritten as:

\[ p_1 \leq (1 - \delta)\theta_H + \delta \theta_L. \quad (20.36) \]

In the incentive compatibility inequality function (20.35) for the type \( \theta_L \), the left hand is his utility of first-period consumption, and the right hand is his utility of second-period consumption (with second-period price \( p_2 = \theta_H \)), the inequality (20.35) can be rewritten as

\[ p_1 \geq \theta_L. \quad (20.37) \]

From the inequality functions (20.36) and (20.37), we have \( \delta \leq 1 \). Thus the necessary condition for existence of “full revelation” equilibrium is that \( \delta \leq 1 \). If \( \delta > 1 \), then the incentive compatibility inequality function (20.36) for type \( \theta_H \) means \( p_1 \leq \theta_L \), then both types of \( \theta_H \) and \( \theta_L \) choose to buy at the first-period, and for the type \( \theta_L \), he will not buy at the second-period (since \( p_2 = \theta_H \)), belonging to “take the money and run”.

Therefore, when \( \delta \leq 1 \) holds, “full revelation” equilibrium exists, and \( p_1^R = (1 - \delta)\theta_H + \delta \theta_L \), the principal’s profit is

\[ \pi^R = \beta[(1 - \delta)\theta_H + \delta \theta_L] + \delta[\beta \theta_H + (1 - \beta)\theta_L] = \beta \theta_H + \delta \theta_L, \]

and the rent for type \( \theta_H \) is \( \delta(\theta_H - \theta_L) \).
“No Revelation” Equilibrium

In the “no revelation” (or pooling) equilibrium, \( x_{H1} = x_{L1} \). Then \( \hat{\beta}(x_1) = \beta > \bar{\beta} \), and \( p_2 = \theta_H \). Although there are two possible “no revelation” equilibria, i.e., \( x_{H1} = x_{L1} = 0 \), and \( x_{H1} = x_{L1} = 1 \), for the principal, the optimal (profit maximizing) “no revelation” equilibrium is \( x_{H1} = x_{L1} = 1 \), therefore the necessary condition for such equilibrium is that the participation constraint should hold, and the first-period price is \( p_1 = \theta_L \).

The principal’s profit is

\[ \pi^p = \theta_L + \delta \theta_H \]

and the rent for type \( \theta_H \) is \( \theta_H - \theta_L \).

Now we compare \( \pi^R \) and \( \pi^p \). For \( \delta \leq 1 \), by

\[ \pi^R - \pi^p = (\beta \theta_H - \theta_L)(1 - \delta), \]

we have \( \pi^R \geq \pi^p \).

“Partial Revelation” Equilibrium

Let \( \rho_i \) be the probability of the purchase for type \( \theta_i \) under the first-period price \( p_1 \). Since type \( \theta_H \) has more incentives to purchase than the type \( \theta_L \), so it must be true that \( \rho_H > \rho_L \), and then the ex post belief is

\[ \hat{\beta}(x_1 = 1) = \frac{\beta \rho_H}{\beta \rho_H + (1 - \beta) \rho_L} \geq \beta > \bar{\beta}, \]

Therefore, \( p_2(x_1 = 1) = \theta_H \). There are two possibilities for the second-period price:

(1) If \( \hat{\beta}(x_1 = 0) \leq \bar{\beta} \), then \( p_2(1) = \theta_H \) and \( p_2(0) = \theta_L \).

By \( \rho_H > \rho_L \geq 0 \), we have

\[ \theta_H - p_1 \geq \delta(\theta_H - \theta_L), \]

which means that \( p_1 \leq (1 - \delta) \theta_H + \delta \theta_L \). If \( \delta > 1 \), both types have incentives to buy in period 1, then \( \rho_H = \rho_L = 1 \), it is not “partial revelation”. If \( \delta \leq 1 \), the principle has two options:
(a) \( p_1 = (1 - \delta)\theta_H + \delta \theta_L \). Then \( \rho_H \in (0, 1], \rho_L = 0 \), and her profit is

\[
\pi^S = \rho_H \beta [(1 - \delta)\theta_H + \delta \theta_L] + \delta [\beta \theta_H + (1 - \beta)\theta_L],
\]

which is equal to that of the “full revelation” case, i.e., \( \pi^S = \pi^R \);

(b) \( p_1 = \theta_L \). Then \( \rho_H = 1, \rho_L \in (0, 1) \), and her profit is \( \pi^S = \pi^p \).

Since \( \pi^R > \pi^p \), the principal chooses \( p_1 = (1 - \delta)\theta_H + \delta \theta_L \), which is the same as the “full revelation” case.

(2) If \( \hat{\beta}(x_1 = 0) \geq \bar{\beta} \), then \( p_2(x_1 = 1) = p_2(x_1 = 0) = \theta_H \).

Since \( \hat{\beta}(x_1 = 0) = \frac{\beta(1 - \rho_H)}{\beta(1 - \rho_H) + (1 - \beta)(1 - \rho_L)} > 0 \), then \( \rho_H < 1 \), and thus it must be that \( p_1 \geq \theta_H \). Indeed, if \( p_1 < \theta_H \), then \( \rho_H = 1 \), and \( \rho_L = 0 \). The first-period price must be \( p_1 = \theta_H \). Due to

\[
\hat{\beta}(x_1 = 0) = \frac{\beta(1 - \rho_H)}{\beta(1 - \rho_H) + (1 - \beta)} \geq \bar{\beta},
\]

we have

\[
\rho_H \leq \hat{\rho}_H = \frac{\beta \theta_H - \theta_L}{\beta(\theta_H - \theta_L)},
\]

and thus the principal’s profit is

\[
\pi^S = \beta \theta_H [\hat{\rho}_H + \delta].
\]

If \( \delta \leq 1 \), then \( \pi^R \geq \pi^p \). We compare \( \pi^R \) and \( \pi^S \).

\[
\pi^S - \pi^R = \beta \theta_H \hat{\rho}_H + \beta \theta_H \delta - \beta \theta_H - \delta \theta_L = \\
\beta \theta_H \hat{\rho}_H + \delta(\beta \theta_H - \theta_L) - \beta \theta_H = \\
\theta_H \frac{\theta_H - \theta_L}{\theta_H - \theta_L} + \delta(\beta \theta_H - \theta_L) - \beta \theta_H = \\
(\beta \theta_H - \theta_L)[\frac{\theta_H}{\theta_H - \theta_L} + \delta] - \beta \theta_H = \\
\beta \theta_H [\frac{\theta_H}{\theta_H - \theta_L} + \delta - 1] - \theta_L [\frac{\theta_H}{\theta_H - \theta_L} + \delta].
\]

Therefore, \( \pi^S \geq \pi^R \) is equivalent to

\[
\beta \geq \beta^{RS} \equiv \frac{\theta_L \theta_H + \delta(\theta_H - \theta_L)}{\theta_H \theta_L + \delta(\theta_H - \theta_L)}.
\]
In the following, we discuss the principal’s contract choice under $\delta > 1$. Since there is no “full revelation” equilibrium, we need to compare $\pi^p$ and $\pi^S$.

\[
\pi^S - \pi^p = \beta \theta_H \hat{\rho}_H + \beta \theta_H \delta - \theta_L - \delta \beta \theta_H \\
= \beta \theta_H \hat{\rho}_H - \theta_L \\
= \frac{\beta \theta_H - \theta_L}{\theta_H - \theta_L} \theta_H - \theta_L \\
= \frac{\theta_H}{\theta_H - \theta_L} [\beta - \frac{\theta_L}{\theta_H} (2 - \frac{\theta_L}{\theta_H})].
\]

Thus $\pi^S \geq \pi^p$ is equivalent to

\[
\beta \geq \beta^{PS} \equiv \frac{\theta_L}{\theta_H} (2 - \frac{\theta_L}{\theta_H}).
\]

From the above discussion, we have the following result.

**Proposition 20.3.1** In the lack of commitment, the optimal contract for the principal entail:

- If $\beta \leq \bar{\beta}$, the principal chooses the same contract as the full commitment case and it can be implemented;
- If $\delta \leq 1$ and $\beta \in (\bar{\beta}, \beta^{RS}]$, the optimal contract for the principal is $p_1 = (1 - \delta)\theta_H + \delta \theta_L$, $p_2(x_1 = 1) = \theta_H$, and $p_2(x_1 = 0) = \theta_L$;
- If $\delta \leq 1$ and $\beta > \beta^{RS}$, the principal chooses the full revelation” equilibrium, the optimal contract is $p_1 = \theta_H$, and $p_2(x_1 = 1) = p_2(x_1 = 0) = \theta_H$;
- If $\delta > 1$ and $\beta \in (\bar{\beta}, \beta^{PS}]$, the principal chooses the “no revelation” equilibrium, the optimal contract is $p_1 = \theta_L$, and $p_2(x_1 = 1) = p_2(x_1 = 0) = \theta_H$;
- If $\delta > 1$ and $\beta > \beta^{PS}$, the principal chooses the “partial revelation” equilibrium, the optimal contract is $p_1 = \theta_H$, and $p_2(x_1 = 1) = p_2(x_1 = 0) = \theta_H$. 

802
20.3.3 Dynamic Contracts with Partial Commitment

Next, we turn to the case of partial commitment, in this case, the principal can revise contract only with the consent of the agent. The timing of this dynamic contracting is the following: (1) principal offers a contract \((p_1; p_2(x_1))\); (2) agent chooses \(x_1\); (3) principal offers a new contract \(p'_2(x_1)\); (4) agent accepts \(p'_2(x_1)\) or rejects and sticks to the original contract; (4) agent chooses \(x_1\).

We will look for the principal’s preferred PBE of this contract. Note that a renegotiation offer \(p'_2(x_1)\) following the agent’s choice \(x_1\) will be accepted by the agent if and only if \(p'_2(x_1) < p_2(x_1)\). Therefore, a long-term renegotiable contract commits the principal against raising the price, but does not commit her against lowering the price. Analysis is simplified by observing that the principal can, without loss of generality, offer a contract that is not renegotiated in equilibrium. We first define renegotiation-proof contract.

**Definition 20.3.1** A Renegotiation-Proof (RNP) contract is one that is not renegotiated in the continuation equilibrium.

The following principle can simplify analysis of the above problem.

**Proposition 20.3.2 (Renegotiation-Proof Principle)** For any PBE outcome of the contract, there exists another PBE that implements the same outcome and in which the principal offers a RNP contract.

**Proof.** If the principal offers \((p_1; p_2(1), p_2(0))\) and in equilibrium renegotiates to \(p'_2(x_1) < p_2(x_1)\) after the agent’s choice of \(x_1 \in \{0, 1\}\), then the contract \((p_1; p'_2(1), p'_2(0))\) is RNP.

Dewatripont (1989) discussed the Renegotiation-Proof Principle in detail. The RNP Principle is similar to the Revelation Principle, i.e., for any equilibrium, one can always find a new equilibrium that has the same equilibrium outcome. By the RNP Principle, the optimal dynamic contract can be implemented by a RNP contract. As such, we can restrict our attention to RNP mechanisms.
When $\beta \leq \bar{\beta}$, with full commitment power, all consumers will purchase at date 1, and thus the consumer’s action will not modify the principal’s belief. As such, a contract with full commitment can be always implementable even though the principal may modify the contract without the consent of the agent, and of course it can be implementable with the consent of the agent. As such, we only need to consider the case of $\beta > \bar{\beta}$.

First, the contract with full commitment is not a RNP contract. This is because, under such a contract, the principal offers $p_1 = p_2(1) = p_2(0) = \theta_H$, and the agent chooses $x_1(\theta_H) = x_2(\theta_H) = 1, x_1(\theta_L) = x_2(\theta_L) = 0$. However, with $x_1 = 0$ observed, the consumer can be identified as $\theta_L$ type, then $p_2(0) = \theta_L$ will be accepted by the consumer with $\theta_L$ and thus this will increase the principal’s profit. As such, she has incentives to modify the contract. But, she cannot do so without the consent of the agent.

Secondly, any short-term contract with no commitment can be implementable with the consent of the agent. This is because, if $p_1^N$ and $(p_2(1)^N, p_2(0)^N)$ are two short-term contracts, it is clear that with $x_1$ observed, the contract $(p_1^N; p_2(1)^N, p_2(0)^N)$ in the second period can be implemented. If the principal provides a new contract $(p_2'(1), p_2'(0))$ that can be accepted by the agent, we then must have $p_2'(1) < p_1^N(1)$ or $p_2'(0) < p_1^N(0)$. But, for the profit maximizing dynamic contract, given $x_1$ and the principal’s belief, $p_1^N(x_1)$ is an optimal choice. As such, the principal does not have incentives to modify the original contract.

Thirdly, do long-term renegotiable contracts offer an advantage relative to short-term contracts? This is true when the ability to commit to a low price is useful. Consider the following contract $(p_1 = \theta_H + \delta \theta_L; p_2(1) = 0, p_2(0) = \theta_L)$. This contract provides a separating equilibrium under the consent of the agent. Indeed, $\theta_H$-consumer’s utility with purchase in the first period is

$$\theta_H - (\theta_H + \delta \theta_L) + \delta \theta_H = \delta(\theta_H - \theta_L) > 0,$$

and his utility without purchase in the first period is

$$\delta(\theta_H - \theta_L).$$
Thus, the incentive compatibility and individual rationality constraints are satisfied for $\theta_H$.

As for $\theta_L$, his utility with purchase in the first period is

$$\theta_L - (\theta_H + \delta \theta_L) + \delta \theta_L = -(\theta_H - \theta_L) < 0,$$

and his utility without purchase in the first period is

$$\delta(\theta_L - \theta_L = 0).$$

Thus, the incentive compatibility and individual rationality constraints are satisfied for $\theta_L$.

In addition, the principal will not provide a new the contract ($p_2'(1)$, $p_2'(0)$) for the second period. This is because, if the agent agrees, then $p_2'(1) < 0$ or $p_2'(0) < \theta_L$, which makes the principal worse off. Hence, the contract with consent of the agent is RNP.

However, the contract without the consent of the agent is not implementable. Indeed, once $x_1$ is observed, the consumer must be $\theta_H$-type, and then the principal will choose $p_2(1) = \theta_H$. As such, the contract without commitment power is not implementable.

### 20.4 Sequential Screening

From the above discussion, we learn that in dynamic mechanism design, the agent has information advantage with different degree over time. In dynamic contracting environments, these information advantages will turn into the information rent for agent. In fact, the agent (consumer) may need some time to learn his future consumption type, and has some private information in different degree over time, the principal (monopolist) use some mechanism to elicit the private information over time, such as dynamic contracting to screen agents over time. Courty and Li (2000) introduce sequential screening to analyze the above dynamic principal-agent problems.
20.4.1 An Example of Sequential Screening

Consider the demand for airplane tickets. Travellers typically do not know their valuations for tickets until just before departure, but they know in advance their likelihood to have high and low valuations. A monopolist can wait until the travellers learn their valuations and charge the monopoly price, and more consumer surplus can be extracted by requiring them to reveal their private information sequentially. An illustration of such monopoly practice is a menu of refund contracts, each consisting of an advance payment and a refund amount in case the traveller decides not to use the ticket. By selecting a refund contract from the menu, travellers reveal their private information about the distribution of their valuations, and by deciding later whether they want the ticket or the specified refund, they reveal what they have learned about their actual valuation.

Suppose that at time $t = 0$, one-third of all potential buyers are leisure travellers (type $L$) whose valuation is uniformly distributed on $\theta_L \in [1, 2]$, and two-thirds are business travellers (type $B$) whose valuation is uniformly distributed on $\theta_B \in [0, 1] \cup [2, 3]$. Intuitively, business travellers face greater valuation uncertainty than leisure travellers. Suppose that cost of flying an additional traveller is 1. Suppose monopolist and travellers are all risk neutral. At time $t = 1$, all travellers know their true value. The monopolist design contract in time $t = 0$.

We first discuss the monopolist’s pricing in $t = 1$. If the seller waits until travellers have privately learned their valuations, she faces a valuation distribution that is uniform on $[0, 3]$, i.e., type distribution function being $F(x) = \frac{x}{3}$, the problem of monopolist is:

$$\max_p (p - 1)(1 - F(p)),$$

the monopoly price is $p^m = 2$, and her profit is $\pi^m = \frac{1}{3}$.

Now we discuss the monopolist’s contract in $t = 0$. Travellers only know their general types, i.e., leisure type or business type. Suppose instead that the seller offers two contracts before the travellers learn their valuation, one with an advance payment of 1.5 and no refund (unchangeable), and the other with
an advance payment of 1.75 and a partial refund of 1 (changeable, but with .75 cancellation fee). Contracts: price being $p_l = 1.5$ without refund. Leisure travellers strictly prefer the contract with no refund, $(p_l, r_l)$. Business travellers are indifferent between the two contracts so we assume that they choose the contract with refund, $(p_b, r_b)$.

Since for the leisure type, if choosing the contract $(p_l, r_l)$, his expected utility is:

$$\int_1^2 (\theta_L - 1.5)d\theta_L = 0 > \int_1^2 (\theta_L - 1.75)d\theta_L.$$ 

For the business type, if choosing the contract $(p_b, r_b)$, his expected utility is:

$$\frac{1}{2} \int_0^1 (1 - 1.75)d\theta_B + \frac{1}{2} \int_2^3 (\theta_B - 1.75)d\theta_B = 0 = \frac{1}{2} \int_0^1 (\theta_B - 1.5)d\theta_B + \frac{1}{2} \int_2^3 (\theta_B - 1.5)d\theta_B.$$ 

Therefore, such screening contracts satisfy the incentive compatibility and participation constraint, and the expected rent for both types are 0. In such contracts, the monopolist’s profit is:

$$\frac{1}{3}(1.5 - 1) + \frac{2}{3}\left[\int_0^1 (1.75 - 1)d\theta_B + \int_2^3 (1.75 - 1)d\theta_B\right] = \frac{2}{3} > \pi^m.$$ 

In time $t = 1$, for business type, if $\theta_B \in [0, 1]$, he will choose refund, and if $\theta_B \in [2, 3]$, he will travel, and the monopolist can get profit 0.75 from each business traveller.

The above example reveals the basic idea of sequential screening: On one hand, with time moving, the agent has more advantage on information. To screen such information, the principal need to give up more information rent, the earlier screening is, more reduction of agent’s information rent would be. On the other hand, with time moving, more information makes increase in efficiency of outcome, earlier screening results in the loss of efficiency. Therefore, in sequential screening, there is a trade off between allocative efficiency and information rent in time.

There are many other example of sequential mechanisms that take different forms such as hotel reservations (cancellation fees), car rentals (free mileage vs.
fixed allowance), telephone pricing (calling plans), public transportation (day pass), and utility pricing (optional tariffs). Sequential price discrimination can also play a role in contracting problems such as taxation and procurement where the agent’s private information is revealed sequentially.

20.4.2 Sequential Screening under Incomplete Information

In this subsection, we consider the problem of designing the optimal menu of refund contracts for two ex ante types of potential buyers.

Consider a monopoly seller of airplane tickets facing two types of travellers, $B$ and $L$, with proportion $\beta_B$ and $\beta_L$ respectively. We can think of type $B$ as the “business traveller” and type $L$ as the “leisure traveller”. There are two periods. In the beginning of period one, the traveller privately learns his type. The seller and the traveller contract at the end of period one. In the beginning of period two, the traveller privately learns his actual valuation $v \in [\underline{v}, \bar{v}]$ for the ticket, and then decides whether to travel. Each ticket costs the seller $c$. The seller and the traveller are risk-neutral, and do not discount. The reservation utility of each type of traveller is normalized to zero. The business type may value the ticket more in the sense of first-order stochastic dominance (FSD): type B’s distribution of valuation $G_B$ first-order stochastically dominates the leisure type’s distribution $G_L$ if $G_B(v) \leq G_L(v)$ for all $v$ in the range of valuations $[\underline{v}, \bar{v}]$. Alternatively, the business type may face greater valuation uncertainty in the sense of mean-preserving-spread (MPS): if $G_B$ dominates $G_L$ by MPS, then for $G_B$ and $G_L$, and for their random variables $v_B$ and $v_L$, there exists $v_\epsilon$ so that the related variable respectively, and $v_B = v_L + v_\epsilon$ with $v_\epsilon$ independent of $v_L$, or equivalently, for all $v \in [\underline{v}, \bar{v}]$, $\int_{\underline{v}}^{\bar{v}} (G_B(u) - G_L(u))du \geq 0$.

Example 20.4.1 Let us consider again the example from the last subsection.
$G_B$ and $G_L$ are described as followed:

$$G_B(v) = \begin{cases} \frac{v}{2}, & \text{if } v \in [0, 1], \\ \frac{1}{2}, & \text{if } v \in [1, 2], \\ \frac{v-1}{2}, & \text{if } v \in [2, 3]. \end{cases}$$

and

$$G_L(v) = \begin{cases} 0, & \text{if } v \in [0, 1], \\ v - 1, & \text{if } v \in [1, 2], \\ 1, & \text{if } v \in [2, 3]. \end{cases}$$

The distribution function of $v_\epsilon$ is:

$$G_\epsilon(v) = \frac{v + 2}{4}, v \in [-2, 2].$$

Thus,

$$\int_v^u (G_B(u) - G_L(u)) du = \begin{cases} \frac{v^2}{4} \geq 0, & \text{if } v \in [0, 1], \\ \frac{1}{4} - \frac{(v-1)(v-2)}{2} \geq 0, & \text{if } v \in [1, 2], \\ \frac{(v-3)^2}{4} \geq 0, & \text{if } v \in [2, 3]. \end{cases}$$

Thus, $G_B$ dominates $G_L$ by MPS.

A refund contract consists of an advance payment $a$ at the end of period one and a refund $k$ that can be claimed at the end of period two after the traveller learns his valuation. Clearly, regardless of the payment $a$, the traveler use it only if he values the ticket more than $k$. The seller offers two refund contracts $(a_B, k_B, a_L, k_L)$. The profit maximization problem can be written as:

$$\max (a_B, k_B; a_L, k_L) \beta_B[a_B - k_B G_B(k_B) - c(1 - G_B(k_B))]$$
$$+ \beta_L[a_L - k_L G_L(k_L) - c(1 - G_L(k_L))]$$

(20.38)
subject to

\[-a_B + k_B G_B(k_B) + \int_{\underline{v}}^{\bar{v}} \max\{v, k_B\} dG_B(v) \geq -a_L + k_L G_B(k_L) + \int_{\underline{v}}^{\bar{v}} \max\{v, k_L\} dG_B(v); \quad (20.39)\]

\[-a_L + k_L G_L(k_L) + \int_{\underline{v}}^{\bar{v}} \max\{v, k_L\} dG_L(v) \geq -a_B + k_B G_L(k_B) + \int_{\underline{v}}^{\bar{v}} \max\{v, k_B\} dG_L(v); \quad (20.40)\]

\[-a_B + k_B G_B(k_B) + \int_{\underline{v}}^{\bar{v}} \max\{v, k_B\} dG_B(v) \geq 0; \quad (20.41)\]

\[-a_L + k_L G_L(k_L) + \int_{\underline{v}}^{\bar{v}} \max\{v, k_L\} dG_L(v) \geq 0. \quad (20.42)\]

The constraints of (20.41) and (20.42) are the participation constraints in period one, and (20.39) and (20.40) are the incentive compatibility constraints in period one.

We can verify that in second-best contracts, only the constraints of (20.39) and (20.42) are binding. Since \(\max\{v, k_L\}\) is concave function of \(v\), by the properties of MPS, we have:

\[\int_{\underline{v}}^{\bar{v}} \max\{v, k_L\} dG_B(v) \geq \int_{\underline{v}}^{\bar{v}} \max\{v, k_L\} dG_L(v),\]

and the constraint of (20.39) is binding, i.e.,

\[-a_B + k_B G_B(k_B) + \int_{\underline{v}}^{\bar{v}} \max\{v, k_B\} dG_B(v) \geq -a_L + k_L G_B(k_L) + \int_{\underline{v}}^{\bar{v}} \max\{v, k_L\} dG_B(v),\]

so that we have

\[-a_B + k_B G_B(k_B) + \int_{\underline{v}}^{\bar{v}} \max\{v, k_B\} dG_B(v) \geq -a_L + k_L G_B(k_L) + \int_{\underline{v}}^{\bar{v}} \max\{v, k_L\} dG_L(v) \geq 0,\]

Thus we get the participation constraint (20.41) for type \(B\).

Let us firstly omit the incentive compatible constraint (20.40) for type \(L\) (below we show it is loosely satisfied). Substituting the equations of (20.39) and (20.42) into (20.38), the profit maximization problem can be rewritten as:

\[\max (k_B, k_L) \int_{k_B}^{\bar{v}} \beta_B(v-c) dv + \int_{k_L}^{\bar{v}} \beta_L(v-c) g_L(v) - \beta_B (G_L(v) - G_B(v)) dv \quad (20.43)\]
In the objective function (20.43), define $S_t(k_t) = \int_{k_t}^{\bar{v}} (v - c) g_L(v)dv$ as the consumers’ surplus for type $t \in \{L, B\}$, and $R_B(k_L) = \int_{k_L}^{\bar{v}} (G_L(v) - G_B(v))dv$ as the information rent for type $B$.

The solution for the problem is satisfied:

\[ k_B = c, \quad (20.44) \]
\[ k_L = \arg \max_k f_L S(k) - f_B R_k. \quad (20.45) \]

The second-best solution is the tradeoff between allocation efficiency and information rent, the same logic with principal-agent problem.

Next we show that under the constraints (20.44) and (20.45), the incentive compatibility constraint (20.40) for $L$ is satisfied.

From the binding of (20.39), it means that $a_L - a_B = \int_{k_B}^{k_L} G_B(v)dv$, and thus

\[ -a_L + k_L G(k_L) + \int_{k_L}^{\bar{v}} vdG_L(v) = -a_B + k_B G(k_L) + \int_{k_B}^{\bar{v}} vdG_L(v) - \int_{k_B}^{k_L} (G_B(v) - G_L(v))dv. \]

Therefore (20.40) is equivalent to $\int_{k_B}^{k_L} (G_B(v) - G_L(v))dv \leq 0$. So we need only to verify that $\int_{k_B}^{k_L} (G_B(v) - G_L(v))dv \leq 0$. When $k_L = c$, it is obviously true. When $k_L \neq c$, suppose by way of contradiction that $\int_{c}^{k_L} (G_B(v) - G_L(v))dv > 0$, and consider a new contract that $k'_{LB} = k'_{L} = c$. We have $S_L(k'_{LB}) = S_L(c) > S_L(k_{LB})$, and the information rent is:

\[ R_B(k'_{LB}) = \int_{c}^{\bar{v}} (G_L(v) - G_B(v))dv \]
\[ = \int_{k_L}^{\bar{v}} (G_L(v) - G_B(v))dv + \int_{c}^{k_L} (G_L(v) - G_B(v))dv \]
\[ \leq \int_{k_L}^{\bar{v}} (G_L(v) - G_B(v))dv = R_B(k_L), \]

which is a contradiction to the second-best contract $(k_B, k_L)$, so it must be that $\int_{c}^{k_L} (G_B(v) - G_L(v))dv \leq 0$. Thus, the incentive compatible constraint (20.40) holds under the binding constraints of (20.39) and (20.42).
In the following, we discuss \( k_L \) when \( k_L \neq c \). When \( k_L > c \), it means that the type \( L \) is rationed, and when \( k_L < c \), it means the type \( L \) is subsidized.

For the buyer’s surplus for type \( L \), \( S_L(k_L) = \int_{k_L}^{\bar{v}} (v - c) g_L(v) dv \), \( \forall k_L \in [\underline{v}, \bar{v}] \), \( S_L(c) \geq S_L(k_L) \), if \( k_L = c \), its surplus is biggest. However for the information rent for type \( H \), \( R_B(k_L) = \int_{k_L}^{\bar{v}} (G_L(v) - G_B(v)) \frac{dL(v)}{dv} dv = \int_{\bar{v}}^{k_L} (G_B(v) - G_L(v)) \frac{dL(v)}{dv} dv \geq 0 \), when \( k_l = \bar{v} \) or \( k_l = \bar{v} \) \( \forall \), \( R_B(k_L) = 0 \), we know in the interior of \([\underline{v}, \bar{v}]\), the rent \( R_B \) has an extreme point.

Consider the following special case where \( R_B(k_L) \) is a single-peaked function, which means there exists \( z \) such that \( \frac{dR_B(k_L)}{dk_L} > 0 \) \( \forall k_L < z \) and \( \frac{dR_B(k_L)}{dk_L} < 0 \) \( \forall k_L > z \). If the density functions \( g_B \) and \( g_L \) are symmetric at point \( z \), we must have \( k_L < c \) (or \( k_L > c \)) if \( c < z \) (or \( c > z \)). To see this, we only discuss the case of \( c < z \) (the case of \( c > z \) is similar). Let \( k_L^{SB} \) be second-best solution.

1. \( k_L^{SB} \notin (c, z] \). If \( k_L \in (c, z] \), then \( \frac{dS_L(k_L)}{dk_L} < 0 \) and \( \frac{dR_B(k_L)}{dk_L} > 0 \).

Thus, if \( k_L \) decreases, then \( S_L(k_L) - R_B(k_L) \) increases.

2. \( k_L^{SB} \notin (z, 2z - c] \). If \( k_L \in (z, 2z - c] \), consider a new refund \( \tilde{k}_L = 2z - k_L \) so that \( R_B(\tilde{k}_L) = R_B(k_L) \). Thus, for \( c \leq \tilde{k}_L < k_L \), we have \( S_L(\tilde{k}_L) > S_L(k_L) \).

3. \( k_L^{SB} \notin 2z - c \). If \( k_L > 2z - c \), consider a new refund \( \tilde{k}_L = 2z - k_L \) so that \( R_B(\tilde{k}_L) = R_B(k_L) \). Since \( k > 2z - c \), for \( \tilde{k} = 2z - k \), we get:

\[- \frac{dS_L(k)}{dk} = (k - c) g_L(k) = (k - c) g_L(\tilde{k}) > (c - \tilde{k}) g_L(\tilde{k}) = \frac{dS_L(\tilde{k})}{dk},\]

in which the second equality is from the symmetric hypothesis of \( g_L \) at \( z \) and the third equality is from \( k + \tilde{k} = 2z > 2c \). We then have

\[ S_L(c) - S_L(k_L) = \int_{c}^{k_L} - \frac{dS_L(k)}{dk} dk > \int_{k_L}^{\tilde{k}_L} \frac{dS_L(\tilde{k})}{dk} dk \]

\[ = S_L(c) - S_L(\tilde{k}_L), \]

\[ 812 \]
Thus, $S_L(\tilde{k}_L) > S_L(k_L)$, which is contradictory to the second-best solution of $k_L$.

From the above discussion, we obtain $k^{SB}_L < c$.

**Example 20.4.2** Return to the example from the last subsection and let $c = 1$. We show the contract given in this example is in fact the second-best screening contract to the principal. Indeed, $S_L(k_L)$ and $R_B(k_L)$ are

$$S_L(k_L) = \int_{k_L}^{\hat{v}} (v - c)g_L(v)dv = \begin{cases} \frac{1}{2}, & \text{if } k_L \in [0, 1], \\ k_L - \frac{k^2}{2}, & \text{if } k_L \in [1, 2], \\ 0, & \text{if } k_L \in [2, 3], \end{cases}$$

and

$$R_B(k_L) = \int_{u}^{k_L} (G_B(u) - G_L(u))du = \begin{cases} \frac{k^2}{4} \geq 0, & \text{if } k_L \in [0, 1], \\ \frac{1}{4} - \frac{(k_L-1)(k_L-2)}{2} \geq 0, & \text{if } k_L \in [1, 2], \\ \frac{(k_L-3)^2}{4} \geq 0, & \text{if } k_L \in [2, 3]. \end{cases}$$

So, $z = 1.5$, $k^{SB}_L = \arg \max_k \frac{1}{3}S_L(k) - \frac{2}{3}R_B(k)$, and therefore $k^{SB}_L = 0$. Its graphic illustration is shown the figure (20.3).

Since the participation constraint (20.42) is binding for type $L$, if $k_L = 0$, $a_L = 1.5$; and since $k_B = c = 1$, for type $B$, its participation constraint (20.39) is binding too. Therefore, we get $a_B = 1.75$. Thus, $(a_L = 1.5, k_L = 0; a_B = 1.75, k_B = 1)$ is the second-best screening contract to the principal.

In Courty and Li (2000), they further discuss the continuous type case, interested reader can find more detail analysis in their paper.

### 20.5 Efficient Budget-Balanced Dynamic Mechanism

In this section we turn to the dynamic efficient mechanism in the framework of general mechanism design with more than one agent. Athey and Segal (Econometrica, 2013) provide an analysis framework of dynamic mechanism design, and construct an efficient dynamic mechanism.
The Vickery-Clark-Groves (VCG) mechanism established the existence of an incentive-compatible and efficient mechanism for a general class of static mechanism design problems. The VCG mechanism provides incentives for truthful reporting of private information under the assumption of private values (other agents’ private information does not directly affect an agent’s payoff) and that preferences are quasilinear so that incentives can be provided using monetary transfers. One shortcoming of VCG mechanism is not budget-balanced ex post. Subsequently, a pair of classic papers, Arrow (1979) and d’Aspremont and Gerard-Varet (1979) (AGV), constructed an efficient and incentive-compatible mechanism, called the expected externality mechanism, in which the transfers were budget-balanced, and thus resulting in Pareto efficient outcomes, using the solution concept of Bayesian Nash equilibrium, under the additional assumption that private information is independent across agents.

In this section, we discuss the efficient dynamic mechanism design, which is also budget-balanced, by omitting the requirement of ex post participation constraints. In a static setting, the AGV mechanism gives every agent an incentive
to report truthfully given his beliefs about opponents’ types and truth telling, by
giving him a transfer equal to the “expected externality” his report imposed on
the other agents. Thus, an agent’s current beliefs about opponents’ types play
an important role in determining his transfer. However, in a dynamic setting,
these beliefs evolve over time as a function of opponent reports and the decisions
those reports induce. If the transfers are constructed using the agents’ prior
beliefs at the beginning of the game, the transfers will no longer induce truthful
reporting after agents have gleaned some information about every other’s types.

If, instead, the transfers are constructed using beliefs that are updated using
earlier reports, this will undermine the incentives for truthful reporting at the
earlier stages. Athey and Segal (2013) construct a mechanism, called balanced
team mechanism, that achieves budget-balance property. Such mechanism sus-
tains an equilibrium in truthful strategies by giving each agent in each period an
incentive payment equal to the change in the expected present value (EPV) of the
other agents’ utilities that is induced by his current report. On the one hand,
these incentive payments cause each agent to internalize the expected externality
imposed on the other agents by his reports. On the other hand, the expected
incentive payment to an agent is zero when he reports truthfully no matter what
reporting strategies the other agents use. The latter property makes the budget
balanced by letting the incentive payment to a given agent being paid by the
other agents without affecting those agents’ reporting incentives.

Before we construct such a mechanism, let us first analyze the static case,
discuss the difficulties in dynamic mechanism, and then we introduce the dynamic
efficient mechanism proposed by Athey and Segal.

Consider a seller (agent 1) and a buyer (agent 2) who engage in a two-period
relationship. In each period $t = 1, 2$, they can trade a contractible quantity
$x_t \in R_+.$

Before the first period, the seller privately observes a random type $\tilde{\theta}_1 \in [1, 2],$
whose realization $\theta_1$ determines his cost. The cost function is given by
$$C(\theta_1, x_t) = \frac{1}{2} \theta_1 x_t^2, \ t = 1, 2.$$
where \( x_t \) is output at time \( t \).

The buyer’s value per unit of the good in period 1 is equal to 1, and in period 2 it is given by a random type \( \hat{\theta}_2 \in [0, 1] \) whose realization she privately observes between the periods.

When information is complete, the optimal problem at period 1 is:

\[
\max_{x_1} x_1 - \frac{1}{2} \theta_1 x_1^2,
\]

giving us \( x_1(\theta_1) = \frac{1}{\theta_1} \), and the optimal problem at period 2 is:

\[
\max_{x_1} \theta_2 x_2 - \frac{1}{2} \theta_1 x_2^2,
\]

giving us \( x_2(\theta_1, \theta_2) = \frac{\theta_2}{\theta_1} \). Thus, an efficient (surplus-maximizing) mechanism must have trading decisions \( x_1 \) and \( x_2 \) determined by the decision rules: \( x_1(\theta_1) = \frac{1}{\theta_1} \) and \( x_2(\theta_1, \theta_2) = \frac{\theta_2}{\theta_1} \).

When information is incomplete, but the agents can observe their own types, although there are trades for two periods, it is a static interaction. At this situation, the problem of designing an efficient mechanism comes down to designing transfers to each agent as a function of their reports. In this simple setting, each agent makes only one report. Let us first consider the AGV mechanism for this problem, where we assume that the seller and buyer observe their type at the same time. Taking this as a static benchmark, we discuss the incentive change under the asynchronous observations and reports.

### 20.5.1 Efficient Budget-Balanced Static Mechanism

Since the AGV mechanism can make agent to internalize the externality induced by his action, and in the efficient decision rule, agents have incentives to reveal their information. In the following, we show how AGV mechanism can implement efficient rule, i.e., \( x_1(\theta_1) = \frac{1}{\theta_1} \) and \( x_2(\theta_1, \theta_2) = \frac{\theta_2}{\theta_1} \).

Let \( \gamma_i(\theta_i) \) be the payoff to agent \( i \), to encourage the buyer to reveal his type,
the transfer to him must be the expected externality to the seller:

\[ \gamma_2(\theta_2) = -E_{\hat{\theta}_1}[\frac{1}{2}\hat{\theta}_1(x_1(\hat{\theta}_1))^2 + \frac{1}{2}\hat{\theta}_1(x_2(\hat{\theta}_1, \theta_2))^2] \]

\[ = -\frac{1}{2}E_{\hat{\theta}_1}[\frac{1}{\hat{\theta}_1}](1 + (\theta_2)^2). \]

We verify that under such transfer, the buyer has the incentive to reveal true type. Let the buyer’s report type as \( \hat{\theta}_2 \), his expected utility is:

\[ \gamma_2(\hat{\theta}_2) + E_{\hat{\theta}_1}[x_1\hat{\theta}_1 + \theta_2x_2(\hat{\theta}_1, \hat{\theta}_2)] = -\frac{1}{2}E_{\hat{\theta}_1}[\frac{1}{\hat{\theta}_1}](1 + (\hat{\theta}_2)^2) + E_{\hat{\theta}_1}[\frac{1}{\hat{\theta}_1}](1 + \theta_2\hat{\theta}_2), \]

and then the first condition for \( \hat{\theta}_2 \) is:

\[ -E_{\hat{\theta}_1}[\frac{1}{\hat{\theta}_1}](\hat{\theta}_2) + E_{\hat{\theta}_1}[\frac{1}{\hat{\theta}_1}]\theta_2 = 0, \]

which gives us

\[ \hat{\theta}_2 = \theta_2. \]

In the similar logic, the transfer to seller is equal to the externality to the buyer:

\[ \gamma_1(\theta_1) = E_{\hat{\theta}_2}[x_1(\theta_1) + \hat{\theta}_2x_2(\theta_1, \hat{\theta}_2)] \]

\[ = \frac{1}{\theta_1}[1 + E_{\hat{\theta}_2}(\hat{\theta}_2)^2]. \]

We can easily check that under such transfer, the seller has incentives to reveal his type.

The above transfer is not budget-balanced. However, due to the independence between \( \gamma_i(\theta_i) \) and \( \theta_j, j \neq i \), when the transfer to agent \( \theta_i \) is

\[ \psi_i(\theta_i, \theta_j) = \gamma_i(\theta_i) - \gamma_j(\theta_j), i \neq j, i, j \in \{1, 2\}, \]

it is a budget-balanced AGV mechanism. So in the static mechanism, the efficient rule is Bayesian implementable and budget-balanced. Actually, we know this result in Chapter 19.
20.5.2 Incentive Problem in Dynamic Environments

Now we turn to our dynamic model, where the buyer (agent 2) observes his type \( \hat{\theta}_2 \) between the time 1 and time 2, and the seller (agent 1) reported her type at time 1. The above AGV mechanism cannot implement the efficient rule \( x_1(\theta_1) = \frac{1}{\theta_1} \) and \( x_2(\theta_1, \theta_2) = \frac{\theta_2}{\theta_1} \). This is because, under the dynamic setting, when the seller reports his true type \( \theta_1 \), the buyer can learn the seller’s type \( \theta_1 \) from the first-period trade \( x_1(\theta_i) \), and then the transfer \( \gamma_2(\theta_2) \) cannot induce the true revelation for the buyer. Indeed, if the buyer announces his type being \( \hat{\theta}_2 \), his expected utility is

\[
\gamma_2(\hat{\theta}_2) + \theta_1[x_1(\theta_1) + \theta_2x_2(\theta_1, \hat{\theta}_2)] = -\frac{1}{2}E_{\theta_1}[\hat{\theta}_1](1 + (\hat{\theta}_2)^2) + \frac{1}{\theta_1} (1 + \theta_2\hat{\theta}_2). \tag{20.46}
\]

In this dynamic setting, the buyer can learn the seller’s private information. For the buyer’s expected utility (20.46), the first-order condition for \( \hat{\theta}_2 \) is

\[
-\frac{1}{\theta_1} \hat{\theta}_2 + \frac{\theta_2}{\theta_1} = 0,
\]

and so

\[
\hat{\theta}_2 = \frac{\theta_2}{\theta_1} \left( E_{\theta_1}[\hat{\theta}_1] \right)^{-1}.
\]

Therefore, when \( \frac{1}{\theta_1} > E_{\theta_1}[\hat{\theta}_1] \), the buyer has incentives to over-report his value; otherwise has incentives to underreport his value. The reason that the buyer has incentives to distort his type, is that it cannot fully internalize the externality induced by his action.

However, if we let buyer bear externality so that it is given by:

\[
\hat{\gamma}_2(\theta_1, \theta_2) = -\frac{1}{2} \theta_1(1 + (\theta_2)^2),
\]

then under this transfer, the buyer’s incentive can be restored. In this case, we have

\[
\hat{\gamma}_2(\theta_1, \hat{\theta}_2) + \theta_1[x_1(\theta_1) + \theta_2x_2(\theta_1, \hat{\theta}_2)] = -\frac{1}{2} \theta_1[\hat{\theta}_1](1 + (\hat{\theta}_2)^2) + \frac{1}{\theta_1} (1 + \theta_2\hat{\theta}_2). \tag{20.47}
\]

The first condition for \( \hat{\theta}_2 \) then is:

\[
-\frac{1}{\theta_1} \hat{\theta}_2 + \frac{\theta_2}{\theta_1} = 0,
\]

818
and thus

\[ \hat{\theta}_2 = \theta_2. \]

Although \( \gamma_1(\theta_1) \) and \( \hat{\gamma}_2(\theta_1, \theta_2) \) are incentive compatible for truthful revelation, they are not budget-balanced. In order to restore budget-balance property, the transfers to the seller and buyer consumer must be:

\[ \psi_1(\theta_1, \theta_2) = \gamma_1(\theta_1) - \hat{\gamma}_2(\theta_1, \theta_2), \]

\[ \psi_2(\theta_1, \theta_2) = \hat{\gamma}_2(\theta_1, \theta_2) - \gamma_1(\theta_1). \]

However \( \hat{\gamma}_2(\theta_1, \theta_2) \) depends on \( \theta_1 \), such transfer \( \psi_1(\theta_1, \theta_2) \) is not incentive compatible for \( \theta_1 \).

### 20.5.3 Efficient Budget-Balanced Dynamic Mechanism

The above difficulty is called the problem of contingent deviation. Athey and Segal (2013) proposed a mechanism that overcomes this difficulty. Similar to the AGV mechanism, their construction proceeds in two steps: (i) construct incentive compatible transfers \( \gamma_1(\theta) \) and \( \gamma_2(\theta) \) to make each agent report truthfully if he expected the other to do so, where \( \theta = (\theta_1, \theta_2) \); (2) charge each agent’s incentive compatible transfer to the other agent, making the total transfer to agent \( i \) equal to \( \psi_i(\theta) = \gamma_i(\theta) - \gamma_j(\theta) \). However, in contrast to AGV transfers, the incentive transfer \( \gamma_i(\theta) \) to agent \( i \) will now depend not just on agent \( i \)’s announcements \( \theta_i \), but also on those of the other agents. How do we then ensure that step (ii) does not destroy incentives? For this purpose, we ensure that even though agent \( i \) can affect the other’s incentive payment \( \gamma_i(\theta_i, \theta_{-i}) \), he cannot manipulate the expectation of the payment given that agent \( i \) reports truthfully. We achieve this by letting \( \gamma_i(\theta_i, \theta_{-i}) \) be the change in the expectation of agent \( i \)’s utility, conditional on all the previous announcements, that is brought about by the report of agent \( i \). (In the general model in which an agent reports in many periods, these incentive transfers would be calculated in each period for the latest report.) No matter what reporting strategy agent \( i \) adopts, if he believes agent
agent $i$ reports truthfully, his expectation of the change in his expected utility due to agent $i$’s future announcements is zero by the law of iterated expectations. Hence agent $i$ can be charged $\gamma_i(\theta_i, \theta_{-i})$ without affecting his incentives.

For the above situation, our construction entails giving the buyer an incentive transfer of

$$\gamma_2(\theta_1, \theta_2) = -\frac{1}{2\theta_1}[(\theta_2)^2 - E_{\hat{\theta}_2}(\hat{\theta}_2)^2].$$

Such transfer is incentive compatible for the buyer. This is because, upon the announcement $\hat{\theta}_2$, the buyer’s utility is

$$\gamma_2(\theta_1, \hat{\theta}_2) + \theta_1[x_1\theta_1 + \theta_2x_2(\theta_1, \hat{\theta}_2)] = -\frac{1}{2\theta_1}[(\hat{\theta}_2)^2 - E_{\hat{\theta}_2}(\hat{\theta}_2)^2] + \frac{1}{\theta_1}(1 + \theta_2(\hat{\theta}_2)).$$

The first condition to $\hat{\theta}_2$ is:

$$-\frac{1}{\theta_1}(\hat{\theta}_2) + \frac{1}{\theta_1}\theta_2 = 0,$$

giving us $\hat{\theta}_2 = \theta_2$, which means the incentive compatibility.

Although $\gamma_2(\theta_1, \hat{\theta}_2)$ depends on $\theta_1$,

$$E_{\hat{\theta}_2}[\gamma_2(\theta_1, \hat{\theta}_2)] = 0, \forall \theta_1.$$ 

Therefore, the seller’s report does not affect the expected transfer to the buyer. As we know, $\gamma_1(\theta_1)$ is the incentive transfer for the seller, so the total transfer for the agents are:

$$\psi_1(\theta_1, \theta_2) = \gamma_1(\theta_1) - \gamma_2(\theta_1, \theta_2),$$

$$\psi_2(\theta_1, \theta_2) = -\psi_1(\theta_1, \theta_2).$$

Hence this transfer scheme are budget-balanced and incentive compatible for the seller, because the seller reports earlier than the buyer, and the transfer to the seller is:

$$E_{\hat{\theta}_2}\psi_1(\theta_1, \hat{\theta}_2) = \gamma_1(\theta_1).$$

In this mechanism proposed by Athey and Segal (2013), the transfer for each agent is equal to the change of other agents’ expected present value, and such
mechanism achieves the incentive compatibility and budget balancedness simulta-
neously.

Athey and Segal (2013) generalize the idea of using incentive payments that
give an agent the change in the expected present value of opponent utilities in-
duced by his report to design an efficient mechanism for a general dynamic model,
interested readers can find more detail in their paper.

Dynamic mechanism design is a hot theoretical topic in recent years. Berge-
mann and Valimaki (Econometrica, 2010) proposed an alternative efficient dy-
namic mechanism, but with new properties, such as satisfying the ex post par-
ticipation constraints and “efficient exit” conditions. Pavan, Segal and Toikka
(Econometrica, 2014) developed a general allocation model and derived the op-
timal dynamic revenue-maximizing mechanism. For interested readers, the last
chapter of the monograph by Borgers (2015) is also a good reference on dynamic
mechanism design.

20.6 References

Textbooks and Monographs

University Press.


Laffont, J. J. and Martimort, D. (2002). The Theory of Incentives: The Principal-
Agent Model, Princeton University Press.

procurement, MIT Press.

University Press, Chapter 14.

Papers:


Part VII

Market Design
The previous part discusses the principal-agent theory and general mechanism design. The last part of this notes will focus on two forefront subfields of modern microeconomics: auction theory and matching theory, which can be regarded as the extensions of the general mechanism design theory. Auction mechanism design and matching mechanism design are the core issues of market design where a realistic, traditional and nature market fails, which is considered to be microeconomic engineering with wide applications in reality. In recent years, the research frontiers have promoted the development of auction mechanism design from single object to multiple objects, and matching mechanism design has been applied to school choice, organ matching and many other aspects.

Traditional economics such new neoclassical general equilibrium theory discusses the interaction between market participants and the equilibrium and welfare outcomes of the interaction in a given market. Since the market usually fails in cases such as the allocation of indivisible goods and the supply of public goods, it is often necessary to rectify the market and hence mechanism design arises.

Similarly, market design does not take market as given, but applies the empirical or experimental results of economics and game theory to the design of market rules. With the rapid development of auction theory that mainly considers mechanism design with transfer payments and matching theory that mainly considers mechanism design without transfer payments, and the establishment of their own analytical frameworks, they have increasingly become independent subfields of modern microeconomic theory.

Chapter 21 discusses basic results of auction theory, including the most fundamental realistic and theoretical auction formats. Auction mechanism design is an important part of market design. Auction theory mainly uses price mechanism characterized by transfer payments to design the allocation of resources in the market. Meanwhile, many results on auction theory can be regarded as the extensions of general mechanism design theory. We will see that many basic conclusions in auction theory can be easily obtained by applying the results in general mechanism design theory introduced in the previous part.
Chapter 22, as the last chapter of the notes, discusses another very important category of mechanism design in market design: design of allocation of indivisible objects without transfers. In most cases (such as human organ transplantation, school choice, and office allocation), transfer payment is not allowed for the allocation of such indivisible objects, so that it is a non-price mechanism. We will first discuss matchings in two-sided market and one-sided market, respectively, and then analyze the applications of matching theory: matching mechanism in school choice and organ transplantation.
Chapter 21

Auction Theory

21.1 Introduction

Auction theory is in the frontier of research in modern economics and has wide applications. It has developed systematic theoretical results, which can be used to analyze and understand the superiority of auction and tender mechanism, and also provided many useful insides for practical operation. The auction theory can actually be considered as a branch or extension of the general mechanism design theory. We will see that by applying the results discussed in the previous chapters on general mechanism design, it is easy to get many fundamental results in the auction theory.

As an effective trading mechanism of objects, auction is applied more and more widely and has penetrated into all aspects of daily economic activities. In reality, many major objects or projects are auctioned or tendered. Objects that are often auctioned include tangible assets such as antiques, paintings, jewelery, used cars, buildings and agricultural products, as well as intangible assets such as land use rights, oilfield exploitation rights, and even some special telephone numbers and license plate numbers. The central banks of many western countries often sell government bonds by auction, and the Ministry of the Interior conducts regular auctions of oil exploration rights. In recent years, the U.S. Department of
Communications adopted auction mechanisms designed by economists to allocate a license to provide personal communications services and has gained unprecedented benefits. It has also brought tremendous revenue to the U.S. Treasury, and similar auction mechanisms have been adopted in European countries.

In China, in order to prevent rent-seeking or corruption, the government clearly stipulates that government procurement, the sale of state property and assets, or construction projects must all be conducted through auction or tender. Every year, local governments in China openly auction large quantities of land to developers and car licenses to consumers. The feature of these auctions is free bid by multiple buyers, where the highest bidder wins the object. Besides, auctions are often used for targeted purchases of goods or services. For example, several companies bid for contracting a project or providing a service, which is referred to as a tender in China. In a tender, usually a company with the lowest price wins the contract, but sometimes the bidder’s reputation, commitment, and quality of the product or service provided also need to be considered.

In fact, the history of auction can be traced back to at least the ancient Babylonian period of 700 BC. Some auction events even influenced the whole course of history. In AD 193, Roman emperor Pertinax was killed by his guards because he wanted to rectify military discipline. Subsequently, the controlling Praetorian Guard auctioned the throne in the barracks. Didius Julianus won the auction by paying almost two kilos of gold to each soldier (the actual payment was 25,000 sesterces), thus getting the support of the Guard and the emperor’s throne. However, it did not last long. Two months later, the insurgent army stormed into Rome, and this politician throned by auction eventually ended up with death.

Although the practice of auction has lasted for thousands of years, some of which even imposed a huge historical impact, the scientific research on auction based on rigorous economic theory did not begin until the 1960s. The first work was Vickrey’s less than 30-page paper titled “Counter speculation, Auctions, and Competitive Sealed Tenders” in the 1961 *Journal of Finance* (William Vickrey,
Four Basic Auction Mechanisms

There are many kinds of auctions, among which four forms are most widely used and studied while others are variations of those four auctions.

The first is ascending-price auction (or English auction). Under this rule, the price of the auctioneer is publicly called, and it will increase until there is only one bidder that bids. Then this bidder wins the object and pays the bid. One variation of this auction is that all the bidders bid and the price increases until there is only one bidder bidding and pays his/her bid. This form is often used in live auction, especially on the cultural relics, calligraphy and painting, second-hand goods and commercial land (in China).

The second is descending-price auction (or Dutch auction). Dutch auction is opposite to the English auction, where the price goes from high to low. The auctioneer calls the price at a high level, if there is no bidder who wants to buy the objects, then the auctioneer reduces the price in accordance with the predetermined range until some bidders want to accept it. When there is bidder who accepts the price, he wins the object by the price. Although the price of this auction goes from high to low, the winner is still the bidder whose bid is the highest. In Europe, especially the Netherlands, the auction on flower always takes this form, and that’s why it is called Dutch auction.

The third is first-price auction. Under this rule, each bidder submits his bid document to the auctioneer in sealed form in the specified time independently and indicates the price he wishes to make. Thus the bid of other bidders cannot be seen. The auctioneer then invites all bidders to open their bids on-site at a
specified time. Then the bidder with the highest bid wins the object by his bid. Thus the first-price auction is also called high price auction, bidding auction or mail auction.

The fourth is the second-price auction, which is also called Vickrey auction. Under this rule, the bid is also sealed, but after opening the bid, the bidder whose bid is the highest wins the object and pays the second highest bid. This form is rarely used in practice, but it has a good theoretical nature. This mechanism has been discussed in Chapter 15 as a special case of VCG mechanism. It was first proposed by Vickrey in 1961, and economists began to make an in-depth study of the auction since then.

In each form of auction, if several bidders bid the same with the highest price, then the auctioneer will randomly select among them as a winner. The above introduction is only for single-unit item auction, and we will also discuss auctions of multi-items. Their auction forms are different, but are basically the variants of the above four mechanisms.

It can be seen from the above introduction of the four basic auction mechanisms that auction mainly adopts two major forms: the first two are open outcry methods, such as antique, calligraphy and painting, second-hand goods (such as used cars) and so on, which require that the bidders collect in the same place, while the latter two are sealed auction, which may be submitted by mail, so a bidder may observe the behavior of other bidders in one format and not in another. In many auctions or bidding activities, the confidentiality of participants’ business secrets should be ensured, so the auction or bidding methods adopted should avoid commercial leaks. Compared with the open outcry methods, sealed price methods have the advantage in this respect, so the use of sealed price bidding is necessary.

However, for rational decision makers, some of these differences are superficial. The Dutch open descending price auction is strategically equivalent to the first-price sealed-bid auction in the sense that they have the same equilibrium strategies. When values are private, the English open ascending auction is also
outcome equivalent to the second-price sealed-bid auction in equilibrium outcome in a weak sense that they have the same equilibrium outcomes.

In addition, sellers often have two restrictions for each auction mechanism. One is to set the base price, which is called the reserve price; the other is the charge for bidding. Under the first-price and second-price auction mechanisms, the bidder must bid higher than or equal to the reserve price, otherwise no transaction will be made. As for the second-price auction, if there is only one bidder bidding and bids above the reserve price, he will get the goods and pays the reserve price. The reserve price has a similar effect in both the ascending-price and descending-price auctions.

There are two basic issues to be discussed in the auction theory: (1) Can the resources be efficiently allocated using these auctions? (2) Which auction mechanism can make the most profit for the auctioneer? To answer these two questions, we need to describe the basic analytical framework of the auction mechanism, including a description of bidders’ preferences and the information structure – private value, interdependent value, or common value (a special case of interdependent value). Then we will analyze their bidding strategies to see if they lead to efficient outcomes and compare the auctioneer’s expected revenue.

In the following, we will first examine auctions in the case of private value and focus mainly on the symmetric, risk-neutral, unlimited liability private value economic environment with single object. We then discuss auctions under the context of interdependent value for single object and auctions for multiple objects. As an auction mechanism is a special case of incomplete information mechanism design, the solution concept used is mainly the Bayesian Nash equilibrium solution. In order to make it easier to understand, we try to use the same terms and notations in Chapters 15-16. We first provide some basic results on stochastic dominance.
21.2 Some Useful Mathematics

21.2.1 Continuous Distributions

Given a random variable $X$, which takes on values in $[0, \omega]$, we define its cumulative distribution function $F : [0, \omega] \rightarrow [0, 1]$ by

$$F(x) = \text{Prob}[X \leq x]$$

the probability that $X$ takes on a value not exceeding $x$. By definition, the function $F$ is nondecreasing and satisfies $F(0) = 0$ and $F(\omega) = 1$ (if $\omega = \infty$, then $\lim_{x \to \infty} F(x) = 1$). In this course, we always suppose that $F$ is increasing and continuously differentiable.

The derivative of $F$ is called the associated probability density function and is usually denoted by the corresponding lowercase letter $f \equiv F'$. By assumption, $f$ is continuous and we will suppose that for all $x \in (0, \omega)$, $f(x)$ is positive. The interval $[0, \omega]$ is called the support of the distribution.

If $X$ is distributed according to $F$, then the expectation of $X$ is

$$E(X) = \int_0^\omega x f(x) dx$$

and if $\gamma : [0, \omega] \rightarrow \mathbb{R}$ is some arbitrary function, then the expectation of $\gamma(X)$ is analogously defined as

$$E[\gamma(X)] = \int_0^\omega \gamma(x)f(x) dx.$$

Sometimes the expectation of $\gamma(X)$ is also written as

$$E[\gamma(X)] = \int_0^\omega \gamma(x) dF(x).$$

The conditional expectation of $X$ given that $X < x$ is

$$E[X|X < x] = \frac{1}{F(x)} \int_0^x t f(t) dt$$

and so

$$F(x)E[X|X < x] = \int_0^x tf(t)dt = xF(x) - \int_0^x F(t) dt$$

which is obtained by integrating the right-hand side of the first equality by parts.
21.2.2 Hazard Rates

Let $F$ be a distribution function with support $[0, \omega]$. The hazard rate of $F$ is the function $\lambda : [0, \omega) \to \mathbb{R}_+$ defined by

$$\lambda(x) \equiv \frac{f(x)}{1 - F(x)}.$$ 

If $F$ represents the probability that some event will happen before time $x$, then the hazard rate at $x$ represents the instantaneous probability that the event will happen at $x$, given that it has not happened until time $x$. The event may be the failure of some component—a lightbulb, for instance—and hence it is sometimes also known as the “failure rate”.

Solving for $F$, we have

$$F(x) = 1 - \exp\left(-\int_0^x \lambda(t)dt\right).$$

This shows that any arbitrary function $\lambda : [0, \omega) \to \mathbb{R}_+$ such that for all $x < \omega$,

$$\int_0^x \lambda(t)dt < \infty, \quad \lim_{x \to \omega} \int_0^x \lambda(t)dt = \infty,$$

is the hazard rate of some distribution.

Closely related to the hazard rate is the function $\sigma : (0, \omega] \to \mathbb{R}_+$ defined by

$$\sigma(x) \equiv \frac{f(x)}{F(x)},$$

sometimes known as the reverse hazard rate or is referred to as the inverse of the Mills’ ratio. Similarly, Solving for $F$, we have

$$F(x) = \exp\left(-\int_x^\omega \sigma(t)dt\right),$$

This shows that any arbitrary function $\sigma : (0, \omega] \to \mathbb{R}_+$ such that for all $x > 0$,

$$\int_x^\omega \sigma(t)dt < \infty, \quad \lim_{x \to 0} \int_x^\omega \sigma(t)dt = \infty,$$

is the “reverse hazard rate” of some distribution.
21.2.3 Stochastic Dominance

First-Order Stochastic Dominance

Definition 21.2.1 (First-Order Stochastic Dominance) Given two distribution functions $F$ and $G$, we say that $F$ first-order stochastically dominates $G$ if for all $z \in [0, \omega], F(z) \leq G(z)$.

The first-order stochastic dominance means that for any outcome $x$, the probability of obtaining at least $x$ under $F(\cdot)$ is at least as high as that under $G(\cdot)$, i.e., $F(\cdot) \leq G(\cdot)$ implies that the probability in lower part under $F(\cdot)$ is smaller than under $G(\cdot)$, or that the probability in higher part under $F(\cdot)$ is larger than under $G(\cdot)$. This is analogous to the monotonicity concept under certainty.

There is another test criterion for $F$ to first-order stochastically dominate $G$: Does every expected utility maximizer with an increasing utility function prefer $F(\cdot)$ over $G(\cdot)$?

The following theorem shows that these two criterions are equivalent.

**Theorem 21.2.1** $F(\cdot)$ first-order stochastically dominates $G(\cdot)$ if and only if for any function $u : [0, \omega] \to \mathcal{R}$ that is (weakly) increasing and differentiable, we have

$$
\int u(z)dF(z) \geq \int u(z)dG(z).
$$

**Proof.** Define $H(z) = F(z) - G(z)$. We need to prove that $H(z) \leq 0$ if and only if $\int u(z)dH(z) \geq 0$ for any increasing and differentiable function $u(\cdot)$.

**Sufficiency:** We prove this by way of contradiction. Suppose there is a $\hat{z}$ such that $H(\hat{z}) > 0$. We choose a weakly increasing and differentiable function $u(z)$ as

$$
u(z) = \begin{cases} 
0, & z \leq \hat{z}, \\
1, & z > \hat{z}, 
\end{cases}
$$

then immediately $\int u(z)dH(z) = -H(\hat{z}) < 0$, a contradiction.

**Necessity:**

$$
\int u(z)dH(z) = [u(z)H(z)]^\infty_{-\infty} - \int u'(z)H(z)dz = 0 - \int u'(z)H(\hat{z})dz \geq 0,
$$

834
in which the first equality is based on the formula of integration by parts, the second equality is based on

\[ F(-\infty) = G(-\infty) = 0, \quad F(\infty) = G(\infty) = 1, \]

while the inequality is based on the assumptions that \( u(\cdot) \) is weakly increasing \( (u'(\cdot) \geq 0) \) and \( H(z) \leq 0 \).

Since a monotone function can be arbitrarily approximated by a sequence of monotone and differentiable functions, the differentiability requirement imposed on \( u \) is not necessary. For any two gambles \( F \) and \( G \), as long as an agent’s utility is (weakly) increasing in outcomes, he prefers the one that first-order stochastically dominates the other one.

**Second-Order Stochastic Dominance**

**Definition 21.2.2 (Second-Order Stochastic Dominance)** Given two distribution functions \( F \) and \( G \) with the same expectation, we say that \( F(\cdot) \) second-order stochastically dominates \( G(\cdot) \) if

\[
\int_{-\infty}^{z} F(r)dr \leq \int_{-\infty}^{z} G(r)dr
\]

for all \( z \).

It is clear that first-order stochastic dominance implies second-order stochastic dominance. The second-order stochastic dominance implies not only monotonicity but also lower risk. To do so, we introduce the notion of “Mean-Preserving Spreads”.

Suppose \( X \) is a random variable with distribution function \( F \). Let \( Z \) be a random variable whose distribution conditional on \( X = x \), \( H(\cdot|X = x) \), is such that for all \( x \), \( E[Z|X = x] = 0 \). Suppose \( Y = X + Z \) is the random variable obtained from first drawing \( X \) from \( F \) and then for each realization \( X = x \), drawing a \( Z \) from the conditional distribution \( H(\cdot|X = x) \) and adding it to \( X \). Let \( G \) be the distribution of \( Y \) so defined. We will then say that \( G \) is a mean-preserving spread of \( F \).
While random variables $X$ and $Y$ have the same mean, namely $E[X] = E[Y]$, variable $Y$ is “more spread-out” than $X$ since it is obtained by adding a “noise” variable $Z$ to $X$. Now suppose $u : [0, \omega] \to \mathcal{R}$ is a concave function. Using Jensen’s inequality

$$E(u(X)) \leq E(u(Y)),$$

we obtain

$$E_Y[u(Y)] = E_X[E_Z[u(X + Z)|X = x]]$$

$$\leq E_X[u(E_Z[X + Z|X = x])]$$

$$= E_X[u(X)].$$

As such, similar to Theorem 21.2.1, we have the following conclusion for the second-order stochastic dominance.

**Theorem 21.2.2** If distributions $F(\cdot)$ and $G(\cdot)$ have the same mean, then the following statements are equivalent.

1. $F(\cdot)$ second-order stochastically dominates $G(\cdot)$;
2. for any nondecreasing concave function $u : \mathcal{R} \to \mathcal{R}$, we have $\int u(z)dF(z) \geq \int u(z)dG(z)$;
3. $G(\cdot)$ is a mean-preserving spread of $F(\cdot)$.

**Proof.** (3)$\Rightarrow$(2): It is obtained by using

$$\int u(z)dF(z) = \int u\left(\int (x + z)dH_x(x)\right)dF(z)$$

$$\geq \int \left(\int u(x + z)dH_z(x)\right)dF(z)$$

$$= \int u(z)dG(z),$$

in which the inequality follows from the concavity of $u(\cdot)$.  

836
(1)⇒(2): For expositional convenience, we set $\omega = 1$. We have
\[
\int u(z)dF(z) - \int u(z)dG(z) = -u'(1) \int_0^1 (F(z) - G(z))dz + \int \left(\int_0^z (F(x) - G(x))dx\right)u''(z)dz \\
= \int \left(\int_0^z (F(x) - G(x))dx\right)u''(z)dz \geq 0,
\]
in which the inequality follows from the definition of second-order stochastic dominance, namely
\[
\int_{-\infty}^{z} F(r)dr \leq \int_{-\infty}^{z} G(r)dr,
\]
and also $u''(\cdot) \leq 0$ for any $z$. We thus have
\[
\int u(z)dF(z) - \int u(z)dG(z) \geq 0.
\]

(1)⇒(3): We just show the case with discrete distributions.

Define
\[
S(z) = G(z) - F(z), \\
T(x) = \int_0^x S(z)dz.
\]

By the definition of second-order stochastic dominance, we have $T(x) \geq 0$ and $T(1) \geq 0$, which imply that there exists some $\hat{z}$ such that $S(z) \geq 0$ for $z \leq \hat{z}$ and $S(z) \leq 0$ for $z \geq \hat{z}$.

Since the random variable follows a discrete distribution, $S(z)$ must be a step function. Let $I_1 = (a_1, a_2)$ be the first interval over which $S(z)$ is positive, and $I_2 = (a_3, a_4)$ be the first interval over which $S(z)$ is negative. If no such $I_1 = (a_1, a_2)$ exists, then $S(z) \equiv 0$ and hence statement (3) is immediate. If $I_1 = (a_1, a_2)$ does exist, then $I_2 = (a_3, a_4)$ must exist as well.

So, $S(z) \equiv \gamma_1 > 0$ for $z \in I_1$, and $S(z) \equiv -\gamma_2 < 0$ for $z \in I_2$. By $T(x) \geq 0$, we must have $a_2 < a_3$. If $\gamma_1(a_2 - a_1) \geq \gamma_2(a_4 - a_3)$, then there exist $a_1 < \hat{a}_2 \leq a_2$ and $\hat{a}_4 = a_4$ such that $\gamma_1(\hat{a}_2 - a_1) = \gamma_2(\hat{a}_4 - a_3)$. If $\gamma_1(a_2 - a_1) < \gamma_2(a_4 - a_3)$, then there exists $a_3 < \hat{a}_4 \leq a_4$ such that $\gamma_1(\hat{a}_2 - a_1) = \gamma_2(\hat{a}_4 - a_3)$. 837
Letting
\[ S_1(z) = \begin{cases} 
\gamma_1, & \text{if } a_1 < z < \hat{a}_2, \\
-\gamma_2, & \text{if } a_3 < z < \hat{a}_4, \\
0, & \text{otherwise.}
\end{cases} \]

If \( F_1 = F + S_1 \), then \( F_1 \) is a mean-preserving spread of \( F \). Letting \( S_1^i = G - F_1 \), then we can similarly construct \( S_2(z) \) and \( F_2 \). Since \( S(z) \) is a step function, then there exists an \( n \) such that \( F_0 = F, F_n = G \), and \( F_{i+1} \) is a mean-preserving spread of \( F_i \). Also, a finite summation of mean-preserving spreads is still a mean-preserving spread.

Though a continuous function can be arbitrarily approximated by step functions, the formal proof is complicated, and Rothschild and Stiglitz (1971) provide a complete proof for the case with continuous distributions.

### 21.2.4 Hazard Rate Dominance

**Definition 21.2.3 (Hazard Rate Dominance)** For any two distributions \( F \) and \( G \) with hazard rates \( \lambda_F \) and \( \lambda_G \), respectively. We say that \( F \) **dominates** \( G \) in terms of the hazard rate if \( \lambda_F(x) \leq \lambda_G(x) \) for all \( x \). This order is also referred in short as **hazard rate dominance**.

If \( F \) dominates \( G \) in terms of the hazard rate, then
\[
F(x) = 1 - \exp \left( - \int_0^x \lambda_F(t)dt \right) \leq 1 - \exp \left( - \int_0^x \lambda_G(t)dt \right) = G(x),
\]
and hence \( F \) stochastically dominates \( G \). Thus, hazard rate dominance implies first-order stochastic dominance.

### 21.2.5 Reverse Hazard Rate Dominance

**Definition 21.2.4 (Reverse Hazard Rate Dominance)** For any two distributions \( F \) and \( G \) with reverse hazard rates \( \sigma_F \) and \( \sigma_G \), respectively. We say that \( F \) **dominates** \( G \) in terms of the reverse hazard rate if \( \sigma_F(x) \geq \sigma_G(x) \) for all \( x \). This order is also referred in short as **reverse hazard rate dominance**.
If $F$ dominates $G$ in terms of the reverse hazard rate, then

$$F(x) = \exp \left( -\int_x^\infty \sigma_F(t) dt \right) \leq \exp \left( -\int_x^\infty \sigma_G(t) dt \right) = G(x),$$

and hence, again, $F$ stochastically dominates $G$. Thus, reverse hazard rate dominance also implies first-order stochastic dominance.

### 21.2.6 Order Statistics

Let $X_1, X_2, \ldots, X_n$ be $n$ independently draws from a distribution $F$ with associated density $f$. Let $Y_1^{(n)}, Y_2^{(n)}, \ldots, Y_n^{(n)}$ be a rearrangement of these so that

$$Y_1^{(n)} \geq Y_2^{(n)} \geq \cdots \geq Y_n^{(n)}.$$

These random variables $Y_k^{(n)}, k = 1, 2, \ldots, n$ are referred to as order statistics.

Let $F_k^{(n)}$ denote the distribution of $Y_k^{(n)}$, with corresponding probability density function $f_k^{(n)}$. When the “sample size” $n$ is fixed and there is no ambiguity, we simply write $Y_k$ instead of $Y_k^{(n)}$, $F_k$ instead of $F_k^{(n)}$ and $f_k$ instead of $f_k^{(n)}$. In auction theory, we will typically be interested in properties of the highest and second highest order statistics, namely $Y_1$ and $Y_2$.

#### Highest Order Statistic

The distribution of the highest order statistic $Y_1$ is easy to derive. The event that $Y_1 \leq y$ is the same as the event: for all $k$, $X_k \leq y$. Since each $X_k$ is an independent draw from the same distribution $F$, we have that

$$F_1(y) = F(y)^n.$$

The associated probability density function is

$$f_1(y) = nF(y)^{n-1}f(y).$$

Observe that if $F$ stochastically dominates $G$, and $F_1$ and $G_1$ are the distributions of the highest order statistics of $n$ draws from $F$ and $G$, respectively, then $F_1$ stochastically dominates $G_1$. 839
Second-Highest Order Statistic

The distribution of the second-highest order statistic $Y_2$ can also be easily derived. The event that $Y_2 \leq y$ is the union of the following disjoint events: (1) all $X_k$’s are less than or equal to $y$; and (2) $n - 1$ of the $X_k$’s are less than or equal to $y$ and one is greater than $y$. There are $n$ different ways in which (2) can occur, so we have that

$$F_2(y) = F(y)^n + nF(y)^{n-1}(1 - F(y))$$

$$= nF(y)^{n-1} - (n - 1)F(y)^n.$$

The associated probability density function is

$$f_2(y) = n(n - 1)(1 - F(y))F(y)^{n-2}f(y).$$

Again, it can be verified that if $F$ stochastically dominates $G$ and also $F_2$ and $G_2$ are the distributions of the second-highest order statistics of $n$ draws from $F$ and $G$, respectively, then $F_2$ stochastically dominates $G_2$.

21.2.7 Affiliation

Affiliation is a basic assumption used to study auction with interdependent values which are non-negative correlated.

**Definition 21.2.5** Suppose the random variables $X_1, X_2, \cdots, X_n$ are distributed on some product of intervals $\mathcal{X} \subseteq \mathbb{R}^n$ according to the joint density function $f$. The variables $X = (X_1, X_2, \cdots, X_n)$ are said to be affiliated if for all $x', x'' \in \mathcal{X}$,

$$f(x' \lor x'')f(x' \land x'') \geq f(x')f(x''), \quad (21.1)$$

in which

$$x' \lor x'' = (\max(x'_1, x''_1), \cdots, \max(x'_n, x''_n))$$

denotes the component-wise maximum of $x'$ and $x''$, and

$$x' \land x'' = (\min(x'_1, x''_1), \cdots, \min(x'_n, x''_n))$$
denotes the component-wise minimum of $x'$ and $x''$. If (21.1) is satisfied, then we also say that $f$ is affiliated.

Suppose that the density function $f : \mathcal{X} \to \mathcal{R}_+$ is strictly positive in the interior of $\mathcal{X}$ and twice continuously differentiable. It is “easy” to verify that $f$ is affiliated if and only if, for all $i \neq j$,

$$\frac{\partial^2}{\partial x_i \partial x_j} \ln f \geq 0.$$  

In other words, the off-diagonal elements of the Hessian of $\ln f$ are nonnegative.

**Proposition 21.2.1** Let $X_1, X_2, \cdots, X_n$ be random variables, and $Y_1, Y_2, \cdots, Y_{n-1}$ be the largest, second largest, ..., smallest order statistics from among $X_2, X_3, \cdots, X_n$. If $X_1, X_2, \cdots, X_n$ are symmetrically distributed and affiliated, then we have

1. variables in any subset of $X_1, X_2, \cdots, X_n$ are also affiliated;
2. $X_1, Y_1, Y_2, \cdots, Y_{n-1}$ are affiliated.

**Monotone Likelihood Ratio Property**

Suppose the two random variables $X$ and $Y$ have a joint density $f : [0, \omega]^2 \to \mathcal{R}$. If $X$ and $Y$ are affiliated, then for all $x' \geq x$ and $y' \geq y$, we have

$$f(x', y)f(x, y') \leq f(x, y)f(x', y') \iff \frac{f(x', y')}{f(x, y)} \leq \frac{f(x', y)}{f(x, y')}$$  \hspace{1cm} (21.2)

and

$$\frac{f(y'|x)}{f(y|x)} \leq \frac{f(y'|x')}{f(y|x')} ,$$

so the likelihood ratio

$$\frac{f(y'|x')}{f(y|x')}$$

is increasing and this is referred to as the monotone likelihood ratio property.
Likelihood Ratio Dominance

Definition 21.2.6 (Likelihood Ratio Dominance) The distribution function $F$ dominates $G$ in terms of the likelihood ratio if for all $x < y$,

$$\frac{f(x)}{f(y)} \leq \frac{g(x)}{g(y)}.$$

We thus have the following conclusion.

Proposition 21.2.2 If $X$ and $Y$ are affiliated, the following properties hold:

1. For all $x' \geq x$, $F(\cdot|x')$ dominates $F(\cdot|x)$ in terms of hazard rate; that is,

$$\lambda(y|x') \equiv \frac{f(y|x')}{1 - F(y|x')} \leq \frac{f(y|x)}{1 - F(y|x)} \equiv \lambda(y|x).$$

Or equivalently, for all $y$, $\lambda(y|\cdot)$ is nonincreasing.

2. For all $x' \geq x$, $F(\cdot|x')$ dominates $F(\cdot|x)$ in terms of the reverse hazard rate; that is,

$$\sigma(y|x') \equiv \frac{f(y|x')}{F(y|x')} \leq \frac{f(y|x)}{F(y|x)} \equiv \sigma(y|x),$$

or equivalently, for all $y$, $\sigma(y|\cdot)$ is nondecreasing.

3. For all $x' \geq x$, $F(\cdot|x')$ stochastically dominate $F(\cdot|x)$; that is,

$$F(y|x') \leq F(y|x),$$

or equivalently, for all $y$, $F(y|\cdot)$ is nonincreasing.

All of these results extend in a straightforward manner to the case where the number of conditioning variables is more than one. Suppose $Y, X_1, X_2, \cdots, X_n$ are affiliated and let $F_Y(\cdot|x)$ denote the distribution of $Y$ conditional on $X = x$. Then, using the same arguments as above, it can be deduced that for all $x' \geq x$, $F_Y(\cdot|x')$ dominates $F_Y(\cdot|x)$ in terms of the likelihood ratio. The other dominance relationships then follow as usual.
21.3 Private Value Auctions for Single Object

21.3.1 Basic Analytical Framework

Here we mainly consider the private value benchmark model of auction theory for single object, which usually includes the following six hypotheses:

1. **Private Value**: The valuation of the item by bidders is private information, depending only on their own type and having nothing to do with the type of others;

2. **Independence**: The types of bidders are independent;

3. **Symmetry**: The types of bidders have the same probability distribution;

4. **Risk Neutrality**: The expected utility function of bidders is risk-neutral;

5. **Unlimited liability**: Bidders have no budget constraints and have the ability to pay the bidding price;

6. **Non-collusion**: All buyers decide their own bidding strategy independently and there is no binding cooperative agreement.

The above description of these economic environments is called Symmetric Independent Private Value (in short, SIPV) model. We mainly consider solving symmetric Bayesian-Nash equilibrium under this model and discuss their properties, although there may be asymmetric solutions under some general utility function.

We will ask two basic questions:

(i) What are symmetric equilibrium strategies in a first-price auction (I) and a second-price auction (II)?

(ii) From the point of view of the auctioneer or seller, which of the two auction formats yields a higher expected selling price in equilibrium?
Consider an auction with an indivisible object owned by the auctioneer and \( n \) risk-neutral bidders. Every bidder \( i \) has a private value of the object, denoted by \( \theta_i \), the type of the bidder. In this way, the auction model is considered a special case of indivisible object model we studied in Chapters 15-16: \( Y = \{ y \in \{0, 1\}^n : \sum_i y_i = 1 \} \), and the payoff of buyer \( i \) can be written as

\[
\theta_i y_i + t_i,
\]

where \( t_i \) is the corresponding transfer payment. The value for buyer \( i \) is a random variable on \([\bar{\theta}_i, \theta_i] \) with independent density \( \varphi_i(\cdot) > 0 \), where \( \bar{\theta}_i < \theta_i \), and the cumulative distribution function is denoted by \( \Phi_i(\cdot) \).

We first examine equilibrium outcomes of these common mechanisms under symmetric scenarios, and discuss the forms and properties of the Bayesian-Nash equilibrium, such as validity. Then we compare the seller’s expected payoff under these auction mechanisms.

In private value auctions, some auction rules are outcome-equivalent. For example, the first-price (sealed-bid) auction and the Dutch auction have the highest price called or accepted as the auction price, and so the two auction results are strategically equivalent. For the second-price (sealed-bid) auction and the English auction, as the auction item is private, other bidders’ withdrawal will not affect the bidder’s value expectation on the object, where the second highest price called or accepted is the auction price, so the two auctions are also outcome-equivalent. Therefore, without loss of generality, we only need to discuss the first-price (sealed-bid) auction and second-price (sealed-bid) auction for the private value environment.

Below we focus on the case of symmetry. Assuming that all bidders’ private values are symmetric, that is, for any \( i \), there is \( \theta_i \sim \varphi(\cdot)[0, w] \), and the corresponding distribution function is \( \Phi(\cdot) \). At the same time we consider the symmetry of Bayesian-Nash equilibrium. In symmetric Bayesian-Nash equilibrium, we focus on examining bidder 1’s strategic choices.
21.3.2 First-Price Sealed-Bid Auction

In a first-price auction, given bidder $i$’s value $\theta_i$, if his bid is $b_i$, other bidder’s bid is $b_j, j \neq i$, then bidder $i$’s utility is $U_i = \theta_i y_i(b_i, b_{-i}) + t_i(b_i, b_{-i})$, where the allocation rule is given by

$$y_i(b_i, b_{-i}) = \begin{cases} 1 & \text{if } b_i > \max_{j \neq i} b_j \\ 0 & \text{if } b_i < \max_{j \neq i} b_j \end{cases}$$  \hspace{1cm} (21.3)$$

and transfer payment is given by

$$t_i(b_i, b_{-i}) = \begin{cases} -b_i & \text{if } b_i > \max_{j \neq i} b_j \\ 0 & \text{if } b_i < \max_{j \neq i} b_j \end{cases}$$  \hspace{1cm} (21.4)$$

If $b_i = \max_{j \neq i} b_j$, the object is randomly assigned by the same probability. Consider a symmetric equilibrium denoted by $\beta^I(\cdot)$, which is a mapping $\beta^I : [0, \omega] \rightarrow [0, \infty)$. Without loss of generality, suppose $i = 1$. Let $\vartheta_1, \ldots, \vartheta_{n-1}$ be the order from the largest to smallest of $\theta_2, \ldots, \theta_n$, or $\theta_2, \ldots, \theta_n$’s 1st, ..., $n-1$th sequential order statistics. Suppose $\beta^I(\cdot)$ is an increasing function, as we will show later.

When other bidders choose strategy $\beta^I(\cdot)$, then only if $b_1 > \beta(\vartheta_1)$, bidder 1 can get the object, otherwise he would not. Since $\vartheta_1$ is a continuous random variable, the probability of $b_1 = \beta(\vartheta_1)$ is zero, so we do not need to consider it. As $\theta_j, j \neq 1$ are all distributed according to $\varphi$’s density distribution (the corresponding density function is $\Phi(\cdot)$), it is easy to obtain that $\vartheta_1$’s distribution function is $\Psi(\cdot) = \Phi^{n-1}(\cdot)$ (the corresponding density function is $\psi$).

Given that other bidders choose strategies $\beta(\cdot) = \beta^I(\cdot)$, the (interim) expected utility of bidder 1 with value $\theta$ and bid $b$ then is:

$$E_{\theta_{-1}} U_1 = \Psi(\beta^{-1}(b))(\theta - b),$$

where $\beta^{-1}(\cdot)$ is the inverse function of $\beta(\cdot)$.

The first order condition for $b$ is:

$$\frac{\psi(\beta^{-1}(b))(\theta - b)}{\beta'(\beta^{-1}(b))} - \Psi(\beta^{-1}(b)) = 0.$$
In equilibrium, \( b = \beta(\theta) \), then we have:

\[
\beta'(\theta) = \frac{\psi(\theta)}{\Psi(\theta)} (\theta - \beta(\theta)).
\] (21.5)

Since \( \beta(0) = 0 \), the solution of first differential equation (21.5) is:

\[
\beta(\theta) = \frac{1}{\Psi(\theta)} \int_0^\theta \vartheta_1 \psi(\vartheta_1) d\vartheta_1 = E[\vartheta_1|\vartheta_1 < \theta].
\]

Obviously \( \beta(\theta) < \theta \). From the differential equation (21.5), we get \( \beta'(\cdot) > 0 \), i.e., \( \beta(\cdot) \) is an increasing function. We then have the following proposition.

**Proposition 21.3.1** With symmetric independent private value, the symmetric Bayesian-Nash strategic equilibrium of first-price auction is

\[
\beta^I(\theta) = \frac{1}{\Psi(\theta)} \int_0^\theta \vartheta_1 \psi(\vartheta_1) d\vartheta_1 = E[\vartheta_1|\vartheta_1 < \theta].
\]

**Proof.** We obtain that strategy is a (first-order) necessary condition for Bayesian-Nash equilibrium from the above argument, and in the following we only need to show the sufficiency of the proposition.

If other bidders all choose \( \beta = \beta^I \), under \( \theta \), bidder \( i \) chooses bid \( \beta(\tilde{\theta}) \), and then his expected utility is:

\[
\bar{U}_i(\theta, \tilde{\theta}) = E_{\theta_{-i}}[\theta y_i(\beta(\tilde{\theta}), \beta_{-i}(\theta_{-i})) + t_i(\beta(\tilde{\theta}), \beta_{-i}(\theta_{-i}))]
\]

\[
= \Psi(\tilde{\theta})[\theta - \beta(\tilde{\theta})]
\]

\[
= \Psi(\tilde{\theta})\theta - \Psi(\tilde{\theta}) E[\vartheta_1|\vartheta_1 < \tilde{\theta}]
\]

\[
= \Psi(\tilde{\theta})\theta - \int_0^\tilde{\theta} \vartheta_1 \psi(\vartheta_1) d\vartheta_1
\]

\[
= \Psi(\tilde{\theta})\theta - \Psi(\tilde{\theta})\tilde{\theta} + \int_0^\tilde{\theta} \Psi(\vartheta_1) d\vartheta_1 \text{ (Integration by parts)}
\]

\[
= \Psi(\tilde{\theta})(\theta - \tilde{\theta}) + \int_0^\tilde{\theta} \Psi(\vartheta_1) d\vartheta_1.
\]

Then, no matter \( \theta \geq \tilde{\theta} \) or \( \theta \leq \tilde{\theta} \), we have:

\[
\bar{U}(\theta, \theta) - \bar{U}(\theta, \tilde{\theta}) = \Psi(\tilde{\theta})(\tilde{\theta} - \theta) + \int_\theta^{\tilde{\theta}} \Psi(\vartheta_1) d\vartheta_1 \geq 0.
\]

846
Using integration by parts, the Bayesian-Nash equilibrium bidding solution can be rewritten as:

\[
\beta^I(\theta) = \theta - \frac{1}{\Psi(\theta)} \int_{0}^{\theta} \Psi(\vartheta) d\vartheta < \theta.
\]  

(21.6)

It is easy to verify that \( \beta^I(\theta) \) is a monotonically increasing function. The equilibrium bidding has a positive correlation with the private value, that is, for high private value, the equilibrium bidding is also high, but the bid is always less than the private value. That is, no bidders have the incentive to show their true type. However, since

\[
\frac{\Psi(\vartheta)}{\Psi(\theta)} = \left[ \frac{\Phi(\vartheta)}{\Phi(\theta)} \right]^{N-1},
\]

when the number of bidders \( N \) increase, which means it is more competitive, the Bayesian-Nash equilibrium solution \( \beta^I(x) \) converges to \( x \).

**Example 21.3.1** Suppose values are exponentially distributed on \([0, \infty)\), and there are only two bidders. If \( \Phi(\theta) = 1 - \exp(-\lambda \theta), \lambda > 0 \), then

\[
\beta^I(\theta) = \theta - \int_{0}^{\theta} \frac{\Phi(\vartheta)}{\Phi(\theta)} d\vartheta
\]

\[
= \frac{1}{\lambda} \frac{\theta \exp(-\lambda \theta)}{1 - \exp(-\lambda \theta)}.
\]

\[\text{(21.7)}\]

**21.3.3 Second-Price Sealed-Bid Auction**

Now we discuss the second-price (sealed-bid) auction equilibrium. The highest price for other bidders is \( \bar{b}_{(i)} \equiv \max_{j \neq i} b_j \), where \( b_j \) is the bid of bidder \( j \). For a bidder \( i \) with private value \( \theta_i \), the utility of bid \( b_i \) is \( U_i = \theta_i y_i(b_i, \bar{b}_{(i)}) + t_i(b_i, \bar{b}_{(i)}) \), where

\[
y_i(b_i, \bar{b}_{(i)}) = \begin{cases} 
1 & \text{if } b_i > \bar{b}_{(i)} \\
0 & \text{if } b_i < \bar{b}_{(i)}.
\end{cases}
\]

(21.7)

\[
t_i(b_i, \bar{b}_{(i)}) = \begin{cases} 
-\bar{b}_{(i)} & \text{if } b_i > \bar{b}_{(i)} \\
0 & \text{if } b_i < \bar{b}_{(i)}.
\end{cases}
\]

(21.8)
The following proposition gives the properties of Bayesian-Nash equilibrium strategy for a second-price sealed-bid auction $\beta^{II}$.

**Proposition 21.3.2** For a second-price (sealed-bid) auction mechanism, all bidders’ Bayesian-Nash equilibrium strategy $\beta^{II}(\theta) = \theta$ is a weakly dominant strategy.

**Proof.** This result can be obtained by applying the Vickery-Clarke-Grove mechanism directly. We present a direct proof of this result.

When $\theta_i > \bar{b}_{(i)}$, bidding $b_i \geq \bar{b}_{(i)}$ and the true value $\theta_i$ bring the same revenue $\theta - \bar{b}_{(i)} > 0$, but when $b_i < \bar{b}_{(i)}$, the opportunity to win is lost, so that it is less than the revenue brought by bidding the true value $\theta_i$. Thus, when $\theta_i > \bar{b}_{(i)}$, $\beta^{II}(\theta) = \theta$ is a weakly dominant strategy.

When $\theta_i < \bar{b}_{(i)}$, bidding $b_i < \bar{b}_{(i)}$ and the true value $\theta_i$ bring the same revenue 0, but when the bidding price $b_i \geq \bar{b}_{(i)}$, the revenue $\theta - \bar{b}_{(i)} < 0$ is smaller than the payoff of bidding $\theta_i$. Thus, when $\theta_i < \bar{b}_{(i)}$, $\beta^{II}(\theta) = \theta$ is also a weakly dominant strategy.

When $\theta_i = \bar{b}_{(i)}$, the payoff of choosing any bid $b_i$ and choosing $\theta_i$ are the same. Therefore, for any $\bar{b}_{(i)}$, truly revealing $\theta_i$ is a weakly dominant strategy. ■

As such, in the second-price sealed-bid auction, all the bidders will bid truthfully, that is, their true values are reported; while in the first-price sealed-bid auction, the bidders’ bid are lower than the true values. The intuition of this conclusion is very simple. When using the first-price sealed-bid auction, if the bidder bids the true value of the object according to his own valuation, then even winning the auction is unprofitable. In order to obtain the potential profits, bidders have the incentive to submit a price lower than its true value, while this issue can be well solved in the second-price sealed-bid auction.
21.3.4 Revenue Comparison of First and Second Auctions

As mentioned above, Vickrey (1961) compared the four auction forms that are most widely used in single-object auctions, and obtained the landmark theory of auction - the “Revenue Equivalence Theorem”. We now compare the expected revenue of auctioneers with the first-price auction and the second-price auction under symmetric conditions.

First of all, for a first-price auction, the interim expected expenditure of bidder 1 with private value vector $\theta$ to the auctioneer is:

$$m^{I}(\theta) \equiv -E_{\theta_{1}}t_{1}^{I}(\theta) = \Psi(\theta)\beta_{1}(\theta) = E[\theta_{1}|\theta_{1} < \theta],$$

i.e., $m^{I}(\theta)$ has the sign difference of interim expected transfer payment $\bar{t}_{1}^{I}(\theta) \equiv E_{\theta_{1}}t_{1}^{I}(\theta)$. Here $\Psi(\theta)$ is the probability that bidder 1 wins the auction.

Therefore, the auctioneer’s ex ante expected revenue from bidder $i$ is:

$$Em^{I}(\theta) = -\int_{0}^{w} \bar{t}_{1}^{I}(\theta)\varphi(\theta)d\theta$$

$$= \int_{0}^{w} (\int_{0}^{\theta} \varphi(\theta)d\theta)\varphi(\theta)d\theta$$

$$= \int_{0}^{w} \psi(\theta)d\theta.$$

Thus, the ex ante expected revenue that an auctioneer receives from $n$ bidders is:

$$E_{\theta}R^{I} = nEm^{I}(\theta)$$

$$= \int_{0}^{w} n(1 - \Phi(\theta_{1}))\varphi(\theta_{1})d\theta_{1}.$$ 

For second-price auction, the revenue of auctioneer is the second highest private value of $n$ bidders, denoted by $\vartheta_{2}^{(n)}$, and its distribution function is

$$\Phi(\vartheta_{2}^{(n)}) + n\Phi^{n-1}(\vartheta_{2}^{(n)})(1 - \Phi(\vartheta_{2}^{(n)})).$$

The density function is:

$$n(n-1)(1 - \Phi(\vartheta_{2}^{(n)}))\Phi^{n-2}(\vartheta_{2}^{(n)})\varphi(\vartheta_{2}^{(n)}) = n(1 - \Phi(\vartheta_{2}^{(n)}))\psi(\vartheta_{2}^{(n)}).$$
Therefore, in second-price auction, the auctioneer’s ex ante expected revenue is:

$$
E_\theta R^{II} = \int_0^\infty n\hat{\vartheta}_2^{(n)}(1 - \Phi(\hat{\vartheta}_2^{(n)}))\psi(\hat{\vartheta}_2^{(n)})d\vartheta_2^{(n)} = E_\theta R^I.
$$

Thus, in a symmetric environment, the expected revenues of auctioneer with private value in the first-price auction and second-price auction are the same. In fact, we can get a more general conclusion that all auction mechanisms that satisfy the same allocation rules and have the same expected utility of the bidder at the lowest private value are equivalent to the principal’s expected revenue.

We further compare the expected revenue for the auctioneer with the first-price auction and the second-price auction.

Consider the payment of bidder 1. When $\theta_1 < \vartheta_1$, no matter what kind of auction formats, the interim expected payment of bidder 1 is zero. When $\theta_1 > \vartheta_1$, under first-price auction, the payment of bidder 1 is:

$$
m^I(\theta_1) = -\bar{t}_1^{I}(\theta) = \Psi(\theta)\beta^I(\theta) = E_\theta[\vartheta_1|\vartheta_1 < \theta_1]
$$

For the second-price auction, bidder 1’s expected payment is (Note that we do not consider bidder 1’s payment because it relies on the bidding price for the first order statistics of the other bidder types, which is a random variable):

$$
m^{II}(\theta_1) \equiv -E_\theta(t^{II}(\theta)|\vartheta_1 < \theta_1) = E_\theta[\vartheta_1|\vartheta_1 < \theta_1],
$$

which means $\beta^{II}$ is $\beta^I$’s mean value spreading. Thus, we have $m^I(\theta_1) = m^{II}(\theta_1)$.

Since interim payment is the same, the ex ante payment would also be the same. The reason is that when $m^I(\theta_1) = m^{II}(\theta_1)$, under first-price auction, the ex ante payment the auctioneer receives from bidder 1 is:

$$
E_\theta(m^I) = E_{\theta_1}[\Psi(\theta_1)m^I((\theta_1))].
$$

Under second-price auction, the ex ante payment the auctioneer receives from bidder 1 is:

$$
E_\theta(m^{II}) = E_{\theta_1}[\Psi(\theta_1)m^{II}((\theta_1))].
$$
Thus, $E_{\theta} (m^I) = E_{\theta} (m^{II})$. Thus we have verified in two ways that the ex ante payment to the auctioneer is the same for both the first-price auction and the second-price auction.

In order to have the specific expression of auctioneer’s expected revenue, from (21.6) we give the first-price auction bid:

$$
\beta^I(\theta) = \theta - \frac{1}{\Psi(\theta)} \int_0^{\theta} \Psi(\vartheta_1) d\vartheta_1 < \theta,
$$

by integration by parts for the second part of the above equation, we have

$$
E \beta^I(\theta) = \int_0^{\theta} \nu(\theta') d\Phi^n(\theta'),
$$

where

$$
\nu(\theta) = \theta - \frac{1 - \Phi(\theta)}{\phi(\theta)}.
$$

In this way, we obtain the specific expression of the auctioneer’s expected revenue:

$$
E_{\theta} (m^I) = E_{\theta} (m^{II}) = \int_0^{\theta} \nu(\theta') d\Phi^n(\theta').
$$

In the auction theory and mechanism design theory, $\nu(\theta)$ given above plays a very important role. A buyer (bidder) considers the object’s value at most $\theta$, which is his personal information, and he will tend to under-report this information when dealing with the seller. How much the buyer can under-report depends on the seller’s knowledge on $\theta$, which is the distribution function of $\theta$. From the seller’s point of view, $\nu(\theta)$ is the highest price the buyer is willing to pay due to information asymmetry and adjustment. The design of the optimal auction mechanism will be discussed below. $\nu(\theta)$ is actually the bidder’s virtual value function, which is equivalent to the seller’s (or auctioneer’s) marginal revenue.

### 21.3.5 Reserve Price in First and Second Price Auctions

In the analysis so far, we have implicitly assumed that the seller can sell the object at any price. In many situations, sellers reserve the right to keep the object if the price by the auction is lower some threshold amount, say $r > 0$. Such a price is called the reserve price. What we will show that it is still true that the revenue equivalence between the first and second price auctions even with reserve price.
Reserve Prices in Second Price Auctions

Since the price at which the object is sold can never be lower than $r$, no bidder with a value $\theta < r$ can make a positive profit in the auction. In a second price auction, a reserve price makes no difference to the behavior of the bidders, it is still a weakly dominant strategy to bid one’s value. The expected payment of a bidder with value $r$ is now just $rG(r)$, and the expected payment of a bidder with value $\theta \geq r$ is

$$m^{II}(\theta, r) = rG(r) + \int_r^\theta \theta' g(\theta') d\theta'$$

since the winner pays the reserve price $r$ whenever the second-highest bid is below $r$.

Reserve Prices in First Price Auctions

If $\beta^I$ is a symmetric equilibrium of the first price auction with reserve price $r$, it must be that $\beta^I(r) = r$. This is because a bidder with value $r$ wins only if all other bidders have values less than $r$ and, in that case, can win with a bid of $r$ itself. In all other respects, the analysis of a first-price auction is unaffected, and in a manner analogous to Proposition 21.3.1 we obtain that a symmetric equilibrium bidding strategy for any bidder with value $\theta \geq r$ is

$$\beta^I(\theta) = E[\max\{\vartheta_1, r\} | \vartheta_1 < \theta]$$

$$= r \frac{G(r)}{G(\theta)} + \frac{1}{G(\theta)} \int_r^\theta \theta' g(\theta') d\theta'$$

The expected payment of a bidder with value $\theta \geq r$ is

$$m^I(\theta, r) = G(\theta) \beta^I(\theta)$$

$$= rG(r) + \int_r^\theta \theta' g(\theta') d\theta'$$

which is the same as in the second price auctions with reserve price. Thus, we conclude that the revenue equivalence theorem can be generalized to the reserve price auctions.
Revenue Effects of Reserve Prices

Let $A$ denote either the first ($A = I$) or second ($A = II$) price auction. In both, the expected payment of a bidder with value $r$ is $rG(r)$. Recall that the expected payment of a bidder with value $\theta \geq r$ is

$$m^A(\theta, r) = rG(r) + \int_r^\theta \theta'g(\theta')d\theta'$$

The ex ante expected payment of a bidder is then

$$E[m^A(\theta, r)] = \int_r^\omega m^A(\theta, r)f(\theta)d\theta$$

$$= r(1 - F(r))G(r) + \int_r^\omega \theta'(1 - F(\theta'))g(\theta')d\theta'$$

Suppose that the seller attaches a value $\theta_0 \in [0, \omega)$. This means that if the object is left unsold, the seller would derive a value $\theta_0$ from its use. Clearly, the seller would not set a reserve price $r$ that is lower than $\theta_0$. Then the overall expected payoff of the seller from setting a reserve price $r \geq \theta_0$ is

$$\Pi_0 = N \times E[m^A(\theta, r)] + \theta_0F(r)^N$$

$$= Nr(1 - F(r))G(r) + N\int_r^\omega \theta'(1 - F(\theta'))g(\theta')d\theta' + F(r)^N\theta_0$$

Differentiating this with respect to $r$, we obtain

$$\frac{d\Pi_0}{dr} = N[1 - F(r) - rf(r)]G(r) + NG(r)f(\theta)\theta_0$$

Now recall that the hazard rate function associated with the distribution $F$ is defined as $\lambda(\theta) = f(\theta)/(1 - F(\theta))$. Thus, we can write

$$\frac{d\Pi_0}{dr} = N[1 - (r - \theta_0)\lambda(\theta)](1 - F(r))G(r)$$

First, notice that if $\theta_0 > 0$, then the derivative of $\Pi_0$ at $r = \theta_0$ is positive, implying that the seller should set a reserve price $r > \theta_0$. If $\theta_0 = 0$, then derivative of $\Pi_0$ at $r = 0$ is 0, but as long as $\lambda(r)$ is bounded, the expected payment attains a local minimum at 0, so a small reserve price leads to an increase in revenue. Thus,
A revenue maximizing seller should always set a reserve price that exceeds his or her value.

The relevant first-order condition implies that the optimal reserve price $r^*$ must satisfy

$$r^*\lambda(r^*) = 1 \iff r^* = \frac{1}{\lambda(r^*)}$$

If $\lambda(.)$ is increasing, this condition is also sufficient and it is remarkable that the optimal reserve price does not depend on the number of bidders.

21.3.6 Efficient Allocation and Revenue Equivalence Principle

It can be seen from the previous discussion that if all the bidders’ bids are given independently in all the four auctions, the auctioneer can obtain the same expected revenue regardless of the auction form. The four auction formats all have some common features in that bidders are asked to give the bidding price they are willing to pay, which determines who will receive the object and how much they will pay. This results in efficient outcomes as the bidder with the highest value wins the object. Yet another notable feature is that while the distribution functions are very different, the bids are varied and even not strategically-equivalent, the auctioneer receives the same expected revenue. As such, to what extent are the results still valid?

The 2007 Nobel laureate in economics Roger B. Myerson used the mechanism design approach to study this issue. On this basis, he popularized Vickrey’s theory and proved that: assuming that the bidders’ evaluations of the object are independent, the bidders only care about their own expected payment, etc., all possible standard auction mechanisms, in which the highest bidder gets the object, will bring the same expected revenue to the auctioneer. Obviously, this conclusion goes beyond the revenue equivalence results of earlier auctions by Vickrey and can be applied in studying all possible auctions. The result is significant, making the auction theory one great step further. Before discussing this
result, we will discuss the efficiency of the first-price and second-price auction mechanisms.

Since the bidder’s Bayesian-Nash equilibrium bidding functions are both monotonous and continuous in the first-price and second-price auction mechanisms, and both are the ones with the highest bids win, the allocation result is efficient and we have the following allocative efficiency theorem.

**Theorem 21.3.1 (Allocative Efficiency Theorem)** In Symmetric Independent Private Value (SIPV) model, the first-price and second-price auction mechanisms result in efficient outcomes of object (the ones with the highest value get the object).

We will see that when the symmetry is not satisfied, the second-price auction is still efficient, but the first-price auction may not be efficient.

Now we discuss the Myerson revenue equivalence theorem. We call an auction mechanism *standard* if the highest bidder gets the object. A non-standard form of auction is to buy a lottery. The chance of a bidder’s winning is related to the ratio of the bid amount to the total bid amount of all the bidders. It is non-standard because it does not necessarily guarantee to obtain the object despite the highest bidding. In addition to these four common standard auction mechanisms, another example of a standard auction mechanism is the auction mechanism for “All Pay”, where every bidder pays according to his bid, but only one bidder gets the object. For example, lobbying for the government to adopt a policy or bribing government officials who hold power belongs to such auction formats. In those cases, are the auction mechanisms still the same for buyers’ payoffs? The general Myerson’s equivalence theorem gives an affirmative answer.

We now examine the question of whether the auctioneer’s expected revenue under the standard mechanism is equivalent. Consider a linear model allocating an indivisible object among $n$ risk-neutral buyers: $Y = \{y \in \{0, 1\}^n : \sum_i y_i = 1\}$, buyer $i$’s revenue is $\theta_i y_i + t_i$, and the value is a random variable with individually independent density $\varphi_i(\cdot) > 0$ defined $[\theta_i, \bar{\theta}_i]$. Under an auction rule (or more
general social choice in a linear environment) \((g(\cdot), t_1, \cdots, t_I)\), the auctioneer’s expected revenue can be written as

\[
- \sum_i E_{\theta} t_i(\theta) = \sum_i E_{\theta}[\theta_i y_i(\theta) - U_i(\theta)] = \sum_i E_{\theta}[\theta_i y_i(\theta)] - \sum_i E_{\theta_i} E_{\theta_{-i}}[U_i(\theta)].
\]

In the above equation, the first item is the expected total surplus for all participants, while the second item is the expected utility for all participants. According to the Bayesian incentive compatibility characterization theorem (Proposition xxx) in Chapter 16, the second item is given by

\[
E_{\theta_i} U_i(\theta_i) = E_{\theta_i} [U_i(\theta_i, \theta_{-i})] + \int_{\theta_i}^{\bar{\theta}_i} E_{\theta_{-i}} y(\tau, \theta_{-i}) d\tau,
\]

so it can be completely determined by efficient (allocation) decision rules \(y(\cdot)\) and the lowest type of interim expected utility \(E_{\theta_{-i}}[U_i(\theta_i, \theta_{-i})]\).

Since the total surplus remains unchanged (all mechanisms implement the same decision rule), we have the following Revenue Equivalence Theorem.

**Theorem 21.3.2 (Myerson’s Revenue Equivalence Theorem)** Suppose that two different auction mechanisms both have Bayesian-Nash equilibrium, and at the equilibrium: (i) The same decision (object allocation) rule \(y(\cdot)\) is implemented; (ii) When its value reaches the lowest point \(\theta_i\), each buyer \(i\) has same interim expected utility. The Bayesian-Nash equilibrium of the two auction mechanisms yields the same revenue to the auctioneer.

This theorem shows that there are many ways to obtain interim expected transfer payment \(\bar{t}_i(\theta_i)\) by using ex post transfer payment \(t_i(\theta)\). In this way, even if the rule of decision and the utility of the lowest-type participant is fixed, the seller has a great freedom in designing the auction scheme.

For example, suppose the buyers are symmetric (that is, they have the same distribution function). The seller wants to implement efficient decision rule \(y(\cdot)\) and minimize the expected utility of the buyer with the lowest value to zero. We can use the second-price sealed-bid auction with dominant strategy to truthfully implement the efficient decision rule or the first-price auction mechanism to implement the efficient decision rule. More generally, consider \(k\)th price sealed-bid auction, where \(1 \leq k \leq n\), the highest bidder gets the object and pays the \(k\)th
highest bid. Suppose the value of every buyer on the object is an independent and identically distributed random variable (i.i.d.), then we can prove that the auction mechanism has equilibrium that is unique and symmetric, and the bid of each participant $b(\theta_i)$ is an increasing function of his value (see Fudenberg- Tirole Game Theory, page 223-225.) Since the bidder with highest biding price gets the object, the auction mechanism implements the efficient outcome and thus is an efficient decision rule. Meanwhile, the probability of buyer with lowest value $\theta$ getting the object is zero, so his expected utility is zero. Thus, by revenue equivalence theorem, for any $k$, $k$th price sealed-bid auction brings the same revenue to sellers. All-pay auction mechanism and $k$th price sealed-bid auction also bring the same revenue to sellers.

For a given bidding scheme, it seems that the auctioneer will get a higher revenue when $k$ is smaller. So how will the above results be established? The answer is that the bidder’s bid would be smaller when $k$ is small. For example, we know that the bidder’s bid is the true value in second-price auction. In first-price auction, the buyer’s bid will be less than his true value because the bid equal to his true value would bring zero expected utility. Myerson’s Revenue Equivalence Theorem shows that for both auction mechanisms, the seller’s expected revenue is equal. In particular, we can obtain this without solving the auction equilibrium. When $k > 2$, the revenue equivalence theorem shows that the bidder’s bid will be greater than his true value (but since the price paid is the $k$th $(1 \leq k \leq n)$ highest price, it will not be greater than his true value).

21.3.7 Applications of Revenue Equivalence Principle

There are many auction formats in reality. In addition to the four common auction forms discussed above, there are also “All Pay” auction, $k$ price auction, auction with uncertain number of bidders and other auction formats. These auction mechanisms have many applications in reality. For example, the “All Pay” auction can be applied to attrition competition in the industry, whoever holds to the final will win. However, these enterprises, including the enterprises
that quit midway, have paid certain cost in the process. In addition, the “All
Pay” auction can also be applied to the lobbying activities of interest groups
in politics and so on. Although $k \geq 3$ price auction is rarely used in reality,
such form of auctions has theoretical significance. We find that bidders may bid
more than their own bidding price in a third-price auction. There are some fine
balances of interests. In addition, the actual auctions in reality, such as online
auctions where the number of bidders is not sure, how will each bidder choose
his own strategy? Therefore, these auction mechanisms have great theoretical
and practical significance. However, in these auction mechanisms, the solution
to the bidding equilibrium tends to be relatively complicated, whereby applying
Myerson’s Revenue Equivalence Theorem provides some shortcuts to the solution
process.

**All-Pay Auction**

“All Pay” auction is similar to the auction mechanism where the bidder with the
highest bid gets the object, but the difference lies in that each bidder, regardless
of whether obtaining the object eventually, pays the auctioneer at his own bidding
price. By the framework of the previous mechanism design, let $\theta = (\theta_1, \ldots, \theta_n)$ be
$n$ types of bidders independently distributed to $\varphi(\cdot)[0, w]$, and let $b = (b_1, \ldots, b_n)$
be the bidding vector, bidder $i$’s utility function is

$$U_i(\theta) = \theta_i y_i(\theta) + t_i(\theta),$$

where:

$$y_i(b_i, b_{-i}) = \begin{cases} 
1 & \text{if } b_i > \max_{j \neq i} b_j \\
0 & \text{if } b_i < \max_{j \neq i} b_j
\end{cases}$$

(21.10)

If $b_i = \max_{j \neq i} b_j$, by the same probability the object is randomly allocated and

$$t_i(b_i, b_{-i}) = -b_i.$$  

(21.11)

As before, $\Psi(\cdot)$ is the distribution of $\vartheta_1$ (the first-order statistics of other
participant types). We focus on the case that symmetric equilibrium $\beta(\cdot) = \beta^{AP}(\cdot)$ is a monotonous function (which will be further proved). When $\theta = 0$, it is obvious that bidders will choose the bidding price 0, at the same time the
bidder with the highest private value wins the object. According to the Revenue Equivalence Theorem, at the equilibrium, this auction has the same expected utility as the first-price auction and the second-price auction. Thus, when bidder i’s type is $\theta$, the interim expected utility under “All Pay” auction mechanism equilibrium is:

$$\Pi_{i}^{AP}(\beta(\tilde{\theta}), \theta) = \Psi(\tilde{\theta})\theta - \beta^{AP}(\theta).$$

The interim expected utility at equilibrium under second-price auction mechanism is:

$$\Pi_{i}^{II}(\beta(\tilde{\theta}), \theta) = \Psi(\tilde{\theta})\theta - \Psi(\tilde{\theta})\int_{0}^{\theta} \vartheta_{1}(\vartheta_{1})d\vartheta_{1}.$$

Thus, making the above two equations equal, we obtain a symmetric and balanced bidding strategy for an “All Pay” auction:

$$\beta^{AP}(\theta) = \int_{0}^{\theta} \vartheta_{1}(\vartheta_{1})d\vartheta_{1},$$

which is obviously a strictly increasing function.

**Third-price Auction**

Consider a third-price auction. First of all we use the previous framework to characterize the third-price auction. Let $\theta = (\theta_{1}, \ldots, \theta_{n})$ be $n$ types of bidders independently distributed to $\varphi(\cdot)[0, w]$, $b = (b_{1}, \ldots, b_{n})$ be the bidding vector, and bidder i’s utility function is $\theta_{i}y_{i}(\theta_{i}) + t_{i}(\theta_{i})$, which satisfies:

$$y_{i}(b_{i}, b_{-i}) = \begin{cases} 1 & \text{if } b_{i} > \max_{j \neq i} b_{j} \\ 0 & \text{if } b_{i} < \max_{j \neq i} b_{j}, \end{cases}$$

and

$$t_{i}(b_{i}, b_{-i}) = \begin{cases} -b_{(3)} & \text{if } b_{1} > \max_{j \neq 1} b_{j} \\ 0 & \text{if } b_{1} < \max_{j \neq 1} b_{j}. \end{cases}$$

where $b_{(3)} = \max\{\{b_{2}, \ldots, b_{n}\} \setminus \{\max_{j \neq 1} b_{j}\}\}$, the third highest bidding price.

In the symmetric Bayesian-Nash equilibrium under the third-price auction, like the common first-price and second-price auctions, the highest bidder obtains the object; meanwhile the bid of bidder with $\theta = 0$ is zero, and utility is zero.
Let $\beta^{III}(\cdot)$ be the third-price Bayesian-Nash equilibrium strategy function. It is assumed to be an increasing function, which will be verified later. In a symmetric environment, consider bidder 1’s strategy choices. It is obvious that only when $\theta_1 = \theta > \vartheta_1$, bidder 1 can win the object. Consider $\vartheta_2$ is the second order statistics of $\{\theta_2, \ldots, \theta_n\}$, at equilibrium, bidder 1’s expected payment for winning an auction object is $E_{\vartheta_2}[\beta^{III}(\vartheta_2)|\vartheta_1 \leq \theta]$. By Revenue Equivalence Theorem, we have:

$$\Psi_1(\theta)^{(n-1)} E_{\vartheta_2}[\beta^{III}(\vartheta_2)|\vartheta_1 \leq \theta] = \int_0^\theta \vartheta_1 \psi_1(\vartheta_1) d\vartheta_1.$$ 

Let $\Psi_k^{(m)}$ be the distribution function of the $k$th order statistics in $m$ independent and identically distributed random variables, and $\psi_k^{(m)}$ is the density function. $\vartheta_2|\vartheta_1 \leq \theta$’s density function is:

$$\psi_2^{(n-1)}(\vartheta_2 = \tau|\vartheta_1 \leq \theta) = \frac{1}{\Psi_1(\theta)^{(n-1)}} (n-1)[\Phi(\theta) - \Phi(\tau)] \psi_1(\tau)^{(n-2)},$$

where $(n-1)[\Phi(\theta) - \Phi(\tau)]$ is the probability of $\vartheta_1$’s value in $[\tau, \theta]$; $\psi_1(\tau)^{(n-2)}$ is the density function of the first order statistics in $n-2$ independent and identically distributed random variables.

Thus, for bidder with $\theta_1 = \theta$, the expected payment is:

$$\Psi_1(\theta)^{(n-1)} \int_0^\theta \beta^{III}(\tau) \psi_2^{(n-1)}(\vartheta_2 = \tau|\vartheta_1 < \theta) d\tau
= \int_0^\theta \beta^{III}(\vartheta_2)(n-1)[\Phi(\theta) - \Phi(\tau)] \psi_1(\tau)^{(n-2)} d\tau.$$ 

We then have:

$$\int_0^\theta \beta^{III}(\tau)(n-1)[\Phi(\theta) - \Phi(\tau)] \psi_1(\tau)^{(n-2)} d\tau = \int_0^\theta \vartheta_1 \psi_1(\vartheta_1) d\vartheta_1.$$ 

Taking differential on both sides of $\theta$:

$$(n-1)\varphi(\theta) \int_0^\theta \beta^{III}(\vartheta_2) \psi_1(\tau)^{(n-2)} d\tau = \theta \psi(\theta),$$

where $\psi(\cdot)$ is $\vartheta_1$’s density function, $\psi(\theta) = (n-1)\Phi(\theta)^{n-2} \varphi(\theta)$, we obtain:

$$\int_0^\theta \beta^{III}(\tau) \psi_1(\tau)^{(n-2)} d\tau = \theta \Phi(\theta)^{n-2}.$$
Taking differential on both sides of $\theta$ again, we get:

$$\beta^{III}(\theta)\psi_1(\theta)^{(n-2)} = \Phi(\theta)^{n-2} + \theta \phi(\theta)^{n-2}. $$

Since $\Psi_1^{(n-2)}(\tau) = \Psi(\tau)^{n-2}$, we have

$$\beta^{III}(\theta) = \theta + \frac{\Phi(\theta)}{(n-2)\varphi(\theta)}. $$

Comparing first-price, second-price and third-price auctions, we find:

$$\beta^I(\theta) < \beta^{II}(\theta) = \theta < \beta^{III}(\theta). $$

That is to say, bidders under first-price auction will bid lower than their own value (otherwise the expected utility is less than or equal to zero), bidders under the second-price auction will bid their own value, and bidders under the third-price auction would bid more than their own value. The reason why bidders under the third-price auction will bid exceed their value is: when other bidders bid with $\beta^{III}$, if bidder 1 bids $b > \theta_1$, compared with bidding $\theta_1$ it would be better in $\beta^{III}(\vartheta_2) < \theta_1 < \beta^{III}(\vartheta_1) < b$, though it would be worse in $\theta_1 < \beta^{III}(\vartheta_2) < \beta^{III}(\vartheta_1) < b$, but under $b - \theta = \epsilon$, when $\epsilon$ is very small, the probability of $\beta^{III}(\vartheta_2) < \theta_1 < \beta^{III}(\vartheta_1) < b$ is $\epsilon^2$ order, while $\theta_1 < \beta^{III}(\vartheta_2) < \beta^{III}(\vartheta_1) < b$ would be $\epsilon^3$ order. There will be incentives for bidders to bid beyond their value.

**Uncertain Number of Bidders**

We now consider auctions with uncertain number of bidders. Suppose $\mathcal{N} = \{1, 2, \ldots, N\}$ denote potential bidders. Let $A \subseteq \mathcal{N}$ be the set of actual bidders. Suppose for each potential bidder, the valuations of the object are independent and identically distributed, then the distribution function is $\Phi(\cdot)$. For participant $i \in A$, assume the probability of facing $n$ other bidders is $p_n$, and the probability does not depend on the identity of the bidder nor on his value. As long as the bidder is consistent with the type and number of opponents he faces, then for the symmetric equilibrium, we can also use the Revenue Equivalence Theorem to obtain the Bayesian-Nash equilibrium bidding function. Suppose the symmetric Bayesian-Nash equilibrium bidding strategy $\beta$ is a monotonically increasing
function and the probability for bidder 1 of facing \( n \) other bidders is \( p_n \). Let \( \vartheta_1^{(n)} \) be the first order statistics of types in \( n \) other bidders, and the distribution function is \( \Psi^{(n)}(\cdot) = \Phi^n(\cdot) \). When bidder1’s type is \( \theta \) and bidding price is \( \beta(\tilde{\theta}) \), the probability of winning the auction is:

\[
\Psi(\tilde{\theta}) = \sum_{n=0}^{N-1} p_n \Psi^{(n)}(\tilde{\theta}),
\]

and the interim expected utility is:

\[
\bar{U}_1(\theta, \tilde{\theta}) = \Psi(\tilde{\theta})\theta + \bar{t}_1(\tilde{\theta}),
\]

where \( \bar{t}_1(\tilde{\theta}) \) is the interim expected transfer payment for bidder 1 when choosing \( b_1 = \beta(\tilde{\theta}_1) \).

Let us consider the first-price and second-price auctions with uncertain number of bidders. First, for the second-price auction, despite the number of opponents, \( \beta(\theta) = \theta \) would always be a weakly dominant strategy for bidder 1. Then the interim transfer payment under second-price auction is

\[
\bar{t}^{II}_1(\tilde{\theta}) = -\sum_{n=0}^{N-1} p_n \Psi^{(n)}(\theta) E[\beta^{II}(\vartheta_1^{(n)})|\vartheta_1^{(n)} < \theta].
\]

For the first-price auction, the expected interim payment in equilibrium is:

\[
\bar{t}^{I}_1(\tilde{\theta}) = -\Psi(\theta)\beta^I(\theta).
\]

According to the Revenue Equivalence Theorem, we get: \( \bar{t}^{II}_1(\tilde{\theta}) = \bar{t}^{I}_1(\tilde{\theta}) \), that is:

\[
\beta^I(\theta) = \sum_{n=0}^{N-1} p_n \frac{\Psi^{(n)}(\theta)}{\Psi(\theta)} E[\beta^{II}(\vartheta_1^{(n)})|\vartheta_1^{(n)} < \theta] = \sum_{n=0}^{N-1} p_n \frac{\Psi^{(n)}(\theta)}{\Psi(\theta)} \beta^{I,(n)}(\theta),
\]

where \( \beta^{I,(n)}(\theta) \) denotes the equilibrium bidding price under first-price auction with his own type \( \theta \) when the number of opponents is \( n \).

Thus, for a first-price auction mechanism with an uncertain number of bidders, the bidding strategy of symmetric equilibrium is a weighted average of the first bidding price under a given number of bidders, whose weight is equal to the probability distribution of the opponent bidder’s number.
21.3.8 Optimal Auction Mechanism Design

Now we discuss the seller’s optimal auction scheme, which means to consider the optimal auction mechanism that satisfies the Bayesian incentive compatibility and participation constraints. Suppose the seller can retain the object, we call the seller with “reserve object” be “participant 0”, the value of the object is denoted by \( \theta_0 \), and the decision whether to retain the object is written as \( y_0 \in \{0, 1\} \). Then we must have \( \sum_{i=0}^{n} y_i = 1 \).

Thus, the seller’s expected revenue can be written as

\[
\theta_0 E_{\theta} y_0(\theta) + \text{Expected Revenue} = \theta_0 E_{\theta} y_0(\theta) + \sum_{i=1}^{n} E_{\theta} [\theta_i y_i(\theta)] - \sum_{i=1}^{n} E_{\theta, \theta_{-i}} [U_i(\theta_i, \theta_{-i})].
\]

As such, the expected payoff of the seller equals the difference between the total surplus and the expected information rent of the participants.

According to the Bayesian incentive compatibility characterization theorem in Chapter 16, the expected utility of the buyer must satisfy

\[
E_{\theta_{-i}} [U_i(\theta_i, \theta_{-i})] = E_{\theta_{-i}} [U_i(\theta_i, \theta_{-i})] + \int_{\theta_i}^{\theta_{max}} E_{\theta_{-i}, \tau} [y_i(\tau, \theta_{-i})] \, d\tau, \forall \theta_i \in \Theta_i.
\]

The interim individual rationality constraint of the buyer is

\[
E_{\theta_{-i}} [U_i(\theta_i, \theta_{-i})] = 0, \forall i.
\]

Given the decision rule \( y(\theta) \), the optimal choice for a rational buyer is to set the transfer payment as

\[
E_{\theta_{-i}} [U_i(\theta_i, \theta_{-i})] = 0, \forall i.
\]

Integration by parts for above equation, we can write buyer \( i \)'s information rent as

\[
E_{\theta} E_{\theta_{-i}} \left[ \frac{1}{h_i(\theta_i)} y_i(\theta) \right],
\]

where \( h_i(\theta_i) = \frac{\varphi_i(\theta_i)}{1 - \Phi_i(\theta_i)} \) is participant \( i \)'s hazard rate. Substituting into the seller’s revenue formula, we can write it as expected virtual surplus:

\[
E_{\theta} \left[ \theta_0 y_0(\theta) + \sum_{i=1}^{n} \left( \theta_i - \frac{1}{h_i(\theta_i)} \right) y_i(\theta) \right].
\]

Finally, for \( i \geq 1 \), let \( \nu_i(\theta_i) = \theta_i - 1/h_i(\theta_i) \), we call it participant \( i \)'s “virtual valuation”. Then let \( \nu_0(\theta_0) = \theta_0 \). The seller’s optimization problem can be
written as

\[
\max_{\theta} E_{\theta} \left[ \sum_{i=0}^{n} \nu_i(\theta) y_i(\theta) \right] \quad \text{s.t.} \quad \sum_{i=0}^{n} y_i(\theta) = 1, \\
E_{\theta_{-i}} y_i(\theta_i, \theta_{-i}) \text{ is non-decreasing on } \theta_i, \forall i \geq 1 \text{ (BM)}.
\]

Ignore for now the above Bayesian monotonicity constraint. For each status \( \theta \), maximize the above expectation. The options for giving the object to the participant of highest virtual value are as follows: for all individuals \( j \neq i \), when \( \nu_i(\theta_i) > \nu_j(\theta_j) \), \( y_i(\theta) = 1 \). The corresponding decision rule is \( y(\theta) = y(\nu_0(\theta_0), \nu_1(\theta_1), \ldots, \nu_I(\theta_I)) \), where \( y(\cdot) \) is the efficient decision rule. Intuitively, since the seller (principal) cannot extract all buyer’s information rent, the principal will adopt the virtual value of the participant rather than the true value. The virtual value of the participant is less than his true value because the latter includes the information rent of the participant and the principal cannot extract the rent. The profit maximization mechanism allocates the object to the participant with the highest virtual value.

Thus, under what conditions can we ignore monotonicity constraint? Notice that when virtual value function \( \nu_i(\theta_i) = \theta_i - 1/h_i(\theta_i) \) is an increasing function\(^{1}\), in the solution to the slack problem, the increase in the value of the participant will make it more likely to acquire the object. Therefore, for all \( \theta_{-i} \), \( y_i(\theta_i, \theta_{-i}) \) is non-decreasing with regard to \( \theta_i \). Thus according to the dominant incentive compatibility characterization theorem (Proposition 16.4.1), optimal allocation rules can be implemented by not only the Bayesian-Nash equilibrium, but also the dominant strategy equilibrium. We can also have a DIC transfer payment by DICFOC integration, and the resulting transfer payment is:

\[ t_i(\theta) = -p_i(\theta_{-i}) y_i(\theta), \]

where

\[ p_i(\theta_{-i}) = \inf \{ \hat{\theta}_i \in [\underline{\theta}_i, \bar{\theta}_i] : y_i(\hat{\theta}_i, \theta_{-i}) = 1 \}. \]

\(^{1}\)Such a sufficiency condition is that the default rate \( h_i(\theta_i) \) is a non-decreasing function. The process of proof is the same as that of a single participant (Notice that for linear utility function \( \nu(y, \theta) = \theta y \), we have \( \nu_{\theta \theta} = 0 \)).
Therefore, for each buyer (agent) \(i\), there exists pricing rule \(p_i(\theta_{-i})\), which is a function of other bidder’s true value. If the participant’s bid is greater than the price \(p_i(\theta_{-i})\), then the bidder gets the object, and pays at price \(p_i(\theta_{-i})\). This means telling the truth (true reveal) is the dominant strategy. The logic of argument is the same as the Vickrey auction mechanism: misreport does not affect the price paid, but only affects the acquisition of the object. Bidders want to get the object just when their valuation \(\theta_1\) is higher than the price \(p_i(\theta_{-i})\).

Suppose for all \(i\), \(\psi_i = \psi\) and \(\nu_i = \nu\). That is, the distribution of individual types is the same, then their virtual value functions are the same, denoted by \(\nu(\cdot)\). Assume the function is increasing. When the principal sell the object, since the individual \(\theta_i\) with the highest value has the highest virtual value \(\nu(\theta_i)\), the object would be sold to him. When \(\nu(\max_{i \geq 1} \theta_i) > \theta_0\), the principal sells the object. Then we have

\[
p_i(\theta_{-i}) = \max\{\nu^{-1}(\theta_0), \max_{j \neq i} \theta_j\}.
\]

Thus the optimal mechanism that is truthfully implementable in dominant strategy is a second price auction with a reserve price \(r^* = \nu^{-1}(\theta_0)\). The optimal reserve price has two characteristics. First, if it is greater than the value of the object for the principal, it is equivalent to monopoly pricing, the price of which is higher than the marginal cost. In this sense, if the seller can determine the optimal reserve price, the standard auction mechanism does not achieve Pareto efficiency, and the object cannot be sold in a socially optimal manner because when the highest buyer’s private value is between \(\nu^{-1}(\theta_0)\) and \(r^*\), the seller will not resell the object to this buyer. The intuition is that principal will reduce his information rent by reducing individual spending. This is similar to the case of a single individual.

Why is it optimal to allocate the object on the optimal reserve price (virtual value)? It is not hard to understand. We give the following explanation. Suppose the seller adopts a take-it-or-leave-it offer to sell the object to buyer at price \(p\). If the seller decides a reserve price \(r\), then only when the buyer’s private value
\( \theta \) is higher than or equal to \( r \), he would buy the object. The probability that the buyer will accept this price level is \( 1 - F(p) \), which is also the probability of the value of buyer exceeding \( p \). If we consider the probability of purchase as the buyer \( i \)’s demand, then the demand function is \( q(p) \equiv 1 - F(p) \), where the inverse demand function is \( p(q) \equiv F^{-1}(1 - q) \). Then, the seller’s revenue function is

\[
p(q) \times q = qF^{-1}(1 - q),
\]

where the derivative on \( q \) is

\[
\frac{d}{dq}(p(q) \times q) = F^{-1}(1 - q) - \frac{q}{F'(F^{-1}(1 - q))}.
\]

Since \( F^{-1}(1 - q) = p \), we have

\[
MR(p) \equiv p - \frac{1 - F(p)}{f(p)} = \nu(p),
\]

which means marginal revenue equals buyer’s virtual value at \( p(q) = p \). Thus, buyer’s virtual value \( \nu(p) \) can be interpreted as marginal revenue. Since \( \psi \) is strictly increasing, the seller can set monopoly price \( r^* \) according to the marginal cost equalling the marginal revenue \( MR(p) = MC \). Since the latter is assumed to be \( \theta_0 \), \( MR(r^*) = \nu(r^*) = \theta_0 \) or \( r^* = \nu^{-1}(\theta_0) \).

When the seller faces different types of buyers, the optimal mechanism is to adopt discriminatory reserve price \( r^*_i = \nu^{-1}_i(\theta_0) \). If no buyer’s value of the object exceeds the reserve price \( r^* \), the seller retains this object. Otherwise, the seller sells the object at marginal revenue, and the buyer who wins the object is asked to pay \( p_i = p_i(\theta_{-i}) \), which is the lowest value of winning the object for the buyer.

The second characteristic of the optimal reserve price is that the optimal reserve price is independent of the number of bidders but depends on the distribution of the bidder’s private value. This result does not hold when the bidder’s private value is not independent or when the bidder is not risk-neutral.

In addition, if the virtual value function is not increasing, (BM) may still be compact, so we need some “ironing” (see discussion on ironing in Chapter 16) scheme to ascertain the domain for the compactness of (BM). In this situation,
as the principal will underestimate the true value of the participants, the object is unlikely to be sold in the socially optimal way.

21.3.9 Factors Affecting Revenue Equivalence and Efficiency

According to Myerson, since all possible forms of auction yield the same expected payoff to the auctioneer, can he choose an arbitrary form for the auction? Unfortunately, although Myerson’s conclusion is theoretically beautiful, the conditions are too strong to hold in reality, it is almost impossible to satisfy the conditions imposed by Myerson. If one of these conditions is not satisfied, Myerson’s Revenue Equivalence Theorem cannot hold. In this sense, Myerson’s conclusion is not a perfect end to the auction theory, but provides a cornerstone and a new starting point for more realistic auction mechanism. Indeed, most of the research on auction theory has based on Myerson’s work. Here we do not provide the detailed discussion, instead, we give the basic conclusions.

Number of Bidders

The bidding price of an object is determined by competition, so the more competitive an auction is, the higher the bidder’s bid and the greater payoff for the auctioneer. In the symmetric independent private value (SIPV) model, this conclusion can easily be proved. The two previous auction mechanisms bring the following maximum average net profit to the seller:

\[ E \beta^I(\theta) = \int_0^\theta I(\theta)d\Phi^n(\theta). \]

By integration by parts we get:

\[ E \beta^I(\theta) = \int_0^\theta (1 - \Phi^n(\theta))I'(\theta)d\theta > 0. \]

Therefore, the auctioneer’s maximum payoff increases with the number of bidders. In a real auction, to attract more bidders to participate in the auction is very important.
Limited Liability

In the preceding discussions, we assume that the bidder is able to pay up to his highest value $\bar{\theta}_i$. Suppose that unlimited liability is not established, the bidder’s budget $\bar{w}$ is random variables distributed on $[0, \bar{w}]$. $w$ is a realization of random variable $\bar{w}$.

It can be proved that the second-price auction has a dominant equilibrium, whose equilibrium bidding function is given by $B^{II}(\theta, w) = \min\{x, w\}$. The Bayesian-Nash equilibrium bidding function of first-price auction mechanism is given by $B^I(\beta(\theta), w)$, and the auctioneer’s expected payoff under first-price auction is greater than the expected payoff under second-price auction.

Asymmetry between Bidders

An important hypothesis in the SIPV model is symmetry. The distribution functions of all bidders are the same. It is hard to imagine that this symmetry is always satisfied in the real auction environment. What impact will the asymmetry between bidders have on the auction?

First of all, since truth-telling in second-price auction mechanism is the dominant equilibrium, regardless of whether the auctioneer’s distribution function is symmetric or not, the highest bidder gets the object, so that the second-price auction mechanism is still valid. However, as the truth-telling may not be a dominant equilibrium in other mechanisms, they are not valid. For example, the first-price auction mechanism is actually not valid. To see this, suppose there are two auctioneers, and the Bayesian equilibrium bidding functions are continuously and monotonically increasing, denoted by $\beta_1$ and $\beta_2$. Assume for some $\theta$, $\beta_1(\theta) < \beta_2(\theta)$. Thus, there is sufficiently small $\epsilon > 0$, and we still have $\beta_1(\theta + \epsilon) < \beta_2(\theta - \epsilon)$. This means that bidder 2 gets the object despite his low true value.

Secondly, for auctioneer, the profits of different auction mechanisms may not be the same. From the previous discussion on the optimal auction mechanism design, $\nu(\theta_i)$ is the virtual value function of bidder $i$, which is equivalent to the
seller (or auctioneer)’s marginal revenue. As such, the seller’s optimal mechanism should sell the object to the bidder with the highest marginal revenue \( \nu'(\theta_i) \), provided that his marginal revenue is not less than the seller’s true value or marginal cost. Thus, when the bidder is not symmetric, the optimal auction mechanism is discriminatory and the bidder who wins the object is not necessarily the one with the highest private value. This also means that the seller’s optimal mechanism is not efficient. Furthermore, the seller will also set different reserve prices for different bidders. This mechanism is similar to monopoly price discrimination, where monopoly manufacturers set different prices according to the different needs of consumers.

Which of the first-price and second-price auction mechanisms is beneficial to the seller under asymmetric conditions? First of all, the asymmetry does not affect the bidder’s strategy in the second-price auction, and bidding his true value is still an dominant strategy. However, in the first-price auction, the weaker bidders may bid more aggressively. Therefore, for the seller, first-price auction may be better than the ascending-price auction or second-price auction. For a detailed discussion, see Krishna (2010).

**Risk Non-neutrality**

For bidders, an auction mechanism may bring risk. When the bidder wins, he gains profits; when he loses, his payoff is zero. If the bidder dislikes the risk or has different degree of risk aversion, his bidding strategy will likely be different.

Under first-price auction mechanism, if a bidder increases his bid, it will increase the probability of winning but also reduce the profit after winning. A risk-averse bidder will be more inclined to raise his bid than a risk-neutral bidder. Therefore, if all the bidders are risk averse, the first-price mechanism will bring higher profits to the seller. See Krishna (2010) for a strict proof.

Under the second-price auction mechanism, whether a bidder is risk-neutral or risk-averse, he always bids his true value. Therefore, if other assumptions in the SIPV model remain unchanged but all bidders are risk-averse, then the
Revenue Equivalence Theorem no longer holds. The first-price auction yields a higher average payoff than the second-price auction or ascending-price auction.

**Collusive Bid**

In many auction cases, bidders often bid together to reduce competition and lower the price paid to the seller. This collusive bidding behavior exist in several ways, depending on auction rules and informational structures. For example, some bidders may first take place pre-auctions together to select the best bidders for the formal auction; sometimes bidders take turns to participate in multiple auctions; in some cases, while all bidders participate in the auction, they make an appointment in advance on low bids, allowing one bidder to win and later (or in advance) sharing the revenue.

The analysis of collusive bidding strategy is complicated. The largest difficulty is that bidders may change their strategy. Thus, in order to prevent participants from changing their strategy, the collusive bidding mechanism must satisfy the collusion-proof incentive compatible constraint. In order to bring more bidders into collusion, but not mandatory, this mechanism must meet the participation constraint. Moreover, collusive bidding is best worked out by letting the bidder who is willing to bid the most win the object.

In the SIPV model, McAfee and McMillan (AER, 1992) demonstrated that there exists a direct display collusion mechanism that includes all bidders, with both efficiency and incentive compatibility. If the seller uses a standard auction mechanism such as the first-price or the second-price auction, every bidder is willing to participate in this collusive mechanism. The basic procedure for such a mechanism is that before participating in the formal auction, all the participants hold their own auction to elect the only bidder for the formal auction. The rule is that each bidder will bid separately and the highest bidder will be selected and he must pay his own bid, whereby the profits are equally divided among all other bidders. The selected bidder participates in the formal auction, and his best bid is equal to the seller’s reserve price, thus winning the auction and paying his
bid. This collusive bidding mechanism is equivalent to the case that every bidder wants to bribe others to quit the competition, and the bidder who is willing to pay the highest bribe wins and shares the profits of bribery with others.

**Participation Costs**

When bidders have participation costs, low-value bidders are more motivated to participate in collusion mechanism than high-value bidders. High-value bidders send out a signal of higher value by denying collusion mechanism, which is more credible. Tan and Yilankaya (2007) proved that when the auctioneer uses the second-price auction mechanism, such information will reduce the participation in the bidding mechanism for other bidders, making it harder to form collusion mechanism. Thus McAfee-McMillan’s conclusion no longer holds. Cao, Hsueh and Tian (2015) proved that when the auctioneer adopts the first-price auction mechanism, McAfee-McMillan’s conclusion remains to fail. In addition, Tan and Yilankaya (2006), Cao and Tian (2010, 2013) and Cao, Tan, Tian and Yilankaya (2014) discussed the existence of equilibrium in first-price and second-price auction mechanism with participation costs. In general, in the presence of participation costs, the Revenue Equivalence Theorem will be challenged.

In addition to the above factors, the interdependent valuation that will be discussed below also affects the Revenue Equivalence Theorem.

### 21.4 Interdependent Value Auctions for Single Object

In the previous section of private value auction, we obtain the expected revenue equivalence result to the auctioneer. For any two auction mechanisms, if the bidder’s expected payments under the lowest private value is the same and the allocation rule is the same, then the auctioneer’s expected revenue is the same. In
a symmetric economic environment, the auction mechanisms we are familiar, such as first-price (sealed-bid) auction, second-price (sealed-bid) auction, the English auction (the auction with ascending price) and the Dutch auction (the auction with descending price), all have the same allocation rule (i.e., the highest bidder gets the object) and the expected payoff (usually zero) at the same minimum private value, so that any such auction mechanism is indifferent to the auctioneer.

However, as mentioned above, the establishment of Myerson’s Revenue Equivalence Theorem relies on a series of assumptions, the most crucial one is that all bidders’ valuations of the auctioned object are given independently and are not interdependent to others. In reality, this assumption is apparently not true, and the bidder’s valuation depends not only on himself but also on that of other bidders. For example, in an auction of artworks, bidders take into account not only their own preference for the artwork, but also the potential profits from resale, which is obviously affected by others’ valuations.

After taking into account the behavior of their competitors, the valuation of the object is called “interdependent valuation”. Myerson’s theory no longer holds when there is “interdependent valuation”, and the auctioneer may be able to improve their expected payoff through the design of trading mechanisms. Paul Milgrom and Weber (1982) were the first to study auctions with interdependent values. They built an analytical framework that deals with information, pricing and auctioneer’s profits when there are “interdependent valuations”.

Based on the observation of auction practice, they found that bidders’ valuations may be interdependent: a higher bid by a bidder can easily increase the rating of other participants. Thus, auction can be understood as any buyer’s quote will not only show his own information about the valuation of object, but also partially reveal the private information of other buyers. In this sense, the revenue of a bidder will depend on the degree of privacy of the information.

Once the information is revealed in the auction, bidders can guess each other’s possible bids, and in order to win the auction, they have to bid a higher price. Therefore, for the auctioneer, the auction that will bring him the highest expected
payoff must be the one that most efficiently diminishes the private information of bidders. In the literature on auction theory, Milgrom and Weber called this case as interdependent values.

In reality, the auctioneer devotes a great deal of energy and cost to designing a revenue-maximizing auction mechanism. By applying the auction theory, many well-known auction theorists such as Larry Ausubel, Ken Binmore, Paul Klemperer, Preston McAfee, John McMillan, Paul Milgrom, and Robert Wilson have brought their’s talents into full play in real world auction designs, such as wireless spectrum in the United States, 3G license in the United Kingdom, etc.

All of these objects are not private value goods, but with interdependent values, which are commonplace with a typical example being oilfields. Although the amount of oilfield storage is unknown until it is mined, it has impact for all bidders on the value of the oilfield. As such, the assumption of private value is not applicable for the auction of such object, and thus the auction mechanism design for such economies with interdependent values will have a big impact on the auction results, which is discussed in this section.

21.4.1 Basic Analytical Framework

In the following, we use the mechanism design approach to discuss the auction mechanisms under the interdependent value, as well as the comparison of revenue under different auction formats. As mentioned above, the auction theory of interdependent value was first proposed by Milgrom and Weber (1982), which has a significant influence on the development of auction theory.

We first provide some notations and definitions. There are \( n \) bidders, bidder \( i \)'s signal for the object value is \( \theta_i \), which is private information. The object value \( V_i \) can be expressed as a function of bidder’s private signal, that is, \( V_i = u_i(\theta_1, \ldots, \theta_n) \). If \( V_i = u_i(\theta_i) \), then the object has the characteristics of private value, and \( \theta_i = E(V_i|\theta_i) \) is the private value of bidder \( i \) as described in the previous section. So, the bidder’s value of the object is entirely determined by the signal he/she knows. If \( V_i = u(\theta_1, \ldots, \theta_n) = V \), then the object has the characteristics
of common value. If \( j \neq i, \theta_j \) affects \( V_i \), then the object has the characteristics of interdependent value, so private value and common value can be seen as special cases of interdependent values.

Suppose bidder \( i \)'s bidding strategy is \( b_i \), the probability for the bidder of getting the object is \( y_i(b_i, b_{-i}) \), and the payment is \( t_i(b_i, b_{-i}) \). As before, the bidder \( i \)'s utility in bidding mechanism can be written as:

\[
U(\theta_i, \theta_{-i}, b_i, b_{-i}) = u_i(\theta_1, \ldots, \theta_n) y_i(b_i, b_{-i}) + t_i(b_i, b_{-i}).
\]

For the four common auction mechanisms with interdependent values, bidder \( i \)'s probability to get the object \( y_i(b_i, b_{-i}) \) and interim expected payment after winning \( t_i(b_i, b_{-i}) \) is the same with the case of private value.

For a first-price sealed-bid auction and the Dutch auction, the \( y_i(b_i, b_{-i}) \) is given by Equation (21.3), and the ex post transfer payment \( t_i(b_i, b_{-i}) \) is given by Equation (21.4); for the second-price sealed-bid auction and the English auction, the \( y_i(b_i, b_{-i}) \) is given (21.7), and the ex post transfer payment \( t_i(b_i, b_{-i}) \) is given by (21.8).

For object with common value, the common value is denoted by \( V \). Suppose the private signal \( \theta_i \) under given common value is independently distributed, which is also an unbiased estimate of value, that is, \( E(\theta_i|V = v) = v \). Consider bidder 1 with signal \( \theta_1 = \theta \), then he evaluates the value of the auctioned object as \( E(V|\theta_1 = \theta) \). Let \( \vartheta_1, \vartheta_2, \ldots, \vartheta_{n-1} \) be the largest, second largest, ..., and smallest signal value of \( \{\theta_2, \theta_3, \ldots, \theta_n\} \), or the first-order, second-order, ..., \( n - 1 \)th order statistics of \( \{\theta_2, \ldots, \theta_n\} \). Suppose the bidders are symmetric, and they choose the same bidding strategy function.

It is worth mentioning that the study of interdependent value auction by Milgrom and Weber also showed the possibility of “winner’s curse”. In auction practice, there often exists a phenomenon that the bidder feels not worthy after winning the auction, which is called “winner’s curse” in auction theory. If all bidders are symmetric and follow the same strategy \( \beta \), then this fact reveals to bidder 1 that the highest of the other \( n - 1 \) signals is less than \( x \). As a result, his
estimate of the value upon learning that he is the winner is \( E(V|\theta_1 = \theta, \vartheta_1 < \theta) \). Obviously, \( E(V|\theta_1 = \theta, \vartheta_1 < \theta) < E(V|\theta_1 = \theta) \). The announcement that he has won leads to a decrease in the estimated value, which is lower than his initial estimate. In this sense, winning brings “bad news”. To put it succinctly, if the successful bid comes from an overly optimistic estimate of the auctioned object, the bidder would regret to win the object, which is called the “winner’s curse”. For example, in reality it is often seen that some bidders will boldly raise their bids to a certain level when they know other competitors’ rising price intention. When they discover that the competitors are not going to follow the price raise anymore, they realize themselves bidding too much, but at this moment, their regret is too late. This is the “curse” with winning (e.g., Didius Julianus paid nearly two kilos of gold to each Praetorian Guard soldier, winning their support, and gaining the emperor’s throne. However, this resulted not only in too high price, but also cost his life; corruption brings promotion, but also lead to prison). In fact, this is caused by bidder’s wrong calculation of his expected payoff according to \( E(V|\theta_1 = \theta, \vartheta_1 < \theta) \).

Obviously, the traditional assumption on all bidders’ independent valuation of the auction cannot explain the possibility of “winner’s curse”, while it becomes easy to understand with the introduction of interdependent values. The winning bidder also receives private information about other bidders’ valuations when winning against other bidders, which causes him to lower his rating of the trophy he has just acquired. In the actual bidding, if the bidder is rational, in order to avoid the “winner’s curse”, he will lower his valuation of the object, and the extent of the reduction will be influenced by the auction rules. Therefore, an appropriate auction mechanism needs to minimize the impact of “winner’s curse” for bidders. This is true based on the following discussions. By correct computation according to \( E(V|\theta_1 = \theta) \), this will not happen at equilibrium.
Basic Assumptions

We further relax the assumption that private signals are independently distributed. Suppose that \( n \) bidders get the private signals \( \theta = (\theta_1, \ldots, \theta_n) \). If the density function of \( \theta \) does not satisfy \( \varphi(\theta) = \prod_i \varphi_i(\theta_i) \), where \( \varphi_i(\cdot) \) is the density function of \( \theta_i \), then the distribution of private signals is not independent. We hereby introduce the concept of “affiliation” between signals.

Definition 21.4.1 The random vector \( \theta = (\theta_1, \ldots, \theta_n) \) defined on \( \Theta \subseteq \mathbb{R}^n \) is affiliated. If for any \( \theta, \theta' \in \Theta \), we all have:

\[
\varphi(\theta \vee \theta') \varphi(\theta \wedge \theta') \geq \varphi(\theta) \varphi(\theta'),
\]

where \( \theta \vee \theta' = (\max\{\theta_1, \theta'_1\}, \ldots, \max\{\theta_n, \theta'_n\}) \), \( \theta \wedge \theta' = (\min\{\theta_1, \theta'_1\}, \ldots, \min\{\theta_n, \theta'_n\}) \).

Let \( \vartheta_1, \ldots, \vartheta_{n-1} \) be the first, second, ..., and \( n-1 \)th order statistics of \( \{\theta_2, \ldots, \theta_n\} \), we know (see proof in Milgrom and Weber (1982)): if \( \{\theta_1, \ldots, \theta_n\} \) is affiliated, the variables in any of their subsets are affiliated, and \( \{\theta_1, \vartheta_1, \ldots, \vartheta_{n-1}\} \) are also affiliated. If \( \theta \)’s density function \( \varphi(\cdot) \) is always positive on \( \Theta \), and is second order continuously differentiable, then \( \{\theta_1, \ldots, \theta_n\} \) are affiliated, which is equivalent to \( \frac{\partial^2 \ln(\varphi)}{\partial \theta_i \partial \theta_j} \geq 0, \forall i \neq j \).

In simple terms, the affiliation between variables means that if the value of a random variable is larger, the probability that any other random variable associated with it have a larger value is also higher. For example, in an auction of oilfields, if a bidder learns in advance that a certain oilfield’s reserve is large, other bidders also have a high probability of obtaining similar (private) information.

Define \( \Psi(\cdot | \theta) \) as the conditional distribution function of \( \vartheta_1 \) given \( \theta_1 = \theta \), and \( \psi(\cdot | \theta) \) as the corresponding conditional density function. The affiliation between \( \theta_1 \) and \( \vartheta_1 \) means that, if \( \theta' > \theta \), then \( \Psi(\cdot | \theta') \) is dominant in the sense of adverse risk rate \( \Psi(\cdot | \theta) \), which is:

\[
\frac{\psi(\vartheta \mid \theta')}{\Psi(\vartheta \mid \theta')} \geq \frac{\psi(\vartheta \mid \theta)}{\Psi(\vartheta \mid \theta)}.
\]

It is easy to verify that for increasing functions \( h(\cdot) \), we have \( E(h(\vartheta_1) | \theta') \geq E(h(\vartheta_1) | \theta) \).
Applying the concept of affiliation, Milgrom and Weber analyzed several standard auction mechanisms. In English auctions, bids from bidders who exit early show their information about the value of the object, and the auction price is affiliated to the valuations of all non-winning bidders, resulting in higher profits. In a second-price auction, the auction price is only affiliated to the bidder who values the second-highest, thus generating lower profits. In the Dutch auction and first-price auction, there will be minimal expected payoff to the auctioneer as there is no affiliation between the prices. This finding by Milgrom provides a good explanation for the prevalence of English auctions in reality, and we will discuss the results below.

The following subsections will discuss the equilibrium solution of these auction mechanisms under interdependent values and the result of difference in expected payoff. In the case of private value, the second-price (sealed-bid) auction and the English auction have equivalent auction rules as private information of other participants does not affect the bidder’s valuation of the auctioned object. This is because a bidder’s valuation relies solely on his own signal, while the exit of others will not affect the bidder’s expected value revision, and the rules that the bidder wins the auction and pays (the second highest price) are also the same. Similarly, first-price (sealed-bid) auction and Dutch auction are also equivalent in auction rules. However, in the case of interdependent value and affiliated information, if the number of bidders is three or more, there is no equivalence between the second-price (sealed-bid) auction and the English auction. This is due to the fact that bidders that exit the auction at different prices will change the bidding behavior for the remaining bidders, in particular the bidder winning the auction object, by providing information on revising the value of the item. Thus, the auction results are different under these two auction rules.

If there are only two bidders, once one bidder exits, the other would win the auction. In this case, the exit of a bidder under a bidding price does not change the winner’s payment, whereby the outcome of the English auction and the second-price (sealed-bid) auction is also equivalent. In addition, auction rules for Dutch
and first-price (sealed-bid) auctions still have no effect on the results because the value and payment to the winner are the same in both cases. As such, in the following, we mainly compare the first-price (sealed-bid) auction, the second-price auction and the English auction, and focus on the case of symmetry. The so-called symmetry has two meanings: one is that bidders’ valuation of the object is symmetric, and the other is that bidder’s signal distribution is symmetric, so that we can focus on the discussion of the symmetric bidding (strategy) equilibrium.

We assume that all bidders’ signals are from the same distribution, \( \theta_i \in [0, w] \). The bidders’ valuations of the object are symmetric. Bidder \( i \)'s value function can be written as \( V_i = u_i(\theta) = u(\theta_i, \theta_{-i}) \), which satisfies: \( u_i(\theta_i, \theta_{-i}) = u(\hat{\theta}_i, \theta_{-i}) \), where \( \hat{\theta}_{-i} \) is an arbitrary rearrangement of \( \theta_{-i} \), such as \( u(\theta_i, \theta_j, \theta_k) = u(\theta_i, \theta_k, \theta_j) \), \( i \neq j \neq k \neq i \). Of course, the object with symmetric values do not necessarily have a common value. The following example illustrates this point.

**Example 21.4.1 (Symmetric Value Function)** Consider three bidders’ value signals for the auctioned object are \( \theta_1, \theta_2, \theta_3 \). Suppose for any bidder \( i \), the value function is \( u(\theta_i, \theta_j, \theta_k) = a\theta_i + b\theta_j + c\theta_k \). Bidders’ valuations are symmetric if and only if \( b = c \). When \( a \neq b = c \), the valuation of bidders are symmetric, but the object does not have a common value. In this example, the object has common value only if \( a = b = c \).

Assume that \( u(\theta_i, \theta_{-i}) \) is a continuous function which is strictly increasing with regard to \( \theta_i \), non-decreasing with regard to \( \theta_j, \forall j \neq i \), and satisfies \( u(0, \ldots, 0) = 0, \forall i \).

In the symmetric case, we only need to consider bidder 1’s choice. First define \( v(\theta, \vartheta) = E(V_1|\theta_1 = \theta, \vartheta_1 = \vartheta) \), which characterizes bidder 1’s interim expected value of the object when his private signal is \( \theta \) and the first order statistics for other bidders’ private signals is \( \vartheta \). Meanwhile we assume that all bidders are risk-neutral, and \( v(0, 0) = 0 \). \( \psi(\vartheta|\theta) \) is the density function of \( \vartheta_1 = \vartheta \) under \( \theta_1 = \theta \).
21.4.2 Second-Price Auction

In a second-price auction, the highest bidder gets the object and pays at the second highest price, and pays zero if he does not get the object. We give the symmetric Bayesian-Nash equilibrium results first.

**Proposition 21.4.1** Symmetric Bayesian-Nash Equilibrium for second-Price Auction is $\beta_{II}(\theta) = v(\theta, \theta)$.

**Proof.** For the auction of object, $\beta_{II}(\theta) = v(\theta, \theta) = E(V_1|\theta_1 = \theta, \vartheta_1 = \theta) = E(V_1|\theta_1 = \theta)$ implies that the bidding price is the bidder’s conditional expectation value of the object under private signals, that is, the bidder will use the interim expected value as the bidding price.

Suppose bidders $j \neq 1$ all choose bidding strategy $\beta = \beta_{II}$, bidder 1’s signal is $\theta$, and the chosen bidding price is $b$, then his interim expected payoff $U(b, \theta)$ is:

$$\bar{U}(b, \theta) = \int_0^{\beta^{-1}(b)} [v(\theta, \vartheta) - \beta(\vartheta)]\psi(\vartheta|\theta)d\vartheta$$

$$= \int_0^{\beta^{-1}(b)} [v(\theta, \vartheta) - v(\vartheta, \vartheta)]\psi(\vartheta|\theta)d\vartheta.$$

Obviously under the given assumptions, $v(\theta, \vartheta)$ increases as $\theta$ increases. Thus, if $\vartheta < \theta$, we have $v(\theta, \vartheta) > v(\vartheta, \vartheta)$; If $\vartheta > \theta$, then $v(\theta, \vartheta) < v(\vartheta, \vartheta)$, so when $\beta^{-1}(b) = \theta$, $U(b, \theta)$ reaches the maximum, which means $\beta_{II}$ is a symmetrical Bayesian-Nash equilibrium of the second-price auction. ■

In the above argument, we find that, unlike the second-price sealed-bid price auction with private values, the bidder selecting $\beta_{II}$ is not a weakly dominant strategy, so $\beta_{II}$ is not a weakly dominant equilibrium but a Bayesian-Nash equilibrium.

**Example 21.4.2** Suppose that there are three bidders with a common value $V$ for the object that is uniformly distributed on $[0, 1]$. Given $V$, bidders’ signals $\theta_i$ are uniformly and independently distributed on $[0, 2V]$. 879
Let \( \boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) \) and \( \bar{\theta} \equiv \max\{\theta_1, \theta_2, \theta_3\} \). The density of \( \theta_i \) conditional on \( V \) is \( 1/2V \) on the interval \([0, 2V]\), so the joint density of \((V, \boldsymbol{\theta})\) is \( 1/8V^3 \) on the set \( \{(V, \boldsymbol{\theta}) | \theta_i \leq 2V, \forall i = 1, 2, 3\} \).

Note that the only information about \( V \) that knowledge of \( \boldsymbol{\theta} \) provides is that \( V \geq \bar{\theta}/2 \). Thus, the joint density of \( \boldsymbol{\theta} \) is

\[
\varphi(\boldsymbol{\theta}) = \int_{\bar{\theta}/2}^{1} \frac{1}{8V^3} dV = \frac{4 - \bar{\theta}^2}{16\bar{\theta}^2}
\]

where \( \bar{\theta} = \max\{\theta_1, \theta_2, \theta_3\} \). Thus, the density of \( V \) conditional on \( \theta \) is the same as the density of \( V \) conditional on \( \bar{\theta} \), and then we have

\[
\varphi(V|\theta) = \varphi(V|\bar{\theta}) = \frac{1}{\varphi(\boldsymbol{\theta})} \times \frac{1}{8V^3} = \frac{1}{8V^3} \times \frac{4 - \bar{\theta}^2}{16\bar{\theta}^2}
\]

on the interval \([\bar{\theta}/2, 1]\). Thus,

\[
E(V|\theta) = E(V|\bar{\theta}) = \frac{2\bar{\theta}}{2 + \bar{\theta}}.
\]

Notice that since \( \vartheta_1 = \max\{\theta_2, \theta_3\} \) and \( \bar{\theta} = \max\{\theta_1, \theta_2, \theta_3\}, \bar{\theta} = \max\{\theta_1, \vartheta_1\} \).

\[
v(\theta_1, \vartheta_1) = E(V|(\theta_1, \vartheta_1)) = E(V|\bar{\theta}) = \frac{2\bar{\theta}}{2 + \bar{\theta}}.
\]

Thus we obtain

\[
\beta^{II}(\theta) = v(\theta, \theta) = \frac{2\theta}{2 + \bar{\theta}}.
\]
21.4.3 English Auction

Under English auction, as the bidding price increases, bidders successively exit from the bidding (they can no longer participate in the bidding after exit), and the remaining bidders are the “winners”. The winner’s bidding price is the price at the time of the last bidder’s exit. The following discussion assumes \( n \geq 3 \). Consider bidder 1’s choice: when there are \( k \geq 2 \) bidders remaining, let \( p_n, p_{n-1}, \ldots, p_{k+1} \) be the price (random variable) when the first bidder, second bidder,..., the \( n-k \) bidder exit, obviously when \( j > i \), then \( p_i \geq p_j \). If bidder 1’s private signal is \( \theta \), the minimum price he can exit is \( \beta^k(\theta, p_{k+1}, p_{k+2}, \ldots, p_n) \), which depends on the remaining number of bidders and on the price at the time of the bidder’s exit. This minimum price is equivalent to the bidding price in this case. Due to different information and different beliefs on the expected value of the auction under different number of remaining bidders, the bidding strategy of bidders in the English auction is a series of bidding functions \( \beta = (\beta^n, \beta^{n-1}, \ldots, \beta^2) \).

Consider the bidding strategy below: when no bidder has yet exited, and the bidder’s private signal is \( \theta \), the bidding strategy is

\[
\beta^n(\theta) = u(\theta, \theta, \ldots, \theta). \quad (21.14)
\]

Suppose bidder \( n \) firstly exits after observing signal \( \theta_n \), the price when he exits is \( p_n \), which means \( \beta^n(\theta_n) = u(\theta_n, \theta_n, \ldots, \theta_n) = p_n \). As \( \beta^n(\theta) \) is continuous strictly increasing function in \( \theta \), \( \theta_n \) in the equation is unique. When bidder \( n \) exits, there are \( n-1 \) bidders in the auction, consider the following bidding function: \( \beta^{n-1}(\theta, p_n) = u(\theta, \ldots, \theta, \theta_n) \). Also, \( \beta^{n-1}(\cdot, p_n) \) is a continuous monotonically increasing function, if a bidder, such as the \( n-1 \)th bidder, exits secondly at bidding price \( p_{n-1} > p_n \). If the private signal he observed is \( \theta_{n-1} \), which satisfies \( u(\theta, \ldots, \theta, \theta_{n-1}, \theta_n) = p_{n-1} \). By reverse recursion, at price \( p_n, p_{n-1}, \ldots, p_{k+1} \) there are \( n-k \) bidders exiting the market in order, and the bidding function of the remaining \( k \) bidders is:

\[
\beta^k(\theta, p_{k+1}, \ldots, p_n) = u(\theta, \theta, \ldots, \theta, \theta_{k+1}, \ldots, \theta_n) = p_k, \quad k \geq 2, \quad (21.15)
\]
where $\theta_j, j \geq k + 1$ satisfies: $\beta_j(\theta_j, p_{j+1}, \ldots, p_n) = u(\theta_j, \theta_j, \ldots, \theta_j, \theta_j, \ldots, \theta_n) = p_j$. Since $\beta_j(\ldots, p_{j+1}, \ldots, p_n)$ is a continuously increasing function, the $\theta_j$ satisfying the above formula is unique.

The following proposition characterizes the symmetrical Bayesian-Nash equilibrium of the English auction.

**Proposition 21.4.2** The symmetrical Bayesian-Nash equilibrium in English auction $\beta = (\beta^n, \beta^{n-1}, \ldots, \beta^2)$ is fully determined by (21.14) and (21.15).

**Proof.** Suppose that all bidders except bidder 1 follow the bidding strategy above. Consider bidder 1’s strategy choice. Let $\vartheta_1, \ldots, \vartheta_{n-1}$ be the order statistics of $\theta_2, \ldots, \theta_n$. Suppose any one column value of realized order statistics $\vartheta_1 \geq \vartheta_2 \geq \ldots \geq \vartheta_{n-1}$. Let $\theta$ be the private signal observed by bidder 1. If bidder 1 also follows the bidding strategy above $\beta$, then only when $\theta > \vartheta_1$, bidders can win the auction, and the price that bidder 1 paid is the price when bidder with signal $\vartheta_1$ exits, denoted by $p_2$. Using (21.15), we can get $p_2 = u(\vartheta_1, \vartheta_1, \vartheta_2, \ldots, \vartheta_n)$. Following the bidding strategy above $\beta$, when $\theta > \vartheta_1$, bidder 1’s utility is:

$$u(\theta, \vartheta_1, \vartheta_2, \ldots, \vartheta_n) - p_2 = u(\vartheta_1, \vartheta_1, \vartheta_2, \ldots, \vartheta_n) - u(\vartheta_1, \vartheta_1, \vartheta_2, \ldots, \vartheta_n) > 0.$$  

Obviously, when $\theta > \vartheta_1$, other bidding strategy cannot bring bidder 1 a utility higher than $\beta$.

When $\theta < \vartheta_1$, and following $\beta$, bidder 1 cannot win the auction. If under other bidding strategies, bidder 1 wins the auction, his payment is $p_2 = u(\vartheta_1, \vartheta_1, \vartheta_2, \ldots, \vartheta_n)$. However, if bidder $\vartheta_1$ exits at price $p_2$, the bidder’s assessment of the value of the object is amended as: $u(\theta, \vartheta_1, \vartheta_2, \ldots, \vartheta_n)$. When $\theta < \vartheta_1$, bidder 1’s utility is:

$$u(\theta, \vartheta_1, \vartheta_2, \ldots, \vartheta_n) - u(\vartheta_1, \vartheta_1, \vartheta_2, \ldots, \vartheta_n) < 0.$$  

So, when $\theta < \vartheta_1$, there are no other bidding strategy that can bring bidder 1 a utility higher than $\beta$. Since $\theta_1$ is continuously distributed, the probability of $\theta_1 = \vartheta_1$ is zero, which has no real impact on the bidder’s expected utility. In
summary, there are no other bidding strategies that can bring bidder 1 a utility higher than $\beta$. ■

It should be remarked that English auction has a nice property that its symmetric strategy equilibrium portrayed by (21.14) and (21.15) does not depend on distribution function $\varphi(\cdot)$. As such, it forms an ex post Bayesian-Nash equilibrium. This means that the equilibrium strategy $\beta$ has an important “no regret” feature: for any realization of the signals the bidders have no cause to regret the outcome even if, after the fact, all signals were to become publicly known. This characteristic makes English auctions different from second-price auctions, which rely on the signal distribution function, and as the first-price auction discussed below, are not ex post Bayesian-Nash equilibrium.

21.4.4 First-Price Auction

In the first-price sealed-bid auction, the highest bidder wins the object and pays the bidding price. Suppose $\beta$ is a symmetric Bayesian-Nash equilibrium, first assume $\beta$ is increasing function (verified later). Given that other bidders all choose this strategy, for bidder 1, his private signal is $\theta$. If his bid is $\beta(z)$, his interim expected utility is:

$$
\bar{U}(z, \theta) = \int_0^z [v(\theta, \vartheta) - \beta(z)]\psi(\vartheta|\theta)d\vartheta,
$$

where $v(\theta, \vartheta) = E(V_1|\theta_1 = \theta, \vartheta)$ and $\psi(\cdot|\theta_1 = \theta)$ is the density function of the first order statistics of other participants’ signals when given bidder’s private signal $\theta$, $\Psi(\cdot|\theta)$ is the corresponding distribution function.

The first order condition on $z$ satisfies:

$$
[v(\theta, z) - \beta(z)]\psi(z|\theta) - \beta'(z)\Psi(z|\theta) = 0.
$$

If $\beta$ is a symmetric Bayesian-Nash equilibrium, then it satisfies the first order (necessary) condition:

$$
\beta'(\theta) = v(\theta, \theta) - \beta(\theta)\frac{\psi(\theta|\theta)}{\Psi(\theta|\theta)}.
$$
since \( v(0,0) = 0 \), then \( \beta(0) = 0 \), solve the above first order differential equation and get:

\[
\beta' I(\theta) = \int_0^\theta v(\vartheta, \vartheta) dL(\vartheta|\theta), \quad (21.16)
\]

where \( L(\vartheta|\theta) = \exp(-\int_0^\theta \frac{\psi(t|t)}{\Psi(t|t)} dt) \).

The following proposition characterizes the symmetric Bayesian-Nash equilibrium of the first-price sealed-bid auction.

**Proposition 21.4.3** The symmetric Bayesian-Nash equilibrium of the first-price sealed-bid auction \( \beta I(\theta) \) is characterized by \( (21.16) \).

**Proof.** First \( L(\vartheta|\theta) \) can be regarded as a distribution function on \([0, \theta]\). By the assumption of affiliations, for any \( t > 0 \), we all have:

\[
\frac{\psi(t|t)}{\Psi(t|t)} > \frac{\psi(t|0)}{\Psi(t|0)}.
\]

So,

\[
-\int_0^\theta \frac{\psi(t|t)}{\Psi(t|t)} dt = -\int_0^\theta \frac{\psi(t|0)}{\Psi(t|0)} dt = -\int_0^\theta \frac{d}{dt} (\ln \Psi(t|0)) dt = \ln \Psi(0|0) - \ln \Psi(\theta|0) = -\infty,
\]

Then \( L(0|\theta) = 0 \), and \( L(\theta|\theta) = 1 \). If \( \vartheta' > \vartheta \), obviously there is \( L(\vartheta'|\theta) \leq L(\vartheta|\theta) \), then \( L(\vartheta|\theta) \) is a distribution function.

Then, for all \( \vartheta \leq \theta \), if \( \theta' > \theta \), then \( L(\vartheta|\theta') \leq L(\vartheta|\theta) \), in other words, \( L(\vartheta|\theta') \) is first order stochastically dominant to \( L(\vartheta|\theta) \). Then, if \( \theta' > \theta \), then

\[
\beta' I(\theta') = \int_0^{\theta'} v(\vartheta, \vartheta) dL(\vartheta|\theta') \geq \int_0^\theta v(\vartheta, \vartheta) dL(\vartheta|\theta) = \beta' I(\theta).
\]

Therefore, we verify that \( \beta' I(\theta) \) is a monotonically increasing function.

Since maximization not only requires first-order conditions but also second-order conditions, we need to prove Proposition 21.4.3.
Suppose other bidders all choose $\beta^I$ as bidding strategy, bidder 1’s signal is $\theta$, if choosing bid $\beta(z) = \beta^I(z)$. Then his interim expected utility is:

$$U(z, \theta) = \int_0^z [v(\theta, \vartheta) - \beta(z)]\psi(\vartheta|\theta)d\vartheta.$$ 

The first order condition on $z$ is:

$$\frac{\partial U}{\partial z} = [v(\theta, z) - \beta(z)]\psi(z|\theta) - \beta'(z)\Psi(z|\theta)$$

$$= \Psi(z|\theta)[(v(\theta, z) - \beta(z))\frac{\psi(z|\theta)}{\Psi(z|\theta)} - \beta'(z)].$$

If $z < \theta$, then $v(\theta, z) > v(z, z)$, meanwhile by relevance, we have:

$$\frac{\psi(z|\theta)}{\Psi(z|\theta)} > \frac{\psi(z|z)}{\Psi(z|z)},$$

we get:

$$\frac{\partial U}{\partial z} > \Psi(z|\theta)[(v(z, z) - \beta(z))\frac{\psi(z|z)}{\Psi(z|z)} - \beta'(z)] = 0.$$ 

Similarly, we can get: if $z > \theta$, then $\frac{\partial U}{\partial z} < 0$. So, when $z = \theta$, $U(z, \theta)$ is maximum.

Let us compare it with the first-price auction under private value. Under private value, $v(\vartheta, \vartheta) = \vartheta$. If the signal is distributed independently, $\Psi(\cdot|\theta)$ does not depend on $\theta$, that is $\Psi(\cdot)$ (distribution function about $\vartheta_1$), then $L(\vartheta|\theta) = \frac{\psi(\vartheta)}{\Psi(\theta)}$. Therefore, the Bayesian-Nash equilibrium for the symmetric private value first price auction is:

$$\beta^{I,p}(\theta) = \int_0^\theta \vartheta \frac{\psi(\vartheta)}{\Psi(\theta)} d\vartheta.$$

### 21.4.5 Revenue Comparisons of Auctions

We now look at the capability of the three forms of auctions above by comparing seller’s expected revenues. The basic finding will be that under the symmetric equilibrium, the English auction outperforms the second-price auction, while the seller’s expected revenue under the second-price auction is not less than the seller’s expected revenue under the first-price auction.
Let us compare the expected revenues under the English auction and the second-price auction at first.

The equilibrium strategy under second-price auction with the symmetric e-equilibrium is given by $\beta^{II}(\theta) = v(\theta, \theta)$, where $v(\theta, \vartheta) \equiv E(V|\theta_1 = \theta, \vartheta_1 = \vartheta)$. If $\theta > \vartheta$, the price paid by bidder 1 (the income of the auctioneer) is:

$$v(\vartheta, \vartheta) = E(u(\theta_1, \vartheta_1, \ldots, \vartheta_{n-1})|\theta_1 = \vartheta, \vartheta_1 = \vartheta)$$

In the second-price auction, the expected revenue is:

$$E(R^{II}) = E(\beta^{II}(\vartheta_1)|\theta_1 > \vartheta_1) = E(v(\vartheta_1, \vartheta_1)|\theta_1 > \vartheta_1) \leq E(E(u(\vartheta_1, \vartheta_1, \ldots, \vartheta_{n-1})|\theta_1 = \theta, \vartheta_1 = \vartheta)|\theta_1 > \vartheta_1) = E(u(\vartheta_1, \vartheta_1, \ldots, \vartheta_{n-1})|\theta_1 = \theta, \vartheta_1 = \vartheta) \leq E(\beta^2(\vartheta_1, \vartheta_1, \ldots, \vartheta_{n-1})) = E(R^{Eng}),$$

where the second last equation is obtained by the definition of $\beta^2$ by (21.15), $R^{II}$ and $R^{Eng}$ are the prices paid by the winning bidders of the second-price auction and the English auction respectively.

Now we compare the prices paid by the winning bidders of the second-price auction and of the first-price auction.

We can also suppose that signal of bidder 1 is the highest, so we only have to consider the expected price paid by bidder 1 under the symmetric Bayesian-Nash equilibrium.

$$E(R^{II}(\vartheta_1)|\theta_1 = \theta, \theta_1 > \vartheta_1) = E(v(\vartheta_1, \vartheta_1)|\theta_1 = \theta, \theta_1 > \vartheta_1) \leq \int_0^\theta v(\vartheta, \theta) dK(\vartheta|\theta),$$

where $K(\vartheta|\theta) = \frac{\psi(\vartheta|\theta)}{\Psi(\theta|\theta)}$, $K(\vartheta|\theta)$ is a distribution function with support $[0, \theta]$. 886
But in the first-price auction, $\beta^I = \int_0^\theta v(\vartheta, \vartheta) dL(\vartheta|\theta)$. Next, we will verify that $L(\vartheta|\theta)$ is first-order stochastically dominated by $K(\vartheta|\theta)$.

Because of relevance, if $t < \theta$

$$\frac{\psi(t|t)}{\Psi(t|t)} \leq \frac{\psi(t|\theta)}{\Psi(t|\theta)}.$$  

So, for any $\vartheta < \theta$, we have:

$$- \int_0^\vartheta \frac{\psi(t|t)}{\Psi(t|t)} dt \geq - \int_0^\vartheta \frac{\psi(t|\theta)}{\Psi(t|\theta)} dt$$

$$= - \int_0^\vartheta \frac{d}{dt}(\ln \Psi(t|\theta)) dt$$

$$= \ln \left( \frac{\Psi(\vartheta|\theta)}{\Psi(\theta|\theta)} \right).$$

we will have

$$L(\vartheta|\theta) = \exp(- \int_0^\vartheta \frac{\psi(t|t)}{\Psi(t|t)} dt) \geq \frac{\Psi(\vartheta|\theta)}{\Psi(\theta|\theta)} = K(\vartheta|\theta),$$

which means that $K(\vartheta|\theta)$ first-order stochastically dominates $L(\vartheta|\theta)$. Since $v(\vartheta, \vartheta)$ is an increasing function, by the equivalence condition of first-order dominance, we have

$$E(R^{II}(\vartheta_1)|\theta_1 = \theta, \theta_1 > \vartheta_1) \geq E(\beta^I(\vartheta_1)|\theta_1 = \theta, \theta_1 > \vartheta_1).$$

Proposition below summarizes comparisons of different forms of auctions in expected revenue.

**Proposition 21.4.4** Under the symmetric equilibria, the expected revenue of the seller under the English auction is the highest, followed by the second-price auction, while the expected revenue of the first-price auction is the lowest.

**21.4.6 Efficiency of Auctions**

Under the symmetric economic environment and symmetric Bayesian-Nash equilibria of all the three forms of auctions above, bidder with the highest private signal bids the highest, winning the object. But, can we say that the allocation
is efficient? The efficiency of the allocation means that the bidder being awarded with the object is the one with the highest value (of course, his value must be higher than value of auctioneer). The example below shows that the allocations of all the auctions above may not efficient.

**Example 21.4.3** Consider a two-bidder auction in which their valuation function is symmetric, \( \theta_1 \) and \( \theta_2 \) are private signal the two bidders have. Suppose their valuation functions are:

\[
\begin{align*}
    u_1(\theta_1, \theta_2) &= \frac{1}{4}\theta_1 + \frac{3}{4}\theta_2; \\
    u_2(\theta_1, \theta_2) &= \frac{1}{4}\theta_2 + \frac{3}{4}\theta_1.
\end{align*}
\]

Obviously, if \( \theta_1 > \theta_2 \), then \( u_1 < u_2 \). Thus, bidder 1 wins no matter the auction is English auction, second-price auction or first-price auction. So the outcomes of all the three forms of auctions are inefficient.

In this example, the signal \( \theta_i \) of each bidder \( i \) has less influence on his own valuation than it does on the other bidder’s valuation. And it includes that the winner of the auction is not the bidder with the highest value. So, we need some extra constraints, such as the single crossing condition.

**Definition 21.4.2 (Single Crossing Condition)** For the valuation system of auctions, we will say that \( u_i, i \in \{1, \ldots, n\} \), satisfies the single crossing condition, if for any \( j \neq i \), and for all \( \theta \), satisfy:

\[
\frac{\partial u_i}{\partial \theta_i}(\theta) \geq \frac{\partial u_j}{\partial \theta_i}(\theta).
\]

Under the symmetric equilibrium, if the single crossing condition is satisfied, can the three forms of auction allocate efficiently? Or can we say that the higher the signal is, the more value for the bidder there is. The answer to this problem is positive.

Suppose \( \theta_i > \theta_j \), where \( \alpha(s) = (1 - s)(\theta_j, \theta_i, \theta_{-ij}) + s(\theta_i, \theta_j, \theta_{-ij}) \). According to the mean value theorem of integrals:

\[
u(\theta_i, \theta_j, \theta_{-ij}) - u(\theta_j, \theta_i, \theta_{-ij}) = \int_0^1 \nabla u(\alpha(s)) \cdot \alpha'(t) dt,
\]
where \( \nabla u(\alpha(s), \alpha'(s)) = \frac{\partial}{\partial \theta_i} u(\alpha(s)) (\theta_i - \theta_j) + \frac{\partial}{\partial \theta_j} u(\alpha(s)) (\theta_j - \theta_i) \); and by the symmetry of the value function, \( \frac{\partial}{\partial \theta_i} u(\alpha(s)) = \frac{\partial}{\partial \theta_j} u(\alpha(s)) \), \( \frac{\partial}{\partial \theta_j} u(\alpha(s)) = \frac{\partial}{\partial \theta_i} u(\alpha(s)) \), by single crossing condition, \( \frac{\partial}{\partial \theta_i} u(\alpha(s)) \geq \frac{\partial}{\partial \theta_j} u(\alpha(s)) \), and as \( \theta_i > \theta_j \), we have

\[ u(\theta_i, \theta_j, \theta_{-ij}) = u(\theta_j, \theta_i, \theta_{-ij}). \]

We already discussed three forms of auctions under the symmetric equilibrium. But for a general equilibrium, such as the non-symmetric equilibrium, what is the feature of efficient allocation (which may not exist) of the interdependent valuation auctions should like? At first, we also need a constraint that is the single crossing condition like above. The example below illustrates that there is no efficient allocation in an auction if the single crossing condition is not satisfied.

**Example 21.4.4** Suppose there are two bidders, whose signal are \( \theta_1 \) and \( \theta_2 \), their valuation functions for indivisible object are:

\[ u_1(\theta_1, \theta_2) = \theta_1, \]
\[ u_2(\theta_1, \theta_2) = \theta_2^2. \]

Suppose \( \theta_1 \in [0, 2] \), the signal of bidder 2 does not affect their valuation to the object. Clearly, the valuations do not satisfy the single crossing condition, such as \( \frac{\partial u_1}{\partial \theta_1} (1, \theta_2) < \frac{\partial u_2}{\partial \theta_1} (1, \theta_2) \).

It can be easily seen that \( u_1 > u_2 \) if and only if \( \theta_1 < 1 \). If \( (y(\theta_1), t(\theta_1)) \) is efficient, \( y(\theta_1) = (y_1(\theta_1), y_2(\theta_1)) \) are the probabilities that bidder 1 and bidder 2 get the object; and \( t(\theta_1) = (t_1(\theta_1), t_2(\theta_1)) \) are the transfer payments of bidder 1 and bidder 2. They need to satisfy that if \( \theta_1 < 1 \), then \( y_1(\theta_1) = 1 \); if \( \theta_1 > 1 \), then \( y_2(\theta_1) = 1 \).

Suppose \( \vartheta_1 < 1 < z_1 \), then efficiency and incentive compatibility together require that when agent 1’s private signal is \( z_1 \), we have:

\[ 0 - t_1(z_1) \geq z_1 - t_1(\vartheta_1); \]

and when his private signal is \( \vartheta_1 \), we have:

\[ \vartheta_1 - t_1(\vartheta_1) \geq 0 - t_1(z_1). \]
The two inequalities above imply $\vartheta_1 \geq z_1$, which is a contradiction. So there is no auction satisfying incentive compatibility which can allocate efficiently.

Moreover, we can in fact show that the single crossing condition is also a necessary condition for an auction to be efficient. If there exists an efficient mechanism, then by the revelation principle, it must be a “truth-telling” mechanism. Suppose that all the bidders other than bidder $i$ have observed an signal $\theta_{-i}$, and the bidder $i$’s private signal is $\theta_i$. If no matter what value of his signal $\theta_i$ is, bidder $i$ either always wins or always loses, then the single crossing condition means nothing for $\theta_i$. So we are only concerned for the situation if buyer $i$’s signal will decide whether he can get the object or not. Thus, we will say that buyer $i$ is pivotal, if there exist signals $\vartheta_i, z_i$, such that $u_i(\vartheta_i, \theta_{-i}) > \max_{j \neq i} u_j(\vartheta_i, \theta_{-i})$ and $u_i(z_i, \theta_{-i}) < \max_{j \neq i} u_j(z_i, \theta_{-i})$. Incentive compatibility requires that when bidder $i$’s signal is $\vartheta_i$, he would not report $z_i$, which means that:

$$u_i(\vartheta_i, \theta_{-i}) - t_i(\vartheta_i, \theta_{-i}) \geq 0 - t_i(z_i, \theta_{-i}).$$

And, when he observes $z_i$ he would not report $\vartheta_i$, which means that:

$$0 - t_i(z_i, \theta_{-i}) \geq u_i(z_i, \theta_{-i}) - t_i(\vartheta_i, \theta_{-i}).$$

From the two inequalities above, we have:

$$u_i(\vartheta_i, \theta_{-i}) \geq u_i(z_i, \theta_{-i})$$

That is to say, bidder $i$’s value when he wins the object must be at least as high as when he does not. In other words, keeping others’ signals unchanged, an increase in bidder $i$’s private signal cannot reduce the probability of winning the object. So, by the incentive compatibility, the efficient mechanism must satisfy the monotonic conditions (just like the Maskin monotonicity). This requires that, at $\theta_i$, if $u_i(\theta_i, \theta_{-i}) = u_j(\theta_i, \theta_{-i})$, we can have:

$$\frac{\partial u_i}{\partial \theta_i}(\theta_i, \theta_{-i}) \leq \frac{\partial u_j}{\partial \theta_i}(\theta_i, \theta_{-i}),$$

so that the monotonic conditions can be satisfied. Thus, if the allocation that satisfies incentive compatibility is efficient, it must satisfy the monotonic conditions.
21.4.7 The Generalized VCG Mechanism

Under an economic environment with symmetric and interdependent values, the symmetric equilibrium of the auction is efficient if the single crossing condition is satisfied. But, can we find out an efficient mechanism under general associated valuation system? In the previous chapters on general mechanism design, we have shown that VCG is an efficient mechanism if the players’ value function only depends on themselves.

However, it will lose efficiency when a bidder’s valuation function is interdependent on others’ signals. Indeed, if bidder \( i \) who gets the object is asked to pay the second highest value of all the reported signals, \( \max_{j \neq i} u_j(\theta_i, \theta_{-i}) \), which is depend on \( \theta_i \), then player \( i \) would have the incentive to report a lower signal in order to reduce the payment. As such, the usual VCG mechanism is not a “truth-telling” mechanism. But, the so-called generalized VCG mechanism that is revised from the VCG is efficient. We first describe generalized VCG mechanism.

Suppose there are \( n \) players in the mechanism \( (\Theta, y(\theta), t(\theta)) \), where \( \theta = (\theta_1, \ldots, \theta_n) \in \Theta \) are the private signals of the bidders, \( y(\theta) = (y_1(\theta), \ldots, y_n(\theta)) \) are the probabilities of bidders allocated to the object, \( t(\theta) = (t_1(\theta), \ldots, t_n(\theta)) \) are the transfer payments of the players.

If the mechanism is efficient, we have:

\[
y^*_i(\theta) = \begin{cases} 
1 & \text{if } u_i(\theta) > \max_{j \neq i} u_j(\theta) \\
0 & \text{if } u_i(\theta) < \max_{j \neq i} u_j(\theta).
\end{cases}
\]

If there are more than one bidders who have the highest value, the object is awarded to each of these players with equal probability. But with continuous distribution, the probability of this situation equals to 0.

The bidder who gets the object pays transfer payments:

\[
t^*_i = -u_i(\tilde{\theta}_i(\theta_{-i}), \theta_{-i}),
\]

where

\[
\tilde{\theta}_i(\theta_{-i}) = \inf \{ \bar{\theta}_i : u_i(\bar{\theta}_i, \theta_{-i}) \geq \max_j u_j(\bar{\theta}_i, \theta_{-i}) \}.
\]
Given other bidders’ signal, it is the lower bound of the signal value when bidder $i$ wins the object in an efficient allocation. The bidder who does not get the object pays nothing. We can see that the transfer payments that one pays do not depend on the signal of himself directly, just as in the private value model.

We call the $(\Theta, y^*, t^*)$ the generalized VCG mechanism, because it adapted Vickrey mechanism under the interdependent values setting. As mentioned above, the usual VCG mechanism is not a “truth-telling” mechanism. In order the winner player $i$ to have incentive to tell the truth, we should ask player $i$ only to pay $\max_{j \neq i} u_j(\varrho_i(\theta_{-i}), \theta_{-i})$ (which does not depend on $\theta_i$ directly), rather than $\max_{j \neq i} u_j(\theta_i, \theta_{-i})$. When $u_i(\varrho_i(\theta_{-i}) > \max_{j \neq i} u_j(\theta_i, \theta_{-i})$, we have $\varrho_i(\theta_{-i}) \leq \theta_i$, so $\max_{j \neq i} u_j(\theta_i(\theta_{-i}), \theta_{-i}) \leq \max_{j \neq i} u_j(\theta_i, \theta_{-i})$.

The proposition below shows that the generalized VCG mechanism is an efficient mechanism under the associated valuation system.

**Proposition 21.4.5** Suppose that the valuations of bidders to the object satisfy the single crossing condition. The generalized VCG mechanism is an efficient dominant incentive compatibility (truth telling) mechanism.

**Proof.** Obviously, if everyone tells the true signal, the generalized VCG mechanism is efficient. Now we have to prove it is dominant incentive compatibility. When $u_i(\varrho_i(\theta_{-i}) > \max_{j \neq i} u_j(\theta_i, \theta_{-i})$, player $i$ who gets the object pays $u_i(\varrho_i(\theta_{-i}), \theta_{-i})$. As the payments do not depend on the signal $\theta_i$ he reported, and $\theta_i \geq \varrho_i(\theta_{-i})$, the expected utility of reporting true signal is: $u_i(\theta_i, \theta_{-i}) - u_i(\varrho_i(\theta_{-i}), \theta_{-i}) \geq 0$. When he reports other signal $z_i$, if $z_i > \varrho_i(\theta_{-i})$, his expected utility does not change; if $z_i < \varrho_i(\theta_{-i})$, he would not get the object and his expected utility will be 0. So, in an efficient allocation, bidder $i$ gets the object, and there is no other strategy that can bring more expected utilities than to tell the truth.

When $u_i(\varrho_i(\theta_{-i}) < \max_{j \neq i} u_j(\theta_i, \theta_{-i})$, if he tells the truth, bidder $i$’s expected utilities equal to 0; if he reports $z_i$ to alter the outcome, by the single crossing condition: $z_i > \varrho_i(\theta_{-i}) > \theta_i$, player $i$ wins the object, but his expected utilities
are \( u_i(\theta_i, \theta_{-i}) - u_i(\theta_i(\theta_{-i}), \theta_{-i}) < 0 \). So we have proved that the generalized VCG mechanism is an efficient dominant incentive compatibility mechanism.

Under the circumstance of private values, the generalized VCG mechanism above returns to the usual VCG mechanism. However, when values are interdependent, the usual VCG mechanism is not an efficient “truth-telling” mechanism. At the same time, it is different from the usual VCG mechanism as the generalized VCG mechanism depends on much more details. It means one can decide the payment \( u_i(\theta_i(\theta_{-i}), \theta_{-i}) \) only when he gets the knowledge of the valuation function of all players. As such, the generalized VCG mechanism is not an anonymous mechanism but depends on the features (such as the preferences) of the players.

21.4.8 Optimal Auction Mechanism Design

We now discuss the optimal mechanism design when values are interdependent. Under the optimal mechanism of independent private value, the object is awarded to the bidder whose value is the highest. In order to motivate the bidders to tell the truth, the bidder with the highest value is able to obtain some positive informational rents. But, this outcome will not hold when values are interdependent. Just like the Full Surplus Extraction Theorem of Cremer-McLean discussed in Chapter 16, we will get the optimal mechanism when values are interdependent and the auctioneer will extract all the surplus. Let’s investigate the reasons behind the conclusion.

In order to explain the Full Surplus Extraction Theorem clearly, we set the space of signal discrete. Define \( \Theta_i = \{\theta_1^i, \theta_2^i, \ldots, \theta_{m_i}^i\} \) as the discrete information space of bidders, where \( \theta_1^i < \theta_2^i < \ldots < \theta_{m_i}^i \). Suppose \( \theta_i \in \Theta_i \) and the signal mold of bidder \( i \) is \( \theta \), and the valuation function is \( u_i = u_i(\theta_i, \theta_{-i}) \). For the discrete values, we can define the single crossing condition as: \( j \neq i \), if \( u_i(\theta_i, \theta_{-i}) \geq u_j(\theta_i, \theta_{-i}) \), then \( u_i(\theta_i', \theta_{-i}) \geq u_j(\theta_i', \theta_{-i}) \) holds for any \( \theta_i < \theta_i' \in \Theta \): at the same time, the conclusion does not change either \( \geq \) or \( > \).

For any bidder \( i \), let \( \Pi_i \) be a matrix with \( m_i \) rows and \( \sum_{j \neq i} m_j \) columns and its element is denoted by \( \pi_i(\theta_{-i} | \theta_i) \). Each row corresponds to bidder \( i \)'s beliefs
regarding the signal distribution of other bidders given $\theta_i$. If the signals are independent, then all rows are the same, which means the rank of $\Pi_i$ is 1. If the signals are correlated, the beliefs represented by different rows are not the same. In the proposition below, we suppose the $\Pi_i$ is of rank $m_i$, which means, the bidder’s beliefs regarding the signal distribution of other bidders will change according to different $\theta_i$. We then have the following Full Surplus Extraction Theorem under interdependent values.

**Theorem 21.4.1 (Full Surplus Extraction Theorem under Interdependent Values)**

Suppose that signals are discrete and the valuations satisfy the single crossing condition. Suppose for any bidder $i$, $\Pi_i$, the matrix of beliefs about other players, is of full rank (that is $m_i$), then there exists a mechanism in which truth-telling is an efficient equilibrium and the expected informational rents (that is expected utility in the event) of every bidder is equal to zero. As a result, all the surplus extraction is awarded to the auctioneer.

**Proof.** First, consider the generalized VCG mechanism $(\Theta, y(\theta), t(\theta))$. From the above discussions, we know that it is a truth telling mechanism and efficient. In this mechanism, the expected utility of bidder $i$ with signal $\theta_i$ at the equilibrium is:

$$\bar{U}_i^*(\theta_i) = \sum_{\theta_{-i}} \pi_i(\theta_{-i}|\theta_i)[y_i(\theta)u_i(\theta) + t_i(\theta)].$$

Let $\bar{u}_i^* = [\bar{U}_i^*(\theta_i^1), \bar{U}_i^*(\theta_i^2), \ldots, \bar{U}_i^*(\theta_i^{m_i})]'$. As the rank of the matrix $\Pi_i$ is full row, there is a column vector $c_i = (c_i(\theta_{-i}))_{\theta_{-i} \in \Theta_{-i}}$ of size $\sum_{j \neq i} m_j$, such that:

$$\Pi_i c_i = -\bar{u}_i^*,$$

for $\forall \theta_i$, and then we have:

$$\sum_{\theta_{-i}} \pi_i(\theta_{-i}|\theta_i)c_i(\theta_{-i}) = \bar{U}_i^*(\theta_i)$$

Compared the generalized VCG mechanism $(\Theta, y(\theta), t(\theta))$ with the Crémer-McLean mechanism $(\Theta, y(\theta), t(\theta) + c_i(\theta_{-i}))$, we find that the allocation rule is
the same, but the transfer payment of every bidder has reduced by $c_i(\theta_{-i})$ which does not depend on bidder $i$’s report. Thus, the Crémer-McLean mechanism is still a truth telling mechanism and it is efficient. At the same time, every bidder’s expected utility in the event is reduced to zero, which means the auctioneer may get all the possible surplus extraction.

This conclusion is robust. If there are private values that are correlated, no matter how close they are, the information rent will be reduced to zero. And this conclusion depends on the interdependence of signal other than that of valuations.

21.4.9 Ex Post Implement

Ex Post Implementation with Distribution Unknown

Standard auction theory assumes the seller knows the distribution $F$ from which the buyers’ valuations are drawn. The optimal auction then sells to buyers in order of their valuations while the virtual valuation $\nu(\theta_i)$ exceeds the marginal cost $c$. But if the distribution $F$ is unknown, the monopoly does not have a demand curve. That is to say, suppose that buyers’ valuations are i.i.d. draws from an unknown distribution $F$ over which the designer has a Bayesian prior. This makes buyers’ signals correlated. What mechanism should be used?

We now consider the mechanism design of ex post implementation only. Under the monotonicity constraints, the maximizing expected profits of the mechanism are:

$$
\max_{x^{(i)}} E_{\theta} \left[ \sum_{i=1}^{n} \nu(\theta_i \mid \theta_{-i}) y_i(\theta) - c \left( \sum_{i=1}^{n} y_i(\theta) \right) \right] \\
\text{s.t. } y_i(\theta_i, \theta_{-i}) \text{ is nondecreasing about } \theta_i, \forall i \geq 1. \quad (DM)
$$

where, $\nu(\theta_i \mid \theta_{-i}) = \theta_i - \frac{f(\theta_i \mid \theta_{-i})}{1-F(\theta_i \mid \theta_{-i})}$ is buyer $i$’s conditional value/marginal revenue function (according to symmetric, this function is the same for all the buyers).

In simple terms, we only consider the situation in which the marginal cost is equal to a constant $c$. So, we can ignore DM. The object is sold to buyer $i$ if and only if $\nu(\theta_i \mid \theta_{-i}) \geq c$. If the valuation function $\nu(\theta_i \mid \theta_{-i})$ is non-decreasing for $\theta_i$, 

895
this mechanism also satisfies DM. Based on the information offered by each buyer to other players, this mechanism will offer an optimal monopoly price $p(\theta_{-i})$ and sell the object to the buyer whose value (reported) is higher than the monopoly price. (The buyers report value of themselves at first, then the mechanism decides the monopoly price.) According to the Large Number Theorem, if the number of player $n \to \infty$, $p(\theta_{-i})$ will converge to the optimal monopoly price, so the profits of the seller will converge to the profits in which the $F$ is known.

What should be pointed out is the mechanism above is totally different from the traditional auction mechanism. In this mechanism, the bidding of buyer $i$ affects the allocation of others directly rather than only through buyer $i$ allocations. This effect is through the the “informational linkage” among buyers, rather than the “technological linkage” in the standard auction through the sellers’ cost function or capacity constraint.

**Ex Post Implementation under Common Value**

In a quasi linear environment with common value, the revenue $v_i(x, \theta) + t_i$ of the player $i$ also depends on $\theta_{-i}$.

In this case, the type-related issue still exists. We can set a class of independent common value, such as in the “Wallet Game”, all the players’ revenues are correlated to $\sum_i \theta_i$. However, one more appropriate model would be the “mineral right” model. In this model, the types of all the players are independent identically distributed information under the unknown real state, so all the types of players are related to a prior distribution. Most of the literatures on common value are empirical. These papers only deal with how to solve and compare the equilibrium solution of some mechanisms, rather than how to find out the optimal mechanism. These researches can be traced back to Milgrom and Weber (1982). Recently, some researches about optimal mechanism have turned their attention to “ex post implement” with a new perspective. The basic idea is to introduce “ex post incentive compatibility” constraints:

$$v_i(x(\theta), \theta) + t_i(\theta_i, \theta_{-i}) \geq v_i(x(\hat{\theta}_i, \theta_{-i}), \theta) + t_i(\hat{\theta}_i, \theta_{-i}), \forall \theta_i, \hat{\theta}_i \in \Theta_i, \forall \theta_{-i} \in \Theta_{-i}.$$
It means, no matter what type of the players is, once they obey their equilibrium strategies (report truthfully), telling truth is the best choice for player $i$. Without this assumption, the other players would report $\theta_i'$ even though their truth types are $\theta_{-i}$, and consequently the utilities of player $i$ are $v_i(x(\hat{\theta}_i, \theta_i'), \theta_i, \theta_{-i}) + t_i(\hat{\theta}_i, \theta_{-i})$. Thus, for player $i$, telling truth may not be optimal. In the private value situation, we do not need ex post incentive compatibility as it is consistent with DIC. Otherwise, ex post implement can also take ex post individual rationality into consideration. For this, readers can refer to Dasgupta and Maskin (QJE 2000).

21.5 Simultaneous Private Value Auctions for Multiple Identical Objects

The rest of this chapter discusses multiunit auctions especially under private values. In reality, most of the auctions are multiunit auctions. Multiunit auctions are of many different forms that depend on the relationship of objects in some degree as well as timing and formats of auctions.

Objects can be either identical or heterogenous. When the objects are identical, a buyer may only need one or some of them. Under this situation, the marginal willingness to pay will decrease with the increase in the number of objects bought. When the objects are heterogenous, the objects are are different but may affect each other. Then, the willingness to pay of the buyer will change with the relationship of these objects. For instance, the value of consuming two objects simultaneously may be higher than the sum of value consuming them separately, which means there exists complements within the objects. Of course, there also may be substitutability. The complements and substitutability are the factors worth paying special attention to when we design an auction mechanism.

Multiunit auctions also depend on the timing and formats of auctions. When multiple objects are to be sold, many options can be chosen by the seller. The seller can sell the objects jointly in a single auction or separately in multiple
auctions. The former case is called simultaneous action, in which all the objects are sold in a single auction at one time, but not necessarily all to the same bidder, and the bids on the various objects collectively influence the overall allocation. The latter case is called sequential auction, in which the objects are sold one at a time in separate auctions in a way that the bids in the auction for one of the objects do not directly influence the outcome of the auction for another.

The seller can also choose different auction formats. As for single object auctions, formats can be sealed-bid such as discriminatory auction, uniform-price auction and Vickrey auction or open-bid actions such as Dutch auction, English auction and Ausubel auction.

We will first consider multiple identical object simultaneous auctions and then consider multiple identical object sequential auctions, followed by considering heterogeneous multiple object auctions in private values.

### 21.5.1 Basic Model

Suppose there are $K$ units of identical objects for sale and $n$ bidders, the set of bidders are $N = \{1, \ldots, n\}$. The marginal value of bidder $i$ to different units of goods are $V_i(\theta) = (V_1^i(\theta), \ldots, V_K^i(\theta))$, where $V_k^i$ is bidder $i$'s value to item $k$.

We often suppose: $V_k^i \leq V_l^i, \forall k > l$, which means the marginal utility decreases with the increase in the number of identical objects; $\theta_i$ is the signal observed by bidder $i$ and $\theta$ are the signal observed by all the bidders. The bidding function of bidder $i$ is $b_i = (b_1^i(\cdot, V_i), \ldots, b_K^i(\cdot, V_i))$, where $b_k^i$ is the bid of bidder $i$ to the $k$-th unit of objects.

As multiunit auctions are relatively complex, we focus on the private value models, $V_i(\theta) = V_i(\theta_i) \equiv V_i$. We will discuss the allocation of objects and transfer payments under different mechanisms. All the formats of auctions obey the rule that objects are given to the $K$ highest of these bids.
21.5.2 Simultaneous Sealed-Bid Auctions for Multiple Identical Objects

In a simultaneous auction environment with \( K \) identical objects to be auctioned, each bidder \( i \) is asked to submit \( K \) bids \( b_i = (b_{1i}, \ldots, b_{Ki}) \), satisfying \( b_1^i \geq b_2^i \geq \ldots \geq b_K^i \), to indicate how much he is willing to pay for each additional unit. If the bidder wants to buy \( k \) units, he has to pay \( b_1^i + \ldots + b_k^i \). Given the bid vector \( b_i \) of bidders, we will get the demand function \( d_i \) of bidder \( i \):

\[
d_i(p) = \max\{k : b_k^i \geq p\},
\]

where \( d_i(p) \) is a non-decreasing function. If the price \( p \) is between \( b_k^i \) and \( b_{k+1}^i \), bidder \( i \) will buy \( k \) units. Given the bid vector of bidders, we can construct the demand function for the objects. In turn, if we already know the demand for the objects, we can get the bid vector of the bidders. The two methods are equivalent.

Suppose the biding set of \( n \) bidders is \( \{b_k^i, i = 1, \ldots, n; k = 1, \ldots, K\} \), \( K \) units of objects are allocated to \( K \) bidders whose price is the highest. By the rules of the discriminatory auction, uniform-price auction and Vickrey auction we discuss below, the allocation of objects are the same, but the transfer payments to win the objects are different in the three auctions.

We will illustrate the allocation and pricing rules of the three auctions through a simple example below.

**Example 21.5.1** Suppose there are six units of identical objects and three bidders in the auction. The bid vectors are:

\[
b_1 = (54, 46, 43, 40, 33, 15)
b_2 = (60, 55, 47, 32, 27, 13)
b_3 = (48, 45, 35, 24, 14, 9)
\]

Thus, the six highest bids are:

\[
(b_2^1, b_2^2, b_1^1, b_3^1, b_2^3, b_1^2) = (60, 55, 54, 48, 47, 46)
\]
As such, bidder 1 gets two units, bidder 2 gets three units and bidder 3 gets one unit.

**Discriminatory Auction**

In a discriminatory auction, each bidder pays an amount equal to the sum of his bids that are among the $K$ highest of the $N \times K$ bids submitted in all. That is, if bidder $i$ is allocated to $k_i$ units of objects, his total payment is $b_1^i + \ldots + b_{k_i}^i$. This amounts to perfect price discrimination relative to the submitted demand functions; hence have the name of the auction.

In Example 21.5.1, bidder 1 gets two units objects, and his payment equals to $54 + 46 = 100$. Bidder 2 and bidder 3 get three units and one unit of objects respectively, and their payments equal to $60 + 55 + 47 = 162$ and $48$ respectively. Thus, the total revenue received by the auctioneer is $100 + 162 + 48 = 310$.

The discriminatory pricing rule can also be described by the residual supply function. At price $p$, the residual supply faced by bidder $i$ denoted as $s_i(p)$, is equal to the total units of objects subtracting the sum of the amounts demanded by other bidders, so:

$$s_i(p) \equiv \max\{K - \sum_{j \neq i} d_j(p), 0\}.$$  

Obviously, as $d_i(p)$ is a non-increasing function, $s_i(p)$ is a non-decreasing function. In the discriminatory auction, the total payment of each bidder equals to the area under his demand function up to the point where it intersects with the residual supply curve. In Example 21.5.1, the demand function of each bidder is:

- $d_1(p) = \begin{cases} 
0 & \text{if } p > 54; \\
1 & \text{if } 54 \geq p > 46; \\
2 & \text{if } 46 \geq p > 43; \\
3 & \text{if } 43 \geq p > 40; \\
4 & \text{if } 40 \geq p > 32; \\
5 & \text{if } 32 \geq p > 15; \\
6 & \text{if } 15 \geq p > 0, 
\end{cases}$

- $d_2(p) = \begin{cases} 
0 & \text{if } p > 60; \\
1 & \text{if } 60 \geq p > 55; \\
2 & \text{if } 55 \geq p > 47; \\
3 & \text{if } 47 \geq p > 32; \\
4 & \text{if } 32 \geq p > 27; \\
5 & \text{if } 27 \geq p > 13; \\
6 & \text{if } 13 \geq p > 0, 
\end{cases}$

- $d_3(p) = \begin{cases} 
0 & \text{if } p > 48; \\
1 & \text{if } 48 \geq p > 45; \\
2 & \text{if } 45 \geq p > 35; \\
3 & \text{if } 35 \geq p > 24; \\
4 & \text{if } 24 \geq p > 14; \\
5 & \text{if } 14 \geq p > 9; \\
6 & \text{if } 9 \geq p > 0. 
\end{cases}$
The residual supply function of bidder 1 is:

\[
s_{-1}(p) = \begin{cases} 
6 & \text{if } p > 60; \\
5 & \text{if } 60 \geq p > 55; \\
4 & \text{if } 55 \geq p > 48; \\
3 & \text{if } 48 \geq p > 47; \\
2 & \text{if } 47 \geq p > 45; \\
1 & \text{if } 45 \geq p > 35; \\
0 & \text{if } 35 \geq p \geq 0. 
\end{cases}
\]

When \( p \in [45, 46] \), \( d_1(p) = s_{-1}(p) = 2 \), the payment of bidder 1 when allocated to the two units of objects equal to

\[
\int_0^\infty \min\{d_1(p), 2\} dp = \int_0^2 (d_1)^{-1}(q) dq,
\]

That is the area under his demand function up to the point where it intersects with the residual supply curve, while \( (d_1)^{-1} \) is the inverse function of \( d_1 \).

When \( K = 1 \), discriminatory auction turns out to be first-price auction, and the discriminatory price is just an extension of the first-price auction with multiple objects.

**Uniform-Price Auctions**

In the uniform-price auction, all the objects are sold to the bidders at a market-clearing price. In the discrete model studied here, we take the highest losing bid that is the \( K + 1 \) highest of all the bids as the market-clearing price, as the price that clears the market. In Example 21.5.1, the seventh-highest bid is 45.

In a general situation, if the bid set of \( n \) bidders is \( \{b_i^k, i = 1, \ldots, n; k = 1, \ldots, K\} \), and the bidder \( i \) gets \( k_i \) units of objects from the \( K \) highest of these bids, then the market-clearing price is:

\[
p = \max_i \{b_i^{k_i+1}\}.
\]

For the above example, \( p = \max_i \{b_i^{k_i+1}\} = \max\{43, 32, 45\} = 45 \). The total revenue received by the auctioneer is \( 6 \times 45 = 270 \).
Suppose \( c_{-i} \) is the \( K \)-vector of competing bids of bidder \( i \) facing other bidders, which is constructed by the \( K \) highest of bids of \( \{b^k_j; j \neq i, k = 1, \ldots, K\} \), as these bids will affect whether bidder \( i \) can get the objects. \( c_{-i} = (c^1_{-i}, \ldots, c^K_{-i}) \) is a price vector for rearranging the component in decreasing order. As such, bidder \( i \) wins exactly \( k^i > 0 \) units if and only if

\[
  b^{k^i}_i > c^{K-k^i+1}_{-i} \quad \text{and} \quad b^{k^i+1}_i < c^{K-k^i}_{-i}.
\]

Thus, the surplus supply function of bidder \( i \) is:

\[
s_{-i}(p) = K - \max\{k : c^k_{-i} \geq p\},
\]

which is the inverse function of \( c_{-i} \). Then, the market-cleaning price can be also determined by

\[
p = \max\{b^{k^i+1}_i, c^{K-k^i+1}_{-i}\},
\]

which is independent of \( i \). In Example 21.5.1, since

\[
  b_1 = (54, 46, 43, 40, 33, 15)
\]

\[
  b_2 = (60, 55, 47, 32, 27, 13)
\]

\[
  b_3 = (48, 45, 35, 24, 14, 9)
\]

and

\[
  c_{-1} = (60, 55, 48, 47, 45, 35)
\]

\[
  c_{-2} = (54, 48, 46, 45, 43, 40)
\]

\[
  c_{-3} = (60, 55, 54, 47, 46, 43),
\]

the market-cleaning price can be obtained from bidder 1: \( p = \max\{b^3_1, c^5_{-1}\} = \max\{43, 45\} = 45 \). We can also obtain the same price \( p = \max\{b^4_2, c^4_{-2}\} = \max\{32, 45\} = 45 \) from bidder 2 or \( p = \max\{b^2_3, c^6_{-3}\} = \max\{45, 43\} = 45 \) from bidder 3.

When \( K = 1 \), the uniform-price auction turns out to be a second-price auction. When \( K > 1 \), the uniform-price auction is different from the second-price auction.
in many ways, especially in the pricing strategy. So we cannot extend the second-price auction to the uniform-price auction directly. The Vickrey auction discussed below is more like an extension of the second-price auction with single-object.

**Vickrey Auctions**

In a second-price auction with single-object, the winner’s price is the highest of all the losing bids. In the Vickrey auction with multiunit objects, a bidder who wins $k^i$ units pays the $k^i$ highest losing bids of the other bidders. The basic principle of the Vickrey auction is the same as the Vickrey-Clarke-Groves mechanism: Each bidder is asked to pay an amount equal to the externality he exerts on other competing bidders.

Suppose bidder $i$ wins $k^i$ units, to get his first unit of objects, it must be true $b^1_i > c^K_{-i}$. The bidder is then asked to pay an amount equal to $c^K_i$ units of externality, where $c^K_i$ is the bid of some other bidder. As such, the bids of bidders only affect the allocation of the objects, but do not affect the payment of the bidders when they win. When the bidder gets the second unit, it must be true that $b^2_i > c^{K-1}_{-i}$, and thus he pays $c^{K-1}_i$ units of externality to other bidders. Thus, if the bidder wins $k^i$ units of objects, he will bring $\sum_{k=1}^{k^i} c^{K-k+1}_{-i}$ units of externality to other bidders.

Thus, under the Vickrey, if bidder $i$ wins $k^i$ units, then the amount that he pays is equal to

$$\sum_{k=1}^{k^i} c^{K-k+1}_{-i}.$$  

In Example 21.5.1, bidder 1 wins two units, $c_{-1} = (60, 55, 48, 47, 45, 35)$, the payment is $45 + 35 = 80$. Bidder 2 gets three units of objects, $c_{-2} = (54, 48, 46, 45, 43, 40)$, and thus the payment is $45 + 43 + 40 = 128$; bidder 3 gets one unit of objects, $c_{-3} = (60, 55, 54, 47, 46, 43)$, and thus the payment equal is 43. The total revenue of the auctioneer is 251.

When $K = 1$, the Vickrey auction is reduced to the second-price auction just like the uniform-price auction. But when $K > 1$, the Vickrey auction is an appro-
appropriate extension of the second-price auction to the case of multiple units. Indeed, as we will show, unlike the uniform price auction, it shares many important properties with the second-price auction while the other auction formats do not. For instance, in the uniform-price auction, the bidder $i$’s payment may depend on his own bid because the market-cleaning price is $p = \max\{b_{i}^{k_{i}+1}, c_{i}^{R-k_{i}+1}\}$, which may be determined by $b_{i}^{k_{i}+1}$.

21.5.3 Simultaneous Open-Bid Auctions for Multiple Identical Objects

In the single-object auction, especially under the private value, each sealed auction has a corresponding open format. For example, the Dutch auction is equivalent to the first-price auction and the English auction is equivalent to the second-price auction. We will have the similar features in multiunit auctions. That is, the Dutch auction is equivalent to the discriminatory auction, the English auction is equivalent to the uniform-price auction, and the Ausubel auction is equivalent to the Vickrey auction in the sense each pari of them result in the same equilibrium outcomes.

Dutch Auction

In the multi-unit Dutch auction (or open descending price), the auctioneer starts the auction by calling out a price that is so high that no bidder wants to accept it. The price then keeps going down until there is a bidder who wants to buy a unit at the current price. The bidder gets one unit and the auction continues. Then, the price continue to go down until there is another bidder who wants to buy a unit at a price. The bidder gets the second unit and the auction continues until the price is reduced to a level at which the bidder wants to buy the $K$th unit.

The multi-unit Dutch auction is outcome equivalent to the discriminatory auction. If the bidder $i$ accepts the price according to his bid vector $b_i$ in the
Dutch auction, he will purchase one unit when the price is reduced to \( p = b_1 \) and buy the second unit when the price is \( p = b_2 \), and so on. But the Dutch auction is not strategically equivalent to the multi-unit extension of the discriminatory auction. This is because a bidder’s value to different units may be interdependent under information asymmetry. By observing one bidder’s willingness to get one unite, other bidders can calculate his willingness to buy other units of objects. But in the discriminatory auction, as the bidders have to report their bids to different amounts of objects at the same time, there is no time for other bidders to adopt their beliefs. So they are not strategically equivalent.

**English Auction**

The process of the multi-unit English auction (or open ascending-price) is opposite to the Dutch auction. The auctioneer changes the price from low to high, each bidder reports the number he wants to buy at each reported price. Usually, the total amount of the objects the bidders want to buy is larger than \( K \) at the initially low price. With the price going up, the bidders will adjust the amounts they want to buy. At a certain price, the amount all the bidders want to buy is equal to \( K \), this price will be the market-clearing price. All the bidders will buy their desired units at this price. Just like the relationship of the Dutch auction and the discriminatory auction, the multi-unit English auction is outcome equivalent to the uniform-price auction, but they are not strategically equivalent.

**Ausubel Auction**

The Ausubel auction is an alternative ascending-price format that is outcome equivalent to the Vickrey auction. Similar to the English auction, the price also changes from low to high. The bidders need to report their demand \( d_i(p) \) at price \( p \), and the demand will decrease as the price goes up. But, being different from the English auction, in an Ausubel auction, the price of the objects is determined by the following procedure. At the beginning price \( p \), if each bidder \( i \)’s demand
$d_i(p)$ is large enough, the residual supply function of each bidder is

$$s_{-i}(p) \equiv \max\{K - \sum_{j \neq i} d_j(p), 0\} = 0.$$  

With the price going up, the demand of each bidder goes down. The price continues to increase until it reaches a level $p'$ such that at least one bidder $i$’s residual supply function satisfies $s_{-i}(p') > 0$. Then sell $s_{-i}(p') > 0$ units to any bidder $i$ for whom $s_{-i}(p') > 0$ at a price of $p'$. The price continues to increase as long as the sale is less than $K$.

When the price rises to $p''$ such that at least one bidder’s residual supply function satisfies $s_{-i}(p'') - s_{-i}(p') > 0$. Then sell $s_{-i}(p'') - s_{-i}(p') > 0$ units to a bidder, say bidder $i$ for whom $s_{-i}(p'') - s_{-i}(p') > 0$ at a price of $p''$. The price continues to go up until the market is cleared, and then the auction is ended.

We still use Example 21.5.1 again to describe the rule of Ausubel auction. Given the strategy of each bidder is unchanged, that is

$$b_1 = (54, 46, 43, 40, 33, 15)$$

$$b_2 = (60, 55, 47, 32, 27, 13)$$

$$b_3 = (48, 45, 35, 24, 14, 9)$$

The surplus supply functions of each bidder are:

$$s_{-1}(p) = \begin{cases} 
6 & \text{if } p > 60; \\
5 & \text{if } 60 \geq p > 55; \\
4 & \text{if } 55 \geq p > 48; \\
3 & \text{if } 48 \geq p > 47; \\
2 & \text{if } 47 \geq p > 45; \\
1 & \text{if } 45 \geq p > 35; \\
0 & \text{if } 35 \geq p \geq 0.
\end{cases}$$

906
\[
s_{-2}(p) = \begin{cases} 
6 & \text{if } p > 54; \\
5 & \text{if } 54 \geq p > 48; \\
4 & \text{if } 48 \geq p > 46; \\
3 & \text{if } 46 \geq p > 45; \\
2 & \text{if } 45 \geq p > 43; \\
1 & \text{if } 43 \geq p > 40; \\
0 & \text{if } 40 \geq p \geq 0.
\end{cases}
\]

\[
s_{-3}(p) = \begin{cases} 
6 & \text{if } p > 60; \\
5 & \text{if } 60 \geq p > 55; \\
4 & \text{if } 55 \geq p > 54; \\
3 & \text{if } 54 \geq p > 47; \\
2 & \text{if } 47 \geq p > 46; \\
1 & \text{if } 46 \geq p > 43; \\
0 & \text{if } 43 \geq p \geq 0.
\end{cases}
\]

Suppose the initial price is \( p^0 \leq 35 \), \( s_{-i}(p^0) = 0 \), \( \forall i \in N = \{1, 2, 3\} \). When the price rises to \( p^1 = 35 + \epsilon \), \( s_{-1}(35 + \epsilon) = 1 \), \( s_{-j}(35 + \epsilon) = 0 \), \( \forall j \neq 1 \). When \( \epsilon \to 0 \), bidder 1 buys the first unit of commodity at price 35.

When the price rises to \( p^2 = 40 + \epsilon \), \( s_{-1}(40 + \epsilon) - s_{-1}(35 + \epsilon) = 0 \), \( s_{-2}(40 + \epsilon) - s_{-2}(35 + \epsilon) = 1 \), \( s_{-3}(40 + \epsilon) - s_{-3}(35 + \epsilon) = 0 \). When \( \epsilon \to 0 \), bidder 2 buys the second unit of commodity at price 40.

When the price rises to \( p^3 = 43 + \epsilon \), \( s_{-1}(43 + \epsilon) - s_{-1}(40 + \epsilon) = 0 \), \( s_{-2}(43 + \epsilon) - s_{-2}(40 + \epsilon) = 1 \), \( s_{-3}(45 + \epsilon) - s_{-3}(40 + \epsilon) = 1 \). When \( \epsilon \to 0 \), bidder 2 buys the third unit and bidder 3 buys the fourth unit at price 45. Total sales add up to 4.

When the price further rises to \( p^3 = 45 + \epsilon \), \( s_{-1}(45 + \epsilon) - s_{-1}(43 + \epsilon) = 1 \), \( s_{-2}(45 + \epsilon) - s_{-2}(43 + \epsilon) = 1 \), \( s_{-3}(45 + \epsilon) - s_{-3}(43 + \epsilon) = 0 \). When \( \epsilon \to 0 \), bidder 1 buys the fifth unit and bidder 2 buys sixth unit at price 45. Total sales amount to 6, and the auction ends.

In the end of this Ausubel auction: bidder 1 gets two units of objects with
a payment equal to $35 + 45 = 80$; bidder 2 gets three units of objects with the payment equal to $40 + 43 + 45 = 128$; bidder 3 gets one unit of objects with the payment equal to 43. The revenue of the auctioneer is $80 + 128 + 43 = 251$, which is the same as the Vickrey auction. Similarly, the Ausubel auction is not strategically equivalent to the Vickrey auction.

In the following we will discuss equilibria of the above six formats of auctions with identical objects and private values. As each sealed auction has a corresponding open form, we will focus on the three sealed auctions. In the single-object auction, when the bidders’ values are private and independently distributed and bidders are risk neutral and have no budget constraints, the symmetric equilibria of both the first-price auction and the second-price auction are efficient. But this conclusion does not hold with the multi-unit auctions. Only the Vickrey auction is efficient, while the discriminatory auction and the uniform-price auction are not.

### 21.5.4 Equilibrium of Multiunit Vickrey Auction

We will discuss the multi-unit Vickrey auction first. Suppose the auctioneer has $K$ units of identical objects and the marginal value of bidder $i$ to different units of objects is $V_i = (V_i^1, \ldots, V_i^K)$, where $V_i^k$ is the marginal value of bidder $i$ to $k$th unit of object that satisfies the law of the diminishing marginal utility, i.e. $V_i^k \geq V_i^{k+1}, \forall k < K$. Suppose the value $V_i$ of bidder $i$ is independent and identically distributed on the value space $\chi = \{v_i \in [0, w]^K : v_i^k \geq v_i^{k+1}, \forall k < K\}$ according to the density function $\varphi(v_i)$. Given the value of the bidder, the true demand function of the bidder is $\delta_i(p)$ and can be defined as:

$$\delta_i(p) = \max\{k : v_i^k \geq p\}.$$  

From the previous subsection, we know the demand function reported by the player is $d_i(p) = \max\{k : b_i^k \geq p\}$. Thus, the true demand equals to the reported demand if and only if the bidder bids according to his true value.
From the precious discussion, we know that, in the multi-unit Vickrey auction, each bidder simultaneously submits a $K$-vector $b_i$ that satisfies $b^k_i \geq b^{k+1}_i$. All the bid vectors of the bidders construct the bid set $\{b^k_i, i \in N, k = 1, \ldots, K\}$, in which the $K$ highest bids win the objects and the bidder $i$ wins $k_i$ units of objects so that $\sum_i k_i = K$. $c_{-i}$ is the $K$-vector of competing bids faced by bidder $i$ that is obtained by rearranging in decreasing order the bids of the bid set $\{b^k_j, j \neq i, j \in N, k = 1, \ldots, K\}$ and choosing the first $K$ of these. In the Vickrey auction, the bidder $i$ getting $k_i$ units of objects means $v^k_{i} > c^{K-k_i+1}_{-i}$ and $v^{k_i+1}_i < c^{K-k_i}_{-i}$. At the same time, the total payment of the bidder $i$ to $k_i$ units of objects is equal to $\sum_{k=1}^{k_i} c^{K+1-k}_{i}$. Now we prove that truth-telling $b_i(v_i) = v_i$ is a weakly dominant strategy in the Vickrey auction.

**Proposition 21.5.1** For any bidder $i$, it is a weakly dominant strategy to bid according to $b_i(v_i) = v_i$ in a multi-unit Vickrey auction.

**Proof.** Let $b_i$ be the truthful pricing strategy of bidder $i$, $c_{-i}$ is a vector of competing bids of any other bidders faced by him. And $k_i$ is the units of objects bidder $i$ wins when he chooses to report truthfully, which means $v^k_{i} > c^{K-k_i+1}_{-i}$ and $v^{k_i+1}_i < c^{K-k_i}_{-i}$ and the total payment is $\sum_{k=1}^{k_i} c^{K-k_i+1}_{-i}$. His expected utility then is

$$\sum_{k=1}^{k_i} [v^k_i - c^{K-k+1}_{-i}].$$

If bidder $i$ were to submit a different bid vector $b'$ and still get the $k_i$ units of objects, his expected utility would be the same. If bidder $i$ were to submit a different bid vector $b'$, but the units of objects he gets were $k'_i \neq k_i$. When $k'_i > k_i$, he gets the $k$ units of goods such that $k'_i \geq k \geq k_i + 1$. His marginal net revenue then is

$$v^k_{i} - c^{K-k+1}_{-i} \leq v^{k_i+1}_i - c^{K-k_i+1}_{-i} < 0.$$ 

As such, his expected utility of bidding $b'$ is lower than the expected utility of bidding $b_i$. If $k'_i < k_i$, he cannot get $k'_i + 1, \ldots, k_i$th unit of object, and his
marginal net revenue of getting \( k < k_i \) th unit of goods is

\[
v^k_i - c^{K-k_i+1} \geq v^{k_i+1}_i - c^{K-k_i+1}_i > 0,
\]

which is unchanged and therefore so would the surplus derived from these. But the surplus from any unit \( k < k_i \) was positive and is now forgone. As such, his expected utility of bidding \( b' \) is lower than the expected utility of bidding \( b_i \). Hence bidder \( i \)'s truthful pricing strategy \( b_i \) is a weakly dominant strategy.

In the Vickrey auction, the payment that bidder \( i \) makes to get each unit of object is equal to the opportunity cost caused by this unit of goods to other bidders. And that is the externality caused by bidder \( i \) to other bidders.

Ausubel and Milgrom (2006) defined the payments of participants in the Vickrey auction by the VCG mechanism. Let \( v_i \) be the truthful value vector of bidder \( i \) to the multiple objects. Let \( \hat{v}_i \) be the reported value vector of bidder \( i \). Let \( k = (k_1, \ldots, k_n) \) be an allocation of the objects and \( k^* \) be the optimal allocation given all the bidders’ value vectors.

\[
k^* = \arg\max_k \sum_i \hat{v}_i(k_i), \text{ s.t. } \sum_i k_i \leq K.
\]

The payment of the bidder who gets \( k_i \) units of objects is

\[
t_i(k_i, \hat{v}) = \sum_{j \neq i} \hat{v}_j(k^*_j) - \bar{v}_{-i},
\]

where \( \bar{v}_{-i} = \max\{\sum_{j \neq i} \hat{v}_j(k_j) : \sum_{j \neq i} k_j \leq K\} \). Hence the Vickery auction is an extension of the VCG mechanism with multiple objects.

In the Vickery auction, as each bidder reports his true value and the objects are given to the \( K \) highest bids, it is an efficient auction.

**Proposition 21.5.2** The Vickrey auction allocates the objects efficiently.

At the same time, Ausubel and Milgrom (2006) showed that in all the identical multi-unit mechanisms, the Vickery auction is the only one where truth-telling
is a dominant strategy and the allocation is efficient and the failing bidders need not to pay.

The conclusions above reveal that the Vickrey auction has nice properties. However, if we are loose about homogeneity, especially when we introduce complementarity, there will be a number of shortcomings in the Vickrey auction which we will discuss in the following sections. One of the shortcomings of the Vickrey auction with identical multiple objects is that the prices of identical objects are not the same. This will cause arbitrage and other problems. The example below showed this that identical objects do not have the same price in the Vickrey auction.

**Example 21.5.2** Suppose there are two units of identical objects, the values of two bidders to them are: \( v_1 = (12, 2) \) and \( v_2 = (8, 1) \) respectively. In the Vickrey auction, each bidder gets one unit of commodity and their payments are 1 and 2 respectively. That is to say, they pay different prices for the same object. If the bidders are two listed companies, such bids and payments will cause external pressure to the company that pays a higher price.

We will discuss the other formats of multi-unit auctions: the uniform-price auction and the discriminatory auction. In these mechanisms, bidders often bid untruthfully and the allocations may not be efficient.

### 21.5.5 Equilibrium of Multiunit Uniform Price Auction

In this subsection we focus on multi-unit uniform-price auction in which there is the feature of underreporting demand.

We have already discussed that given the bid vector \( b = (b_i)_{i \in N} \) in the uniform-price auction, how the price is determined and how the objects are allocated. Suppose there are \( K \) units of identical objects, and the bidder’s \( K \) dimension bid vector is \( b_i \). The (marginal) value vector of the bidder to the object is \( v_i \). Let \( c_{-i} \) be the \( K \)-vector of competing bids of the other bidders faced
by the bidder. If the player $i$ wins $k_i$ units of objects, we must have:

$$b_i^{k_i} > c_{-i}^{K-k_i+1} \text{ and } b_i^{k_i+1} < c_{-i}^{K-k_i}.$$  

The price of the uniform-price auction is then given by:

$$p = \max\{b_i^{k_i+1}, c_{-i}^{K-k_i+1}\}.$$  

We cannot get the explicit equilibrium solution in the uniform-price auction. As such, we try to depict some features of the equilibrium strategy of this auction.

The equilibrium strategy of the uniform-price auction has two important features: one is that the bid will not exceed the marginal value; the other is that the bid for the first unit is equal to its marginal value.

**Fact 21.5.1** $\forall i, k$, $b_i^k \leq v_i^k$; i.e., the bids cannot exceed marginal values.

**Proof.** Suppose by way of contradiction that the bid $b_i$ of bidder $i$ satisfies $b_i^k > v_i^k$. We show that it will be weakly dominated by the bidding strategy that satisfies $b_i^k = v_i^k$, $b_i^{k'} = b_i^{k'}$, $k' \neq k$. We discuss it in four cases. 1) $b_i^k = p$. The bidder $i$ can get $k-1$ units of objects under both bidding strategies. But when he chooses $b_i^{k'} = v_i^k$, the $p$ may decrease. 2) $b_i^k < p$. The two bidding strategies make no difference to bidder $i$. 3) $b_i^k > p > v_i^k$. The bidder will suffer a loss from winning the $k$th unit as the price exceeds the marginal value. But he cannot get the $k$th unit of object if he chooses $b_i'$. 4) $b_i^k > v_i^k > p$. The two bidding strategies make no difference. Hence in the uniform-price auction, the bidder’s bid will not exceed marginal value. ■

**Fact 21.5.2** For all $i$, $b_i^1 = v_i^1$; i.e., the bid on the first unit must be the same as its value.

**Proof.** If the bidder bids the first unit according to his marginal value, that is, $b_i^1 = v_i^1$, then this bid is a weakly dominant strategy. Or we can say $b_i^1 < v_i^1$ is a weakly dominated strategy. We will discuss it in three cases: 1) $p \geq v_i^1 > b_i^1$. The bidder can get nothing under the two strategies, so they make no difference
for him. 2) \( v_i^1 \geq p \geq b_i^1 \). If he bids the first unit truthfully, the bidder \( i \) will win it and get positive revenue. But if he chooses \( b_i^1 \), he can get nothing and no revenue. 3) \( v_i^1 > b_i^1 > p \). The two bidding strategies make no difference to bidder \( i \).

Thus, in the uniform-price auction, the bidder’s strategy satisfies that the bid will not exceed the true value. Despite that the bid to the first unit is equal to its value, the bids to the other units are less than the true value. This phenomenon is called strategic demand reduction.

Let us use a simple example to describe the logic behind the strategic demand reduction of the uniform-price auction. Suppose there are two units of identical objects, that is, \( K = 2 \). Suppose there are two bidders whose marginal value vector \( V_i \) is distributed on the value space \( \chi = \{ v_i \in [0, w]^2 : v_i^1 \geq v_i^2 \} \) according to the density function \( \varphi(.) \) and the distribution function \( \Phi(.) \). Let \( \Phi_1(.) \) and \( \Phi_2(.) \) be the marginal distribution functions of \( V_i^1 \) and \( V_i^2 \) respectively. We will focus on bidder 1’s strategies of uniform-price auction in the symmetric equilibrium (superscript omitted).

Suppose the marginal values of bidder 1 are \( (v_1^1, v_1^2) \equiv (v^1, v^2) \), bids are \( (b_1^1, b_1^2) \equiv (b^1, b^2) \), the competing bids faced by him are \( (b_2^1, b_2^2) = (c_1^1, c_2^1) \equiv (c^1, c^2) \). In the symmetric equilibrium, as \( b_i(v_i) = \beta(v_i) \), \( b_i \) is random and its distribution function is \( \Phi(.) \) (its density function is \( h(.) \)), \( H_1(.) \) and \( H_2(.) \) are the marginal distribution functions of \( b_i^1 \) and \( b_i^2 \) respectively, so they are marginal distribution functions of \( c^1 \) and \( c^2 \). Thus, \( H_1(b_2^1) = \text{Prob}[C^1 < b_2^1] \) is the probability that bidder 1 will defeat both competing bids and win two units of objects. Similarly, \( H_2(b_1^1) = \text{Prob}[C^2 < b_1^1] \) is the probability that he will defeat the lower competing bid, and win at least one unit. The probability that he will win exactly one unit is then the difference \( H_2(b_1^1) - H_1(b_2^1) > 0 \). Also, \( H_2(b_2^1) - H_1(b_2^1) = \text{Prob}[C^2 < b_2^1 < C^1] \) is the probability that the highest losing bid at which the units are sold is \( b_2^1 \).
Using these facts, bidder 1’s expected utility when he bids $b_1 = (b^1, b^2)$ is:

$$Eu_1(b, v) = \int_{\{c : c^1 < b^2\}} (v^1 + v^2 - 2c^1)h(c)dc + \int_{\{c : c^1 > b^2, c^2 > b^1\}} (v^1 - \max\{c^2, b^2\})h(c)dc$$

$$= H_1(b^2)(v^1 + v^2) - 2\int_0^{b^2} c^1 h_1(c^1)dc + [H_2(b^1) - H_1(b^2)]v^1 - [H_2(b^1) - H_1(b^2)]b^2$$

$$- \int_{b^1}^{b^2} c^2 h_2(c^2)dc.$$

We know that the bid for the first unit of object satisfies $b^1 = v^1$. The first-order condition for maximizing $b^2$ results in:

$$\frac{dEu_1}{db^2} = h_1(b^2)(v^2 - b^2) - [H_2(b^2) - H_1(b^2)] = 0.$$

Since $H_2(b^2) - H_1(b^2) > 0$, we have

$$b^2 = v^2 - \frac{H_2(b^2) - H_1(b^2)}{h_1(b^2)} < v^2.$$

In the symmetric equilibrium of the uniform-price auction, there is strategic demand reduction except for the first unit. In a more general case, Ausubel and Cramton (2002) showed that as long as $K > 1$, in the equilibrium of the uniform-price auction, the bid of each bidder satisfies $\beta_{i1}(v^1_i) = v^1_i, \beta_{ik}(v^k_i) < v^k_i, \forall k > 1$.

Milgrom (2004) relates the strategic demand reduction of the uniform-price auction with the properties of the demand of the buyer monopoly. Consider a simple example. Suppose $K = 2$, bidder 1’s demand is unit demand, that is $V^2_1 = 0$. Suppose his value to one unit of commodity is $V^1$, which is distributed on $[0,1]$ according to the uniform distribution. According to the features of the uniform-price auction, we get $b_1(v^1) = v^1$. Bidder 2’s marginal value vectors are $v^1$ and $v^2$ that satisfy $1 > v^1 \geq v^2 > 0$, each of which is distributed on $[0,1]$ according to the uniform distribution. At bidding, he will take $b^1_2 = v^1$ and $b^2_2 = b^2$. If $b^2 > v^1$, the bidder will win two units of objects. If $0 \leq b^2 < v^1$, he will win one unit. So the expected utility of bidder 2 choosing $b^2$ is:

$$Eu_2(b^2, v^1, v^2) = \int_{v^1 < b^2} (v^1 + v^2 - 2v^1)dv^1 + \int_{v^1 > b^2} (v^1 - b^2)dv^1 = v^1 - b^2(1 - v^2).$$
Thus the optimal bidding strategy is $b^2 = 0$. This means that no matter what value $v^2$ takes, bidder 2’s optimal bid for the second unit is equal to zero.

The reason of strategic demand reduction is similar to the logic behind the buyer monopoly market. In a buyer monopoly market, the monopolist pays $TE(q) = qp(q)$ to buy $q$ units of objects, where $p(q)$ is the market supply function. The marginal expenditure of the monopolist buying the $q$th unit is $TE'(q) = p(q) + qp'(q) > p(q)$ (as the market supply function satisfies $p'(q) > 0$). Thus, there are two kinds of effects of buying one more unit of commodity, one is the expenditure $p(q)$ of buying the object, the other is the increased marginal expenditure $qp'(q)$. The latter reduces the demand incentive of the monopolist.

Suppose the value of the monopolist to the object is $v(q)$, the amount of goods purchased to achieve maximum utility is:

$$v'(q) - TE'(q) = v'(q) - p(q) - qp'(q) = 0,$$

which means the demand $q$ of the monopolist satisfies $v'(q) - p(q) > 0$. That is the logic behind the strategic demand reduction.

In the example from Milgrom (2004), when the bid $b_2 = 0$, bidder 2 will get one unit of goods for free. But if he wants to get the second unit, not only does the acquisition cost $v^1$ of the second unit have to be paid, but the cost of the first unit also increases. Then, even if $v^2 > v^1$, bidder 2 does not have the incentive to get the second unit. Thus, multi-unit uniform-price auction is not an efficient mechanism. However, if every bidder is of unit demander, by Fact 21.5.2, all the bidders report truthfully, and uniform-price auction is also efficient.

Summarizing our discussions, we have the following propositions.

**Proposition 21.5.3** When $K > 1$, at a undominated equilibrium of the uniform-price auction, the bid on the first unit is equal to the value of the first unit, but bids on other units are lower than the respective marginal values.

**Proposition 21.5.4** When $K > 1$, every undominated equilibrium outcome of the uniform price auction is inefficient.
21.5.6 Equilibrium of Multiunit Discriminatory Auction

Now we discuss the equilibrium of the discriminatory auction with multiple objects. In this auction, the bidder pays his bids when he wins some units of the objects. Let \( b_i \) be the bid vector of the bidder \( i \). When he wins \( k_i \) units of objects, his total payment is \( \sum_{k=1}^{k_i} b_i^k \). As the general close-form solution is impossible to get, like the situation of the uniform-price auction, we focus on the natures of the symmetric equilibrium here.

To have the nature of equilibrium bids, consider a simple economy with two units of objects and two bidder. Suppose the bidders’ marginal value vector is \( v_i = (v_1^i, v_2^i) \). Consider the symmetric equilibrium \((\beta^1, \beta^2)\), i.e. \( b_i^1 = \beta^1(v_i) \) and \( b_i^2 = \beta^2(v_i) \).

First, if the highest bid on the second unit is \( \max_v \beta^2(v) \), a bidder does not bid more than \( \max_v \beta^2(v) \) on the first unit. This is because if \( b_i^1 \) on the first unit is greater than \( \max_v \beta^2(v) \), the bidder will win with probability 1 and he could do better by reducing it slightly. Indeed, let \( \bar{b} = \max_v \beta^2(v) \). For the bidder \( i \), \( b_i^1 > \bar{b} \) is strictly dominated by \( b_i^1 = b_i^1 - \epsilon > \bar{b} \), where \( \epsilon > 0 \). As \( b_i^2 \leq \bar{b} \), for the bidder \( i \), \( b_i^1 \) and \( b_i^1 \) are both successful bids, but paying \( b_i^1 \) will lower the acquisition cost.

Thus, we have the first nature that for any symmetric equilibrium bidding strategy, it satisfies

\[
\max_v \beta^1(v) = \max_v \beta^2(v).
\]

Second, consider a bidder, say bidder 1, and let \( c_{-1} = (b_2^1, b_2^2) = c_{-1} = (c^1, c^2) \) denote the competing bids of bidder 1, that is, the bids of bidder 2. Similarly, let \( H_1(.) \) and \( H_2(.) \) be the marginal distribution functions of \( c^1 \) and \( c^2 \), respectively.

Thus,

\[
H_k(c) = \text{Prob}[\beta^k(V) \leq c].
\]

Since, for all \( v \), \( \beta^1(v) \geq \beta^2(v) \), it is clear that the distribution \( H_1 \) stochastically dominates the distribution \( H_2 \). As usual, let \( h_1 \) and \( h_2 \) denote the corresponding densities.
Suppose bidder 1 has values \((v_1^1, v_2^1) = (v_1, v_2)\) and bids \((b_1^1, b_1^2) = (b_1, b_2)\). He wins both units if \(c^1 < b_2\) and the probability of this event is \(H_1(b_2)\). But if \(c^2 < b_1\) and \(c^1 > b_2\), he may get one unit with the probability \(H_2(b_1) - H_1(b_2)\). Thus, given bidder 1’s marginal value \(v_1 = (v_1^1, v_2^1) = (v_1, v_2)\) and bid vector \(b_1 = (b_1^1, b_1^2) = (b_1, b_2)\), his expected utility is:

\[
Eu_1(b_1, v_1) = H_1(b_2)(v_1 + v_2 - b_1 - b_2) + [H_2(b_1) - H_1(b_2)](v_1 - b_1)
= H_2(b_1)(v_1 - b_1) + H_1(b_2)(v_2 - b_2).
\]

The bidder’s optimization problem is choose \(b_1\) to maximize \(Eu_1(b_1, v_1)\) subject to the constraint:

\[b_1 \geq b_2.\]

When the constraint does not bind at the optimum, that is \(b_1 > b_2\), solving the first-order conditions with respect to \(b_1, b_2\), we have

\[
h_2(b_1)(v_1 - b_1) = H_2(b_1) \tag{21.17}
\]

\[
h_1(b_1)(v_2 - b_2) = H_1(b_2). \tag{21.18}
\]

Thus, the bids are independent, which means \(\beta_1\) does not depend on \(v_2\), and \(\beta_2\) does not depend on \(v_1\). When the constraint binds at the optimum, that is \(b_1 = b_2 = b\), the first-order conditions are:

\[
h_2(b)(v_1 - b) + h_1(b)(v_2 - b) = H_2(b) + H_1(b).
\]

In this case, the bidder submits a “flat demand” function, i.e. bidding the same amount for each unit. The examination of the first-order conditions above shows that if the bidder submits the bid \(b\) for the value vector \(v = (v_1, v_2)\) and \(v' = (v'_1, v'_2)\), then for any \(\lambda \in [0, 1]\), he will submit the same bid \(b\) for the value \(\lambda v + (1 - \lambda)v'\).

### 21.5.7 Efficient Multiunit Auctions

For the uniform-price auction, we know that it is not efficient except that all the bidders’ demand is unit demand. Is the discriminatory auction efficient? We will
offer a conclusion on how to judge whether a multi-unit auction is efficient before talking about the efficiency of discriminatory auction.

As there are different forms of auctions, we will focus on the standard auctions in which the objects are given to the bidders whose bids are the $K$ highest. For a standard auction, there are $K$ units of objects. Let $\beta = (\beta_1, \ldots, \beta_n)$ be the bidding equilibrium, where $\beta_i = (\beta_1^i, \ldots, \beta_K^i)$ is the bid vector of the bidder $i$. Suppose the bidder $i$’s marginal value vector is $v_i$. If $\beta_k^i(v_i)$ is a winning bid, it must belong to the $K$ highest bids. If the auction is efficient and $v_k^i$ belongs to the $K$ highest marginal value, the bidder $i$ at least wins the $k$th unit of objects. If the auction is efficient, the equilibrium of this auction must satisfy the condition below:

$$\beta_k^i(v_i) > \beta_l^i(v_j) \text{ if and only if } v_k^i > v_l^j.$$  

Then, it has the following features:

1. The bid of any bidder $i$ to the $k$th unit only depends on his marginal value $V_i^k$ to $k$th unit, that is

$$\beta_k^i(v_i) = \beta_k^i(v_i^k).$$

If the bid on the $k$th unit depends on marginal value of the other units, say $v_i^{k'}$, then $v_k^i = v_j^l$ and $v_k^{k'} \neq v_j^{l'}$, implies $\beta_k^i \neq \beta_j^l$. That means that if $v_k^i$ and $v_j^l$ belong to the $K$ highest marginal values of bidders, but $\beta_k^i$ and $\beta_j^l$ might not belong to $K$ highest bids, then this auction is not efficient. Thus, in order for us to have an efficient auction, bidders’ bidding strategies for different units must be independent.

2. To have an efficient auction, the bidding strategy must be the same for any bidder and any object, that is,

$$\beta_k^i(.) = \beta_j^l(.) = \beta(.) \text{, } \forall i, j, k, l.$$ 

Otherwise, we may have $b_k^i \neq b_j^l$ when $v_k^i = v_j^l$, which will make the allocation of the auction inefficient.
In review of the three multi-unit auctions above, the Vickrey auction is efficient as all the bidders report their true value, that is, $b_i^k = \beta_i^k(v_i) = v_i^k$. For the uniform-price auction, suppose that at least one bidder’s demand is not unit demand. Then, each bidder’s bid on the first unit is true, but their bids on the other units are less than the true marginal value, so the uniform-price auction is not efficient.

For the discriminatory multi-unit auction, we consider the bidding strategy equilibrium of two units of object and two bidders. In case of $b_1^1 > b_2^1$, by the first-order condition (21.17) and (21.18), despite the bids on different units are independent, they are not the same bidding strategies unless $\frac{H_2}{h_2} = \frac{H_1}{h_1}$. And if $\beta_1(\cdot) = \beta_2(\cdot)$, that means $H_1(\cdot)$ will first-order stochastically dominate $H_2(\cdot)$, where $\frac{H_2}{h_2} \neq \frac{H_1}{h_1}$. But if $b_1^1 = b_2^1$, then $v_1^1 > v_2^1$, which is not efficient, either. Thus, of all the three static multi-unit auctions, only the Vickrey auction is efficient.

21.5.8 Revenue Equivalence Principle in Multiunit Auction

For single-object auction, if it is private value, independent, symmetric, without external (budget) constraints and with risk neutrality, the standard auctions, such as the first-price sealed auctions, second-price sealed auctions, English auctions, Dutch auctions, and all-pay auctions, are not only efficient but also interim revenue equivalent to the auctioneer. When we extend the single-object auction to the multi-unit auction, as discussed above, even if all these assumptions are satisfied, not all these auctions are efficient. The question here is whether the revenue equivalence principle also holds in a multi-unit auction. We will discuss it in this subsection.

The revenue equivalence principle of single-object auction applies to auctions resulting in the same equilibrium outcome. In the multi-unit auction, the discussion below also applies to the auctions resulting in the same equilibrium allocation. Despite that these auctions may not be efficient, we cannot exclude the
possibility that outcomes of these auctions may be the same in some economic environments. Now we investigate that whether the revenues of two auctions are the same when the allocations of the auctions are identical.

Suppose there are $K$ units of identical objects to be sold to $n > K$ bidders. The set of bidders is $N = \{1, \ldots, n\}$. Each bidder’s marginal value (random) vector is $V_i$ with $V_i \in \chi \equiv \{v \in [0, w]^K : v^k \geq v^{k+1}, \forall k\}$. They are independent but do not need to be identically distributed. Let $\Phi_i(.)$ be the distribution function and $\varphi_i(.)$ be the density function.

We can use the framework of mechanism design to discuss revenue equivalence principle of the multi-unit auction, especially, apply the Bayesian incentive compatibility characterization theorem (Proposition xxx) in Chapter 16 to obtain the revenue equivalence principle. Here, instead, we give a direct proof on the principle. Let $(\beta_1, \ldots, \beta_n)$ be the bidding equilibrium strategy of an auction, say, auction A. In equilibrium, the bidder with $v_i$ will choose $\beta_i(v_i)$ as his strategy. When the bidder $i$ with marginal value $v_i$ report $\hat{v}_i$, which means he chooses $\beta_i(\hat{v}_i)$ as strategy, his probability of getting the $k$th unit of objects is $y^k_i(\hat{v}_i)$, and his transfer is $t_i(\hat{v}_i)$. So bidder $i$’s expected utility is:

$$u_i(\hat{v}_i, v_i) = y_i(\hat{v}_i)v_i + t_i(\hat{v}_i),$$

where $y_i = (y^1_i, \ldots, y^K_i)$ is the probability vector of bidder $i$ getting the objects.

In equilibrium, for any $\hat{v}_i$, we have

$$y_i(v_i)v_i + t_i(v_i) \geq y_i(\hat{v}_i)v_i + t_i(\hat{v}_i). \quad (21.19)$$

Let $U_i(v_i) \equiv u_i(v_i, v_i) = \max_{\hat{v}_i} y_i(\hat{v}_i)v_i + t_i(\hat{v}_i)$. (21.19) can be written as:

$$U_i(v_i) \geq U_i(\hat{v}_i) + y_i(\hat{v}_i)(v_i - \hat{v}_i). \quad (21.20)$$

In (21.20), $y_i(\hat{v}_i)$ can be seen as the subgradient of the convex function $U_i(.)$ at the point $\hat{v}_i$.

Compared to the single-object auction, there are multidimensional private values in the multi-unit auction, i.e. the private marginal value vector. In the
single-object auction, we are able to integrate to obtain \( U_i \). By reducing the multidimensional problem to a single dimension, we can also express \( U_i \) as a form of integration.

For any point \( v_i \) of the \( K \) dimensional space, define a one dimensional function \( W_i : [0, 1] \to \mathbb{R} \):

\[
W_i(s) = U_i(sv_i).
\]

Thus, \( W_i(0) = U_i(0) \) and \( W_i(1) = U_i(v_i) \). As \( U_i(.) \) is a continuous convex function, \( W_i(.) \) is also continuous and convex.

The convex function of one variable is absolutely continuous, and thus it is differentiable in the interior of its domain. Moreover, since any absolutely continuous function is the integral of its derivative, we have:

\[
W_i(1) = W_i(0) + \int_0^1 \frac{dW_i(s)}{ds} ds. \tag{21.21}
\]

By the (21.20), we have

\[
W_i(s + \Delta) - W_i(s) = U_i((s + \Delta)v_i) - U_i(sv_i)) \geq y_i(sv_i)\Delta v_i.
\]

When \( \Delta > 0 \), we get:

\[
\frac{W_i(s + \Delta) - W_i(s)}{\Delta} \geq y_i v_i.
\]

As \( \Delta \downarrow 0 \), we obtain:

\[
\frac{dW_i(s)}{ds} \geq y_i(sv_i)v_i.
\]

When \( \Delta < 0 \), we get:

\[
\frac{dW_i(s)}{ds} \leq y_i(sv_i)v_i.
\]

Since \( W_i(s) \) is differentiable, we have:

\[
\frac{dW_i(s)}{ds} = y_i(sv_i)v_i.
\]

Thus, by (21.21), we have:

\[
U_i(v_i) = U_i(0) + \int_0^1 y_i(sv_i)v_i ds. \tag{21.22}
\]
Hence, at $v_i$, the equilibrium expected utility $U_i(.)$ can be determined by $y_i$, and thus
\[ t_i(y_i) = U_i(y_i) - y_i(v_i)v_i \]
the same as long as outcome $y_i$ is the same. Then, we have the following proposition that summarizes the revenue equivalence in multi-unit auction.

**Theorem 21.5.1** When the expected utilities $U_i(0)$ are the same, the interim expected utility and transfer payment of any bidder in any two multi-unit auctions that have the same allocation rule are the same, and therefore, the revenue of auctioneer is the same for these two auctions.

### 21.5.9 Application of Revenue Equivalence Principle

In the uniform-price auction, if the demands of the bidders are unit demand, every bidder will bid truthfully. Then it is reduced to Vickrey auction. For multi-unit auctions, although the uniform-price auction and discriminatory auction in general are not efficient, they may be efficient in some cases. Usually, solving the equilibrium of the uniform-price auction and discriminatory auction is very hard. But the revenue equivalence principle can be used to derive their equilibrium bidding strategies.

Consider an economic environment that there are three units of identical objects and two bidders, each of whom wants at most two units, which means $V_i^3 = 0, i = 1, 2$. Bidders’ value vector is $V_i = (V_i^1, V_i^2), i = 1, 2$. Suppose that in the support $\chi = \{v \in [0, 1]^2 : v_1 \geq v_2\}$, the marginal valuation functions are identically and independently distributed according to the density function $\varphi(.)$ and the distribution function $\Phi(.)$. Let $\varphi_1(.)$ and $\varphi_2(.)$ be the marginal density function of $V_i^1$ and $V_i^2$ respectively, the distribution functions be $\Phi_1(.)$ and $\Phi_2(.)$ respectively.

We show that the allocation of the uniform-price auction and the discriminatory auction is as efficient as the Vickrey auction.
We will first show that the allocation of the uniform-price auction is efficient. Recall that, for the uniform-price auction, each bidder bids truthfully on the first unit, that is \( b^1_i = v^1_i \), \( i = 1, 2 \). As there are three units of identical objects and two bidders, each of whom wants at most two units, each bidder wins at least one unit of object. If there exists a symmetric and increasing equilibrium bidding strategy \( \beta^2(.) \) on the second unit, then \( \beta^2(v^2_i) > \beta^2(v^2_j) \), \( i \neq j \) which means \( v^2_i > v^2_j \), so the bidder \( i \) will win the second unit. Suppose that the equilibrium bidding strategy is symmetric and increasing. Then for any \((v^1_1, v^2_1)\) and \((v^1_2, v^2_2)\), if \( v^2_1 > v^2_2 \), bidder 1 will win two units and bidder 2 will win one unit. The opposite is similar. Obviously, now the uniform-price function is efficient.

Then we show the allocation of the discriminatory auction is efficient. Once again, each bidder is assured of winning at least one unit. Suppose the bid vector of one bidder is \((b^1, b^2)\) such that \( b^1 > b^2 \). Reducing the bid on the first unit to \( b^1 - \epsilon > b^2 \) does not affect a bidder winning the first unit (as each bidder will get at least one unit). This does not affect the bidder winning the second unit either, but will reduce his payoff. As such, there must be \( b^1 = b^2 \) at the equilibrium, which means each bidder will report a “flat demand”. Moreover, the equilibrium bidding strategy is determined only by the marginal value \( v^2 \) of the second unit.

In the discriminatory auction, if the bid on the second unit is an increasing function \( \beta^2 : [0, 1] \rightarrow \mathbb{R}_+ \), each bidder will choose the bid vector \((\beta^2(v^2)^i, \beta^2(v^2)^j)\) in symmetric equilibrium. As each bidder gets at least one unit, \( \beta^2(v^2)^i > \beta^2(v^2)^j \) implies \( v^2_i > v^2_j \), and bidder \( i \) will get two units. Obviously, the discriminatory auction is efficient.

Next, we will use the revenue equivalence principle to derive the symmetric equilibrium strategy on the second unit for the uniform-price auction and the discriminatory auction.

In the three auction formats above, the bidder with \( v_i = 0 \) will not win the object, and \( U_i(0) = 0 \). In the case above, the three auctions are efficient. Thus the conditions of the revenue equivalence principle are satisfied.

First, we calculate the expected revenue in the Vickrey auction. As each
bidder will choose the true bids $b_i^k = v_i^k$, $i = 1, 2, k = 1, 2$ in the Vickrey auction, the competing bid vector faced by bidder $i$ is $c_i = (v_j^1, v_j^2, 0)$. If $v_i^2 > v_j^2$, bidder $i$ will win two units and pay $-t_i = 0 + v_j^2$. If $v_i^2 < v_j^2$, bidder $i$ will win one unit and pay $-t_i = 0$, then the expected payment of bidder $i$ with value $v_i$ is:

$$-Et_i^v(v_i) = \int_0^{v_i^2} v_2 \varphi_2(v_2) dv_2. \quad (21.23)$$

We then want to solve the symmetric equilibrium bidding strategy $\beta^{u2}(v_i^2)$ of the uniform-price auction on the second unit. In the equilibrium, if $v_i^2 > v_j^2$, bidder $i$ wins two units and pays $2\beta^{u2}(v_j^2)$. If $v_i^2 < v_j^2$, bidder $i$ wins one unit and pays $\beta^{u2}(v_i^2)$. So in the uniform-price auction, the expected payment of the bidder $i$ with value $v_i$ is:

$$-Et_i^u(v_i) = \int_0^{v_i^2} 2\beta^{u2}(v_2)\varphi_2(v_2) dv_2 + (1 - \Phi_2(v_i^2))\beta^{u2}(v_i^2). \quad (21.24)$$

According to the revenue equivalence principle, $Et_i^v(v_i) = Et_i^u(v_i)$, that is:

$$\int_0^{v_i^2} v_2 \varphi_2(v_2) dv_2 = \int_0^{v_i^2} 2\beta^{u2}(v_2)\varphi_2(v_2) dv_2 + (1 - \Phi_2(v_i^2))\beta^{u2}(v_i^2).$$

Differentiating with respect to $v_i^2$, we have:

$$v_i^2 \varphi_2(v_i^2) = \beta^2(v_i^2)\varphi_2(v_i^2) + (1 - \Phi_2(v_i^2))\beta^{u2}(v_i^2),$$

and thus we obtain the differential equation:

$$\beta^{u2}(v_i^2) = (v_i^2 - \beta^2(v_i^2))\lambda_2(v_i^2),$$

where

$$\lambda_2(v_i^2) = \frac{\varphi_2(v_i^2)}{1 - \Phi_2(v_i^2)}.$$

Then, by the initial condition $\beta^2(0) = 0$, the symmetric equilibrium bidding strategy on the second unit is:

$$\beta^{u2}(v_i^2) = \int_0^{v_i^2} v_2 \lambda_2(v_2) dL(v_2|v_i^2),$$

where $L(v_2|v_i^2) = \exp(-\int_{v_2}^{v_i^2} \lambda_2(s) ds)$. 924
Obviously $\beta^{u2}(.)$ is an increasing function. In the uniform-price auction, the equilibrium bidding strategy of the bidder whose marginal value is $v_i$ is $(v_i^1, \beta^{u2}(v_i^2))$.

In the end, we now solve the symmetric equilibrium bidding strategy of the discriminatory auction on the second unit. We need to show $b^1_i = b^2_i$. Let $\beta^{d2}(.)$ be the symmetric equilibrium strategy. Then, in the equilibrium, if $v_i^2 \geq v_j^2$, bidder $i$ will win two units and pay $2\beta^{d2}(v_i^2)$. If $v_i^2 < v_j^2$, bidder $i$ will win one unit and pay $\beta^{d2}(v_i^2)$. Hence the expected payment of the bidder whose value is $v_i$ is:

$$Et^d_i(v_i) = \beta^{d2}(v_i^2) + \Phi^{d2}(v_i^2). \quad (21.25)$$

According to the revenue equivalence principle, we have

$$\int_0^{v_i^2} v_2\varphi_2(v_2)dv_2 = \beta^{d2}(v_i^2) + \Phi^{d2}(v_i^2).$$

Then, the equilibrium bidding strategy on the second unit is:

$$\beta^{d2}(v_i^2) = \frac{1}{1 + \Phi^{d2}(v_i^2)} \int_0^{v_i^2} v_2\varphi_2(v_2)dv_2.$$ 

It is easy to verify that $\beta^{d2}(.)$ is an increasing function. Thus, in the discriminatory auction, the equilibrium bidding strategy of the bidder whose marginal value is $v_i$ is $(\beta^{d2}(v_i^2), \beta^{d2}(v_i^2))$.

Through the example above, we find that by using the revenue equivalence principle, we can easily obtain the equilibrium bidding strategy of some auctions that are hard to solve. But we have to ensure the decision rule (allocations) of the two auctions are the same when we use this method.

### 21.6 Sequential Private Value Auctions for Multiple Identical Objects

In this section we study the sequential auctions of multiple identical objects. A sequential auction is an auction in which the units are sold one at a time in
separate auctions that are conducted sequentially till all objects are sold. In particular, we mainly consider two formats of auctions: the first-price sealed-bid auctions and the second-price sealed-bid auctions. In these auctions, at every unit of objects, all bidders simultaneously submit sealed bids, and then the price at which it is sold – the winning bid – is announced.

The sequential auction of multiple objects is very different from the simultaneous multiunit auction discussed earlier, in which all the units are sold at one time, while in the former situation the units are sold one at a time in separate auctions that are conducted sequentially. Sequential auctions bring some new strategic issues where bidders need to choose different bidding strategies at different stages, and this dynamic consideration can make the analysis of the bidder’s strategy quite complicated. For the sake of discussion, we focus our attention on the simplest situations in which each bidder has a single-unit demand. That is, for any bidder $i$, we have $V^k_i = 0, \forall k \geq 2$.

We first discuss the sequential first-price sealed-bid auctions, and then the sequential second-price sealed-bid auctions.

### 21.6.1 The Sequential First-Price Sealed-bid Auctions

Consider an economic environment in which $K$ identical items are sold to $n > K$ bidders, and each bidder has a single-unit demand. We assume that bidder $i$ has private value $V_i$ and that each bidder’s value $V_i$ is drawn independently from the same distribution $\Phi(\cdot)$, corresponding density $\phi_i(\cdot)$ on $[0, \omega]$. A particular bidder may be still active in the $K$ round auction (if he wins an object, he will stop bidding). So a bidding strategy for a bidder consists of $K$ functions, denoted by $\beta^{I1}(\cdot), \cdots, \beta^{IK}(\cdot)$, with the superscript $I$ standing for the first-price auction, and the subscript $k$ representing the strategy of the $k$ round auction.

We look for a symmetric equilibrium that is an equilibrium where all bidders use the same strategy. As in each round, there is different information, in $k \leq K$ round, the bidding strategy is $\beta^{Ik}(v; p^1, \cdots, p^{k-1})$, where $p^1, \cdots, p^{k-1}$ were the prices in the previous $k - 1$ auctions respectively. Moreover, the sequential
auctions are assumed to be held in a short time, so we disregard the time discount issue.

If the equilibrium strategies $\beta^k(v; \cdot)$ for $k \leq K$ are increasing functions of $v$, the first item will go to the bidder with the highest value, and the $k$th item to the bidder with the $k$th highest value. In this sense, the sequential auction mechanism is efficient. Before discussing the sequential auction of $K$ units of objects, we begin by considering the simplest situation in which only two units are sold so as to get some basic characteristics of equilibrium bidding strategy in the sequential first price auction.

The Sequential Auctions for Two Units of Identical Objects

We begin by looking at an economy in which only two units are sold. Let $(\beta^1, \beta^2)$ (the superscript $I$ is omitted here) be the symmetric equilibrium bidding strategy. In the first round, the bidding strategy is $\beta^1(\cdot) : [0, w] \rightarrow \mathbb{R}^+$. If it is a strictly increasing function, there is an inverse function, which is the price paid in the first auction, $p^1$.

Before the start of the second round of bidding, the bidder can infer that the private value of the bidder who won the auction in the first round according to the inverse function $(\beta^1)^{-1}$ at $p^1$. As it is a symmetric equilibrium, we can simply consider only the bidding strategy of bidder 1. Denote by $w_1 \equiv w_1^{(n-1)}$ the highest of $n - 1$ stochastic variables $V_2, \ldots, V_n$, and by $w_2 \equiv w_2^{(n-1)}$ the second-highest of $n - 1$ stochastic variables $V_2, \ldots, V_n$. We assign them as $w_1$ and $w_2$.

Let $\psi_1$ and $\psi_2$ be the corresponding densities, and $\Psi_1$ and $\Psi_2$ be the distributions of $W_1 \equiv W_1^{(n-1)}$ and $W_2 \equiv W_2^{(n-1)}$, respectively, where $\Psi_1 = \Phi(n-1)(\cdot)$. We assume that the considered concept of bidding equilibrium is a subgame perfect equilibrium, which requires starting from the first bidding price and then the second round of bidding strategy should be a Baysian-Nash equilibrium. Below we start with the analysis of the second round of auction.

In the second round of auction, suppose the auction price of the first period is $p^1$. If bidder 1 does not get the auctioned goods in the first round, he can
infer that \( w_1 = (\beta^1)^{-1}(p^1) \). Suppose all the other bidders follow the equilibrium strategy \( \beta^2(\cdot; w_1) \). Since \( W_2 < w_1, \beta^2(\cdot; w_1) \) is an increasing function, it makes no sense for bidder 1 to bid an amount greater than \( \beta^2(w_1; w_1) \). Noting that the private value of bidder 1 is \( v_1 \), if he bids \( \beta^2(\hat{v}_1; w_1) \) where \( \hat{v}_1 \leq w_1 \), his interim expected utility in the second auction is:

\[
Eu_1(\hat{v}_1, v_1; w_1) = \Psi_2(\hat{v}_1|w_1)[v_1 - \beta^2(\hat{v}_1; w_1)].
\]

Differentiating with respect to \( \hat{v}_1 \), we obtain the first-order condition in equilibrium:

\[
\psi_2(v_1|w_1)[v_1 - \beta^2(v_1; w_1)] - \Psi_2(v_1|w_1)\beta^2(v_1; w_1) = 0,
\]

Rearranging this results in the differential equation and then we have

\[
\beta^2(v_1; w_1) = \frac{\psi_2(v_1|w_1)}{\Psi_2(v_1|w_1)}[v_1 - \beta^2(v_1; w_1)], \tag{21.26}
\]

where the initial value condition satisfies \( \beta^2(0; w_1) = 0 \).

Given the winner of the first round, the highest of \( n - 1 \) values, except that of bidder 1, equals \( w_1 \). The probability that bidder 1 will win the second auction, if he chooses \( \hat{v}_1 \), is equivalent to \( \text{prob}(W_1^{(n-2)} < \hat{v}_1) \), where the \( W_1^{(n-2)} \) is the private value of the second-highest of \( n - 2 \) values, except that of bidder 1 and the winner at the first round. Actually, \( W_2^{(n-1)} \equiv W_1^{(n-2)} \).

\[
\Psi_2(v_1|w_1) = \Psi_1^{(n-2)}(v_1|W_1^{(n-2)} < w_1) = \frac{\Phi(v_1)^{n-2}}{\Phi(w_1)^{n-2}}, \tag{21.27}
\]

and thus we have

\[
\psi_2(v_1|w_1) = \frac{(n-2)\Phi(v_1)^{n-1}\phi(v_1)}{\Phi(w_1)^{n-2}}. \tag{21.28}
\]

Useing (21.27) and (21.28), the differential equation (21.26) can be written as

\[
\beta^2(v_1; w_1) = \frac{(n-2)\phi(v_1)}{\Phi(v_1)^n}[v_1 - \beta^2(v_1; w_1)],
\]

or equivalently,

\[
\frac{\partial}{\partial v_1}(\Phi(v_1)^{n-2}\beta^2(v_1; w_1)) = (n-2)\Phi(v_1)^{n-3}\phi(v_1)v_1.
\]
Together with the initial conditions, the solution to the differential equation is:

\[ \beta^2(v_1) = \frac{1}{\Phi(v_1)^{n-2}} \int_0^{v_1} vd(\Phi(v)^{n-2}) = E[W_1^{(n-2)}|W_1^{(n-2)} < v_1] = E[W_2|W_2 < v_1 < W_1]. \tag{21.29} \]

Thus, for bidder 1, the complete bidding strategy for the second period is to bid \( \beta^2(v_1) \) following (21.29) if \( v_1 \leq w_1 \) and to bid \( \beta^2(w_1) \) if \( v_1 > w_1 \). In case of \( v_1 > w_1 \), if bidder 1 chooses an equilibrium strategy in the first auction, he should win the auction in the first round. Even though this represents off-equilibrium behavior on the part of bidder 1 himself, as per the concept of a strategy, we must count in this possibility.

Now we discuss the first-round equilibrium strategy \( \beta^1(\cdot) \). Again let us assume that the private value of bidder 1 is \( v_1 \) and that all the other bidders follow the first-period strategy \( \beta^1(\cdot) \). Further, suppose that all bidders, including bidder 1, will follow the equilibrium bidding strategy (21.29) in the second period. We now study the conditions that the equilibrium calls on bidder 1 to bid \( \beta^1(\cdot) \) in the first stage.

In equilibrium, bidder 1 should bid \( \beta^1(v_1) \) in the first stage, but consider what is his expected utility if he decides to bid \( \beta^1(\hat{v}_1) \) instead. If \( \hat{v}_1 \geq v_1 \), he wins the first auction when \( \hat{v}_1 > W_1 \). But when \( \hat{v}_1 < W_1 \), he can not get the item in the first round, and he will enter the second round of bidding. When \( W_2 < v_1 \leq \hat{v}_1 < W_1 \), he loses the first auction but wins the second auction and cannot win the auction in any other cases. The expected utility of choosing \( \beta^1(\hat{v}_1) \) is:

\[ Eu_1(\hat{v}_1, v_1) = \Psi_1(\hat{v}_1)[v_1 - \beta^1(\hat{v}_1)] + (n - 1)(1 - \Phi(\hat{v}_1))\Phi(v_1)^{n-2}[v_1 - \beta^1(v_1)]. \tag{21.30} \]

On the other hand, if \( \hat{v}_1 < v_1 \), when \( \hat{v}_1 > W_1 \), he wins the first auction with a bid. When \( \hat{v}_1 < W_1 \), if \( W_2 < v_1 < W_1 \), he loses the first auction but wins
the second at the price of $\beta^2(v_1)$. If $\hat{v}_1 < W_1 < v_1$, he will pay $\beta(W_1)$, and his expected utility is:

$$Eu_1(\hat{v}_1, v_1) = \Psi_1(\hat{v}_1)[v_1 - \beta^1(\hat{v}_1)] + [\Psi_2(v_1) - \Psi_1(v_1)][v_1 - \beta^2(v_1)]$$
$$+ \int_{\hat{v}_1}^{v_1} [v_1 - \beta^2(w_1)]\psi_1(w_1)dw_1. \tag{21.31}$$

Differentiating equations (21.30) and (21.31) with respect to $\hat{v}_1$, we obtain the first-order conditions in the two cases

$$0 = \psi_1(\hat{v}_1)[v_1 - \beta^1(\hat{v}_1)] - \Psi_1(\hat{v}_1)\beta^1(\hat{v}_1) - (n - 1)\phi(\hat{v}_1)\Phi(v_1)^{n-2}[v_1 - \beta^1(v_1)], \tag{21.32}$$

$$0 = \psi_1(\hat{v}_1)[v_1 - \beta^1(\hat{v}_1)] - \Psi_1(\hat{v}_1)\beta^1(\hat{v}_1) - \psi_1(\hat{v}_1)[v_1 - \beta^1(v_1)]. \tag{21.33}$$

Since $\Psi_1(v) = \Phi(v)^{n-1}$, $\psi_1(v) = (n - 1)\phi(v)\Phi(v)^{n-2}$. Therefore, the two equations (21.32) and (21.33) are the same. In equilibrium it is optimal to bid $\beta^1(\cdot)$ and setting $\hat{v}_1 = v_1$ in either first-order condition results in the differential equations ((21.32) and (21.33)). From the equation (21.32) and the equation (21.33), we have

$$\beta^1(v_1) = \frac{\psi_1(v_1)}{\Psi_1(v_1)}[\beta^2(v_1) - \beta^1(v_1)]. \tag{21.34}$$

Together with the initial value condition $\beta^1(0) = 0$, we can get the solution to the above differential equation:

$$\beta^1(v_1) = \frac{1}{\Psi_1} \int_0^{v_1} \beta^2(w_1)\psi_1(w_1)dw_1$$
$$= E[\beta^2(W_1)|W_1 < v_1]$$
$$= E[E[W_2|W_2 < W_1]|W_1 < v_1]$$
$$= E[W_2|W_1 < v_1]. \tag{21.35}$$

Thus, together with the equations (21.35) and (21.29), we obtain the following proposition:
**Proposition 21.6.1** Suppose all the bidders have a single-unit demand, the symmetric equilibrium strategies for the sequential first-price sealed-bid auction with two units are

\[ \beta^I(v) = E[W_2|W_1 < v], \]
\[ \beta^{I2}(v) = E[W_2|W_2 < v < W_1], \]

where \( W_1 \equiv W_1^{(n-1)} \) is the highest, and \( W_2 \equiv W_2^{(n-1)} \) is the second highest, of independent and stochastic variables in \( n - 1 \) identical distributions.

Obviously \( \beta^I(v_1) \) and \( \beta^{I2}(v_1) \) are strictly increasing functions in \( v_1 \). In the following we give two features of the dynamic bidding equilibrium for the case of \( K = 2 \). First, from (21.35), we have

\[ \beta^I(v_1) = E[\beta^2(W_1)|W_1 < v_1], \]

which means that if bidder 1 with \( v_1 \) is the highest value participant, whether he chooses to win the item in the first round or in the second round, his interim expected payment will be the same. To be precise, the (random) dynamic payment price is martingale (i.e., the conditional expected value of the next observation, given all the past observations, is equal to the most recent observation). At this point, the bidder with the highest value has no incentive to delay winning the item, thus equilibrium results can prevent strategic delays.

Second, from the (21.34) and the fact that \( \beta^I(\cdot) \) is a strictly increasing function, we have:

\[ \beta^2(v_1) - \beta^I(v_1) > 0, \]

which means that the bidding strategy in the second round will be more active than in the first round. This is because, after the second round, the bidder will no longer have the opportunity to win the auction, the bidding in the latter stages will be more intense. If we extend \( K = 2 \) to a more general case, the above two features will still exist.
The Sequential Auctions for \( K \) Units of Identical Objects

Consider a more general economy in which the \( K > 1 \) identical items are sold to \( n > K \) bidders. Assume that \( K \) units of identical objects are sold in a sequence of first-price sealed-bid auctions. Also, assume that, at round \( k \), the prices in the first \( k - 1 \) auctions are \( p_1, \ldots, p_{k-1} \). Let \( W_k = W_k^{(n-1)} \) be the \( k \)th highest independent and stochastically drawn variables from \( n-1 \) identical distributions, \( \Psi_k(\cdot) \) be the distribution of \( W_k \), and \( \psi_k(\cdot) \) be the corresponding density. We will derive symmetric equilibrium bidding strategies \( (\beta_1, \ldots, \beta_K) \), and similarly, only the strategies of bidder 1 are considered.

We now derive symmetric equilibrium bidding strategies by backward from the last auction. So first consider the \( K \)th auction conducted in the last period. Using the similar inference process as before, the equilibrium bidding strategy \( \beta_K \) in the last period is

\[
\beta^K(v_1) = E[W_K|W_K < v_1 < W_{K-1}].
\] (21.36)

We then consider the auction for the \( k < K \) round. Again let us take the perspective of bidder 1. Suppose that all bidders will follow symmetric equilibrium bidding strategies \( \beta^{k+1}, \ldots, \beta^K \) in the subsequent auctions. When all other bidders choose the bidding strategy \( \beta^K \) in the round \( k \), we examine the strategy choice of bidder 1. Suppose all the other bidders choose the equilibrium bidding strategy. If \( v_1 > W_{(n-1)}^1 \), bidder 1 wins the first round auction. If \( W_{(n-1)}^{k-1} < v_1 < W_{(n-1)}^k, k \leq K \), he wins the \( k \)-th round auction. If \( v_1 < W_{(n-1)}^K \), he can not get the round item in equilibrium. Meanwhile, if \( W_{(n-1)}^k > v_1, k \leq K \), he wins the \( k \)th round auction with a bid of \( W_{(n-1)}^k \). If \( W_{(n-1)}^k < v_1, k \leq K - 1 \), he wins the \( k+1 \)th round auction with a bid of \( W_{(n-1)}^k \). In addition, other bidders can not get the item. Therefore, if the symmetric equilibrium bidding strategy \((\beta^1, \ldots, \beta^K)\) is a strictly increasing function, the sequential auction is efficient.

Besides, in the first \( k - 1 \) round auctions, if the prices are \( p_1, \ldots, p_{k-1} \), the bidders can infer that the bidders’ private value of winning the first \( k - 1 \) round auctions is: \( w_1 = (\beta^1)^{-1}(p_1), \ldots, w_{k-1} = (\beta^{k-1})^{-1}(p^{k-1}) \).
The equilibrium calls on bidder 1 with private value $v_1$ to bid $\beta^k(v_1)$ in the $k$th stage but consider what happens if he decides to bid slightly higher, say, $\beta^k(v_1 + \Delta)$. (His choice of a slightly lower tender price $\beta^k(v_1 - \Delta)$ is similar. For the previous case $K = 2$, in the analysis of the first round of bids, we know that the first-order conditions are the same regardless of whether the bid price is higher or lower.) If $v_1 > W_k \equiv W_k^{(n-1)}$, he gets the item in the $k$ round and the expected payment increases by

$$\Psi_k(v_1|w_{k-1})[\beta^k(v_1 + \Delta) - \beta^k(v_1)].$$

If $v_1 < W_k < v_1 + \Delta$, he would have lost the $k$th auction with his equilibrium bidding strategy whereas bidding higher $\beta^k(v_1 + \Delta)$ results in his winning. Now there are two subcases.

Case 1. When $W_{k+1} < v_1 < W_k < v_1 + \Delta$, in equilibrium, he would have won the $k + 1$st auction.

Case 2. When $v_1 < W_{k+1} < W_k < v_1 + \Delta$, however, in equilibrium, he would have lost both the $k$th and the $k + 1$st auctions, and possibly won a later auction, for $l > k + 1$. When $\Delta$ is small, however, the probability of $v_1 < W_{k+1} < W_k < v_1 + \Delta$ is very small – it is of second order in magnitude (proportional to $\Delta^2$). The probability of $W_{k+1} < v_1 < W_k < v_1 + \Delta$ is very small – it is of first order in magnitude (proportional to $\Delta$). Therefore, when $\Delta$ is very small, the main effect is case 1, when bidder 1’s bidding strategy changes from $\beta^k(v_1)$ to $\beta^k(v_1 + \Delta)$, his expected utility increases approximately by:

$$[\Psi_k(v_1 \pm \Delta|w_{k-1}) - \Psi_k(v_1|w_{k-1})][(v_1 - \beta^k(v_1)) - (v_1 - \beta^{k+1}(v_1))].$$

Thus the total expected change in utility is approximately:

$$\Psi_k(v_1|w_{k-1})[\beta^k(v_1 + \Delta) - \beta^k(v_1)]
+ [\Psi_k(v_1 + \Delta|w_{k-1}) - \Psi_k(v_1|w_{k-1})] \times [(v_1 - \beta^k(v_1)) - (v_1 - \beta^{k+1}(v_1))]. \quad (21.37)$$

Equation (21.37), divided by $\Delta$, and taking the limit as $\Delta \to 0$, we obtain

933
the differential equation:

\[
\beta^k(v_1) = \frac{\psi_k(v_1|w_{k-1})}{\Psi_k(v_1|w_{k-1})}[\beta^{k+1}(v_1) - \beta^k(v_1)], \tag{21.38}
\]

where

\[
\Psi_k(v_1|w_{k-1}) = \Psi^{n-k}_1(v_1|w_{k-1}) = \frac{\Phi(v_1)^{n-k}}{\Phi(w_{k-1})^{n-k}},
\]

because of \(W^{(n-1)}_k = W^{(n-k)}_1\).

Together with the initial condition \(\beta^k(0) = 0\), we can obtain the solution to the differential equation (21.38):

\[
\beta^k(v_1) = \frac{1}{\Phi^{n-k}(v_1)} \int_0^{v_1} \beta^{k+1}(v)d(\Phi^{n-k}(v))
= E[\beta^{k+1}(W_k^{(n-k)})|W_k^{(n-k)} < v_1]
= E[\beta^{k+1}(W_k)|W_k < v_1 < W_{k-1}]. \tag{21.39}
\]

The equation (21.39) is a first-order recursion equation, given the bidding equilibrium equation of the last round (21.36), we can obtain the equilibrium bidding strategy in the \(K - 1\)th auction:

\[
\beta^{K-1}(v_1) = E[\beta^K(W_{K-1})|W_{K-1} < v_1 < W_{K-2}]
= E[E[W_K|W_{K} < W_{K-1}]|W_{K-1} < v_1 < W_{K-2}]
= E[W_K|W_{K-1} < v_1 < W_{K-2}]. \tag{21.40}
\]

By backward from the last auction proceeding inductively in this fashion results in the equilibrium bidding strategy for all \(k < K\) rounds:

\[
\beta^k(v_1) = E[\beta^{k+1}(W_k)|W_k < v_1 < W_{k-1}]
= E[E[W_{K}|W_{K+1} < W_k]|W_k < v_1 < W_{k-1}]
= E[W_{K}|W_k < v_1 < W_{k-1}]. \tag{21.41}
\]

Let \(W_0 = \infty\), we get a symmetrical equilibrium of sequential first-price sealed-bid auction mechanism.
Proposition 21.6.2. Suppose that $K$ identical items are sold to $n > K$ bidders by sequential first-price sealed-bid auctions. Suppose each bidder has a single-unit demand and private values. Then, symmetric equilibrium strategies $(\beta^{1}(v), \cdots, \beta^{K}(v))$ are given by

$$
\beta^{ik}(v) = E[W_{K}|W_{k} < v < W_{k-1}], k = 1, \cdots, K,
$$

where $W_{k} \equiv W_{k}^{(n-1)}$ is the $k$th highest of independent and stochastically drawn variables from $n - 1$ identical distributions.

Again, two features of the sequential auctions are worth noting. First, as the auction goes on, the bidding prices become more and more active, that is, $\beta^{k}(v) > \beta^{m}(v)$. It can be shown that $\beta^{Ik}(v)$ is an increasing function and can also be induced by the equation (21.38).

Secondly, the equilibrium price path is a martingale. Suppose that in equilibrium, bidder 1 with value $v_{1}$ wins the $k < K$th round auction. Then, there must be that $W_{K} < \cdots < W_{k} < v_{1} < W_{k-1} < \cdots < W_{1}$. We know that the (equilibrium) realized price in period $k$ is $p^{k} = \beta^{k}(v_{1})$. Moreover, the price in the $k + 1$th round is a random variable $P^{k+1} = \beta^{k+1}W_{k}$ and from (21.41), we have:

$$
E(P^{k+1}|p^{k}) = E[\beta^{k+1}(W_{k})|W_{k} < v_{1} < W_{k-1}]
$$

$$
= \beta^{k}(v_{1})
$$

$$
= p^{k}.
$$

This establishes that the price path in sequential auctions is a martingale. The logic behind this conclusion is very simple: if $E(P^{k+1}|p^{k}) \neq p^{k}$, for example, $E(P^{k+1}|p^{k}) < p^{k}$, the bidder who wins the auction in the $k$ round would be unwilling to use the equilibrium strategy of $\beta^{k}$, as the expected auction price will be lower and the expected utility for his winning the auction in the next round will be greater. In the model, we assume that the bidder is risk neutral, but the conclusion will be similar even for risk aversion.
21.6.2 The Sequential Second-Price Sealed-bid Auctions

In this subsection we discuss another kind of sequential auctions: the items are sold in a series of second-price sealed-bid auctions. If there is a symmetric increasing bidding strategy \((\beta_{II1}, \ldots, \beta_{IIK})\) in the sequential second-price sealed-bid auctions, then such second-price sealed-bid auctions will also be efficient. This means that the sequential first-price and second price auctions are revenue equivalent. We will find the equilibrium strategies by making use of this revenue equivalence principle.

Specifically, let \(m_I(v)\) and \(m_{II}(v)\) denote the (interim) expected payments by a bidder with value \(v\) in \(k\) sequential first- and second-price forms, respectively. Now define \(m^k_I(v)\) to be the expected payment made in the \(k\)th auction by a bidder when the items are sold by means of the first-price sealed-bid auctions. So \(m_I(v) = \sum_{k=1}^{K} m^k_I(v)\). Define \(m_{II}(v) = \sum_{k=1}^{K} m^k_{II}(v)\) in an analogous fashion for the sequential second-price sealed-bid auctions. We first use the revenue equivalence principle to get a strong version of revenue equivalence:

\[
m^k_I(v) = m^k_{II}(v), \forall k \in \{1, \cdots, K\}. \tag{21.42}
\]

Equation (21.42) implies that each round of auction satisfies the revenue equivalence principle. We verify by induction, starting with the \(K\)th auction. Prior to the last auction, as the the equilibrium strategies of two auction mechanisms are monotonous, the remaining \(n - K + 1\) bidders can infer the private value of the previous winning bidders. For instance, due to symmetry, we assume that bidder 1 knows his own value \(v_1 = v\), while his competitors have values \(W_K, \cdots, W_{n-1}\). At the round \(K - 1\), the winner's value is \(W_{k-1} = w_{k-1}\). The revenue equivalence principle implies that \(m_I(v) = m_{II}(v)\). Since bidders with a private value of \(v\) does not win the auction in the previous \(K - 1\) rounds, \(m^k_I(v) = m^k_{II}(v) = 0, \forall k \leq K - 1\), so we have:

\[
m^K_I(v) = m^K_{II}(v).
\]

Now consider the start of auction \(K - 1\). The remaining \(n - K + 2\) bidders bid the remaining items. Suppose that bidder 1 knows his own value \(v_1 = v\) and his
competitors have values $W_{K-1}, \ldots, W_{n-1}$. At the round of $K - 2$, the winner’s value is $W_{k-2} = w_{k-2}$. Once again, the revenue equivalence principle implies that

$$m_{I}^{K-1}(v) + m_{I}^{K}(v) = m_{II}^{K-1}(v) + m_{II}^{K}(v).$$

And since $m_{I}^{K}(v) = m_{II}^{K}(v)$, we have $m_{I}^{K-1}(v) = m_{II}^{K-1}(v)$. Proceeding inductively in this way we will get:

$$m_{I}^{k}(v) = m_{II}^{k}(v), \forall k \in \{1, \ldots, K\}.$$  

With this strong version of revenue equivalence principle, we are ready to find the equilibrium bidding strategies of sequential second-price auctions.

Now notice that for the previous $K - 1$ rounds, bidder 1 does not win the auction. He wins the $K$th auction with a private value of $v_1 = v$, then apparently in the $K$th auction his (weakly dominant) bidding strategy must be:

$$\beta^{IIK}(v) = v.$$  

Suppose that in the $k < K$ round, the bidder wins the auction, then at this time there must be:

$$W_K < \cdots < W_k < v < W_{k-1} < \cdots < W_1.$$  

In the second-price auctions, his payment is $\beta^{IIk}(W_k)$, and for the bidder with a private value of $x$, his expected payment in the $k$ round is:

$$m_{I}^{k}(v) = \text{Prob}[W_k < v < W_{k-1}] \times E[\beta^{IIk}(W_k) | W_k < v < W_{k-1}].$$

On the other hand, in the sequential first-price auctions, the payment for a winning bidder with a private value of $v$ in the $k$ round is:

$$m_{I}^{k}(v) = \text{Prob}[W_k < v < W_{k-1}] \times \beta^{Ik}(v)$$

$$= \text{Prob}[W_k < v < W_{k-1}] \times E[\beta^{I(k+1)}(W_k) | W_k < v < W_{k-1}], \quad (21.43)$$

where the second line of the equation (21.43) comes from $\beta^{Ik}(v) = E[\beta^{I(k+1)}(W_k) | W_k < v < W_{k-1})]$. According to the revenue equivalence results for each period we mentioned above, that is, $m_{I}^{k}(v) = m_{II}^{k}(v)$, we have

$$E[\beta^{IIk}(W_k) | W_k < v < W_{k-1}] = E[\beta^{I(k+1)}(W_k) | W_k < v < W_{k-1}].$$
Differentiating both sides of the equality with respect to \( v \) results in the identity:

\[
\beta^{II}(v) = \beta^{I(k+1)}(v).
\]

Thus, we have

\[
\beta^{II}(v) = E[W_K | W_{k+1} < v < W_k].
\]

Obviously, \( \beta^{II}(\cdot) \) is an increasing function. Summarizing the above discussion, we have the following proposition.

**Proposition 21.6.3** Suppose that \( K \) identical items are sold to \( n > K \) bidders by sequential second-price sealed-bid auctions. Suppose each bidder has a single-unit demand and private values. Then, symmetric equilibrium strategies are given by

\[
\beta^{II_K}(v) = v
\]

and for all \( k < K \),

\[
\beta^{II_k}(v) = \beta^{I(k+1)}(v)
\]

where \( \beta^{I(k+1)}(v) \) is the \( k+1 \)st period equilibrium bidding strategy in the sequential first-price auction format, derived in Proposition 21.6.2.

Comparing the strategies for the sequential first-price auctions with the strategies for the sequential second-price auctions, we will find that due to \( \beta^{I(k+1)}(v) > \beta^{Ik}(v) \), we have \( \beta^{II_k}(v) > \beta^{Ik}(v) \). That is, every bidder bids more actively in sequential second-price auctions than in sequential first-price auctions. In addition, as with sequential first-price auctions, the equilibrium price process in a sequential second-price auction is also a martingale. Suppose the equilibrium price process in a sequential second-price auction is \( P^{III}, \ldots, P^{IIK} \). Since \( \beta^{II_k}(v) = \beta^{I(k+1)}(v) \), we only need to verify that the prices between the last two periods have the property of a martingale.

Suppose bidder 1 with a value of \( v \) wins in round \( K - 1 \), so we have \( W_{k_1} < v < W_{k-2} \). Thus, the auction (random) price of the \( K - 1 \) round is \( P^{K-1} = \beta^{II(K-1)}(W_{K-1}) \). Let the realized price be \( p^{K-1} = \beta^{II(K-1)}(w_{K-1}) \), where \( w_{K-1} \)
is the realized value of $W_{K-1}$. In the last round the bidder with a value of $w_{K-1}$ will win and his payment will be $P^K = \beta^{\Pi K}(W_K) = W_K$. Now we have:

$$E[P^{\Pi K}|p^k] = E[W_K|W_K < w_{K-1}] = \beta^{\Pi (K-1)}(w_{K-1}) = p^{K-1}.$$ 

Thus, the (random) equilibrium price process in a sequential second-price auction is also a martingale. No participant would like to win the auction through a strategic delay.

### 21.7 Combinatorial Private Value Auctions for Heterogeneous Objects

We have so far only discussed homogeneous objects in considering multi-item auctions. In reality, objects are often heterogeneous: they can be substitutable, complementary, or sometimes not related, and the relevance of items to different bidders may also vary. Thus, there is a new problem with multi-item auctions, where the auction of different combinations of items creates new complications. For example, if there is complementarity between two items, the bidder would be cautious about the bidding if it was anticipated that the complementary items would not be obtained in the auction, which is known as exposure problem. Many auction mechanisms are dedicated to solving such incentive issues in multi-item auctions. As the issue of combinatorial auctions of heterogeneous objects will be more complex, we mainly introduce some basic conclusions in this section. We focus on the case of private values.
21.7.1 The Basic Model

Let $K = \{a, b, c, \cdots\}$ be a finite set of distinct objects to be auctioned and $N = \{1, 2, \cdots, n\}$ be the set of bidders. Bidder $i$'s private value of combination $S \subseteq K$ is $v_i(S)$, and bidder $i$'s value vector of all combinations is $v_i = (v_i(S))_{S \subseteq K}$. Suppose $v_i(\emptyset) = 0$ and $v_i(S) \leq v_i(T), \forall S \subseteq T \subset K$. The set of all value vectors for bidder $i$ is $\chi_i$, which is a nonnegative closed convex set with $0 \in \chi_i$.

For bidder $i$, there may be some specific relationship between objects, such as substitutability or complementarity.

**Definition 21.7.1 (Substitutable Multiple Items)** Objects are said to be **substitutable** for bidder $i$ if for any $a \in K, a \notin T, S \subseteq T$,

$$v_i(S \cup a) - v_i(S) \geq v_i(T \cup a) - v_i(T), \forall v_i \in \chi_i. \quad (21.44)$$

It can be shown that (21.44) is equivalent to:

$$v_i(S) + v_i(T) \geq v_i(S \cup T) + v_i(S \cap T), \forall v_i \in \chi_i. \quad (21.45)$$

If (21.45) holds, it implies bidder $i$’s value has **submodular**. If $S \cap T = \emptyset$, the inequality (21.45) reduces to

$$v_i(S) + v_i(T) \geq v_i(S \cup T), \forall v_i \in \chi_i. \quad (21.46)$$

In this case, we say that bidder $i$’s value is **subadditive**.

The following defines the complementarity of objects.

**Definition 21.7.2 (Complementary Multiple Items)** Objects are said to be **complementary** for bidder $i$ if for any $a \in K, a \notin T, S \subseteq T$, we have

$$v_i(S \cup a) - v_i(S) \leq v_i(T \cup a) - v_i(T), \forall v_i \in \chi_i. \quad (21.47)$$

Then analogously, bidder $i$’s value has **supermodular** if it satisfies

$$v_i(S) + v_i(T) \leq v_i(S \cup T) + v_i(S \cap T), \forall v_i \in \chi_i. \quad (21.48)$$
Moreover, bidder \( i \)'s value is said to be superadditive, if it satisfies:

\[
v_i(S) + v_i(T) \leq v_i(S \cup T), \forall v_i \in \chi_i, S \cap T = \emptyset.
\] (21.49)

In multi-item auctions, let \( y_i \in K \) denote the set of items obtained by bidder \( i \) in allocation \( y \). The allocation \( y = (y_1, \ldots, y_n) \) is feasible, if \( y_i \cap y_j = \emptyset, \forall i \neq j \), and \( \bigcup_i y_i = K \). A feasible allocation is actually a partition of the set of all objects \( K \). Let \( Y \) denote the set of all feasible allocations.

Next we use the framework of mechanism design to discuss multi-item auctions. Suppose the value vector of all bidders is \( v = (v^1, \ldots, v^n) \). Define the allocation rules of an auction mechanism as \( y(v) = y_1(v), \ldots, y_n(v) \), where \( y_i(v) \) is the set of items obtained by participant \( i \) at a value of \( v \). We consider the direct mechanism, where each bidder reports his value \( \hat{v}_i, i \in N \). Let \( y_i(\hat{v}_i) = (y_i(S, \hat{v}_i)_{S \subseteq K}) \), where \( y_i(S, \hat{v}_i) \) is the probability that bidder \( i \) wins the combination \( S \) when reporting his values of \( \hat{v}_i \). \( y_i(\hat{v}_i) \) is the probability vector that bidder \( i \) wins the item combinations, the payment of which is \( m_i(\hat{v}_i) \) or in term of transfer, it is \( -t_i(\hat{v}_i) \).

A direct mechanism \( (y(\cdot), t(\cdot)) \) is said to be incentive compatible, if it satisfies

\[
U_i(v^1) = y_i(v_i).v_i + t_i(v_i)
= \max_{\hat{v}_i} y_i(\hat{v}_i).v^1 + t_i(\hat{v}_i).
\]

Similarly, we have the following revenue equivalence theorem for combinatorial auctions in the following.

**Theorem 21.7.1** Suppose that the allocation rule \( y(\cdot) \) of two auctions are the same, and every bidder’s expected utility is zero at the lowest value. Then the revenue of the auctioneer is the same.

The proof is similar to that of Proposition 21.5.1. The only difference is that reported value vectors are of size \( 2^{|K|} \) instead of size \( K \), thus we have the following equation:

\[
U_i(v_i) = U_i(0) + \int_0^1 y_i(s v_i) v_i ds. \tag{21.50}
\]
21.7.2 A Benchmark Combinatorial Auction: VCG Mechanism

In the following we discuss a benchmark mechanism of combinatorial auctions, i.e., Vickrey-Clarke-Groves (VCG) auction mechanism. In private value settings, similar to the cases of single item and homogeneous multiple-object auction, the heterogeneous multiple-object combinatorial VCG auction is also an efficient mechanism.

An allocation rule is efficient if for any value vector \( v \) of any combination of participants, the allocation \( y(v) \) maximizes social welfare, that is,

\[
y(v) \in \arg\max_{y_1, \ldots, y_n} \sum_{i \in N} v_i(y_i).
\] (21.51)

Given the value vector \( v \), the social welfare from an allocation \( y(v) \) is

\[
W(v) = \sum_{i \in N} v_i(y_i(v)).
\] (21.52)

The social welfare of individuals other than \( i \) from an allocation \( y(v) \) is

\[
W_{-i}(v) = \sum_{j \neq i} v_j(y_j(v)).
\] (21.53)

Now we define the VGG mechanism of combinatorial auctions, which is an efficient mechanism. Given the value vector \( v \), and the allocation rule \( y(v) \) as defined by (21.51), the transfer rule is

\[
t_i(v_i, v_{-i}) = \sum_{j \neq i} v_j(y(v)) - \sum_{j \neq i} v_j(y(v_{-i})) = W_{-i}(v) - W_{-i}(v_{-i}).
\] (21.54)

In fact, from Chapter 18, this mechanism is also called the pivotal or Clark mechanism, which is a special case of a general VCG mechanism. We can interpret the transfers of the VCG mechanism defined by the formula (21.54) as the externality that bidder \( i \) exerts on the other bidders. When bidder \( i \) does not bid, the social allocation rule is \( y(v_{-i}) \). When bidder \( i \) bids, the social allocation rule
is \( y(v) \), while the change in others’ welfare is characterized by (21.54). Like the usual VCG mechanism, the transfer payment \( t_i(v_i, v_{-i}) \) of bidder \( i \) is independent of \( v_i \). Thus, truth-telling is a weakly dominant strategy in the VCG mechanism. The following proposition describes the VCG mechanism under the combinatorial auctions.

**Proposition 21.7.1** Under combinatorial auction VCG mechanism, truth-telling is a weakly dominant strategy to every bidder. The VCG mechanism is an efficient auction mechanism when each bidder truly reports his own value type.

### 21.7.3 Properties of VCG Mechanism

VCG mechanism is an efficient combinatorial auction mechanism, under which it is a weakly dominant strategy of each bidder to report his true value. The VCG mechanism can strengthen the robustness of the conclusion of efficiency, as the efficiency of VCG mechanism does not depend on the distribution of the value types of participants so that it is ex post implementable. As we mentioned in Chapter 18, Green and Laffont (1979) and Holmstrom (1979) showed that, under relatively weak assumptions, VCG mechanism is the unique weakly dominant mechanism with truthful reporting, efficient outcomes, and zero utility by losing bidders who do not win the object (see Proposition 18.7.3). However, this conclusion needs one extra assumption on the bidder’s value type, that is, smooth path connection. This assumption that is weaker than differentiability used in the proof of Proposition 18.7.3 has been used in the proof of revenue equivalence earlier. Here’s a formal definition.

**Definition 21.7.3** The set of bidder’s value functions is said to be **smoothly path connected** if, for any two functions \( v_i(\cdot), \hat{v}_i(\cdot) \in \chi_i, \forall i \in N \), there is a family of values \( v_i(\cdot, s) \in \chi_i, s \in [0, 1] \) with a parameter \( s \), such that \( v_i(\cdot, 0) = v_i(\cdot), v_i(\cdot, 1) = \hat{v}_i(\cdot), \) and \( v_i(\cdot, s) \) is differentiable in \( s \) and the derivative satisfies

\[
\int_0^1 \sup_{S \subseteq K} \left| \frac{\partial v_i(S, s)}{\partial s} \right| < \infty.
\]
Under the above conditions, Ausubel and Milgrom (2006) proved the uniqueness of the VCG mechanism.

**Proposition 21.7.2** If the set of all value functions is smoothly path connected and $0 \in \chi_i, \forall i \in N$, then the VCG mechanism is the unique direct revelation mechanism in which truthful reporting is a weakly dominant strategy equilibrium, the resulting equilibrium outcomes are always efficient, and there is no transfer payment to losing bidders.

**Proof.** Given any value combination for the bidders besides bidder 1 and consider any mechanism satisfying the above assumptions. If bidder 1 reports $\hat{v}_i = 0$, then his VCG allocation is zero and his transfer payment is also zero. Suppose that bidder 1 reports some value $v_1(\cdot, 1)$ and let $v_1(\cdot, 0) = 0, \{v_1(\cdot, s)|s \in [0, 1]\}$ be a family of smooth value functions. Denote the socially optimal allocation when bidder 1 reports $v_1(\cdot, s')$ by $y(s)$, and let $U_1(s) = \max_{s'}\{v_1(y_1(s'), s) - t_i(s')\}$, where $y_i(s') = y(v_i(\cdot, s'), v_{-i})$ and $t_i(s') = t_i(v_i(\cdot, s'), v_{-i})$ is the allocation and transfer payment received by bidder $i$ under the VCG mechanism. By the envelope theorem in integral form (Milgrom and Segal (2002)),

$$U(1) - U(0) = \int_0^1 \frac{\partial v_1(y(s), s)}{\partial s} ds.$$  

Let $\hat{t}_i(s)$ be the transfer payment made under any direct revelation mechanism for which truthful reporting is a dominant strategy equilibrium, the resulting equilibrium outcomes are always efficient, and there are no payments by or to losing bidders. Let $\hat{U}_1(s) = \max_{s'}\{v_1(y_1(s'), s)+\hat{t}_i(s')\}$. By the envelope theorem,

$$\hat{U}_1(1) - \hat{U}_1(0) = \int_0^1 \frac{\partial v_1(y(s), s)}{\partial s} ds = U_1(1) - U_1(0).$$

Since $U(0) = \hat{U}(0) = 0$, we have

$$v_1(y(1), 1) + t_i(1) = U_1(1) = [\hat{U}_1(1) = v_1(y(1), 1) + \hat{t}_i(1),$$

Hence, $t_i(1) = \hat{t}_i(1)$, so the VCG mechanism is the direct revelation mechanism for which truthful reporting is a dominant strategy, the outcomes are efficient, and there are no payments by or to losing bidders. ■
Another important property of the VCG mechanism is that when the solution concept of implementation is Bayesian-Nash equilibrium, combined with the previous revenue equivalence theorem 21.7.1, we can obtain that the expected revenues under VCG mechanism are not less than that from any other efficient (static) combinatorial auction mechanism that satisfies incentive compatibility.

21.7.4 Defects of Combinatorial Auction VCG Mechanism

Despite the above advantageous properties of VCG mechanism, it is seldom adopted in practical multiple objects auctions. There are many reasons for this. First, VCG mechanism is complicated. For instance, the US Federal Communications Commission’s auction of spectrum No.31 in 1994 contained 12 different frequency spectrums and bidders were required to report a value of possible combinations of $2^{12} = 4,096$. This will not only entail a significant computational burden to the auctioneer, but as the number of items to be auctioned increases, the calculations required for the auction show an exponential growth.

Second, the VCG mechanism involves a considerable amount of private information, and such information disclosure may affect future bidding. Bidders may rationally be reluctant to report their true values, fearing that the information they revealed will later be used against them.

Third, the VCG mechanism sometimes may trigger controversy because two bidders may pay different prices for identical items. In addition to these issues mentioned above, this subsection also discusses VCG mechanism from the perspective of auction revenues: excessively low seller’s revenues, non-monotonicity of the seller’s revenues in the set of bidders and the bids, vulnerability to collusion by a coalition of bidders and vulnerability to shill bidding. This subsection mainly refers to Ausubel and Milgrom (2002, 2006).

Next we discuss these defects of combinatorial VCG mechanism by a series of examples.
Example 21.7.1  Consider an auction of two items \{a, b\} to three bidders. The values of the three bidders are:

\[ v_1 = (v_a^1, v_b^1, v_{\{a,b\}}^1) = (0, 0, 2), \]
\[ v_2 = (v_a^2, v_b^2, v_{\{a,b\}}^2) = (2, 2, 2), \]
\[ v_3 = (v_a^3, v_b^3, v_{\{a,b\}}^3) = (2, 2, 2), \]

and \( V_i^\emptyset = 0, \forall i \). It is easy to see that these two items are complements to bidder 1 but substitutes to bidders 2 and 3. At this point, there are two efficient allocations: one is \((y_1, y_2, y_3) = (\emptyset, \{a\}, \{b\})\); the other is \((y_1, y_2, y_3) = (\emptyset, \{b\}, \{a\})\).

In the VCG mechanism, the transfers of bidders 2 and 3 both zero regardless of which allocation is. The reason is that:

\[ t_2(v) = W - 2(v) - W - 2(v - 2) = 2 - 2 = 0. \]

The same conclusion applies symmetrically to \( t_3(v) = 0 \). The total auction revenues are zero in the multi-item VCG mechanism.

Notice that, in this example, if the auctioneer sells the two items as bound sale, and every bidder values the combined items \{a, b\} as 2, then the seller revenues are 2. The multi-item VCG mechanism brings excessively low seller revenues.

In this example, we also find that for multi-item VCG mechanism, the seller’s revenue is not an increasing function of the number of bidders.

Example 21.7.2  The following example is a variant of Example 21.7.1. Suppose bidder 3 does not participate in bidding, that is, only bidders 1 and 2 participate. Also consider the VCG mechanism where there are four possible efficient allocations and corresponding VCG transfer payments:

1. \((y_1, y_2) = (\{a, b\}, \emptyset), \) and \( t_1 = -2, t_2 = 0; \)
2. \((y_1, y_2) = (\{a\}, \{b\}), \) and \( t_1 = 0, t_2 = -2; \)
3. \((y_1, y_2) = (\{b\}, \{a\}), \) and \( t_1 = 0, t_2 = -2; \)
4. \((y_1, y_2) = (\emptyset, \{a, b\}), \) and \( t_1 = 0, t_2 = -2. \)

Under the VCG mechanism where only bidder 1 and bidder 2 participate, the seller’s revenue is 2. The more bidders at this time may actually bring less auction revenues.
In addition, the multi-item VCG mechanism is prone to bidder collusion. The following example depicts the VCG mechanism’s collusion incentive.

**Example 21.7.3** Consider a variant of Example 21.7.1. Suppose that the value of bidder 1 is unchanged, but bidders 2 and 3’s value of the item is only 0.5. That is, the values of the three bidders of the item are:

\[
v_1 = (v_1^a, v_1^b, v_1^{a,b}) = (0, 0, 2),
\]
\[
v_2 = (v_2^a, v_2^b, v_2^{a,b}) = (0.5, 0.5, 0.5),
\]
\[
v_3 = (v_3^a, v_3^b, v_3^{a,b}) = (0.5, 0.5, 0.5).
\]

Then the only efficient allocation is \((y_1, y_2, y_3) = (\{a, b\}, \emptyset, \emptyset)\). In the VCG mechanism, \(t_1 = -1, t_2 = t_3 = 0\), and the auction revenue is 1.

However, if bidders 2 and 3 collude, and the values they report are \(\tilde{v}_2 = \tilde{v}_3 = (2, 2, 2)\), according to the analysis of Example 21.7.1, the auction allocation becomes \((\emptyset, \{a\}, \{b\})\) or \((\emptyset, \{b\}, \{a\})\), and \(t_2 = t_3 = 0\). Thus, the VCG mechanism is not efficient under collusion, and the auctioneer’s revenue turns into zero again.

At the same time, the multi-item VCG mechanism is not a mechanism that can prevent shill bidding. The so-called “shill” in the auction mechanism means that the bidder participates in bidding by introducing a dummy bidder to obtain higher profits. Here, the dummy bidder is called the shill of the bidder. The example below shows that in multi-item VCG mechanism, there will be a dummy bidder.

**Example 21.7.4** Consider a variant of Example 21.7.3. Suppose there are two bidders for the auction of two items \(\{a, b\}\). Bidder 1 has the same value as Example 21.7.3, and bidder 2 can be considered as the sum of bidder 2 and and bidder 3 in 21.7.3. That is, the values of bidders 1 and 2 are

\[
v_1 = (v_1^a, v_1^b, v_1^{a,b}) = (0, 0, 2),
\]
\[
v_2 = (v_2^a, v_2^b, v_2^{a,b}) = (1, 1, 1).
\]
In the VCG mechanism, the allocation rule is \((y_1, y_2) = (\{a, b\}, \emptyset), t_1 = -1, t_2 = 0\), and the auction revenue is 1.

With a dummy bidder 3, suppose the values of bidder 2 and dummy bidder 3 are reported as \(\hat{v}_2 = \hat{v}_3 = (2, 2, 2)\). As in the case of Example 21.7.3, the auction allocation outcome become \((\emptyset, \{a\}, \{b\})\) or \((\emptyset, \{b\}, \{a\})\) under VCG mechanism, and \(t_2 = t_3 = 0\). Bidders can obtain higher profits through dummy bidder 3, but the auction allocation is no longer efficient, and at the same time the auctioneer’s revenue has been reduced to zero.

In addition, the VCG auction mechanism may prevent some efficient decisions. The following example reveals that the VCG mechanism will inhibit the adjustment of efficient organizational structure.

**Example 21.7.5** This example is a variant of Example 21.7.1. The values of three bidders are the same as 21.7.1. We can consider bidders 2 and 3 as two firms. If the two firms merge and assume the merged firm as bidder 4, and then suppose bidder 4 has the value of \(v_4 = (v_4^a, v_4^b, v_4^{a,b}) = (4 + x, 4 + x, 4 + x)\). \(x\) can be seen as the benefits of the merger, it is clear that as long as \(x > 0\), the merger can produce efficiency improvement (to the value of the item). However, if the two firms merge, the merged firm gets the item in the VCG mechanism and needs to pay 2 at the same time. The merged firm’s net income is \(2 + x\). If two firms do not merge, bidders 2 and 3 can get a total of 4 net gains. Thus, if and only if \(x \geq 2\), the two firms will have an incentive to merge. This shows that VCG mechanism may distort the merger decision before the auction.

Note that in the five examples above, bidder 1’s value of the auctioned item is not substitutable. If the value of bidder 1 is also substitutable as that of bidders 2 and 3, e.g., the value of bidder 1 is changed to:

\[v_1 = (v_1^a, v_1^b, v_1^{a,b}) = (1, 1, 2),\]

the above discussion of VCG defects may change. In Example 21.7.1, we can verify that the auctioneer can get revenue 2 under the VCG mechanism. In the
meantime, we can verify that the auctioneer’s revenue is an increasing function of the number of bidders in the cases from 21.7.2~ to 21.7.5, and the bidders have no conspiracy nor incentive to find a dummy bidder and will not distort the pre-auction merger decision. Thus, the substitutability of auction items is very important and affects the operational efficiency of the auction mechanism. When the assumption of substitutability of bidder’s value of auction items is no longer satisfied, we need to introduce other multi-item auctions. Ausubel and Milgrom (2002) devised a series of ascending combinatorial auctions. When the substitutability is satisfied, the ascending combinatorial auctions are similar to the VCG mechanism, and when the substitutability assumption is not satisfied, they have better properties.

21.7.5 Mechanism of Ascending Combinatorial Auctions

This subsection focuses on the ascending combinatorial auction mechanism of Ausubel and Milgrom (2002). We first introduce this mechanism: ascending combinatorial bidding proceeds in multiple rounds, and in each round, bidders place bids in terms of a certain amount of money, say(ɛ), (i.e. the bids are discrete); in any round, say s, each bidder can place a bid \( b_i(s) \) on any package \( S \subseteq K \), \( b_i(s) \) is the bids of bidder \( i \) in the round \( s \), and within a round, all bids are placed simultaneously. At the end of a round, the auctioneer designates a set of provisionally winning bids that maximize the total revenue and the allocation corresponding to the set of winning bids: given that the bids of all bidders in the round \( s \) is \( b \), the allocation is \( y^* = (y_1^*, \ldots, y_n^*) \) which corresponds to the winning bids designated by the auctioneer, \( y^* \) is feasible allocation, and,

\[
y^* = \arg\max_y \sum_{i \in N} b_i^*(y_i).
\] (21.55)

Then if \( y_i^* \neq \emptyset \), bidder \( i \) will be the provisionally winning bidder in the round \( s \). As long as there is a higher bid on a certain package in round \( s + 1 \), i.e. there is \( b_i^{s+1}(S) > \max_{j \in N} b_j^s(S) \), the auctioneer will re-select a new set of provisionally winning bids and its corresponding allocation in round \( s + 1 \). If no higher bids on
any package appear in round $s^* + 1$, the auction comes to an end in the round $s^*$.

Let $y_{s^*}$ denote the allocation that corresponds to the set of winning bids, then if

\[ u_{i}^{s^*} = v_i(y_{i}^{s^*}) - b_{i}^{s^*}(y_{i}^{s^*}). \]

If $i$ is not the winner, his utility is zero.

The ascending combinatorial auction mechanism of Ausubel and Milgrom (2002) uses proxy bidding, being a direct mechanism in which each bidder submits a value vector to a proxy agent who then bids in the bidder’s interest. The goal of the ascending proxy auction mechanism of Ausubel and Milgrom (2002) is to ensure that each bidder has no incentive to lie to the proxy agent. The following is a discussion of the operation of the ascending proxy auction mechanism: suppose the true value of the bidder $i$ is $v_i$ and the value reported to the proxy agent is $\hat{v}_i$.

1. In round 0, the auctioneer sets an initial price vector as the bidding price vector $b_0$; at the same time, let $b_0^i(S) = b_0^0(S), \forall i \in N, S \subseteq K$, the provisional winner of this round be the seller, and $y_0^0 = K, y_i^0 = \emptyset, \forall i \in N$. Here superscript 0 denotes the seller. Define $\hat{N} = N \cup \{0\}$, i.e. the set of all the bidders and the sellers.

2. Let $b_{s-1}$ be bidder’s bidding price vector in round $s - 1$. The minimum bid that $i$ can place in round $s$ is $b_i^s(S) = b_i^{s-1}(S) + \epsilon$ if $S \neq S_{s-1}^i$; otherwise, $b_i^s(S) = b_i^{s-1}(S)$. The $\epsilon$ here is the minimum bidding increment required if the bidder wants to change the provisional allocation of the previous round.

The proxy $i$ determines the optimal package $\hat{y}_i$ based on $\hat{y}_i$ given by

\[ \hat{y}_i = \arg\max_S v_i(S) - b_i^s(S). \]

Meanwhile, the proxy $i$ places a bid of $\beta_i^s(S|\hat{v}_i)$, that is

\[
\beta_i^s(S|\hat{v}_i) = \begin{cases} 
  b_i^s(S) & \text{if } S = \hat{y}_i \text{ and } \hat{v}_i(S) \geq b_i^s; \\
  b_i^{s-1}(S) & \text{otherwise}
\end{cases}
\]
At the same time, at the end of the round, the auctioneer chooses a set of provisionally winning bids that maximize the total revenue, that is, (21.55), and the allocation $y_t$ that corresponds to the set of winning bids.

If no proxy places a new bid in the next round following $s^*$, the winner of round $s^*$ will pay for the package according to the bidding price placed by his proxy. If next round a proxy places a new bid, then repeat (2) until the auction ends. As the bid price is ascending, the auction will end in finite rounds. It is noted that in some cases, some auction items are held by the seller unless the auctioneer sets the initial bid price vector as zero.

When the auction items are substitutable for each other, the ascending proxy auction mechanism has the same result as the VCG mechanism. Before the formal proof of this conclusion, we need to introduce a definition and a lemma.

Given a value vector $v$ and a coalition of all the buyers and seller $I \subset N$, define

$$w(I) = \max_{y} \sum_{i \in I} V_i(y_i).$$

(21.56)

to be the maximum surplus attainable by the coalition $I$ from an optimal allocation of all the objects.

**Definition 21.7.4** The coalitional value function is bidder-submodular if for all $\{0\} \in I \subseteq J \subseteq \tilde{N}$ and $i \notin J$, and all coalitions satisfy

$$w(I \cup i) - w(I) \geq w(J \cup i) - w(J).$$

Describing the coalitional value function as bidder-submodular means that the surplus marginal contribution of each individual to the coalition decreases as the set of the coalition gets larger. In the definition, symbol $\{0\}$ denotes the auctioneer. If the coalition does not include the buyer, it is clear that the value of the coalition is zero.

The property of coalitional value function in the auction is closely related to the relationship between the goods, and the following is a lemma proved by Ausubel and Milgrom (2002).
Lemma 21.7.1 If the items are substitutes for all the bidders, the value function of coalition is submodular.

The prove of the lemma is relatively complicated, and readers interested can refer to Ausubel and Milgrom (2002).

Next, we use the coalitional value function to characterize the expected utility of the bidder in a VCG auction. Given the bidder’s value of items as $v$, based on the definition of $W(\cdot)$, i.e. the equation (21.52), we can get: $w(\tilde{N}) = W(v)$ and $w(\tilde{N} \setminus i) = W(v_{-i})$. By the definition of transfer payment in the VCG mechanism, that is, the equation (21.54), the expected utility of bidder $i \in N$ under the VCG mechanism can be written as

$$\bar{u}_i = w(\tilde{N}) - w(\tilde{N} \setminus i).$$ (21.57)

Ausubel and Milgrom (2002) showed that the result of ascending proxy auction is equivalent to that of the VCG mechanism when the items are substitutes.

The main results of this subsection are discussed below. As discussed earlier, the so-called ex post Nash equilibrium is such a strategic combination that when all participants in the game know the private information of other participant types, the participants will not change their strategic choices. In the incomplete information game, compared with Bayesian–Nash equilibrium, the ex post Nash equilibrium is a more robust equilibrium that does not depend on the distribution of participant types. Obviously, the ex post Nash equilibrium is Bayesian–Nash equilibrium, but the reverse is not necessarily true.

We then have the following the proposition. To prove it, we first need to establish that truthful reporting in the ascending proxy auction will lead to VCG outcomes, and then show that truthful reporting is indeed an ex post equilibrium.

**Proposition 21.7.3** When the items to be auctioned are substitutes for all bidders, truthful reporting to the proxy is an ex post equilibrium of the ascending proxy auction, in which the equilibrium outcome is the same as in the VCG mechanism.
Proof. Suppose all bidders report their values truthfully to the proxy, we claim that the utility of the ascending proxy auction is the same as in the VCG mechanism. Let \( s^* \) be the round where the ascending proxy auction ends, and the utility obtained by each bidder is \( u^*_i \). We claim that for any \( i \), there is \( u^*_i \geq \bar{u}_i \) (up to \( \epsilon \)). Suppose by way of contradiction that in some round \( s \leq s^* \), \( u^*_i < \bar{u}_i \). Then bidder \( i \) must be a provisional winner of round \( s \). Let \( \hat{I} = I \setminus 0 \) be the set of provisional winners in round \( s \). If \( i \notin I \), then the seller’s provisional revenue in round \( s \) is

\[
\begin{align*}
    w(I) - \sum_{j \in I} u^*_j < & w(I) - \sum_{j \in I} u^*_j + \bar{u}_i - u^*_i \\
    = & w(I) - \sum_{j \in \hat{I} \cup i} u^*_j + w(\hat{N} \setminus i) - w(\hat{N} \setminus i) \\
    \leq & w(I) - \sum_{j \in \hat{I} \cup i} u^*_j + w(I \cup i) - w(I) \\
    = & w(I \cup i) - \sum_{j \in \hat{I} \cup i} u^*_j.
\end{align*}
\]

Here the inequality in the third line follows the lemma of 21.7.3. Thus, for the auctioneer, including \( i \) in the set of provisional winners in round \( s \) will generate larger provisional revenue. However, in round \( s \), the utility of bidder \( u^*_i < \bar{u}_i \) shows that the proxy of bidder \( i \) didn’t bid in a way that maximizes bidder \( i \)’s interest, unless \( \bar{u}_i - u^*_i < \epsilon \). Hence, we have that for all \( i \), \( u^*_i \geq \bar{u}_i \) (up to \( \epsilon \)).

Now suppose that there is a bidder \( i \in \hat{N} \) such that \( u^*_i > \bar{u}_i \). Then

\[
w(\hat{N}) = u^*_0 + u^*_i + \sum_{j \neq i} u^*_j > u^*_0 + w(\hat{N}) - w(\hat{N} \setminus i) + \sum_{j \neq i} u^*_j,
\]

which implies that

\[
u^*_0 + \sum_{j \neq i} u^*_j < w(\hat{N} \setminus i).
\]

(21.58)
On the other hand, the seller’s revenue

\[ u_0^* = \max_y \sum_{j \in N} \theta_j^*(y_j|v_j) \]

\[ = \max_y \sum_{j \in N} \max(v_j(y_j) - u_j^*, 0) \]

\[ = \max \max_{I \subseteq N} \sum_{j \in I} (v_j(y_j) - u_j^*) \]

\[ = \max \max_{I \subseteq N} \sum_{j \in I} (v_j(y_j) - u_j^*) \]

\[ = \max_{I \subseteq N} (w(I) - \sum_{j \in I} u_j^*). \]

and so, in particular, for \( I = \tilde{N}\setminus i \), this implies that

\[ u_0^* + \sum_{j \neq i} u_j^* \geq w(\tilde{N}\setminus i). \quad (21.59) \]

Obviously, the two inequations (21.58) and (21.59) are contradicting. We have thus argued that with truthful reporting, ascending proxy mechanism will have the equivalent outcome as the VCG mechanism.

In order to show that truthful reporting constitutes an ex post equilibrium, consider a bidder \( i \) and suppose that all bidders \( j \neq i \) report truthfully, and then we discuss the reporting incentives of bidder \( i \). We know from (21.59) that independent of what bidder \( i \) submits as his value vector, the seller’s revenue satisfies

\[ u_0^* > w(\tilde{N}\setminus i) - \sum_{j \neq i} u_j^*, \]

The right-hand side is a lower bound on the seller’s revenue because it can be obtained by ignoring bidder \( i \) and including all other bidders in the set of provisionally winning bidders. At the same time, since the total payoff of all the participants—the bidders and the seller—can never exceed \( w(\tilde{N}) \), which means

\[ u_0^* \leq w(\tilde{N}) - \sum_{j \in N} u_j^*. \]
For bidder $i$, his arbitrary reporting will not make his utility exceed $w(\tilde{N}) - w(\tilde{N}\setminus i) = \bar{u}_i$. The right hand of the equation is the expected utility obtained in ascending proxy auction if bidder $i$ chooses to report truthfully, so bidder $i$ will have the incentive to report truthfully and thus truthful reporting constitutes an ex post equilibrium.

21.7.6 Sun-Yang Auction Mechanism for Multiple Complements

Ausubel and Milgrom (2002, 2006) further discussed when the auction items are not substitutes for each other, the ascending proxy auction mechanism will be better than the VCG mechanism and can overcome the drawbacks and problems faced by the VCG mechanism mentioned earlier.

The ascending combinatorial auction mechanism is similar to the dynamic adjustment mechanism in general equilibrium. When the goods are overdemanded, the price will rise. The slight difference is that in ascending combinatorial auctions there is only one-way price change (i.e. increment), the equilibrium price of ascending combinatorial auction is actually the general equilibrium price (of discrete commodity), and the demand corresponding to the aggregate of bidding prices is consistent with the supply corresponding to the auctioneer’s optimal income. However, when the items are complementary, the ascending proxy mechanism will face an exposure problem, that is, there is no general equilibrium price. We now discuss the exposure problem in multiple objects auctions by an example (adapted from Milgrom (2000, p. 257, Footnote 12)).

Example 21.7.6 (Exposure problem in auction for complements) Suppose that the set of objects is $K = \{a, b, c\}$, there are three bidders $N = \{1, 2, 3\}$, the objects are mutual complements for all bidders (value function is superadditivity),
and the values attached by the bidders to these objects (package) are

\[ v_1(a) = 1, v_1(b) = 1, v_1(c) = 0, v_1(\{a, b\}) = 3, v_1(\{a, c\}) = 1, v_1(\{b, c\}) = 1, v_1(\{a, b, c\}) = 3; \]
\[ v_2(a) = 0, v_2(b) = 0.9, v_2(c) = 1, v_2(\{a, b\}) = 1, v_2(\{a, c\}) = 1, v_2(\{b, c\}) = 3, v_2(\{a, b, c\}) = 3; \]
\[ v_3(a) = 1, v_3(b) = 0, v_3(c) = 1, v_3(\{a, b\}) = 1, v_3(\{a, c\}) = 3.5, v_3(\{b, c\}) = 1, v_3(\{a, b, c\}) = 3.5. \]

If a competitive equilibrium does exist, its allocation would be efficient. In this example, there is an efficient allocation

\[ y_1 = \{b\}, y_2 = \emptyset, y_1 = \{a, c\}. \]

to support the efficient allocation, the prices must satisfy

\[ p_b \leq 1, p_a + p_b \geq 3, p_b + p_c \geq 3, p_a + p_c \leq 3.5. \]

However, these together imply that \( p_a + 2p_b + p_c \geq 6 \) and \( p_b \leq 1 \), which is inconsistent with \( p_a + p_c \geq 4 \) and \( p_a + p_c \leq 3.5 \). So, there is no competitive equilibrium.

The problem associated with the existence of the general equilibrium of multi-item auctions is called the exposure problem. The new problems appeal when auction items are complements. If bidders submit their bids according to actual needs, they may be exposed to a possible risk. When the bidder is included in the list of provisional winners by the auctioneer, he will face the risk of not getting some complementary items, and he will not be willing to pay for the other complementary items according to the bidding price. It is quite difficult to design an efficient auction mechanism for complements.

This subsection briefly discusses an auction mechanism for multiple complements. Sun and Yang (JPE, 2014) constructed a dynamic auction mechanism to solve the design of an efficient auction mechanism for multiple complements. This dynamic mechanism is called Sun–Yang auction mechanism.

Sun–Yang auction mechanism introduces the concept of nonlinear general equilibrium, where the general equilibrium of nonlinearity exists when the items
are complementary to all the bidders, although the general linear equilibrium may not exist.

Suppose there is a set of objects $K = \{a, b, \cdots\}$ and a group of bidders $N = \{1, 2, \cdots, n\}$. Bidder $i \in N$ attaches a monetary value to the objects, which is $u_i : 2^K \to \mathbb{Z}_+$, where $2^K$ denotes the set of all subsets of items $K$ and $\mathbb{Z}_+$ is the set of all nonnegative integers. The seller (denoted by superscript 0) has a reserve price function $u_0 : 2^K \to \mathbb{Z}_+$ with $u_i(\emptyset) = 0, \forall i \in N \cup \{0\}$. Let $\hat{N} = N \cup 0$ represent the set of all agents.

Suppose that the bidder’s utility function is quasi-linear and satisfies the superadditivity for the value function of the item. Here we first introduce the concept of non-linear general equilibrium.

A function $p : 2^K \to \mathbb{R}_+$ is a pricing function that satisfies $p(\emptyset) = 0$ is feasible if $p(S) \geq u_0(S), \forall S \subset K$. For the given pricing function $p$, bidder’s demand correspondence (which may be a set) is

$$D_i(p) = \text{argmax}_{S \subseteq K} u_i(S) - p(S).$$

We call $\pi = \{\pi^1, \cdots, \pi^m\}, m \leq n$, a partition of the set of items $K$. $y = (y_0, y_1, \cdots, y_n)$ is a feasible allocation, that is, for any $y_i$, or $y_i = \emptyset$, or $y_i \in \pi$, $y_i \cap y_j = \emptyset, i \neq j, \bigcup_{i \in N} y_i = K$. The corresponding partition of allocation $y$ is $\pi = \{y_i, i \in \hat{N} : y_i \neq \emptyset\}$. Given the price function $p$, the seller’s supply correspondence is

$$Y(p) = \text{argmax}_\pi \sum_{S \in \pi} p(S).$$

An allocation $y^* = (y_0^*, y_1^*, \cdots, y_n^*)$ is efficient, if for every allocation $y$, we have

$$\sum_{i \in N} u_i(y_i^*) \geq \sum_{i \in N} u_i(y_i)).$$

**Definition 21.7.5 (Nonlinear general equilibrium)** A nonlinear general equilibrium consists of a price function $p^*$ and an allocation $y^*$ such that

1. for the seller, $y^* \in Y(p^*)$;
2. for every bidder $i \in N$, $y_i^* \in D_i(p^*)$. 

957
In the above definition, similar to general equilibrium, condition (1) characterizes the supply under the nonlinear equilibrium price function; condition (2) depicts the demand under the nonlinear equilibrium price function, in which at the equilibrium price $p^*$, the quantity supplied equals the quantity demanded.

The following example on complements (from Sun and Yang) show that a nonlinear general equilibrium exists, but the general equilibrium does not exist.

**Example 21.7.7** Suppose that objects $K = \{a, b, c\}$ are complements for three bidders $N = \{1, 2, 3\}$. Bidders’ values and seller’s reserve prices are given as follows

$v_1(a) = 2, v_1(b) = 2, v_1(c) = 0, v_1(\{a, b\}) = 7, v_1(\{a, c\}) = 4, v_1(\{b, c\}) = 4, v_1(\{a, b, c\}) = 7;

v_2(a) = 2, v_2(b) = 0, v_2(\{a, b\}) = 2, v_2(\{a, c\}) = 3, v_2(\{b, c\}) = 6, v_2(\{a, b, c\}) = 3, v_2(\{a, b, c\}) = 6;

v_3(a) = 0, v_3(b) = 2, v_3(\{a, b\}) = 2, v_3(\{a, c\}) = 4, v_3(\{b, c\}) = 4, v_3(\{a, b, c\}) = 6, v_3(\{a, b, c\}) = 7;

$u_0(a) = 1, u_0^0(b) = 2, u_0(\{a, b\}) = 1, u_0(\{a, c\}) = 3, u_0(\{b, c\}) = 4, u_0(\{a, b, c\}) = 5.$

In this example, there are two efficient allocations: $y_1 = \{a, b\}, y_2 = \{c\}, y_3 = \emptyset$, and $y'^1 = \{a, b\}, y'^2 = \emptyset, y'^3 = \{c\}$. At the same time, there is a nonlinear general equilibrium price:

$p^*(a) = p^*(b) = p^*(c) = 2, p^*(\{a, b\}) = p^*(\{a, c\}) = p^*(\{b, c\}) = 6, p^*(\{a, b, c\}) = 7.$

On the demand side, $D_1(p^*) = \{a, b\}, \{c\} \in D_2(p^*), \emptyset \in D_3(p^*)$; on the supply side, $(\{a, b\}, \{c\}, \emptyset) \in Y(p^*)$. Also, price $p^*$ and $y'$ also constitute a nonlinear general equilibrium.

However, there is no general equilibrium under the linear price. Suppose there is a price vector $p(a), p(b), p(c)$. If the price $p^*$ and the allocation $y'$ consist an equilibrium, then:

for the seller, we must have $p(a) \geq 1, p(b) \geq 2, p(c) \geq 1$;

for bidder 1, we must have: $p(a) + p(b) \leq 7$;

for bidder 2, we must have: $p(c) \leq 2, p(a) + p(c) \geq 6$;

for bidder 3, we must have: $p(c) \geq 2, p(b) + p(c) \geq 6$.

958
From $p(c) = 2$, combining the inequalities $p(a) + p(c) \geq 6$ and $p(b) + p(c) \geq 6$, we have $p(a) \geq 4$ and $p(b) \geq 4$, which contradicts $p(a) + p(b) \leq 7$.

Sun and Yang (2014) showed that if the items are complementary to all bidders and the bidders have quasi-linear utility functions, there always is a nonlinear general equilibrium. For readers who are interested in the proof of the theorem, see the Proof Appendix of Sun and Yang’s Working Paper (2012).

Sun and Yang (2014) first constructed an ascending auction mechanism when bidders choose to bid honestly (which will be defined below). An elaborate dynamic auction mechanism was then constructed to ensure that each bidder has the incentive to bid honestly, and more precisely speaking, honest bidding was a refined ex post Nash equilibrium.

**Definition 21.7.6** In a dynamic auction, we say that bidder $i$ bids honestly, if at every round and for any price function $p^t(\cdot)$, bidder $i$ reports demand

$$d^t_i \in D_i(p^t) = \text{argmax}_{S \subseteq K} \{u_i(S) - p^t(S)\}$$

with $d^t_i = \emptyset$ when $\emptyset \in D_i(p^t)$.

When bidders choose to bid honestly, Sun and Yang constructed the following ascending auction mechanism, by which the final result of the auction is a nonlinear general equilibrium.

**Definition 21.7.7 (Ascending auction under truthful reporting)** There are several steps:

1. The seller announces his reserve price $u_0$ and the auctioneer sets the initial pricing functions $p^0 = u_0$.
2. In each round $t = 0, 1, 2, \cdots$, given the the initial pricing functions $p^t$ at the round $t$, the auctioneer chooses a supply set $y_t \in Y(p^t)$, and every bidder $i$ reports his demand $d^t_i \in D_i(p(t))$ against prices $p^t$. When there is at least one $i, j$ with $i \neq j$ such that $d^t_i = d^t_j = S, S \subseteq K$ or there is only one $i$ such that $d^t_i = S$ but no $k$
that makes $S \neq y_k^t$, then we call this bundle $S$ over demanded. For any over demanded bundle $S$, the auctioneer adjusts the price function through the formulas $p^{t+1}(S) = p^t(S) + 1$ and repeats (2) into round $t + 1$. If no bundle at the $t$ round is overly demanded, written as $t^*$.

(3) In round $t^*$, the auctioneer allocates $d_i^* = y_i^*$ to bidder $i$ and asks him to pay the price $p^*(y_i^*)$. If there is no $i$ of bundle $S$ such that $S = d_i^*$ and $p^*(S) = u_0(S)$, the bundle $S$ is left to the seller. If $p^*(S) > u_0(S)$, the bundle $S$ is left to the bidder who finally gives up the choice of bundle $S$. Then the auction ends.

Here we show that the results of the ascending auction under truthful reporting are non-linear general equilibria.

**Example 21.7.8** Given $p^t = (p^t(a), p^t(b), p^t(c), p^t(\{a, b\}), p^t(\{a, c\}), p^t(\{b, c\}), p^t(\{a, b, c\}))$.

In the initial round, the price, supply and demand are as follows:

$$p^0 = (1, 2, 1, 3, 3, 4, 5), y_0 = \{\{b, c\}, a\}, d_1^0 = \{a, b\}, d_2^0 = \{a, c\}, d_3^0 = \{a, b, c\}.$$  

In rounds 1~6, the price of supply and demand are as follows:

$$p^1 = (1, 2, 1, 4, 4, 4, 6), \pi^1 = \{\{a, c\}, b\}, d_1^1 = \{a, b\}, d_2^1 = \{a, c\}, d_3^1 = \{b, c\};$$  

$$p^2 = (1, 2, 1, 5, 4, 5, 6), \pi^2 = \{\{a, b\}, c\}, d_1^2 = \{a, b\}, d_2^2 = \{a, c\}, d_3^2 = \{b, c\};$$  

$$p^3 = (1, 2, 1, 5, 5, 6, 6), \pi^3 = \{\{a, c\}, b\}, d_1^3 = \{a, b\}, d_2^3 = \{a, c\}, d_3^3 = \{a, b, c\};$$  

$$p^4 = (1, 2, 1, 6, 5, 6, 7), \pi^4 = \{\{a, b\}, c\}, d_1^4 = \{a, b\}, d_2^4 = \{a, c\}, d_3^4 = \{c\};$$  

$$p^5 = (1, 2, 2, 6, 6, 6, 7), \pi^5 = \{\{a, c\}, b\}, d_1^5 = \{a\}, d_2^5 = \{a\}, d_3^5 = \emptyset;$$  

$$p^6 = (2, 2, 2, 6, 6, 6, 7), \pi^6 = \{\{a, b\}, c\}, d_1^6 = \{a, b\}, d_2^6 = \emptyset, d_3^6 = \emptyset.$$  

Then no bundle is overdemanded, $t_* = 6$, and the equilibrium prices are $p^* = (2, 2, 2, 6, 6, 6, 7)$. Bidder 1 gets $\{a, b\}$ and pays 6. Bidder 3 gets $\{c\}$ and pays 2.

Comparing with example 21.7.7, the results of an ascending auction are non-linear general equilibria.
Sun and Yang (2014) built a dynamic Sun–Yang auction mechanism that induces each bidder to report truthfully. Since this mechanism involves much technical details and symbols, here we only briefly discuss the basic structure of it.

In the incentive-compatible dynamic auction mechanism, Sun and Yang constructed \( n+1 \) markets and each corresponds to two price functions: the first price function and the second price function. All participants, including the seller, have different prices for each market at each round. Let \( p_{t-0}^l \) and \( p_0^l \) denote the sellers’ the first price function and the second price function in each round \( t \) of market \( M_{-0} \) with the set \( N \) of bidders. Let \( p_{t-i}^l \) and \( p_i^l \) denote bidder \( i \)’s the first price function and the second price function in each round \( t \) of market \( M_{-i} \) that stands for the market without bidder \( i \).

Each bidder \( i \) reports a demand bundle against the second price function \( p_i^l \), and the auctioneer announces a revenue-maximizing supply set of each market \( n+1 \) at the first price function \( p_{-k}(k \in \bar{N}) \). For bidders, because of different prices in each market, the prices faced in making demand decisions are also different. When the market overly demands certain bundle, the seller adjusts the price depending on how much the bidders influence the demand for the bundle, so the adjustment of each market and each bidder is different. If bidder \( i \) is the crucial demander of a bundle (i.e. the bidder who reports the highest price), bidder \( i \) will face a price increase of 1 unit while the price of the others will be unchanged. This way of price adjustment can internalize the externality of price changing by bidder ’s demand report.

Sun and Yang (2014) proved that under this dynamic mechanism, the bidder’s payment is a generalized VCG payment. Sun–Yang auction mechanism has many elaborate designs: the over demand is divided into first over demand and second overly demanded; the price adjustment process is bidirectional, raising the price when overly demanded and decreasing the price when oversupplied; the bidder is allowed to withdraw some of their past bids, but will be punished if they nullify their bids for too many times; the Sun–Yang auction can always end within a finite time and so on. If you are interested in more details, please directly refer
to their papers.

The Sun–Yang auction mechanism has many favorable features. For example, it requires bidder to report only one demand against the market price so it is privacy preserving and informationally efficient; it can solve the exposure problem easily that is caused by complementarity; it tolerates mistakes made by bidders in the bidding process and so on.

For the exposure problem, a more complex situation is that a bundle of objects can be complements to one bidder but substitutes to another. In this case, the nonlinear general equilibrium may not exist either. Sun and Yang (2009, 2015) put forward some other efficient auction mechanisms according to the substitution and complementarity of items.
Reference

Books:


Papers:


Chapter 22

Matching Theory

22.1 Introduction

This chapter discusses another subfield of market design—matching theory, which is the core part of market design. The path-breaking contribution on matching theory was made by David Gale and Lloyd S. Shapley who published a paper in *The American Mathematical Monthly* in 1962 discussing the matching problem in the marriage market. This mathematical paper, without using any math formulas, opens up a whole new area in economics. Roth (2015) revealed that the price mechanism usually did not play a dominant role in matching issues, and especially, no transfer payments are made in many cases.

For example, college admission is not given to the highest bidder, and universities do not adjust the demand for university places (student applications) through changing tuition fees. Employers in certain occupations do not rely entirely on wage levels to select employees but look for employees who have professional skills important to the position rather than who ask for low pay. Also, in the organ transplantation market, matching organs for transplantation is through non-price mechanisms, since buying and selling organs are illegal in most countries. Another typical pheromone are the office allocation of employees without transfer payments.
Matching has become an important mechanism and tools in non-price mechanisms to deal with economic environments where traditional/natural market fails in the sense it results in Pareto inefficient allocations. Thus in this chapter, we focus on two topics: how the agents can achieve an efficient allocation of resources and how the incentives promote the interaction among agents. Depending on whether agents in one side have preferences, there are two matching models: the two-sided matching (agents on both sides of the match have preferences) and the one-sided matching (agents of only one side have preferences and its agents are often matched to objects, quotas, positions, etc. on the opposite side). We will also discuss some common matching mechanisms in different matching models and their properties.

The matching theory has a wide range of applications to specific matching designs. For example, Alvin Roth, one of the recipients of the 2012 Nobel Prize in Economics Sciences, participated in the design of the United States-based National Resident Matching programme in 1995; he and his collaborators (Abdulkadiroglu, Pathak and Sonmez) were involved in the Boston Public School matching design; he also collaborated with other economists in the design of the New England Kidney Exchange mechanism. The above successful design projects fully demonstrate the great potential of market design theory in solving real-life problems.

This chapter is organized as follows: Section 1.2 and 1.3 discuss the two-sided and one-sided market matching models, respectively. Section 1.4 describes the applications of matching theory in the university admission and organ transplant processes.

### 22.2 Two-Sided Market Matching

The matching theory is considered as the theoretical basis of market mechanism design. It exerts a great influence on economics and provides guidance on practical applications. As such, in 2012, the Nobel Prize in Economics was awarded to
Shapley and Roth for their fundamental contributions on market design in general and matching theory in particular. The two-sided matching can be divided into one-to-one, many-to-one and many-to-many matchings. In addition, depending on whether trading media or contracts are involved, the two-sided matching can also be divided into pure matching and conditional matching (involving the use of currency or contract). In this section, we will begin with the one-to-one matching, a basic matching model in two-sided matching, and discuss matching properties and then move to many-to-one and many-to-many matchings as well as conditional matchings.

22.2.1 One-to-One Matching

Some symbols and concepts need to be introduced first. Assume that there are two groups of agents $M = \{m_1, \ldots, m_n\}$ and $W = \{w_1, \ldots, w_k\}$ with $M \cap W = \emptyset$, say one for men (M) and the other for women (W). The preference of member $i$, expressed as $\succsim_i$, is defined on the opposite set and himself/herself. We mainly consider the case of strict preference $\succ_i$, that is, excluding indifferent outcomes. Matchings with weak preference $\succeq_i$ are much complicated and the research in it shows an increase in recent years.

We first introduce the basic concept of matching.

**Definition 22.2.1 (Matching)** We call a bijective function $\mu : M \cup W \rightarrow M \cup W$ a matching if it satisfies:

1. $\mu(m) \notin W$ implies $\mu(m) = m, \forall m \in M$;
2. $\mu(w) \notin M$ implies $\mu(w) = w, \forall w \in W$;
3. $\mu(m) = w$ if and only if $\mu(w) = m, \forall m \in M, w \in W$.

A matching in the above definition refers to the matching of each member of a group with a member of the other group or with himself/herself. Thus a matching is an idempotent function $\mu : M \cup W \rightarrow M \cup W$, i.e., $\mu^2(x) = x$ for all $x \in M \cup W$. 

969
In the marriage matching, man is matched with a woman or with himself (i.e., \( \mu(m) = m \), being single, or equivalently, denoted by \( \mu(m) = \emptyset \)), and women are alike. A man \( m \) is acceptable to \( w \) if \( m \succ_w \emptyset \), and a woman \( w \) is acceptable to \( m \) if \( w \succ_m \emptyset \). For a man \( m \in M \), a woman \( w \) is unacceptable if \( m \succ_m w \), where \( \succ_m \) is the preference of the man.

We now introduce the concept of Pareto-efficient matching. For the convenience of discussion, under strict preference \( \succ_i \), we define the weak preference as \( \succeq_i : \varphi(i) \succeq_i \mu(i) \) if and only if \( \mu(i) \succ_i \varphi(i) \) does not hold. (In Chapter 3, Subsection 3.4.5, we regard \( \succeq_i \) as a completion of \( \succ_i \) such that \( \succeq_i \) is a weak binary relation with reflexivity and completeness.)

**Definition 22.2.2 (Pareto-efficient Matching)** A matching \( \mu \) is called Pareto-efficient, if there is no other matching \( \varphi \) such that \( \varphi(i) \succeq_i \mu(i) \), \( \forall i \in W \cup M \) and there is at least one \( i \in W \cup M \), s.t. \( \varphi(i) \succ_i \mu(i) \).

A matching \( \mu \) is a weak Pareto-efficient matching, if there is no other matching \( \varphi \) such that \( \varphi(i) \succ_i \mu(i) \), \( \forall i \in W \cup M \).

In addition, the concept of Pareto-efficient matching also applies to a group of members but not both groups.

**Definition 22.2.3** A matching \( \mu \) is said to be:

- Pareto-efficient for the group \( M \) (or \( W \)) if there is no other matching \( \varphi \) such that \( \varphi(i) \succeq_i \mu(i) \), \( \forall i \in M \) (or \( W \)), and there is at least one \( i \in M \) (or \( W \)) such that \( \varphi(i) \succ_i \mu(i) \);
- weak Pareto-efficient for the group \( M \) (or \( W \)), if there is no other matching \( \varphi \) such that \( \varphi(i) \succ_i \mu(i) \), \( \forall i \in M \) (or \( W \)).

We then have the following notion of stable matching.

**Definition 22.2.4 (Stable matching)** A matching \( \mu \) is stable if it satisfies:

(i) there is no matching pair \( (m, w) \in M \times W \) such that \( w \succ_m \mu(m) \) and \( m \succ_w \mu(w) \);

(ii) for \( i \in M \cup W \), if \( \mu(i) \neq i \), then \( \mu(i) \succ_i i \).
Condition (ii) above means that if a mate is not himself/herself, then the mate surely is acceptable to him/her. This condition is called \textit{individual rationality}. If a matching does not satisfy this condition, then it will be individually blocked by an agent; at the same time, Condition (i) means that there are no two individuals of the opposite groups who would prefer each other over their own matching mates. If a matching does not satisfy this condition, it will be blocked by a certain matching pair, a condition being referred to as “pair-blocked”. Obviously, \textit{any stable matching is Pareto-efficient}, because, if not Pareto-efficient, it will inevitably fail to satisfy individual rationality or is blocked by a certain matching pair.

Here we use a simple example (see Roth and Sotomayor, 1990) to discuss stable and unstable matchings.

\textbf{Example 22.2.1} Consider the matching between three men and three women. Suppose their (strict) preferences are as follows:
\begin{align*}
p(m_1) &= w_2, w_1, w_3, \quad p(w_1) = m_1, m_3, m_2; \\
p(m_2) &= w_1, w_3, w_2, \quad p(w_2) = m_3, m_1, m_2; \\
p(m_3) &= w_1, w_2, w_3, \quad p(w_3) = m_1, m_3, m_2,
\end{align*}
where \(p(m_1) = w_2, w_1, w_3\) is man \(m_1\)'s rankings of women in order of preference from the highest to the lowest, i.e. \(w_2 \succ m_1 w_1 \succ m_1 w_3 \succ m_1 m_1\).

Consider the following two matchings \(\mu^1\) and \(\mu^2\).
\[
\mu^1 = \begin{pmatrix} w_1 & w_2 & w_3 \\ m_1 & m_2 & m_3 \end{pmatrix},
\]
which lists the matched pair of two people in one column, such as the pair of \((w_1, m_1)\). By the definition of stability of matching, \(\mu^1\) is unstable since \((m_1, w_2)\) is a blocking pair.
\[
\mu^2 = \begin{pmatrix} w_1 & w_2 & w_3 \\ m_1 & m_3 & m_2 \end{pmatrix}.
\]
We can verify that \(\mu^2\) is a stable matching.
Deferred Acceptance Algorithms

To discuss matching, two important issues are asked: whether a stable matching exists and how to find such a matching. Gale and Shapley (1962) answered these two issues by proposing a specific algorithm called deferred acceptance (DA) algorithm, which could not only determine whether a matching was stable or not, but also provide an algorithm for finding such a stable matching. Moreover, the stable matching obtained by the deferred acceptance procedure is also optimal for agents on the proposing side. That is, every agent on the proposing side is at least as well off under the assignment given by the deferred acceptance procedure as he would be under any other stable assignment. We now discuss the deferred acceptance algorithm in the following, which coincides with practical proposing ways.

Individuals were divided into two groups, the proposing group and the proposed group. Agents in the former group make proposals to the agents in the latter group, who can choose to reject the proposal or hold it without commitment. The deferred acceptance algorithm goes through several phases, each containing two steps. We will start with the first stage.

Step 1: Each agent in the proposing group (such as a male agent) proposes to his most preferred choice (such as a female agent) in the proposed group. If no one is acceptable, he makes no propose, which means his matching mate is himself;

Step 2: Each agent in the proposed group rejects unacceptable agents and if other proposals remains, chooses the most preferred and rejects the rest.

In general, at stage $k$, an agent in the proposing group who was rejected at stage $k - 1$ by an agent in the proposed group proposes to his/her most preferred acceptable agent who has not yet rejected him/her. If no acceptable choice remains, he/she makes no proposal. Each agent in the proposed group
selects the most preferred agent from among the retained proposals in the previous stage and the new proposals (if any) in the current stage and rejects the rest.

Since no agents proposes twice to the same agent, this algorithm always terminates in finite stages. The algorithm terminates when there are no more rejections. In this stage, an agent in the proposed group is matched to another agent in the waiting list of the opposite group. If an agent in the proposing group makes no proposals or receives no acceptance or an agent in the proposed group receives no proposals or rejects all, he/her is matched to himself/herself.

Gale and Shapley (1962) proved the following theorem.

Theorem 22.2.1 A stable matching exists for every marriage market.

Proof. We only discuss the matching market with strict preference. Because only finite agents are involved, the algorithm will terminate in finite stages. So we only need to verify that the result of the deferred acceptance algorithm, expressed as $\mu_{da}$, is a stable matching. We assume that men are proposing. First, the matching produced by the deferred acceptance algorithm satisfies the individual rationality, because each proposing man only proposes to an acceptable woman while each proposed woman only chooses an acceptable man. Secondly, no blocking pairs form in this matching. Suppose that a pair of man and woman $(m, w)$ is a blocking pair, which means $w \succ_m \mu_{DA}(m)$ and $m \succ_w \mu_{DA}(w)$. According to the deferred acceptance algorithm, $m$ first proposes to $w$ and only proposes to $\mu_{DA}(m)$ after being rejected by $w$ ($\mu_{DA}(m) = m$ presents the matching after the man is rejected by all acceptable women). However, by the deferred acceptance algorithm, if $w$ rejects the proposal from $m$ in some stage, it means that she has a more preferred man, but $w$ will ever more prefer the proposing offers that she holds in later steps, which contradicts the hypothesis that $m \succ_w \mu_{da}(w)$. Thus, the matching produced by the deferred acceptance algorithm is stable.

We use an example below to illustrate the process of the deferred acceptance algorithm.

Example 22.2.2 (An Example of deferred acceptance algorithm) Consider
the marriage matching between three men and three women. Each agent’s preference for agents in the opposite side is as follows (any unacceptable man/woman is removed from the preferences lists):

\[ p(m_1) = w_2, w_1; \quad p(w_1) = m_1, m_3, m_2; \]
\[ p(m_2) = w_1, w_2, w_3; \quad p(w_2) = m_2, m_1, m_3; \]
\[ p(m_3) = w_1, w_2; \quad p(w_3) = m_1, m_3, m_2. \]

Suppose men are proposing. In stage 1, every man makes a proposal to his most preferred woman, i.e. \( m_2 \) and \( m_3 \) both propose to \( w_1 \) and \( m_1 \) to \( w_2 \); the outcome is that \( w_1 \) rejects \( m_2 \) and holds the offer from \( m_3 \), and \( w_2 \) holds the offer from \( m_1 \):

In stage 2, \( m_2 \) proposes to \( w_2 \), \( w_2 \) rejects \( m_1 \) and holds the offer from \( m_2 \);

In stage 3, \( m_1 \) proposes to \( w_1 \), \( w_1 \) rejects \( m_3 \) and holds the offer from \( m_1 \);

In stage 4, \( m_3 \) proposes to \( w_2 \), \( w_2 \) rejects \( m_3 \) and holds the offer from \( m_2 \).

Then no further proposals will be made.

In the process above, no man proposes to \( w_3 \), so \( w_3 \) is matched to herself; \( m_3 \) is rejected by all acceptable women, so he is also matched to himself. The final matching outcome is:

\[ \mu_{DA} = \begin{pmatrix} w_1 & w_2 & w_3 & m_3 \\ m_1 & m_2 & m_3 & m_3 \end{pmatrix}. \]

The following table (22.2.2) characterizes the process of the deferred acceptance algorithm, where the underlined agent refers to the man whose proposal is being held without commitment by some woman, and a man without underline means that he is rejected by some woman.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( w_1 )</td>
<td>( w_2 )</td>
<td>( w_3 )</td>
</tr>
<tr>
<td>( m_2, m_3 )</td>
<td>( m_1 )</td>
<td></td>
</tr>
<tr>
<td>( m_2, m_1 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m_1, m_3 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m_1 )</td>
<td>( m_2 )</td>
<td>( m_3 )</td>
</tr>
</tbody>
</table>
where \( \mu(m_3) = \emptyset \), which is equivalent to \( \mu(m_3) = m_3 \), a situation of being matched with himself.

Multiple stable matchings tend to exist in a matching. The following example shows that matching outcomes may be different for men-proposing and women-proposing.

**Example 22.2.3** Consider the marriage matching between three men and three women. Each agent’s preference for the opposite set’s agents is as follows:

\[
P(m_1) = w_1, w_2, w_3; \quad P(w_1) = m_1, m_2, m_3;
\]

\[
P(m_2) = w_1, w_2, w_3; \quad P(w_2) = m_1, m_3, m_2;
\]

\[
P(m_3) = w_1, w_3, w_2; \quad P(w_3) = m_1, m_2, m_3.
\]

When men are proposing, the matching outcome is:

\[
\mu^{DA}_M = \begin{pmatrix}
m_1 & m_2 & w_3 \\
1 & 2 & 3
\end{pmatrix}.
\]

When women are proposing, the matching outcome is:

\[
\mu^{DA}_W = \begin{pmatrix}
m_1 & m_2 & w_3 \\
1 & 2 & 3
\end{pmatrix}.
\]

From the men group \( M \)'s point of view, the matching \( \mu^{DA}_M \) is better than \( \mu^{DA}_W \); on the contrary, from the women group \( W \)'s point of view, the matching \( \mu^{DA}_W \) is better than \( \mu^{DA}_M \). As such, in the following, we introduce the notion of optimal stable matching for groups.

**Definition 22.2.5** A stable matching is men-optimal, denoted by \( \mu_M \), if for any stable matching \( \mu \), we have \( \mu_M \succeq_m \mu \), i.e. \( \forall m \in M, \mu_M(m) \succeq_m \mu(m) \). The women-optimal stable matching \( \mu_W \) can be defined in the same way.

Gale and Shapley (1962) proved the following theorem.
Theorem 22.2.2 When all agents have strict preferences, the matching produced by the men-proposing (or women-proposing) deferred acceptance algorithm is optimal among all possible stable matchings for the men (or women) group, i.e.

\[ \mu_M = \mu_M^{DA}, \]

and

\[ \mu_W = \mu_W^{DA}. \]

Proof. The following is a brief proof why the matching by the men-proposing deferred acceptance algorithm is optimal among all possible stable matchings for the men group.

We first define the concept of the achievable matching. We call a matching pair \((m, w)\) as an achievable matching if there exists a stable matching rule under which \((m, w)\) is one of its matching pairs. We use mathematical induction to prove the above statement. By the men-proposing deferred acceptance algorithm, each man proposes in decreasing order of preference. Suppose that in all previous stages, no man has been rejected by his acceptable women. But in the current stage, some man \(m\) is rejected by the proposed woman \(w\).

We only have to prove that \((m, w)\) is not a stable matching outcome. If \(w\) rejects \(m\), it must be because she receives a proposal from another man \(m'\) that is more preferred to \(m\). For \(m'\), \(w\) is his most preferred woman of all the remaining acceptable mates except the women who reject him. If there exists a matching rule \(\mu\) that can make \(\mu(m) = w\) and \(\mu(w) = m\), obviously, such an outcome will be overturned by \((m', w)\). Thus, \(\mu\) is not stable. In this way, we show that the matching by the men-proposing deferred acceptance algorithm is the optimal stable one among all possible stable matchings for the men group. Similarly, the matching produced by the women-proposing deferred acceptance algorithm is the optimal stable one among all possible stable matchings for the women group.

Regarding stable matchings for different groups, Knuth (1976) further revealed that the optimal matching for one side is the worst for the other side.
Theorem 22.2.3 Suppose that all agents have strict preferences. On the set of all stable matchings, the interests of two groups are opposite, that is, if $\mu$ and $\varphi$ are two stable matchings, $\mu \succ_M \varphi$ if and only if $\varphi \succ_W \mu$.

Proof. Suppose that $\mu$ and $\varphi$ are two stable matchings, and $\mu \succ_M \varphi$. Suppose by way of contradiction that $\varphi \succ_W \mu$ does not hold. Then there exists a $w \in W$ such that $\mu(w) \succ_w \varphi(w) \succeq_w w$. Let $m = \mu(w)$. Since $w = \mu(m) \succ_m \varphi(m)$, then $(m, w)$ forms a blocking pair to $\varphi$, which contradicts the fact that $\varphi$ is a stable matching. Hence, $\mu \succ_M \varphi$ implies $\varphi \succ_W \mu$. Similarly, if $\varphi \succ_W \mu$, then we have $\mu \succ_M \varphi$. ■

The above two theorems require that all agents have strict preferences, without which the two theorems may not hold. The following example is from Roth and Sotomayor (1990), which shows that if an agent’s preference is not strict, the optimal stable matching might not exist and the attitudes of both sides are not opposite in stable matchings.

Example 22.2.4 Suppose the preference orders of 3 men and 3 women are as follows:

$m_1$’s preferences: $p(m_1) = [w_2 \sim w_3, w_1]$; $p(w_1) = m_1, m_2, m_3$;

$m_2$’s preferences: $p(m_2) = w_2, w_1$; $p(w_2) = m_1, m_2$;

$m_3$’s preferences: $p(m_3) = w_3, w_1$; $p(w_3) = m_1, m_3$.

In the deferred acceptance algorithm, the matching of $m$-proposing or $w$-proposing is not unique, but has two stable matchings:

$$
\mu_{DA} = \begin{pmatrix} m_1 & m_2 & m_3 \\
                          w_2 & w_1 & w_3 
\end{pmatrix}; \quad \mu_{DA}' = \begin{pmatrix} m_1 & m_2 & m_3 \\
                                                   w_3 & w_2 & w_1 
\end{pmatrix}.
$$

However, no optimal stable matching exists for $m$ or $w$ because $\mu_{DA}(m_3) \succ_{m_3} \mu_{DA}'(m_3)$, but $\mu_{DA}'(m_2) \succ_{m_2} \mu_{DA}(m_2)$; at the same time, $\mu_{DA}(w_2) \succ_{w_2} \mu_{DA}'(w_2)$, but $\mu_{DA}'(w_3) \succ_{w_3} \mu_{DA}(w_3)$.

A further result on the deferred acceptance algorithm matching given by Roth-Sotomayor (1990) reveals that $\mu_M^{DA}$ is weakly Pareto-efficient among all individ-
ually rational matchings for the group $M$ and the conclusion also holds for the group $W$.

**Theorem 22.2.4** There is no individually rational matching (stable or unstable) $\mu$ such that $\mu \succ_M \mu_M^{DA}$.

**Proof.** Suppose by way of contradiction that there is an individually rational matching $\mu$ such that $\mu(m) \succ_m \mu_M^{DA}(m), \forall m \in M$. Then for any $m$ that is matched to a woman $w$, if she rejected $m$ before in the deferred acceptance algorithm, she intends to accept another $m'$ that is more preferred to $m$ by her (i.e., if $w = \mu(m)$, we must have $\mu_M(w) \succ_w \mu(w)$). Thus, all women in $\mu(M) \subseteq W$ are matched in $\mu_M^{DA}$, i.e. $\forall w \in \mu(M), \mu_M^{DA}(w) \neq w$. Thus, all men are matched in $\mu_M^{DA}$ and $\mu_M^{DA}(M) = \mu(M)$ (since $\mu_M(\mu_M(i)) = i, \forall i \in M \cup W$). Since all men in $\mu_M^{DA}(M)$ are matched to women, any woman who receives a proposal in the last stage of the algorithm must have never rejected an acceptable man. It means that the deferred acceptance algorithm stops when each woman in $\mu_M^{DA}(M)$ receives a proposal from an acceptable man. Thus, the women who receives a proposal in the last stage must remain single in the end of matching $\mu$ (this is because when $w = \mu(m) \succ_m \mu_M^{DA}(m)$ and with the stable $\mu_M, \mu_M^{DA}(w) \succ_w \mu(w)$), which contradicts the fact that $\mu_M^{DA}(M) = \mu(M))$ ■

**Properties of Stable Matchings**

The deferred acceptance algorithm provides two stable matchings $\mu_M^{DA}$ and $\mu_W^{DA}$. However, in many cases, there may be other stable matchings. Here we discuss some basic results on matching stability.

First, we introduce the mathematical concept of lattice.

**Definition 22.2.6 (Lattice)** $(X, \succeq)$ is called a lattice if $\succeq$ satisfies reflexivity, antisymmetry and transitivity, and $\sup_{\succeq}\{x, x'\} \in X$ and $\inf_{\geq}\{x, x'\} \in X$ for any $x, x' \in X$.  

978
Suppose all preferences are strict. If \( \mu \) and \( \mu' \) are stable matchings, then \( \overline{\mu} = \mu \lor M \mu' \) and \( \overline{\mu'} = \mu \lor W \mu' \) are also stable matchings.

**Proof.** We only prove that \( \overline{\mu} \) is a stable matching. The proof of \( \overline{\mu'} \) is similar. The proof consists of two steps. The first step is to prove that \( \overline{\mu} \) is a matching, and the second is to prove that it is stable.

To prove that \( \overline{\mu} \) is a matching, we only need to verify that \( \overline{\mu}(m) = w \) if and only if \( \overline{\mu}(w) = m \). By the stability of \( \mu \) and \( \mu' \), we can derive its necessity: if \( \overline{\mu}(m) = w \), then \( \overline{\mu}(w) = m \). Suppose by way of contradiction that \( \overline{\mu}(w) \neq m \). When \( w = \overline{\mu}(m) = \mu(m) \succ_m \mu'(m) \), then \( \mu(w) = m \succ_w \overline{\mu}(w) = \mu'(w) \), and thus \( (m, w) \) forms a blocking pair for \( \mu' \). Similarly, when \( w = \overline{\mu}(m) = \mu'(m) \succ_w \mu(m) \), \( (m, w) \) will form a blocking pair for \( \mu \).

**Sufficiency:** Let \( M' = \{ m : \overline{\mu}(m) \in W \} = \{ m : \mu(m) \in W, \text{ or } \mu'(m) \in W \} \). By the necessity we just proved, \( \overline{\mu}(M') \subseteq W' = \{ w : \overline{\mu}(w) \in M \} = \{ w : \mu(w) \in M, \text{ and } \mu'(w) \in M \} \). Since \( \overline{\mu}(m) = \overline{\mu}(m') = w \), then \( m = m' = \overline{\mu}(w) \), and by \( |\overline{\mu}(M')| = |M'| \geq |\mu W| = |W'| \) where \( |\cdot| \) is a function on the number of elements in a set, we have \( \overline{\mu}(M') = W' \). Thus, for any \( w \in W' \), there exists \( m \in M' \) such that \( m = \overline{\mu}(w) \). If \( w \notin W' \), then surely we have \( \overline{\mu}(w) = w \).

With the previous necessity condition, if \( \overline{\mu}(w) = m \), then \( \overline{\mu}(m) = w \).
Next, we need to prove that $\bar{\mu}^M$ is stable. Suppose that $(m, w)$ is a blocking pair for $\bar{\mu}^M$, then $w \succ_m \bar{\mu}^M(m)$, which means $w \succ_m \mu(m)$ and $w \succ_m \mu'(m)$. Also, $m \succ_w \bar{\mu}^M(w)$ means $m \succ_w \mu(w)$ or $m \succ_w \mu'(w)$, contradicting the fact that $\mu$ and $\mu'$ being stable. ■

In addition, the agents who are matched in a stable matching have the following properties.

**Theorem 22.2.6 (McVitie-Wilson, 1970)** In all stable matchings, the set of matched men and women is the same.

**Proof.** Let $\mu_M$ be the optimal stable matching for $M$ and $\mu$ be any other stable matching. Thus, all men prefer $\mu_M$ and all women prefer $\mu$. Define $MM(\mu_M) = \{m : \mu_M(m) \neq m\}$ as the set of the men who are matched in $\mu_M$ and $WM(\mu_M) = \{w : \mu_M(w) \neq w\}$ as the set of the women who are matched in $\mu_M$. Define $MM(\mu)$ and $WM(\mu)$ in a similar way. Since all men prefer $\mu_M$, we have $MM(\mu) \subseteq MM(\mu_M)$ and hence $|MM(\mu_M)| \geq |MM(\mu)|$; since all women prefer $\mu$, we have $WM(\mu_M) \subseteq WM(\mu)$ and thus $|WM(\mu_M)| \leq |WM(\mu)|$. With $|MM(\mu_M)| \geq |WM(\mu_M)|$, we have $MM(\mu) = MM(\mu_M)$ and $WM(\mu) = WM(\mu)$. ■

What follows is a discussion on the relationship between the core and the stable matching. An allocation in a cooperative game is said to have the core property if it cannot be blocked by a coalition of some agents, that is, the coalition cannot make each member better off under another allocation. In a matching, members in different groups can form coalitions to block matchings. We now define the notion of core of the marriage market (See Roth-Sotomayor, 1990).

**Definition 22.2.7 (Core of Marriage Market)** An outcome in a matching $\mu$ of the marriage market is in the core, if there is no other matching $\mu'$ and coalition $A \subseteq W \cup M$ such that

\begin{align*}
\mu'(m) &\in A, \forall m \in A; \\
\mu'(w) &\in A, \forall w \in A; \\
\mu'(m) &\succ_m \mu(m), \forall m \in A; \\
\mu'(w) &\succ_w \mu(w), \forall w \in A,
\end{align*}

(22.1) (22.2) (22.3) (22.4)
where conditions (22.1) and (22.2) mean that coalition $A$ consists of agents who are matched to other agents in the new matching $\mu'$; and conditions (22.3) and (22.4) mean that for the coalition members, the outcomes of the new matching $\mu'$ are better than that of $\mu$, that is, $\mu'$ dominates $\mu$ under coalition $A$. In other words, the outcomes of matching $\mu$ have the core properties when there is no other coalition that can produce matching outcomes $\mu'$ dominant over $\mu$.

The following theorem given by Roth-Sotomayor (1990) reveals the relationship between the core and the stable matching.

**Theorem 22.2.7** The core of the marriage market coincides with the set of stable matchings.

**Proof.** First, we want show that the core of each marriage market is a stable matching. If a matching $\mu$ does not satisfies individual rationality, i.e. there exists $i$ such that $i \succ_i \mu(i)$, then $\mu$ is blocked by the coalition consisting of $A = \{i\}$, thus $\mu'(i) = i \succ_i \mu(i)$. If there exists a blocking pair $(m, w)$ that can make $m \succ_w \mu(w)$ and $w \succ_m \mu(m)$, then $\mu$ is blocked by the coalition consisting of $A = \{m, w\}$ and for coalition $A$, and thus $\mu'(m) = w$ and $\mu'(w) = m$ dominate $\mu$. Hence, unstable matching cannot serve as the core of the marriage market.

Next, we show that if a matching is not in the core of the marriage market, then it is unstable. Suppose that $\mu$ is not in the core of the marriage market, then there exists a coalition $A$ with matching outcomes $\mu'$ that dominates $\mu$. Suppose $\mu$ is a stable matching, then $\mu$ is individually rational, which means $\mu'(w) \in M, \forall w \in A$. Let $w \in A$, and $m = \mu'(w) \in A$ means $\mu'(m) = w \succ_m \mu(m)$. Thus, $(m, w)$ forms a blocking pair for $\mu$, contradicting the hypothesis that $\mu$ is a stable matching. ■

**Strategic Behaviors of Matchings**

In reality, agents' preferences are usually private information, which may lead agents to engage in strategic behavior of providing false information, that is, to
misrepresent their true preferences to achieve a better matching outcome. We now focuses on the issue of strategic behavior.

We use the mechanism design approach to study the incentive issues for agents in a matching. Suppose that agent $i \in W \cup M$ has strict preference $\succ_i$. Let $M = \{m_1, \ldots, m_K\}, W = \{w_1, \ldots, w_L\}$ and let the preferences of all agents be $\succ = (\succ_{m_1}, \ldots, \succ_{m_K}, \succ_{w_1}, \ldots, \succ_{w_L})$. Suppose that the set of all possible preferences is $\mathcal{P}$ and the set of all possible matchings is $\mathcal{M}$. A direct mechanism is $(f, \mathcal{P}, \mathcal{M})$, where $f : \mathcal{P} \to \mathcal{M}$, and it determines a matching $f(\succ) = \mu \in \mathcal{M}$ for each preference $\succ$.

**Definition 22.2.8** A mechanism $f$ is stable, if for any $\succ \in \mathcal{P}$, $\mu = f(\succ)$ is a stable matching; a mechanism $f$ is (weakly) Pareto-efficient, if for any $\succ \in \mathcal{P}$, $\mu = f(\succ)$ is (weakly) Pareto-efficient; a mechanism $f$ is individually rational, if for any $\succ \in \mathcal{P}$, $\mu = f(\succ)$ is an individually rational matching.

Next, we introduce the notion of incentive-compatibility mechanism.

**Definition 22.2.9** A mechanism $f$ is strategy-proof, if it satisfies:

$$f(\succ_i, \succ_{-i})(i) \succ_i f(\succ_i', \succ_{-i})(i), \text{ for } \forall i \in M \cup W, \forall \succ_i, \succ_i' \in \mathcal{P}_i, \succ_{-i} \in \mathcal{P}_{-i}.$$  

Roth (1982) proved the following theorem.

**Theorem 22.2.8 (Impossibility Theorem (Roth))** In one-to-one matching, there is no mechanism that is stable and strategy-proof.

**Proof.** The proof is simple since we only need to find a counterexample. The following example comes from Abdulkadiroglu and Sonmez (2013), which reveals that no mechanism satisfies the stability and strategy-proofness.

Consider the case of two men and two women, with preferences given by:

$$w_1 \succ_{m_1} w_2 \succ_{m_1} m_1;$$
$$w_2 \succ_{m_2} w_1 \succ_{m_2} m_2;$$
$$m_2 \succ_{w_1} m_1 \succ_{w_1} w_1;$$
$$m_1 \succ_{w_2} m_2 \succ_{w_2} w_2.$$
We can verify that there are only two stable matchings: $\mu_M$ with stable matching outcomes $(m_1, w_1), (m_2, w_2)$ and $\mu_W$ with stable matching outcomes $(m_1, w_2), (m_2, w_1)$.

Next we need to prove that these two stable matchings can potentially be manipulated. Let $\varphi$ be a stable matching:

(1) When $\varphi = \mu_M$, if $w_1$ misstate its preference as $m_2 \succ_w w_1 \succ_w m_1$, then $w_1$ could be better off at $\mu_M$, $\mu_M(w_1| \succ_w m_1) = m_2$;

(2) When $\varphi = \mu_W$, if $m_1$ misstate its preference as $w_1 \succ m_1 \succ m_1 \succ w_2$, then $m_1$ could be better off at $\mu_W$, $\mu_W(m_1| \succ m_1) = w_1$.

The corollary below follows the Impossibility Theorem.

**Corollary 22.2.1** No stable matching mechanism exists such that truthtelling their preferences consist of a Nash equilibrium.

Roth’s Impossibility Theorem reveals that strategy-proofness is in conflict with stability. The following theorem reveals that it is also in conflict with Pareto-efficiency and individual rationality.

**Theorem 22.2.9 (Alcalde-Barbera,1994)** For any matching, there is no mechanism that is Pareto-efficient, individually rational and strategy proof.

**Proof.** Consider an economy with two men and two women, $M = \{m_1, m_2\}$ and $W = \{w_1, w_2\}$. Suppose $f$ is any Pareto-efficient and individually rational matching mechanism. We want to show that it is not strategy-proof.

Consider the following five cases of preferences:

**Case 1:**
\[
\begin{align*}
p^1(m_1) &= w_1, w_2; & p^1(w_1) &= m_2, m_1; \\
p^1(m_2) &= w_2, w_1; & p^1(w_2) &= m_1, m_2.
\end{align*}
\]

**Case 2:**
\[
\begin{align*}
p^1(m_1) &= w_1, w_2; & p^1(w_1) &= m_2; \\
p^1(m_2) &= w_2, w_1; & p^1(w_2) &= m_1, m_2.
\end{align*}
\]

**Case 3:**
\[
\begin{align*}
p^1(m_1) &= w_1, w_2; & p^1(w_1) &= m_2; \\
p^1(m_2) &= w_2, w_1; & p^1(w_2) &= m_1.
\end{align*}
\]
Case 4:
\[ p^1(m_1) = w_1; \quad p^1(w_1) = m_2, m_1; \]
\[ p^1(m_2) = w_2, w_1; \quad p^1(w_2) = m_1, m_2. \]

Case 5:
\[ p^1(m_1) = w_1; \quad p^1(w_1) = m_2, m_1; \]
\[ p^1(m_2) = w_2; \quad p^1(w_2) = m_1, m_2. \]

All corresponding individually rational and Pareto-efficient matchings for the above five cases are:

\[ \mu_1^1 = \begin{pmatrix} m_1 & m_2 \\ m_2 & w_1 \\ w_2 & w_1 \end{pmatrix}, \quad \mu_1^2 = \begin{pmatrix} m_1 & m_2 \\ w_2 & w_1 \end{pmatrix}; \]
\[ \mu_2^1 = \begin{pmatrix} m_1 & m_2 & w_1 \\ m_1 & w_2 & w_1 \end{pmatrix}, \quad \mu_2^2 = \begin{pmatrix} m_1 & m_2 \\ w_2 & w_1 \end{pmatrix}; \]
\[ \mu_3 = \begin{pmatrix} m_1 & m_2 \\ w_2 & w_1 \end{pmatrix}; \]
\[ \mu_4^1 = \begin{pmatrix} m_1 & m_2 \\ w_1 & w_2 \end{pmatrix}, \quad \mu_4^2 = \begin{pmatrix} m_1 & m_2 & w_2 \\ m_1 & w_1 & w_2 \end{pmatrix}; \]
\[ \mu_5 = \begin{pmatrix} m_1 & m_2 \\ w_1 & w_2 \end{pmatrix}. \]

Suppose \( f(p^1) = \mu_1^1 \). If \( f(p^2) = \mu_2^2 \), then under preference \( p^1(w_1) \), \( w_1 \) has the incentive to misstate its preference as \( p^2(w_1) \); if \( f(p^2) = \mu_1^2 \), then under preference \( p^2(w_2) \), \( w_2 \) has the incentive to misstate its preference as \( p^3(w_2) \). Suppose \( f(p^1) = \mu_2^1 \). If \( f(p^2) = \mu_1^2 \), then under preference \( p^1(m_1) \), \( m_1 \) has the incentive to misstate its preference as \( p^4(m_1) \); if \( f(p^2) = \mu_1^2 \), then under preference \( p^2(m_2) \), \( m_2 \) has the incentive to misstate its preference as \( p^5(m_2) \).

As seen in the example above, we have exhausted all possible matchings that are individually rational and Pareto-efficient, and they all provide incentives for agents to manipulate their preferences. ■
However, if we only consider the incentive within a single group, stability can be compatible with strategy proofness from the following theorem.

**Theorem 22.2.10 (One-Sided Stable Strategy-Proof Possibility Theorem)**
The men-optimal stable matching makes it a weakly dominant strategy for each man to tell the truth (state his true preference); the women-optimal stable matching makes it a weakly dominant strategy for each woman to tell the truth (state her true preference);

Before proving the above theorem, we first introduce “Blocking Lemma” proved by Hwang, Gale-Sotomayor (1985), which has many applications in the matching theory. By the lemma, if the set of men who prefer individually rational matching $\mu$ to men-optimal stable matching $\mu_M$ is not empty, then there must be a matching pair $(m, w)$ that blocks the individually rational matching $\mu$, where $m$ prefers the optimal stable matching. The lemma goes as follows:

**Lemma 22.2.1 (Blocking Lemma)** Let $\mu$ be an individually rational matching under preference $\succ$, let $\mu_M$ be the men-optimal stable matching, and let $M'$ be the set of the men in $M$ who prefer $\mu$. If $M' \neq \emptyset$, then there exists a pair $(m, w)$ that blocks $\mu$, where $m \in M - M' \equiv M \setminus M'$ and $w \in \mu(M')$.

**Proof.** Since $\mu \succ_{M'} \mu_M$, in the set $M'$, the agents who are matched in $\mu_M$ will also be matched in $\mu$. Thus, we can consider the following two cases.

Case 1: $\mu(M') \neq \mu_M(M')$. Then, there exists $w \in \mu(M') - \mu_M(M')$. Let $m' = \mu(w)$ and $m = \mu_M(w)$. Since $m' \in M'$, we have $w \succ_{m'} \mu_M(m')$, which means $m \succ_w m'$, otherwise $(m', w)$ will form a blocking pair for $\mu_M$, contradicting $\mu_M$. Since $w \notin \mu_M$, $m \notin M'$ means $w \succ_m \mu(m)$. Hence, $(m, w)$ forms a blocking pair for $\mu$.

Case 2: $\mu(M') = \mu_M(M')$. Let $w$ be the last woman who receives a proposal from $m'$ in group $M'$ in the deferred acceptance algorithm, $m' = \mu_M(w)$. Since all women in $W'$ have previously rejected a proposal from agents in group $M'$ in the deferred acceptance algorithm, we may as well suppose $m$ to be the man held by $w$ without commitment when she receives a proposal from $m'$. We need to
show that \((m, w)\) forms a blocking pair for \(\mu\). First, \(m \notin M'\). Suppose by way of contradiction that \(m \in M'\). After being rejected by \(w\), he will propose to \(w'\) in group \(W'\), leading to a contradiction with \(w\) being the last one who receives a proposal from group \(M'\).

However, \(m\) being rejected by \(w\) in the deferred acceptance algorithm means \(w \succ_m \mu_M(m)\). Since \(m \notin M'\), which means \(\mu_M(m) \succ_m \mu(m)\), we have \(w \succ_m \mu(m)\). In addition, since \(m\) is the last man rejected by \(w\), then \(w\) must have rejected \(\mu(w) \neq \mu_M(w) = m'\) before rejecting \(m\). Thus, \((m, w)\) forms a blocking pair for \(\mu\).

Using Blocking Lemma above, we can have a stronger result than Theorem 22.2.10.

**Theorem 22.2.11 (Demange-Gale-Sotomayor, 1987)** Let \(p\) be a preference profile (not necessarily strict) of agents, and let \(\bar{p}\) differ from \(p\) in that there is a subset \(A\) of \(M \cup W\) such that under \(\bar{p}\) agents have the incentives to misstate their preferences. Then there is no stable matching \(\mu\) under \(\bar{p}\) such that for any stable matching \(\hat{\mu}\), \(\mu \succ_A \hat{\mu}\).

**Proof.** Suppose by way of contradiction that there is a non-empty subset \(A = M \cup W \subseteq M \cup W\) in which agents have the incentives to misstate their true preferences, and they are better off in a stable matching \(\mu\) under \(\bar{p}\) than in all other matchings under \(p\). If \(\mu\) is not individually rational under \(p\), which may have \(i \succ_p(i)\mu(i)\), then \(i\) will surely misstate his preferences and \(i \in M\), contradicting the fact that \(i\) is better off in \(\mu\) than in all stable matchings under \(p\).

Suppose \(\mu\) is individually rational under \(p\) and obviously, \(\mu\) is unstable under \(p\). Now we construct a profile of strict preferences \(p'\): if any agent \(i\) weakly prefers a matching mate \(j\) under \(p\), then under \(p'\), \(i\) strictly prefers \(j\). Thus, we have: if \((m, w)\) forms a blocking pair for \(\mu\) under \(p'\), then \((m, w)\) forms a blocking pair for \(\mu\) under \(p\). Each stable matching under \(p'\) is also stable under \(p'\). Let \(\mu_M\) and \(\mu_W\) be the optimal stable matchings for men and women under \((M, W, p')\). By the assumption of \(\mu\) (i.e. \(A\) is better off in \(\mu\) than in any other stable matching
under $p$), we have:

$$\mu(m) \succ_m \mu_M(m), \forall m \in \tilde{M}, \quad (22.5)$$

$$\mu(w) \succ_w \mu_W(w), \forall w \in \tilde{W}. \quad (22.6)$$

If $\tilde{M}$ is not empty, by Blocking Lemma, in the matching market $(M, W, p')$, we have $\tilde{M} \subseteq M'$ in (22.5). Thus $(m, w)$ forms a blocking pair for $\mu$ under $p'$, and it is also a blocking pair for $\mu$ under $p$, making $\mu_M(m) \succ_m \mu(m)$ and $\mu_M(w) \succ_w \mu(w)$, which means $m, w \notin A$ and they will not misstate their preferences under $p$. So, it is also a blocking pair for $\mu$ under $\tilde{p}$, contradicting the stability of $\mu$. When $\tilde{M} = \emptyset$, the proof is similar under $\tilde{W} = \emptyset$.

The corollary below follows Theorem 22.2.11.

**Corollary 22.2.2** (Dubins-Freedman, 1981; Roth, 1982) *Let $p$ be a profile of preferences (not necessarily strict) of the agents, and let $\tilde{p}$ differ from $p$ in that there is a subset $A$ of $M$ under $\tilde{p}$ such that agents have the incentives to misstate their preferences. Then there is no stable matching $\mu$ for $\tilde{p}$ such that $\mu \succ_A \mu_M$.***

Corollary 22.2.2 reveals that there is no subset of men in the matching $\mu_M$ that can achieve better matching outcomes by misstating their preferences. Thus, Corollary 22.2.2 implies Theorem 22.2.10. Theorem 22.2.10 reveals that men-optimal stable matching $\mu_M$ makes it a weakly dominant strategy for all men to state their true preferences and the women-optimal stable matching $\mu_W$ makes it a weakly dominant strategy for all women to state their true preferences. By Roth’s Impossibility Theorem 22.2.8, we have the following more specific corollary:

**Corollary 22.2.3** *In the men-optimal stable matching $\mu_M$, there is at least one woman who misstates her preference; in the women-optimal stable matching $\mu_W$, there is at least one man who misstates his preference.*

### 22.2.2 Many-to-One Matching

The matchings in the real world are mostly between institutions and individuals, which are usually many-to-one matchings, such as college admissions, hospital-
s/residents, etc. While many results on many-to-one matchings can be viewed as direct extensions of one-to-one matching, there are other conclusions that are different from those in one-to-one matchings. Here we set up a many-to-one matching model based on college admissions and hospitals/residents.

Suppose there are two sets of agents: one is the set of institutions $H = \{h_1, \ldots, h_n\}$, and the other is the set of individuals $D = \{d_1, \ldots, d_m\}$. In the example of college admissions, $H$ is the colleges and $D$ is the students and in the example of medical graduates, $H$ is the hospitals and $D$ is the medical graduates. Institution $h \in H$ will select some agents from $D$. Suppose the upper bound for the number of the agents that one institution can choose is $q_h$, called the quota of institution $h$; Individual $d \in D$ can choose only one institution or doesn’t choose at all. When $q_h = 1, \forall h \in H$, it becomes a one-to-one matching.

Next we discuss the preferences of institutions and individuals. The individual’s (strict) preference $\succ_d, d \in D$ is the same as that in one-to-one matching, or equivalently, we use $p(d)$ to presents the preference order of institutions that are acceptable to $d$, arranging from the highest to the lowest. However, institutions’ preferences are more complicated in the many-to-one matching. Because institutions demand multiple individuals, the relationship between individuals may affect the preferences of institutions. The preference of an institution is expressed as $\succ_h, h \in H$, and $p(h)$ presents an order list of individuals that are acceptable to $h$, arranging from the highest to the lowest.

We first define the notion of many-to-one matching.

**Definition 22.2.10** The set of institutions $H$ and the set of individuals $D$ constitute a matching $\mu : H \cup D \rightarrow 2^{H \cup D}$, if it satisfies the following properties:

- $\mu(h) \subseteq D \cup \{h\}$, satisfying: $|\mu(h)| \leq q_h, \forall h \in H$;
- $\mu(d) \in H \cup \{d\}, \forall d \in D$;
- $d \in \mu(h)$ if and only if $\mu(d) = h, \forall h \in H, d \in D$.

Next is the definition of many-to-one stable matching.
Definition 22.2.11 A matching $\mu$ between institutions $H$ and individuals $D$ is a stable matching, if the following requirements are satisfied:

1. if $\mu(h) \neq \{h\}$, then $\mu(h) \succ_h \{h\}$;
2. if $\mu(d) \neq d$, then $\mu(d) \succ_d d$;
3. there does not exist any coalition $A \in H \cup D$ of institutions and individuals and matching $\mu'$ on the coalition such that:

\[
\begin{align*}
\mu'(d) &\in A, \forall d \in A; \\
\text{if } d' \in \mu'(h), \text{ then } d' &\in A \cup \mu(h), \text{ and } \mu'(h) \leq |q_h|, \forall h \in A; \\
\mu'(d) &\succ_d \mu(d), \forall d \in A; \\
\mu'(h) &\succ_h \mu(h), \forall h \in A.
\end{align*}
\]

A stable matching, first, should satisfy individual rationality and second, can prevent profitable deviations by either group. The profitable deviation of a group is defined as follows: when $\mu' \in A$ is a profitable deviation for matching $\mu$, Condition (22.7) means that the matching mate of individual $d$ in the coalition is an institution in the coalition; Condition (22.8) means that in the deviation, the institution’s matching mate is either from the coalition or the original matching mate; Condition (22.9) means that the individuals in the coalition are better off in the new matching; Condition (22.10) means that the institutions in the coalition are better off in the new matching, and if the matching mates of the institutions are made more profitable in coalition $A$, then they stay the same in the original matching $\mu(h)$. Thus, such a coalition is feasible as well as makes its members more profitable.

However, a stable matching may not even exist if institution preferences are not restricted.

Example 22.2.5 (Roth and Sotomayor, 1990) Consider a matching with two institutions $H = \{h_1, h_2\}$, $q_{h_1} = q_{h_2} = 2$ and three individuals $D = \{d_1, d_2, d_3\}$. 

989
Suppose the preferences of institutions and individuals are as follows:

\[
p(h_1) = \{d_1, d_3\}, \{d_1, d_2\}, \{d_2, d_3\}, \{d_1\}, \{d_2\};
p(h_2) = \{d_1, d_3\}, \{d_1, d_2\}, \{d_2, d_3\}, \{d_3\}, \{d_1\}, \{d_2\};
p(d_1) = h_2, h_1;
p(d_2) = h_2, h_1;
p(d_3) = h_1, h_2.
\]

Now we consider all individually rational matchings. There are five such matchings altogether where all individuals are matched:

\[
\mu_1 = \begin{pmatrix}
h_1 & h_2 \\
\{d_1, d_3\} & \{d_2\}
\end{pmatrix},
\]
which is blocked by coalition \( A = \{h_2, d_1\} \) and \( \mu'(h_2) = \{d_1, d_2\}, \mu'(d_1) = h_2; \)

\[
\mu_2 = \begin{pmatrix}
h_1 & h_2 \\
\{d_1, d_2\} & \{d_3\}
\end{pmatrix},
\]
which is blocked by coalition \( A = \{h_2, d_1\} \) and \( \mu'(h_2) = \{d_1, d_3\}, \mu'(d_1) = h_2; \)

\[
\mu_3 = \begin{pmatrix}
h_1 & h_2 \\
\{d_2, d_3\} & \{d_1\}
\end{pmatrix},
\]
which is blocked by coalition \( A = \{h_2, d_2\} \) and \( \mu'(h_2) = \{d_1, d_3\}, \mu'(d_1) = h_2; \)

\[
\mu_4 = \begin{pmatrix}
h_1 & h_2 \\
\{d_1\} & \{d_2, d_3\}
\end{pmatrix},
\]
which is blocked by coalition \( A = \{h_1, d_3\} \) and \( \mu'(h_1) = \{d_1, d_3\}, \mu'(d_3) = h_1; \)

\[
\mu_5 = \begin{pmatrix}
h_1 & h_2 \\
\{d_2\} & \{d_1, d_3\}
\end{pmatrix},
\]
which is blocked by coalition \( A = \{h_1, d_3\} \) and \( \mu'(h_1) = \{d_2, d_3\}, \mu'(d_3) = h_1. \)
In addition, if agent $d \in D$ is not able to be matched, i.e. $\mu(d) = d$, then there exists at least one $h \in H$, $|\mu(h)| \leq 1$. In this case, $(d, h)$ forms a profitable deviation coalition $\mu'(h) = \mu(h) \cup \{d\}$ and $\mu'(d) = h$ forms a profitable deviation.

For each possible matching above that satisfy individual rationality, we can always find a profitable deviation coalition. Thus, there exists no stable matchings.

In order to overcome the non-existence of stable matchings in the many-to-one matching, Roth (1985) introduced a concept of responsiveness for institutions’ preferences.

**Definition 22.2.12 (Responsive Preference)** Institutions $H$ have responsive preferences, if for $\succ_h$, $\forall h \in H$, the following conditions are satisfied:

(i) for any $J \subseteq D$, $|J| < q_h$, $d \in D \setminus J$, we have:

$$\{d\} \cup J \succ_h J \iff \{d\} \succ_h \emptyset;$$

(ii) for any $J \subseteq D$, $|J| < q_h$, $d, d' \in D \setminus J$, we have:

$$\{d\} \cup J \succ_h \{d'\} \cup J \iff \{d\} \succ_h \{d'\}.$$

Obviously, Example 22.2.5 does not have responsive preferences, since for $h_2$, though $\{d_1, d_2\} \succ_h \{d_2, d_3\}, \{d_3\} \succ_h \{d_1\}$.

When preferences are responsive, the definition of stable matching is simpler, which is equivalent to Definition 22.2.11.

**Definition 22.2.13** Matching $\mu$ between $H$ and $D$ with responsive preferences is stable, if the following conditions are satisfied:

(i) for $\mu(h) \neq \{h\}$, $\mu(h) \succ_h \{h\}$;

(ii) for $\mu(d) \neq d$, $\mu(d) \succ_d d$;

(iii) there is no blocking pair $(h, d)$ such that:

$$h \succ_d \mu(d); \quad (22.11)$$

$$\exists d' \in \mu(h) \text{ with } \{d\} \succ_h \{d'\}, \text{ or } |\mu(h)| < q_h \text{ and } \{d\} \succ_h \emptyset. \quad (22.12)$$
We show that Definition 22.2.11 and Definition 22.2.13 are equivalent. To do so, we only need to show that the four conditions ((22.7), (22.8), (22.9), and (22.10)) of $A$ and $\mu'$ in Definition 22.2.11 are equivalent to the two conditions (22.11) and (22.12) of $(h,d)$ in Definition 22.2.13.

Obviously, if there exists a blocking pair $(h,d)$ for $\mu$, the corresponding coalition is $A = \{h,d\}$. Then $\mu'(h) = \{d\} \cup \mu(h) \setminus \{d'\}$, and $\mu'(d) = h$ satisfies conditions ((22.7), (22.8), (22.9) and (22.10)).

If $\mu$ is blocked by $A$ and $\mu'$, by $\mu'(h) \succ_h \mu(h)$ and the responsiveness of preferences, there exist $d \in \mu'(h) - \mu(h)$ and $d' \in \mu(h) - \mu'(h)$ such that $\{d\} \succ_h \{d'\}$. Otherwise, for all $d' \in \mu(h) - \mu'(h)$ and $d \in \mu'(h) - \mu(h)$ that satisfy $\{d'\} \succeq_h \{d\}$, we have $\mu(h) \succeq_h \mu'(h)$, and then $(h,d)$ forms a blocking pair of $\mu$.

From the above equivalent definition, we find that the conditions for stable many-to-one matching are very similar to those for stable one-to-one matching. The deferred acceptance algorithm can also be applied to many-to-one matchings. We give the institution-proposing deferred acceptance algorithm in the following.

**Definition 22.2.14 (Deferred Acceptance Algorithm with Institution-Proposing)**

Step 1 Each institution $h \in H$ proposes to its $q_h$ most preferred individuals, and each individual who receives a proposal accepts its most preferred institution and rejects the rest;

Step $k$: Assuming that in Step $k - 1$, institution $h$ is rejected by individual $d$, and $h$ proposes to its $q_h$ most preferred individuals of those who have never rejected it. Each individual who receives a new proposal has to choose between the new proposal and the one it holds from previous step, accepts the optimal one and rejects the rest.

When no further proposals are made, the matching ends. Each individual is matched with the institution that proposes to the individual in the last step.

The individual-proposing deferred acceptance algorithm is similar to the one above. Each individual proposes to its most preferred institutions that have never rejected it in the previous steps. Institution $h$ chooses the most $q_h$ preferred
proposals among its acceptable individuals from whom it receives proposals and rejects the rest.

**Example 22.2.6** Consider a matching with 2 institutions and 2 individuals, $H = \{h_1, h_2\}$ and $D = \{d_1, d_2\}$ whose preferences are given by:

- $p(h_1) = \{d_1, d_2\}, \{d_1\}, \{d_2\}$;
- $p(h_2) = \{d_1\}, \{d_2\}$;
- $p(d_1) = h_1, h_2$;
- $p(d_2) = h_2, h_1$.

The matching produced by the institution-proposing deferred acceptance algorithm is:

$$\mu_H = \begin{pmatrix} h_1 & h_2 \\ d_1 & d_2 \end{pmatrix}.$$ 

It can be verified that the matching $\mu_D = \mu_H$ produced by the individual-proposing deferred acceptance algorithm is the only stable matching.

We can directly apply the conclusion of one-to-one matching on the stability of deferred acceptance algorithm to many-to-one matchings:

**Theorem 22.2.12** The matching produced by the institution-proposing (or individual-proposing) deferred acceptance algorithm is stable.

In fact, many conclusions on one-to-one matchings can be extended to the case of many-to-one matchings. In the following, we introduce a transformation of a many-to-one matching to a one-to-one matching, and then prove the stability of the transformed one-to-one matchings is the same as the stability of the many-to-one matchings.

**Definition 22.2.15** A many-to-one matching can be equivalently transformed to a one-to-one matching, if:
(1) divide the institution $h_k$ with quota $q_{hk}$ into branches $h_k^1, \ldots, h_k^{q_{hk}}$ such that the quota for each branch is 1, and the preference of each branch $h_k^i$ satisfies: \{d\} $\succ h_k^i$ \{d'\} if and only if \{d\} $\succ_{h_k}$ \{d'\}; $d$ $\succ h_k^i \emptyset$ if and only if \{d\} $\succ_{h_k} \emptyset$.

(2) the order of preference for each individual $d \in D$ toward the branches of the same institution $h_k$ is the same: $h_k^1 \succ \ldots \succ h_k^{q_{hk}}$, and its preferences for different institutions stay the same. Thus, for the old preference $\succ_d$, if $h_k \succ_d h_l$, then the order of preference for the corresponding branches of the two institutions satisfies $h_k^i \succ_d h_l^j, \forall i \leq q_{hk}, j \leq q_{hl}$.

The transformation of many-to-one matching to one-to-one matching is basically to divide the institution $h_k$ with quota $q_{hk}$ into $q_{hk}$ branches and make all individuals share a common order of preference for the institution’s branches. By doing so, the matching between institutions and individuals has been transformed to a one-to-one matching between branches and individuals. Now, we introduce a lemma first.

**Lemma 22.2.2** A many-to-one matching between institutions and individuals is stable if and only if the corresponding transformed one-to-one matching is stable.

**Proof.** Let $\mu$ be a many-to-one matching of $(H,D,p)$ and let $\bar{\mu}$ be the corresponding matching after the transformation such that:

$$\mu(d) = h_k \text{ if and only if there is } j \leq q_{hk}, \bar{\mu}(d) = h_k^j,$$

$$\bar{\mu}(h_k^j) \in \mu(h_k), \text{ or } \bar{\mu}(h_k^j) = \emptyset, \forall j \leq q_{hk}.$$

We first prove that $\mu$ is not stable means that $\bar{\mu}$ is not stable. If $(h,d)$ is a blocking pair for $\mu$, which means $h \succ_d \mu(d)$ and there exists $d' \in \mu(h)$ such that \{d\} $\succ h$ \{d'\}, then there exists $j \leq q_h$ such that $(h_k^j, d)$ forms a blocking pair for $\bar{\mu}$ or $|\mu(h)| < q_h$ and \{d\} $\succ h \emptyset$. Therefore, there exists $j \leq q_h$ such that $\bar{\mu}(h_k^j) = \emptyset$, which makes $(h_k^j, d)$ a blocking pair for $\bar{\mu}$. If $\mu$ does not satisfy the condition of
individual rationality, then $\emptyset \succ_h \mu(h)$ or $\emptyset \succ_d \mu(d)$. In these two cases, obviously, all the corresponding matchings $\bar{\mu}$ are not individually rational.

In a similar way, we can prove that $\bar{\mu}$ is not stable means that $\mu$ is not stable.

On the basis of Lemma 22.2.2, many results on one-to-one matchings can be extended directly to the many-to-one matchings, and some of the important ones are listed below.

**Corollary 22.2.4** We have the following corollaries:

1. There exists an optimal stable matching $\mu_D$ for individuals, that is, for every individual $d \in D$, $\mu_D$ is better than any other stable matching outcome and the matching by the individual-proposing deferred acceptance algorithm is also $\mu_D$.

2. There exists an optimal stable matching $\mu_H$ for institutions, that is, for every institution $h \in H$, $\mu_H$ is better than any other stable matching outcome and the matching by the institution-proposing deferred acceptance algorithm is also $\mu_H$.

3. The individual-optimal stable matching $\mu_D$ is the least preferred stable matching by institutions; similarly, the institution-optimal stable matching $\mu_H$ is the least preferred stable matching by individuals.

4. For any two stable matchings $\mu$ and $\nu$, the set of matched individuals and the set of matched institutions are the same, i.e.

$$\{d : \mu(d) \neq d\} = \{d : \nu(d) \neq d\}, \forall d \text{ and } |\mu(h)| = |\nu(h)|.$$  

5. Stable matchings have lattice property: if two matchings $\mu$ and $\mu'$ are stable, then $\lambda = \mu \vee_H \mu', \varphi = \mu \vee_D \mu'$ are also stable.

6. There is no individually rational (unstable) matching $\nu$ such that $\nu(d) \succ_d \mu_D(d), \forall d \in D$.  

995
However, not all results on one-to-one matchings can be extended to many-to-one matchings. The following example reveals that for institutions, there may exist an individually rational matching that is Pareto-dominant over the optimal stable matching.

**Example 22.2.7** Consider a matching with two institutions and two individuals, \( H = \{h_1, h_2\} \) and \( D = \{d_1, d_2\} \) with the same preferences as in Example 22.2.6:

\[
\begin{align*}
p(h_1) &= \{d_1, d_2\}, \{d_1\}, \{d_4\}; \\
p(h_2) &= \{d_1\}, \{d_2\}; \\
p(d_1) &= h_1, h_2; \\
p(d_2) &= h_2, h_1.
\end{align*}
\]

The matching produced by the institution-proposing deferred acceptance algorithm is:

\[
\mu_H = \begin{pmatrix} h_1 & h_2 \\ d_1 & d_2 \end{pmatrix}.
\]

But the following (unstable) matching \( \nu \)

\[
\nu = \begin{pmatrix} h_1 & h_2 \\ d_2 & d_1 \end{pmatrix},
\]

is Pareto-dominant over \( \mu_H \) for institutions.

To prove the consistency theorem in many-to-one matchings, we first need to introduce two lemmas (see Knuth, 1976). The first is Decomposition Lemma, which has many applications in the matching theory.

**Lemma 22.2.3 (Decomposition Lemma)** Let \( \mu \) and \( \mu' \) be two stable matchings in the marriage market \( (M, W, p) \). Denote by \( M(\mu') = \{ m : \mu'(m) \succ_m \mu(m) \} \) the set of men who prefer \( \mu' \), and \( W(\mu) \) the set of women who prefer \( \mu \). Then \( \mu \) and \( \mu' \) are one-to-one mappings from \( M(\mu') \) to \( W(\mu) \).
Suppose In a stable matching have corresponding stable matchings after the transformation. If for things between institutions and individuals, and suppose Lemma 22.2.4 (Roth-Sotomayor, 1989) Suppose that there exists such that \( h = \mu'(m) \). \( \mu' \) being stable means \( \mu'(m) \succ_m \mu(m) \), and then \( m \in M(\mu') \). Thus we have \( \mu'(M(\mu')) \subseteq W(\mu) \).

Equivalently, \( w \in W(\mu) \) means \( \mu(w) \succ_w \mu'(w) \). Let \( \mu(w) = m \). \( \mu' \) is stable implies \( \mu'(m) \succ_m \mu(m) \), and then \( m \in M(\mu') \). Thus we have \( \mu(W(\mu)) \subseteq M(\mu') \). Therefore, \( \mu \) and \( \mu' \) are one-to-one mappings from \( M(\mu') \) to \( W(\mu) \).

The second lemma describes the consistency of preferences among an institution’s branches under the transformation.

**Lemma 22.2.4 (Roth-Sotomayor, 1989)** Suppose \( \mu \) and \( \mu' \) are two stable matchings between institutions and individuals, and suppose \( \mu \) and \( \mu' \) are two corresponding stable matchings after the transformation. If for \( h \in H \), \( \mu(h^i) \succ_h \mu'(h^i), i \leq q_h \), then \( \mu(h^j) \succeq_h \mu'(h^j), \forall j \leq q_h \).

**Proof.** Let \( \mu(h^i) \succ_h \mu'(h^i) \). We only need to show that \( \mu(h^j) \succ_h \mu'(h^j), \forall j > i \). Suppose that there exists \( j > i \) such that \( \mu'(h^j) \succeq_h \mu(h^j) \). By Conclusion (4) of Corollary 22.2.4, we have \( \mu'(h^j) = d' \subseteq D \). By Lemma 22.2.3, then \( h^i = \mu'(d') \succ_d \mu'(d') \). Moreover, since \( \mu'(h^j) \succ_h \mu'(h^j) \) (otherwise \( (h^i, \mu'(h^j)) \) forms a blocking pair for \( \mu' \)), we have \( d' \succ_h \mu'(h^j) \succ_h \mu(h^j) \). By Lemma 22.2.3, \( \mu'(d') \neq h^j \), and so \( h^j \succ_d \mu'(d') \), which means \( (d', h^j) \) forms a blocking pair for \( \mu' \), contradicting \( \mu' \).

Thus we have the following consistency theorem for many-to-one matchings.

**Theorem 22.2.13 (Roth, 1986)** In a stable matching \( \mu \), if \( h \) is not fully matched to its quota, i.e., \(|\mu(h)| < q_h\), then for any other stable matching \( \nu \), we have \( \mu(h) = \nu(h) \).

**Proof.** Suppose \( h \in H \) satisfies \(|\mu(h)| < q_h\). By Conclusion (4) of Corollary 22.2.4, \( \mu(h^k) = \emptyset \) if and only if \( \mu'(h^k) = \emptyset \). If there exists \( i < q_h \) such that \( \mu(h^i) \neq \mu'(h^i) \), and, say, \( \mu(h^i) \succ_h \mu'(h^i) \), then for all \( j \) satisfying \( q_h \geq j > i \), we have \( \mu(h^j) \succ_h \mu'(h^j) \). However, it contradicts the fact that \( \mu(h^{q_h}) = \mu'(h^{q_h}) = \emptyset \).
Theorem 22.2.14 (Roth-Sotomayor, 1989) Let \( \mu \) and \( \nu \) be two stable matchings between institutions and individuals. If for \( h \in H, \mu(h) \succ_h \nu(h) \), then for any \( d \in \mu(h) \setminus \nu(h) \) and \( d' \in \nu(h) \setminus \mu(h) \), we have \( d \succ h d' \).

**Proof.** Consider the transformation to one-to-one matching between institutions and individuals. \( \bar{\mu} \) and \( \bar{\nu} \) are the stable matchings corresponding to \( \mu \) and \( \nu \), respectively. If for \( h \in H, \mu(h) \succ_h \nu(h) \), then by Theorem 22.2.13, there must be \( |\mu(h)| = |\nu(h)| = q_h \) and also \( \nu(h) \setminus \mu(h) \neq \emptyset \). Thus, there exist \( d' \in \nu(h) \setminus \mu(h), h^i, \) and \( d' = \bar{\nu}(h^i) \) such that \( \bar{\nu}(h^i) \neq \bar{\mu}(h^i) \). By Lemma 22.2.4, there exists \( h^i \) such that \( \bar{\nu}(h^i) \succ_h \bar{\mu}(h^i) \) and by Lemma 22.2.3, we have \( h^i \succ_{d'} \bar{\mu}(d') \), which implies \( h \succ_{d'} \mu(d') \). Since \( \mu \) is stable and \( \forall d \in \mu(h) - \nu(h) \), we have \( d \succ_h d' \). 

**Incentive Problem**

In the following we discuss the incentives (of revealing preferences) of matchings between individuals and institutions. Since a many-to-one matching between individuals and institutions can be equivalently transformed to a one-to-one matching, the following conclusion is a direct extension of the incentive-theorem drawn from the marriage market.

**Corollary 22.2.5 (Corollaries on Incentives on Many-to-One Matching)**

The following corollaries hold:

1. In a many-to-one matching between individuals and institutions, there exists no mechanism that is both strategy-proof and stable.

2. In a many-to-one matching between individuals and institutions, there exists no mechanism that is Pareto-efficient, individually rational and strategy-proof.

3. The matching by the individual-proposing deferred acceptance algorithm is the optimal stable matching for individuals and makes it a weakly dominant strategy for each individual to state its true preference.
However, there’s no such corollary like (3) in Corollary 22.2.5 for institutions. The following is the theorem regarding the impossibility for institutions to truthfully reveal their preferences.

**Theorem 22.2.15** There does not exist any stable matching (even for the optimal stable matching for institutions) such that it is a weakly dominant strategy for institutions to truthfully reveal their preferences.

To prove this theorem, we only need to find one counterexample.

**Example 22.2.8** Still consider the matching with two institutions and two individuals $H = \{h_1, h_2\}$ and $D = \{d_1, d_2\}$ with the same preferences as in Example 22.2.6:

\[
\begin{align*}
p(h_1) &= \{d_1, d_2\}, \{d_2\}, \{d_1\}; \\
p(h_2) &= \{d_1\}, \{d_2\}; \\
p(d_1) &= h_1, h_2; \\
p(d_2) &= h_2, h_1.
\end{align*}
\]

According to the analysis in Example 22.2.6, the matching by the individual-proposing deferred acceptance algorithm is the only stable matching:

\[
\mu_H = \begin{pmatrix} h_1 & h_2 \\ d_1 & d_2 \end{pmatrix}.
\]

Consider $h_1$. If it manipulates by stating preference as $p'(h_1) = \{d_2\}$, which means only individual $d_2$ is better than its own matching mate, then for $p' = (p'(h_1), p(h_2), p(d_1), p(d_2))$, the only stable matching is:

\[
\mu'_H = \begin{pmatrix} h_1 & h_2 \\ d_2 & d_1 \end{pmatrix}.
\]

Then, $\mu'_H(h_1) \succ_{h_1} \mu_H(h_1)$. Thus we have shown that stability and strategy-proofness are in conflict in the case of institutions.
Since each institution is matched to more than one individuals, another kind of manipulation by institution is via misrepresenting their capacities (while the institution states its true preferences).

**Definition 22.2.16** Matching mechanism \( f \) is called **non-manipulable via capacities** if there is no \( h \in H \) that can profitably manipulate its capacity, that is:

\[
f(p, q'_h, q_{-h})(h) \succ_h f(p, q_h, q_{-h}), \text{ for any } q'_h < q_h, \text{ with } q_h \text{ being the true capacity of } h.
\]

The next theorem characterizes the incentive of institutions to reveal its capacity.

**Theorem 22.2.16** Suppose there are at least two institutions and two individuals in a matching. Then the institution-optimal stable rule is not non-manipulable via capacities.

Now we verify this theorem with the following example.

**Example 22.2.9** Consider again the matching with two institutions and two individuals in Example 22.2.8, \( H = \{h_1, h_2\} \) and \( D = \{d_1, d_2\} \). According to the analysis in Example 22.2.6, the only stable matching obtained by the institution-proposing deferred acceptance algorithm is:

\[
\mu_H = \begin{pmatrix}
h_1 & h_2 \\
d_1 & d_2
\end{pmatrix}.
\]

Consider \( h_1 \). If it manipulates by underreporting its capacity as \( q'_{h_1} = 1 < q_{h_1} \), then we can verify that the matching obtained by the institution-proposing deferred acceptance algorithm is:

\[
\mu'_H = \begin{pmatrix}
h_1 & h_2 \\
d_2 & d_1
\end{pmatrix},
\]

Thus we have \( \mu'_H(h_1) \succ_{h_1} \mu_H(h_1) \), which shows that the institutions-optimal matching is in conflict with non-manipulation via capacities.
In the above example, the matching obtained by the individual-proposing deferred acceptance algorithm is non-manipulable via capacities, since we have $\mu_D' = \mu_H = \mu_D$. Sonmez (1997) provided a more general result on non-manipulation via capacities.

**Theorem 22.2.17 (Non-manipulation via Capacities Theorem)** Suppose there are at least two hospitals and three interns in matching. Then there exists no matching rule that is stable and non-manipulable via capacities.

We prove this theorem by providing a counterexample.

**Example 22.2.10** Consider a matching with two institutions and three individuals, $H = \{h_1, h_2\}$ and $D = \{d_1, d_2, d_3\}$ with their preferences given by:

- $p(h_1) = \{d_1, d_2, d_3\}, \{d_1, d_2\}, \{d_1\}, \{d_2, d_3\}, \{d_2\}, \{d_3\}$;
- $p(h_2) = \{d_1, d_2, d_3\}, \{d_2, d_3\}, \{d_1, d_3\}, \{d_3\}, \{d_1, d_2\}, \{d_2\}, \{d_1\}$;
- $p(d_1) = h_2, h_1$;
- $p(d_2) = h_1, h_2$;
- $p(d_3) = h_1, h_2$.

Suppose each institution has 2 possible capacities $q_h \in \{1, 2\}, h = h_1, h_2$. Then we need to verify that institutions have an incentive to manipulate via capacities in each possible stable matching.

Consider the following three matchings:

- $\mu_1 = \begin{pmatrix} h_1 & h_2 \\ \{d_2, d_3\} & \{d_1\} \end{pmatrix}, \mu_2 = \begin{pmatrix} h_1 & h_2 \\ \{d_1, d_2\} & \{d_3\} \end{pmatrix}, \mu_3 = \begin{pmatrix} h_1 & h_2 \\ \{d_1\} & \{d_3\} \end{pmatrix}$.

We can verify that when capacities are reported as $q_{h_1} = 2, q_{h_2} = 2$, the only stable matching is $\mu_1$; when $q_{h_1} = 2, q_{h_2} = 1$, there are 2 stable matchings $\mu_1, \mu_2$; when $q_{h_1} = 1, q_{h_2} = 1$, the only stable matching is $\mu_3$.

In this case, the stable matchings are:

$$f(p, q_1, q_2) = \mu_1, f(p, q_1 = 1, q_2 = 1) = \mu_3, f(p, q_1 = 2, q_2 = 1) \in \{\mu_1, \mu_2\}$$. 

1001
When \( f(p, q_1 = 2, q_2 = 1) = \mu_1, \) \( f(p, q_1 = 1, q_2 = 1)(h_1) = \mu_3(h_1) = \{d_1\} \) and \( f(p, q_1 = 2, q_2 = 1)(h_1) = \mu_1(h_1) = \{d_2, d_3\} \), we also have

\[
f(p, q_1 = 1, q_2 = 1)(h_1) \succ_h \ f(p, q_1 = 2, q_2 = 1)(h_1).
\]

Thus, when \( q_{h_2} = 1 \), if the true capacity of \( h_1 \) is \( q_{h_1} = 2 \), then it has an incentive to misreport its capacity as \( q'_{h_1} = 1 \).

When \( f(p, q_1 = 2, q_2 = 1) = \mu_2, \) \( f(p, q_1 = 2, q_2 = 1)(h_2) = \mu_2(h_2) = \{d_3\} \) and \( f(p, q_1 = 2, q_2 = 2)(h_2) = \mu_1(h_2) = \{d_1\} \), we also have

\[
f(p, q_1 = 2, q_2 = 1)(h_2) \succ_h \ f(p, q_1 = 2, q_2 = 2)(h_2).
\]

Thus, when \( q_{h_1} = 2 \), if the true capacity of \( h_1 \) is \( q_{h_2} = 2 \), then it has an incentive to misreport its capacity as \( q'_{h_2} = 1 \).

In summary, stable matching mechanisms can not prevent institutions from manipulating via capacities. When there exists at least 2 institutions, we can set other institutions’ capacities to 0; when there exists at least 3 individuals, we can set other individuals’ preferences as preferring only himself or herself. We can extend the above example to the matching between over 2 institutions and over 3 individuals.

However, when the number of individuals are sufficiently large and individuals’ preferences satisfy extended max-min criterion, Jiao and Tian (2015b) proved that institutions-proposing deferred acceptance algorithm is weakly Pareto-efficient and strategy proof. The following quota-saturability is to characterize a sufficiently large number of individuals.

**Definition 22.2.17 (Quota-Saturability)** Let \( \tilde{S} \) be a subset of \( S \) such that every institution \( c \in C \) and every individual \( s \in \tilde{S} \) are acceptable to each other. We say the quota-saturability condition holds if there exists some \( \tilde{S} \subseteq S \) such that \( |\tilde{S}| \) is large enough to satisfy \( |\tilde{S}| \geq \sum_{c \in C} q_c \), i.e., the number of available and acceptable individuals is not less than the aggregate quota of colleges.
In other words, quota-saturability means that there are enough available and acceptable individuals such that each institution $c$ can be assigned $q_c$ acceptable individuals if it wants. Thus, the quota-saturability condition presents the relative size of the agents on the two-sided matching market. An intuitive interpretation of quota-saturability would be a certain “excess supply” of individuals in the market that allows all institutions to fill their quotas if they want.

**Definition 22.2.18 (Extended Max-Min Criterion)** The preference relation of $c \in C$ is said to satisfy the *extended max-min criterion* if for any two sets of acceptable students $G_1, G_2 \in 2^S$ with $|G_1| \leq q_c$ and $|G_2| \leq q_c$,

(i) The strict preference relation $\succ_c$ is defined as: $G_1 \succ_c G_2$ if and only if either $G_2$ is a proper subset of $G_1$, or $|G_1| \geq |G_2|$ and $\min(G_1) \succ_c \min(G_2)$ (i.e., $h$ prefers the least preferred individual in $G_1$), where $\min(G_i)$ denotes $h$’s least preferred individual in $G_i$;

(ii) The weak preference relation $\succeq_c$ over groups of individuals is defined as: $G_1 \succeq_c G_2$ if and only if $G_1 \succ_c G_2$ or $G_1 = G_2$.

We say that the *extended max-min criterion* is satisfied if the preference of every college satisfies the extended max-min criterion.

The extended max-min criterion implies that institutions are always ready to hire more acceptable individuals, which, apparently, is a reasonable assumption, since the facilities and resources of the institution should be fully utilized. For instance, when colleges decide how many students they wish to admit, they should make full use of their resources. Here “max” indicates that colleges always want to match with as many students as possible, and “min” indicates that colleges focus on the worst student in ranking different sets of students and colleges would like to match with as preferable students (in the sense of preferences over individual student) as possible.

We then have the following theorem:
Theorem 22.2.18 (Jiao and Tian, 2015b) Suppose the quota-saturability condition and the extended max-min criterion are satisfied. Then, the institutions-proposing deferred acceptance algorithm is weakly Pareto-efficient and strategy-proof for the institution. (i.e. it is a weakly dominant strategy for each institution to truthfully reveal its preferences).

As the proof of the theorem above is long, interested readers can refer to Jiao and Tian (2015b) for a fully proof.

22.2.3 Many-to-Many Matching

Now we discuss many-to-many matching. In matchings between institutions and individuals in real life, individuals might accept more than one institution’s proposals at the same time. For instance, one person can have more than one job or work full- and part-time, like serving as adjunct or special-term professor in a university. The classic examples of many-to-many matching are between advisers and companies as well as between listed companies and independent directors. Many-to-many matching can be viewed as an extension of many-to-one matching, though it may have its own properties. The purpose of this subsection is to discuss these new properties.

First we will introduce an analysis framework for many-to-many matching. There are two type of agents in our economic environment: one is called institution $H = \{h_1, \cdots, h_n\}$, and the other one is called individual $D = \{d_1, \cdots, d_m\}$. For agent $a \in H \cup D$, its strict preference is denoted by $\succ_a$, its weak preference by $\succeq_a$, its preference order by $p(a)$, and its capacity by $q_a$. For institution $h \in H$, its preference $\succ_h$ is a preference ordering relationship such that each subset $S \subseteq D$ satisfies $|S| \leq q_h$, where $|\emptyset| = 0$ and $\emptyset \succ_h S, \forall |S| > q_h$; similarly, for an individual, $d \in D$, its preference $\succ_d$ is a preference ordering relationship such that each subset $S \subseteq H$ satisfies $|S| \leq q_d$. Denote by $p = p(a)_{a \in H \cup D}$ a preference ordering. When $|q_d| = 1, \forall d \in D$, a many-to-many matching is reduced down to a many-to-one matching.
Now we define a many-to-many matching relationship $\mu$ between $H$ and $D$.

**Definition 22.2.19 (Many-to-Many Matching)** Institutions $H$ and individuals $D$ constitute a many-to-many matching $\mu : H \cup D \rightarrow 2^{H \cup D}$, if the following properties are satisfied:

\[
\begin{align*}
\mu(h) &\subseteq D \cup \{h\} \text{ and } |\mu(h)| \leq q_h, \forall h \in H; \\
\mu(d) &\subseteq H \cup \{d\} \text{ and } |\mu(d)| \leq q_d, \forall d \in D; \\
d &\in \mu(h) \text{ iff } h \in \mu(d), \forall d \in D, h \in H.
\end{align*}
\]

Similar to the previous analysis, to study many-to-many matchings, we remain our focus on stability, efficiency and strategy problems.

In many-to-many matchings, we can also define individual rationality of matching.

**Definition 22.2.20** Matching $\mu$ is individually rational, if for any $a \in H \cup D$, and for any $A \subseteq \mu(a)$, we have $\mu(a) \succeq_a A$, or:

\[
\mu(a) = C_a(\mu(a), \succ_a), \forall a \in H \cup D,
\]

where $C_a$ is the rational choice of agent $a$.

The concept of individual rationality above is stronger than the previous one, since in the previous definition of individual rationality, we only require $\mu(a) \succeq_a \emptyset, \forall a \in H \cup D$. Also, we can define a blocking pair $(h, d)$ for $\mu$, if it satisfies

\[
h \notin \mu(d), h \in C_d(\mu(d) \cup \{h\}, \succ_d), d \in C_h(\mu(h) \cup \{d\}, \succ_h).
\]

There are various stability concepts in the many-to-many matching. First we introduce the concept of pairwise stable.

**Definition 22.2.21 (Pairwise Stability)** A many-to-many matching $\mu$ between $H$ and $D$ is pairwise stable, if $\mu$ is individually rational and there exists no blocking pair $(h, d)$ for $\mu$. Denote $S(p)$ as the set of all pairwise stable matchings under preference ordering relationship $p$.

Next we introduce another stability concept, the concept of core.
Definition 22.2.22 (Core of Many-to-Many Matching) A many-to-many matching $\mu$ between $H$ and $D$ is in core, if there exists no coalition $A \subseteq H \cup D, A \neq \emptyset$ and matching $\mu'$ that satisfy:

$$
\mu'(a) \subseteq A, \forall a \in A;
$$

$$
\mu'(a) \succeq_a \mu(a), \forall a \in A;
$$

$$
\mu'(a) \succ_a \mu(a), \exists a \in A.
$$

Denote by $C(p)$ the set of all cores under preference ordering relationship $p$.

In the previous many-to-one matching, Roth (1985) introduced responsive preference to ensure the existence of stable matching in the many-to-one matching, which is:

Definition 22.2.23 (Responsive Preference) Institutions $H$ and Individuals $D$ have responsive preferences, if

1. for $h \in H$, it satisfies:

   for any $J \subseteq D, |J| < q_h, d \in D \setminus J$, we have:

   $$
   \{d\} \cup J \succ_h J \iff \{d\} \succ_h \emptyset
   $$

   for any $J \subseteq D, |J| < q_h, d, d' \in D \setminus J$, we have:

   $$
   \{d\} \cup J \succ_h \{d'\} \cup J \iff \{d\} \succ_h \{d'\};
   $$

2. for $d \in D$, we require similar conditions.

Under responsive preference, many-to-one pairwise stable matching is a core allocation. However, this conclusion cannot be extended to many-to-many matchings. The following example is from Roth and Sotomayor (1990, p.177).

Example 22.2.11 Consider the matching between three companies $\{h_1, h_2, h_3\}$ and three workers $\{d_1, d_2, d_3\}$, with each agent’s capacity being 2 and their preferences given by:

$$
p(h_1) = \{d_1, d_2\}, \{d_2, d_3\}, \{d_1\}, \{d_2\}, \{d_3\};
$$
\[ p(h_2) = \{d_2, d_3\}, \{d_1, d_3\}, \{d_2\}, \{d_1\}, \{d_3\}; \]
\[ p(h_3) = \{d_1, d_3\}, \{d_1, d_2\}, \{d_3\}, \{d_1\}, \{d_2\}; \]
\[ p(d_1) = \{h_1, h_2\}, \{h_2, h_3\}, \{h_1\}, \{h_2\}, \{h_3\}; \]
\[ p(d_2) = \{h_2, d_3\}, \{h_1, d_3\}, \{h_2\}, \{h_1\}, \{h_3\}; \]
\[ p(d_3) = \{h_1, h_3\}, \{h_1, h_2\}, \{h_3\}, \{h_1\}, \{h_2\}. \]

The only pairwise stable matching in this situation is:

\[ \mu(h_i) = d_i, i = 1, 2, 3. \]

However, \( \mu \) is strictly dominated by the following \( \mu' \), which means \( \mu'(a) \succ \mu(a), \forall a \in H \cup D, \)

\[ \mu'(h_1) = \{d_2, d_3\}, \mu'(h_2) = \{d_1, d_3\}, \mu'(h_3) = \{d_1, d_2\}; \]
\[ \mu'(d_1) = \{h_2, h_3\}, \mu'(d_2) = \{h_1, h_3\}, \mu'(d_3) = \{h_1, h_2\}. \]

In the example above, the pairwise stable matching is not in the core and the core is not a pairwise stable matching. In the following example (from Echenique and Oviedo, 2006), we will see that a core allocation may not even individually rational.

**Example 22.2.12** Consider a matching with three companies \( \{h_1, h_2, h_3\} \) and three workers \( \{d_1, d_2, d_3\} \). Suppose that each agent’s capacity is 2 and their preferences (non-responsive) are given by:

\[ p(h_1) = \{d_3\}, \{d_2, d_3\}, \{d_1, d_3\}, \{d_1\}, \{d_2\}; \]
\[ p(h_2) = \{d_1\}, \{d_1, d_3\}, \{d_1, d_2\}, \{d_2\}, \{d_3\}; \]
\[ p(h_3) = \{d_2\}, \{d_1, d_2\}, \{d_2, d_3\}, \{d_3\}, \{d_1\}; \]
\[ p(d_1) = \{h_3\}, \{h_2, h_3\}, \{h_1, h_3\}, \{h_1\}, \{h_2\}; \]
\[ p(d_2) = \{h_1\}, \{h_1, h_3\}, \{h_1, h_2\}, \{h_2\}, \{h_3\}; \]
Consider the following matching \( \mu : \mu(d_1) = \{h_2, h_3\}, \mu(d_2) = \{h_1, h_3\}, \mu(d_3) = \{h_1, h_2\}. \) This matching is a core allocation, because if there exists one coalition, \( A \), and another matching \( \mu' \), then when \( h_1 \in A \), and \( \mu'(h_1) \succ h_1 \mu(h_1) \), it means \( w_3 \in A \) and \( w_2 \notin \mu'(h_1) \subseteq A \). If the well-being of \( w_3 \in A \) does not decrease in \( \mu' \), then we have \( h_2 \in A \); if the well-being of \( h_2 \) does not decrease in \( \mu' \). Then we have \( d_1 \in A \) and thus \( h_3 \in A \), which contradicts the fact that \( w_2 \in A \). By the symmetric preferences of each agent from \( \{h_1, h_2, h_3\} \) and \( \{d_1, d_2, d_3\} \), and the symmetric matching \( \mu \), it’s easy to verify that for any \( a \in H \cup D \), when \( a \in A \) and \( \mu'(a) \succ a \mu(a) \), we will have similar contradictions.

However, matching \( \mu \) doesn’t satisfy the previous definition of individual rationality because

\[
\mu(h_1) = \{d_2, d_3\} \neq C_{h_1}(\mu(h_1), \succ h_1) = \{d_3\},
\]

Here, in the many-to-many matching, individual rationality requires that institutions can choose their employees on their own, which means that \( h_1 \) can fire \( d_2 \) and only keep \( d_3 \).

To make sure that a matching can still satisfy individual rationality while keeping the core property, we introduce another concept of stability.

**Definition 22.2.24 (Individually-Rational Core)** A many-to-many matching \( \mu \) is an **individually-rational core**, if it satisfies individual rationality and the core property. Denote by \( IRC(p) \) the set of all individually rational core matchings under the preference relationship \( p \).

There is another important concept of stability in many-to-many matching, which is setwise stable matching.

**Definition 22.2.25 (Setwise Stable Matching)** A many-to-many matching \( \mu \) is a **setwise stable matching**, if it satisfies individual rationality and there...
exists no blocking set \((A, \mu')\), where \(\emptyset \neq A \subseteq H \cup D\), satisfying:

\[
\mu'(a) \setminus \mu(a) \subseteq A, \ \forall a \in A; \\
\mu'(a) \succ_a \mu(a) \subseteq A, \ \forall a \in A; \\
\mu'(a) = C_a(\mu'(a), \succ_a), \ \forall a \in A.
\]

(22.13) (22.14) (22.15)

Denote by \(SW(p)\) the set of all set-wise stable matchings under preference relationship \(p\).

A setwise stable matching and an individually-rational core matching are different in the following two aspects: First, the blocking set (22.13) only requires that the new member in \(\mu'\) comes from Coalition \(A\), and does not require that members in the matching all come from the coalition. Second, the new matching needs to be individually rational (22.15), which means those who join the coalition \(A\) won’t have an incentive to leave the coalition.

Echenique and Oviedo (2006) introduced another new concept of stability, which is called strictly pairwise stability. Denote by \(SS(p)\) the set of all strong pairwise stable matchings under preference relationship \(p\).

**Definition 22.2.26 (Strongly Pairwise Stable Matching)** A many-to-many matching \(\mu\) is strongly pairwise stable, if it is individually rational and there exists no \((\hat{D}, h)\) that blocks \(\mu\), \(\hat{D} \subseteq D, h \in H\) such that:

\[
\hat{D} \cap \mu(h) = \emptyset, \ \hat{D} \subseteq C_h(\mu(h) \cup \hat{D}, \succ_h); \\
h \in C_d(\mu(d) \cup \{h\}, \succ_d), \ \forall d \in \hat{D}.
\]

(22.16)

Apparently strongly pairwise stable matching is pairwise stable matching.

Echenique and Oviedo (2006) introduced a fixed-point algorithm to characterize the properties of matching stability.

We say \(v = (v_H, v_D)\), where \(v_H : H \to 2^D\) and \(v_D : D \to 2^H\), is a pre-matching, and the set of all pre-matchings is denoted by \(V\). Based on pre-matching \(v\), we can define two sets:

\[
U(h, v) = \{d \in D : h \in C_d(v(d) \cup \{h\}, \succ_d)\}; \\
V(d, v) = \{h \in H : d \in C_h(v(h) \cup \{d\}, \succ_h)\}.
\]

(22.16) (22.17)
$U(h, v)$ is the set of all individuals who are willing to choose $h$ from $v(d)$ and $V(d, v)$ is the set of all institutions who are willing to choose $d$ from $v(h)$. With these two sets, we define a fixed-point algorithm $T : \mathcal{V} \rightarrow \mathcal{V}$:

$$Tv(a) = \begin{cases} C_a(U(a, v), \succ_a) & \text{if } a \in H; \\ C_a(V(a, v), \succ_a) & \text{if } a \in D. \end{cases}$$

An intuitive explanation for this algorithm is: $(Tv)(h)$ is the set of individuals most preferred by $h$ among all the individuals who are willing to choose institution $h$. Similarly, $(Tv)(d)$ is the set of institutions most preferred by $d$ among all the institutions who are willing to choose individual $d$.

If agents' preference relationship is given as $p$, then the set of all fixed-points of algorithm $T$ can be denoted as $\mathcal{E}(p) = \{v : v = TV\}$. Two important pre-matchings are $v_0$ and $v_1 : v_0(h) = v_1(d), v_0(d) = H, v_1(h) = W$.

As we have discussed in the many-to-one matching, stable matchings do not always exist for all preferences. To make sure that the existence of stable matching, Echenique and Oviedo (2006) introduced the conditions for substitutability and strong substitutability, which also characterize the relationship between stable matchings.

**Definition 22.2.27 (Substitutability)** A preference relationship $p$ satisfies **substitutability**, if for any $a \in H \cup D$, and any $S \subseteq S'$, $b \in C_a(S' \cup b, p(a))$, we have $b \in C_a(S \cup b, p(a))$.

**Definition 22.2.28 (Strong Substitutability)** A preference relationship $p$ satisfies **strong substitutability**, if for any $a \in H \cup D$, and for any $S$ and $S'$ with $S' \succ_a S$ and $b \in C_a(S' \cup b, p(a))$, we have $b \in C_a(S \cup b, p(a))$.

We can verify that strong substitutability implies substitutability: Suppose $S \subseteq S'$, $b \in X = C_a(S' \cup \{b\}, p(a))$, which means $X = C_a(X, p(a))$. Since $S \cup b \subseteq S' \cup b$, so $X \succeq_a S \cup b$. When $X = S \cup b$, we have $b \in S \cup b = C_a(S \cup b, p(a))$; when $X \succ_a S \cup b$, by strict substitutability, we have $b \in C_a(S \cup b, p(a))$.

Next we discuss the existence of stable matchings under substitutable and strongly substitutable conditions and the relationship between all stabilities.
Theorem 22.2.19 The following conclusions hold.

(1) we have:

\[ \mathcal{E}(p) \subseteq SS(p) \subseteq S(p), \]

and further if \( p \) is substitutable, then \( \mathcal{E}(p) \neq \emptyset, \mathcal{E}(p) = S(p); \)

(2) we have:

\[ SW(p) \subseteq \mathcal{E}(p), \]

and further if \( p(D) \) is substitutable and \( p(H) \) is strictly substitutable, then \( SW(p) = \mathcal{E}(p) \subseteq IRC(p). \)

The proof of the above theorem is long, interested readers can refer to Echenique and Oviedo (2006) for more details.

Although many-to-many matching is an extension of many-to-one matching, any different tiny detail between many-to-one matching and many-to-many matching would greatly change the matching outcomes. The following example explains this phenomenon.

Example 22.2.13 Suppose the set of institutions is \( H = \{h_1, \cdots, h_K, \bar{h}\} \), and the set of individuals is \( D = \{\bar{d}, d_1, \cdots, d_{2K}\} \), with individuals’ preferences given by:

\[
\begin{align*}
p(d_k) & : h_1, \cdots, h_K, \bar{h}, \quad k = 1, \cdots, 2K; \\
p(\bar{d}) & : \{h_1, \bar{h}\}, \{\bar{h}\}, \{h_1\}. 
\end{align*}
\]

and the institutions’ preferences are given by:

\[
\begin{align*}
p(h_k) & : \{d_{2k-2}, d_{2k-1}\}, \{d_{2k-1}, d_{2k}\}, \{d_{2k}, \bar{d}\}, \quad k = 2, \cdots, K; \\
p(h_1) & : \{d_1, \bar{d}\}, \{d_1, d_2\}; \\
p(\bar{h}) & : \bar{d}, \quad h_1, \cdots, h_K.
\end{align*}
\]

Since \( |q_d| = 2 \) and it is a many-to-many matching, if \( |q_d| = 1 \), the corresponding preference is: \( p'(\bar{d}) : \{\bar{h}\}, \{h_1\} \), then the matching becomes a many-to-one
matching. It’s easy to verify that, when the preference relationship is \((p'(\bar{d}), p_{-\bar{d}})\), the only many-to-one stable matching is 
\[ \mu(p'(\bar{d}), p_{-\bar{d}}) = \mu' \]
with
\[
\mu' = \begin{pmatrix}
 h_1 & h_2 & \cdots & h_k & \cdots & h_K & \tilde{h} \\
\{d_1, d_2\} & \{d_3, d_4\} & \cdots & \{d_{2k-1}, d_{2k}\} & \cdots & \{d_{2K-1}, d_{2K}\} & \bar{d}
\end{pmatrix}.
\]

When the preference relationship is \(p\), the only many-to-many pairwise-stable matching is 
\[ \mu(p(\bar{d}), p_{-\bar{d}}) = \mu \]
with
\[
\mu = \begin{pmatrix}
 h_1 & h_2 & \cdots & h_k & \cdots & h_K & \tilde{h} \\
\{\bar{d}, d_1\} & \{d_2, d_3\} & \cdots & \{d_{2k-2}, d_{2k-1}\} & \cdots & \{d_{2K-2}, d_{2K-1}\} & \bar{d}
\end{pmatrix}.
\]

For the incentive issues in many-to-many matchings, obviously under general preferences, the impossibility results on many-to-one matchings still holds for many-to-many matchings. For instance, there exists no incentive-compatible and stable matching mechanism for institutions. However, if we impose restrictions on preference and add some other certain conditions, there might exist some incentive-compatible and stable matching mechanism for institutions. Indeed, under the same assumptions, Jiao and Tian (2017) proved that the institution-proposing deferred acceptance mechanism is strategy-proof for institution in many-to-many matching. To prove this result, Jiao and Tian (2015b) first extended the Blocking Lemma 22.2.1 to many-to-one matchings under the extended max-min preference criterion, then Jiao and Tian (2017) proved that the Blocking Lemma holds for many-to-many matchings under the extended max-min preference criterion and quota-saturability condition.

The following theorem is from Jiao and Tian (2017).

**Theorem 22.2.20 (Jiao and Tian, 2017)** Suppose that the institutions satisfy quota-saturability condition, and the preference of each player (each institution and each individual) satisfies the extended max-min criterion. Then the optimal stable matching for institution is weakly Pareto-efficient and strategy-proof (i.e., it is a weakly dominant strategy for institutions to truthfully reveal their preferences).
For the details of the proof for the above theorem, please check the original paper.

22.2.4 Matching with Transfers

In reality, salary is usually taken as a condition for the matching process between firms and workers, but it apparently is not the only one. Because for any side, the matching object is heterogeneous. Then if we introduce transfers into matching between individuals and institutions, it’ll broaden the scope of application of matching between individuals and institutions. For example, in the case of college admission, some universities or research agencies will provide scholarship to influence the matching between colleges and students. Kelso and Crawford (1982) were the first to discuss matching with transfers. Now we focus on the Kelso and Crawford (1982)’s dynamic salary adjusting process.

In the following model, each firm (institution) and each worker (individual) not only cares about who it is matched to, but also cares about transfers between them, which is worker’s salary. In this way, the outcome of matching includes not only the matching between firms and workers, but also the allocation of the revenue between firms (profit) and workers (salary) in the matching.

Suppose the set of firms is \( H = \{h_1, \cdots, h_n\} \) and the set of workers is \( D = \{d_1, \cdots, d_m\} \). We characterize the preferences and utilities of matchings of firms and workers with transfers.

The utility that worker \( d \in D \) receives a wage of \( s_d \) in matching with institution \( h \in H \) is denoted by \( u_{dh}(s_d) \). If worker \( d \) does not have a job, then his or her utility is \( u_{d0} \). Denote by \( s_{dh} \) the lowest wage that \( d \) is willing to accept in matching with firm \( h \), i.e., \( u_{dh}(s_{dh}) = u_{d0} \). Suppose the wage range is nonnegative integer and the smallest difference between two different wages (i.e., the minimum rate of wage adjustment) is 1. Denote by \( s_d = (s_{d1}, \cdots, s_{dn}) \) the vector of the lowest wages that \( d \) is willing to accept.

For convenience, we assume that there is no hiring capacity for \( h \in H \), or we can say its capacity for employees is \( m \), the set of employed workers is \( A \subseteq D \), the
revenue brought about by employees to $h$ (measured through money) is $y^h(A)$.

The following are some assumptions on firm’s revenues from matchings:

\begin{align*}
  y^h(\emptyset) &= 0, \forall h \in H, \quad (22.18) \\
  y^h(A \cup \{d\}) - y^h(A) &\geq s_d, \forall d \notin A, \forall A \subseteq D, \forall h \in H. \quad (22.19)
\end{align*}

(22.19) means that, for each firm $h$, increase in the marginal revenue for hiring one more worker is more than his or her minimum acceptable salary. This assumption does not impose strong restrictions and is just for convenience of the analysis. If a worker’s marginal product to the firm is lower than his or her minimum acceptable salary, then the firm hiring this worker can be seen as the worker being free from the control of the firm and given no salary, or we can say, the worker is only nominally employed by the firm.

The above setting does not include the case of unemployment, since when a worker joins a firm, it is a Pareto improvement for both sides. Suppose there is a matching $\mu$, where $\mu(d) \in H$, $\mu(h) \subseteq D$. Its matching outcome is $(\mu, s, \pi)$. $\mu(h)$ denotes the set of workers employed by $h$, and $(\mu(h_1), \ldots, \mu(h_n))$ is one division of $D$; the salary of $d \in \mu(h)$ is $s_d$, and the profit of $h$ is: $\pi_h = y^h(\mu(h)) - \sum_{d \in \mu(h)} s_d$.

Similarly, the matching outcome $(\mu, s, \pi)$ is individually rational, if

\begin{align*}
  s_d &\geq s_{d\mu(d)}, \forall d \in D; \pi_h \geq 0, \forall h \in H.
\end{align*}

Next we define the notion of stable matching. Kelso and Crawford (1982) introduced the stability by the core.

**Definition 22.2.29 (Core Allocation with Transfers)** A matching $(\mu, s, \pi)$ is a core allocation, if there is no $(h, A, r)$, $h \in H, A \subseteq D$, where $r$ is a nonnegative integer vector of wage level, such that:

\begin{align*}
  \pi_h &< y^h(A) - \sum_{d \in A} r_d, \quad (22.20) \\
  u_{d\mu(d)}(s_d) &< u_{dh}(r_d). \quad (22.21)
\end{align*}
If there exists a matching that satisfies (22.20) and (22.21), then \((h, A, r)\) is called a blocking to the matching \((\mu, s, \pi)\). (22.20) means that hiring workers in \(A\) through \(r\) can make more profit; (22.21) means that receiving wage \(r\) from the firm \(h\) brings higher utility than in matching \((\mu, s, \pi)\).

Similar to those previous matchings between institutions and individuals, if we do not impose restrictions on firms’ outputs, then there might exist no stable matchings. The following example characterizes the case where there is no stable matchings.

**Example 22.2.14** Consider an economy consisting of two firms and two workers. The product function for the two firms are:

\[
\begin{align*}
y^{h_1}(\{d_1\}) &= 4, & y^{h_2}(\{d_1\}) &= 8; \\
y^{h_1}(\{d_2\}) &= 1, & y^{h_2}(\{d_2\}) &= 5; \\
y^{h_1}(\{d_1, d_2\}) &= 10, & y^{h_2}(\{d_1, d_2\}) &= 9.
\end{align*}
\]

Suppose that a worker only cares about the salary rather than his or her firm. Then for worker \(d\) and his or her firm \(h\), the utility function for worker is \(u_{dh}(s) = s_d\), \(s_{dh} = 0\). All the possible matching outcomes, without considering unemployment, are as follows:

\[
\begin{align*}
\mu_1 &= \begin{pmatrix} h_1 & h_2 \\ \{d_1, d_2\} & \emptyset \end{pmatrix}, \\
\mu_2 &= \begin{pmatrix} h_1 & h_2 \\ \emptyset & \{d_1, d_2\} \end{pmatrix}, \\
\mu_3 &= \begin{pmatrix} h_1 & h_2 \\ \{d_1\} & \{d_2\} \end{pmatrix}, \\
\mu_4 &= \begin{pmatrix} h_1 & h_2 \\ \{d_2\} & \{d_1\} \end{pmatrix}.
\end{align*}
\]

Now we need to verify that there exists no allocation \((s, \pi)\) that can make \((\mu_i, s, \pi)\) a core allocation.

Consider \(\mu_1\): \(s = (s_{d_1}, s_{d_1}), \pi_{h_1} = 10 - s_{d_1} - s_{d_2}, \pi_{h_2} = 0\). Then \((h_2, \{d_1\}, r_{d_1} = s_{d_1} + 1)\) or \((h_2, \{d_2\}, r_{d_2} = s_{d_2} + 1)\) will form a blocking to \((\mu_1, s, \pi)\). Otherwise, \(0 = \pi_{h_2} \geq 8 - s_{d_1} - 1\) and \(0 = \pi_{h_2} \geq 5 - s_{d_2} - 1\), which means \(s_{d_1} + s_{d_1} \geq 11\), and thus \(\pi_{h_1} = 10 - s_{d_1} - s_{d_1} \leq -1\), contradicting the individual rationality of \((\mu_i, s, \pi)\).

Consider \(\mu_2\): \(s = (s_{d_1}, s_{d_1}), \pi_{h_1} = 0, \pi_{h_2} = 9 - s_{d_1} - s_{d_2}\). If \((\mu_1, s, \pi)\) wants to avoid blocking by \((h_2, \{d_1\}, r_{d_1} = s_{d_1} + 1)\) or by \((h_2, \{d_2\}, r_{d_2} = s_{d_2} + 1)\), it needs to
satisfy $9 - s_{d_1} - s_{d_2} = \pi_{h_2} \geq 8 - s_{d_1} - 1$, and at the same time, $9 - s_{d_1} - s_{d_2} = \pi_{h_2} \geq 5 - s_{d_2} - 1$, so we have $s_{d_1} + s_{d_2} \geq 7$. However, $(h_1, \{d_1, d_2\}, r_{d_1} = s_{d_1} + 1, r_{d_2} = s_{d_2} + 1)$ forms a blocking. Otherwise we need to satisfy:

$$10 - s_{d_1} - s_{d_2} - 2 \leq \pi_{h_1} = 0,$$

and thus $s_{d_1} + s_{d_2} \geq 8$, which still leads to a contradiction.

Consider $\mu_3$: $s = (s_{d_1}, s_{d_1})$, $\pi_{h_1} = 4 - s_{d_1}$, $\pi_{h_2} = 5 - s_{d_2}$. Then $(h_1, \{d_1, d_2\}, r_{d_1} = s_{d_1} + 1, r_{d_2} = s_{d_2} + 1)$ or $(h_2, \{d_1\}, r_{d_1} = s_{d_1} + 1)$ forms a blocking to $(\mu_1, s, \pi)$. To avoid this, it needs to satisfy:

$$4 - s_{d_1} = \pi_{h_1} \geq 10 - s_{d_1} - s_{d_2} - 2 \leq \pi_{h_1},$$

$$5 - s_{d_2} = \pi_{h_2} \geq 8 - s_{d_1} - 1 \leq \pi_{h_1},$$

thus $s_{d_2} \geq 4, s_{d_1} \geq 6$, then we have $\pi_{h_1} = 4 - s_{d_1} \leq -2$, which contradicts the individual rationality of $(\mu_i, s, \pi)$.

Consider $\mu_4$: $s = (s_{d_1}, s_{d_1})$, $\pi_{h_1} = 1 - s_{d_2}$, $\pi_{h_2} = 8 - s_{d_1}$. Then $(h_1, \{d_1, d_2\}, r_{d_1} = s_{d_1} + 1, r_{d_2} = s_{d_2} + 1)$ or $(h_2, \{d_2\}, r_{d_1} = s_{d_1} + 1)$ forms a blocking to $(\mu_1, s, \pi)$. To avoid this, it needs to satisfy:

$$1 - s_{d_2} = \pi_{h_1} \geq 10 - s_{d_1} - s_{d_2} - 2 \leq \pi_{h_1},$$

$$8 - s_{d_1} = \pi_{h_2} \geq 5 - s_{d_2} - 1 \leq \pi_{h_1},$$

thus $s_{d_2} \geq 3$. Then, $\pi_{h_1} = 1 - s_{d_2} \leq -2$, contradicts the individual rationality of $(\mu_i, s, \pi)$.

In the above example, the output of firm $h_1$ reflects the increasing returns to scale, or we can say that two employees in firm $h_1$ are complementary, which damages the matching stability. Firm $h_1$ can not prevent its workers from being poached by the other firm $h_2$ and make profit at the same time. To ensure the existence of stable matching outcome between firms and workers, Kelso and Crawford (1982) introduced the substitutability condition.

Define $M^h(s)$ be the set of solutions for the optimization problem: $\max_{A \subseteq D}[g^h(A) - \sum_{d \in A} s_d]$ under the wage level $s$. Let $C^h(s) \subseteq M^h(s)$. 

1016
Definition 22.2.30 (Gross Substitutability) For any firm $h \in H$, $h$’s employment has gross substitutability, if whenever $C^h(s) \subseteq M^h(s)$, $s' \geq s$, there exists some $C^h(s') \subseteq M^h(s')$, and $C^h(s') \subseteq C^h(s)$.

The gross substitutability of employment means that, if worker $d$ is employed by firm $h$ with a wage of $s$, then when the wage level of other worker $d'$ is raised to $s'_{d'} > s_{d'}$, worker $d$ under wage level $s'$ will still belongs to the choice set of firm $h$. In Example 22.2.14, when the wage level is $s = (4, 4)$, $C^{h_1}(s) = \{d_1, d_2\}$, and when it is $s' = (7, 4)$, $C^{h_1}(s') = \emptyset$.

When all firms satisfy gross substitutability in employment, Kelso and Crawford (1982) used an algorithm, similar to the deferred acceptance algorithm, to analyze the dynamic process of matching between firms and workers and salary adjustment and found the outcome to be a core allocation.

Stage 1: Firm $h$ chooses the initial wage level $s_h(0) = (s_{dh}(0) = s_{dh})_{d \in D}$ and each firm proposes to all workers (since the worker’s minimum acceptable salary is lower than its marginal output).

Stage $t$: At the wage level $s_h(t) = (s_{dh}(t))_{d \in D}$ of stage $t$, the firm proposes to some worker in $C^h(s_h(t))$. Among the set of workers who can bring the equal profit level, firm $h$ makes its choice under the following rules: The contracts that are not rejected at stage $t - 1$ will be repeated at stage $t$. Doing so will not cause damage to the firm with the assumption of gross substitutability. This is because, when other workers’ wage does not decrease with time, if worker $d$ does not reject the contract proposed at stage $t - 1$, then at the wage lever at stage $t - 1$, it is always rational for the firm to choose worker $d$.

At stage $t$, worker $d$ will choose his or her most preferred proposal from all the proposals he or she received (institutions and the wage offers both considered) and reject the rest. When a worker receives more than one optimal proposal, the worker randomly picks one of them. The proposals that are not rejected at stage $t - 1$ will still be efficient at stage $t$. Dynamic salary adjustment is:

If worker $d$ rejects the proposal from firm $h$ at stage $t - 1$, then $s_{dh}(t) = s_{dh}(t - 1) + 1$, otherwise $s_{dh}(t) = s_{dh}(t - 1)$. 

1017
At stage \( t^\ast \), no proposal is rejected and the matching ends. The matching outcome and the salary level at \( t^\ast \) is the final outcome of this algorithm.

Kelso and Crawford (1982) proved the following theorem.

**Theorem 22.2.21 (Nonempty Core)** If all firms satisfy gross substitutability, then the core is non-empty.

**Proof.** We only need to verify that the outcome of the algorithm used by Kelso and Crawford (1982) is a core.

Obviously, since the adjustment of salary is discrete, the algorithm of Kelso and Crawford will end in finite time. Let \( s^\ast = (s^\ast_{d_1}, \ldots, s^\ast_{d_m}) \) be the salary level at stage \( t^\ast \). The matching between firms and workers is characterized by \( \mu \) and the profit of firm \( h \) is

\[
\pi^\ast_h = y^h(\mu(h)) - \sum_{d \in \mu(h)} s^\ast_d.
\]

In the algorithm of Kelso and Crawford (1982), we obviously have \( s^\ast_d \geq s_{d\mu(d)}, \forall d \in D \), and \( \pi^\ast_h \geq 0, \forall h \in H \).

At stage \( t^\ast \), workers in firm \( h \) make choices as follows:

\[
C^h(s^\ast) \in \arg\max_{A \subseteq D} y^h(A) - \sum_{d \in A} s^\ast_d,
\]

which means

\[
\pi^\ast_h \geq y^h(A) - \sum_{d \in A} s^\ast_d, \forall A \subseteq D. \tag{22.22}
\]

If there exists a blocking \( (h, A, r) \) to \( (\mu, s, \pi^\ast) \), then

\[
u_{dh}(r_d) > u_{d\mu(d)(s_d)}, \forall d \in A, \tag{22.23}
\]

\[
y^h(A) - \sum_{d \in A} r_d > \pi^\ast_h. \tag{22.24}
\]

By the algorithm of Kelso and Crawford (1982) and the inequality (22.23), worker \( d \in A \) does not receive more than \( r_d \) wage firm \( h \), which means that, for \( d \in A \), there must be \( s_{dh}(t^\ast) \leq r_d \). Thus, by the inequality (22.24), we have

\[
\pi^\ast_h < y^h(A) - \sum_{d \in A} r_d \leq y^h(A) - \sum_{d \in A} s^\ast_d,
\]

contradicting (22.22). \( \blacksquare \)
22.2.5 Matching with Contract

A general framework is developed by Hatfield and Milgrom (2005) to analyze different kinds of two-sided matchings, which covers one-to-one, many-to-one and many-to-many matchings, the matching with or without transfer and the matching between bidders and objects in multi-goods-combination auction. For instance, the increasing multi-unit combinatorial auction of Ausubel and Milgrom (2002) in the last chapter uses deferred acceptance algorithm. All such problems can be characterized as for matching with contracts. In this subsection, we will discuss the issues of matching with contracts and its relative properties.

Suppose there are two types of agents: institutions \( H = \{h_1, \ldots, h_n\} \) and individuals \( D = \{d_1, \ldots, d_m\} \). The set of all contracts is denoted by \( X \) and a contract \( x \) should at least include the two sides \( x_D \in D, x_H \in H \) in the framework of matching, where \( x_D \) refers to individuals in the contract \( x \), and \( x_H \) refers to institutions in the contract \( x \). In the contract between firms and workers, \( X \equiv H \times D \times S \). How should we understand the function of contract in the matching? The following example can be used to explain the difference between a matching with no conditions and a matching with contract: in the marriage markets, a pure matching is where a man and a woman choose each other as the partner without conditions; when the contract is involved, then a man and a woman can make a contract before the matching about such problems as “who is going to be taking care of household tasks after getting married”, “whose name should appear on the proof of ownership of property”, etc.

\( \emptyset \in X \) denotes the contracts that have not yet been accepted. In the matching with contracts, \( \succ_d \) and \( \succ_h \) represent the preferences of individual \( d \) and institution \( h \) towards the matching, respectively. Here we mainly focus on the many-to-one matching. The matching with contracts requires that each individual \( d \) can sign at most one contract, and for the contract to be individually rational, it requires that \( x \succ_d \emptyset \); each institution can choose multiple contracts and each of them represents the contract between the institution and one individual.

Now we discuss the choices of individuals and institutions over the contracts.
Suppose $X' \subseteq X$ is the set of available contracts in the market. Denote the choice of individual $d$ over contracts as $C_d(X')$, with its definition given by:

$$C_d(X') = \begin{cases} \emptyset, & \text{if } \{ x \in X' : x \succ_d \emptyset \} = \emptyset; \\ \max_{x \rightarrow_d \emptyset} \{ x \in X' : x_D = d \}, & \text{otherwise.} \end{cases} \quad (22.25)$$

The choice of institutions over contracts is denoted by $C_h(X')$, with its definition given by:

$$C_h(X') = \max_{\succ_h} \{ C \subseteq X' : C \succeq_h A, \forall A \subseteq X' \}, \quad (22.26)$$

such that

$$(\forall h \in H, \forall x, x' \in C_h(X')), x \neq x' \Rightarrow x_D \neq x_D'.$$

Let $C_D(X') = \bigcup_{d \in D} C_d(X')$, which is the set of all the contracts chosen by individuals. Denote by $R_D(X') = X' - C_D(X')$ the contracts that no individual $d \in D$ would accept or the contracts that are rejected by $D$; similarly, we denote $C_H(X') = \bigcup_{h \in H} C_h(X')$, and $R_H(X') = X' - C_H(X')$.

Our focus is whether the matching with contracts would be stable. If one allocation is stable, then there exists no matching between some institution and some individual that can make them better off. One equivalent condition is that there exists no such allocation of institutions and individuals that can make at least one member better off while other members’ profits stay the same. The following is the definition of stability of matching with contracts.

**Definition 22.2.31** A matching with contracts $X' \subseteq X$ is stable, if:

$$C_D(X') = C_H(X') = X'; \quad (22.27)$$

$$\exists h \in H, X'' \neq C_h(X'), \text{ such that } X'' = C_h(X' \cup X'') \subseteq C_D(X' \cup X''). \quad (22.28)$$

If Condition (22.27) fails, it means that one institution or some individual wants to reject some contracts. In this case, the institution or the individual blocks this contract allocation. If Condition (22.28) fails, then it implies that there exist some other contracts that can make institution $h$ better off while
individuals’ well-being stay the same. The following theorem characterizes the properties of a stable matching with contracts.

**Theorem 22.2.22 (Hatfield and Milgrom, 2005)** Suppose \((X_D, X_H) \subseteq X^2\) is the solution for equations:

\[
X_D = X - R_H(X_H); \tag{22.29}
\]
\[
X_H = X - R_D(X_D). \tag{22.30}
\]

Then \(X' = X_H \cap X_D\) is a stable matching with contracts that satisfies \(X_H \cap X_D = C_D(X_D) = C_H(X_H)\). Conversely, for any stable matching with contracts \(X' \subseteq X\), there exists \((X_D, X_H)\) such that \(X' = X_H \cap X_D\), which also satisfies equations (22.29) and (22.30).

**Proof.** We can regard \(X_D\) and \(X_H\) as opportunity sets for individuals and institutions, respectively, which satisfy \(X = X_H \cup X_D\). When \(X'\) is stable, we have \(X' \subseteq X_H, X' \subseteq X_D\), and \(X'\) will not be rejected by individuals or firms.

Suppose that \((X_D, X_H)\) is the solution for equations (22.29 and (22.30). Then we have \(X_D \cap X_H = X_D \cap (X - R_D(X_D)) = X_D - R_D(X_D) = C_D(X_D)\). Similarly, we can obtain \(X_D \cap X_H = C_H(X_H)\). Let \(X' = X_H \cap X_D\). Then \(X' = C_D(X_D) = C_H(X_H)\). By the revealed preference property \((A \subseteq B, A = C\succcontinues(B) \Rightarrow A = C\succcontinues(A))\), we will have \(X' = C_D(X') = C_H(X')\), which meets the stability conditions in (22.27).

Now consider institution \(h\) and contract set \(X''\) such that \(X'' \subseteq C_D(X' \cup X'')\). Since \(X' = C_D(X_D)\), by the revealed preference property, \(X'' \cap X_D \subseteq C_D(X_D)\) holds, which means \(X'' \cap R_D(X_D) = X'' \cap X_D \cap R_D(X_D) \subseteq C_D(X_D) \cap R_D(X_D) = \emptyset\). Then there is \(X'' \subseteq X - R_D(X_D) = X_H\). If \(X'' \neq C_h(X')\), by the revealed preference of institution \(h\), we will have \(X'' \succ_h C_H(X_H) = C_H(X')\) and \(X'' \neq C_h(X' \cup X'')\). Thus, the stability condition in (22.28) is also satisfied. Therefore, \(X' = X_H \cap X_D\) is a set of stable matching contracts.

Suppose \(X'\) is a set of stable matching contracts, and \(X' = C_D(X_D)\) and \(X' = C_D(X') = C_H(X')\). Let \(X_H\) be the contract set weakly preferred by individual
set $D$ to $X'$, that is, $X_H \succeq_D X'$, and let $X_D$ be the set of contracts that less preferred to $X'$ (weakly) by $D$, that is $X' \succeq_D X_D$. So we can say that the set $X$ can be divided into \{X’, X_H−X’, X_D−X’\}. Suppose there exists $h \in H$ such that $C_h(X_H) \neq C_h(X')$. Then $X'' = C_h(X_H)$, which contradicts the stability condition (22.28). Thus we must have $C_h(X_H) = C_h(X')$, $\forall h \in H$, which means $C_H(X_H) = C_H(X') = X'$ and $X − R_H(X_H) = X − (X_H − X') = X − (X − X_D) = X_D$.

Similarly, we have $X − R_D(X_D) = X − (X_D − X') = X − (X − X_H) = X_H$. Thus, $(X_D, X_H)$ is the solution for equations (22.29) and (22.30) and $X' = C_D(X') = C_H(X')$ holds.

Similar to many-to-one matching, we also need to introduce constraints on institutions’ preferences in the matching with contracts, that is, substitutability of preferences.

**Definition 22.2.32** For institution $h$, the contracts in $X$ is *substitutable* if for any subset $X' \subseteq X'' \subseteq X$, the relation $R_h(X') \subseteq R_h(X'')$ always holds.

Under the assumption of substitutable preferences, Hatfield and Milgrom (2005) proposed a new algorithm that is similar to the deferred acceptance algorithm. The individuals-proposing matching with contracts goes as follows:

First, each individual $d$ offers contracts to institutions. Assume that no contracts are offered by individuals or refused by institutions before the matching starts, which means $X_D(0) = X$ and $X_H(0) = \emptyset$. Let $X_H(t)$ be the cumulative set of contracts offered by individuals to institutions until step $t$, let $X_D(t)$ be the set of contracts that are not rejected by institutions until step $t$ and let $X_H(t) \cap X_D(t)$ be the set of contracts held in step $t$. The iteration between step $t$ and step $t+1$ can be defined as:

\[
X_D(t) = X − R_H(X_H(t − 1)), \quad (22.31)
\]
\[
X_H(t) = X − R_D(X_D(t)). \quad (22.32)
\]

$X_H(t − 1)$ denotes the cumulative set of contracts offered to all institutions before step $t$. Each institution $h$ accepts the best offers for itself from the contracts.
it has received and rejects the rest. Let the cumulative set of contracts rejected by institutions be denoted by $R_H(X_H(t-1))$. The contracts held in the end of step $t$ is in $X_D(t) = X - R_H(X_H(t-1))$. If the contracts offered by $h$ in previous steps before step $t$ are held, then the last contract he proposed, $\max_{\succ_h}(X_D(t))$, is the best one for him in $X_D(t)$; however, if the contracts proposed by $h$ are all rejected before step $t$, then he will have to choose the best contract for himself from $X_D(t)$ in step $t$, which is $\max_{\succ_h}(X_D(t))$. Thus, in step $t$, the contracts rejected by all individuals are in $R_D(X_D(t))$. Then in the end of step $t$, the cumulative contracts institutions have been offered are in $X_H(t) = X - R_D(X_D(t))$. By Theorem 22.2.22, when $X_D(t), X_H(t)$ is the solution of equations (22.29) and (22.30) or is a fixed point of equations (22.31) and (22.32) after iterations, $X_H(t) \cap X_D(t)$ is one Hatfield-Milgrom stable matching with contracts.

Next we will use an example from Hatfield and Milgrom (2005) to illustrate Hatfield-Milgrom’s iterative algorithm.

**Example 22.2.15** Consider a matching with contracts between two individuals $D = \{d_1, d_2\}$ and two institutions $H = \{h_1, h_2\}$ and $X = D \times H$, with their order of preference given by:

- $h_1 \succ_{d_1} h_2$,
- $h_1 \succ_{d_2} h_2$,
- $\{d_1\} \succ_{h_1} \{d_2\} \succ_{h_1} \emptyset$,
- $\{d_1, d_2\} \succ_{h_2} \{d_1\} \succ_{h_2} \{d_2\} \succ_{h_2} \emptyset$.

$X = \{(d_1, h_1), (d_1, h_2), (d_2, h_1), (d_2, h_2)\}$ and the iterative algorithm initiates with $X_D(0) = X$ and $X_H(0) = \emptyset$.

Step 1: $X_D(1) = X$. Individual $d_1$ chooses contract $(d_1, h_1)$ and individual $d_2$ chooses contract $(d_2, h_1)$. In this step, the rejected contracts by individuals are in $R_D(X_D(1)) = \{(d_1, h_2), (d_2, h_2)\}$ and the set of contracts institutions received is $X_H(1) = X - R_D(X_D(1)) = \{(d_1, h_1), (d_2, h_1)\}$. Institution $h_1$ holds contract $(d_1, h_1)$ and rejects $(d_2, h_1)$.
Step 2: The set of contracts that are not rejected by institutions is $X_D(2) = X - R_H(X_H(1)) = \{(d_1, h_1), (d_1, h_2), (d_2, h_2)\}$. Individual $d_1$ makes the same choice as in step 1 by choosing $(d_1, h_1)$ and individual $d_2$ chooses $(d_2, h_2)$. In step 2, the rejected contracts are in $R_D(X_D(2)) = (d_1, h_2)$. Therefore, the cumulative set of contracts institutions have received until step 2 is $X_H(2) = X - R_D(X_D(2)) = \{(d_1, h_1), (d_1, h_2), (d_2, h_2)\}$. Institution $h_1$ holds contract $(d_1, h_1)$ and rejects $(d_2, h_1)$ and institution $h_2$ holds contract $(d_2, h_2)$.

Step 3: The set of contracts not rejected by institutions is $X_D(3) = X_D(2) = \{(d_1, h_1), (d_1, h_2), (d_2, h_2)\}$. We know that $(X_D(2), X_H(2))$ is already an iterative fixed point of equations (22.31) and (22.32).

The institutions-proposing matching process is similar but $X_D(t), X_H(t)$ will have different meanings: $X_D(t)$ denotes the cumulative set of contracts that individuals have received before step $t$ and $X_H(t)$ denotes the cumulative set of contracts that individuals have not rejected until step $t$. In step $t$, the contracts that individuals hold are in $X_D(t + 1) \cap X_H(t)$. The initial value in the institutions-proposing matching is changed to be $X_D(0) = \emptyset$ and $X_H(t) = X$.

There must exist an iterative fixed point in Hatfield-Milgrom’s deferred acceptance algorithm. In order to illustrate the iterative algorithm, we define an extended form of mapping first, $F : X \times X \rightarrow X \times X$:

\[
F_1(X') = X - R_H(X'),
\]

\[
F_2(X') = X - R_D(X'),
\]

\[
F(X_D, X_H) = (F_1(X_H), F_2(F_1(X_H))).
\]

The fixed point of mapping $F$ is equivalent to that in equations (22.31) and (22.32). A preference relation $\succ$ is defined on $X \times X$ as:

\[(X_D, X_H) \succ (X_D', X_H')\]

if and only if

\[X_D' \subseteq X_D, X_H \subseteq X_H'.\]
For \((X \times X, \succ)\), \(X \times X\) can be divided into lattice with \(\succ\). With this preference ordering, if contracts are substitutable for institutions, then \(R_H\) is isotone, that is, when \(X' \subseteq X''\), \(R_h(X') \subseteq R_h(X'')\) holds. By the revealed preference principle, \(R_D : X \rightarrow X\) is isotone, too. Then it is easy to verify that the mapping \(F : X \times X \rightarrow X \times X\) defined by (22.33), (22.34) and (22.35) is isotone as well. In Example 22.2.15, \(X_H(t)\) grows larger (or stays unchanged) while \(X_D(t)\) grows smaller (or stays unchanged). Using the fixed-point theorem for finite lattices (Tarski’s Fixed-point Theorem), a fixed point exists at the minimum initial value \((X_D, X_H) = (\emptyset, X)\) or at the maximum initial value \((X_D, X_H) = (X, \emptyset)\) on \(X \times X\).

The following theorem illustrates the existence of stable matching with contracts and the optimal stable matching with contracts for individuals and institutions.

\textbf{Theorem 22.2.23} If contracts are substitutable for institutions, then:

1. In the mapping \(F\) defined by (22.33), (22.34) and (22.35) on \(X \times X\), there exists a fixed point. Also, there is a minimal fixed point \((X_D, X_H)\) and a maximal fixed point \((\bar{X}_D, \bar{X}_H)\);

2. Starting from the maximum initial value \((X_D, X_H) = (X, \emptyset)\), Hatfield-Milgrom deferred acceptance algorithm converges monotonically to the maximum fixed point \((\bar{X}_D, \bar{X}_H)\). The stable matching with contracts \(\bar{X}_D \cap \bar{X}_H\) is the optimal stable matching for individuals;

3. Starting from the minimum initial value \((X_D, X_H) = (\emptyset, X)\), Hatfield-Milgrom deferred acceptance algorithm converges monotonically to the minimum fixed point \((X_D, X_H)\). The stable matching with contracts \(X_D \cap X_H\) is the optimal stable matching for institutions.

When contracts are not substitutable, there may be no stable matching with contracts. Moreover, in the structure of Hatfield and Milgrom’s matching with contracts, if we introduce the law of aggregate demand (If the opportunity set
expands, the total number of contracts chosen by institutions rises), then we can have similar conclusions on incentives as in previous studies. In this way, the matching with contracts becomes a more general framework to analyze matching problems, which can be applied in various fields, such as combinatorial auctions.

Roth and Sotomayor (1990) is a classical textbook in the literature of two-sided matching. But study on two-sided matching is still developing and a more comprehensive review can be found in Abdulkadiroglu and Sonmez (2013) or Kojima (2015). Meanwhile, when individuals are complementary, for example, when a couple look for jobs, there are no stable matchings in general and to guarantee the existence of a stable matching, we have to introduce more restrictions on preferences. More about this topic can be found in Kojima, Pathak and Roth (2013) or Jiang and Tian (2014).

22.3 One-Sided Matching

So far, we talked about two-sided matchings such as men and women, students and colleges, workers and firms, etc. The main matching mechanism that has several desired properties is the deferred acceptance algorithm. However, in real life, many matchings are not two-side, but one-sided. For instance, when we discuss allocations of students to schools. Do schools have preferences? The answer may be no. Many public schools do not prefer one child over another. School seats are sometimes viewed as objects to be allocated.

The so-called one-sided matching is concerned with the allocation of resources and one side of the matching is indivisible objects, which do not have preferences. The one-sided matching has various application, such as school choice and organ transplants (allocation). In literature, the one-sided matching problem is further divided into housing market and house allocation. The only difference between these two problems is that the former assumes that house is private property and the player owns the house while the latter assumes that house is collectively or publicly owned. We will discuss both problems. First we will deal with the
housing market problem.

22.3.1 Housing Market Problem

The housing market model was originally introduced and studied in Shapley and Scarf (1974). We will first consider the exchange of indivisible heterogeneous goods, such as houses and offices. Suppose that there is a group of people and each of them owns a house. The houses are different and each person has his own preference for houses. If we do not have to consider the monetary factor, then what would be a stable housing matching outcome that can maximize the welfare of different people?

Firstly we will introduce the basic model. Let $N = \{1, \ldots, n\}$ be the set of players. For player $i$, its initial endowment is $h_i$ and $H$ is the set of initial endowments of all players. The number of players is the same as that of houses, $|H| = |N|$. Let the preference of player $i$ be $\succ_i$ and for convenience, assume $\succ_i$ is strict preference, which means that player $i$ is not indifferent between any two items. In this way we rule out the possibility of a tie. For player $i$, $p_i$ is a ranking in the order of preferences from the highest to the lowest for the goods in the set of $H$. If $|\{h' \in H| h' \succ_i h\}| = k - 1$, which means for player $i$, only $k - 1$ goods are preferred to $h$ in the set of $H$, then $h$ ranks $k$th in player $i$’s preference ranking, which is expressed as $h = p_i(k)$.

House exchange is a form of house matching. Let $\mu$ be a matching from $N$ to $H$ such that:

$$\mu(i) \in H, \mu(i) \neq \mu(j), \forall i, j \in N, i \neq j.$$  

We call $\nu$ Pareto dominates $\mu$, if $\nu(i) \succeq_i \mu(i)$ for all $i \in N$, and $\nu(j) \succ_j \mu(j)$ for some $j \in N$. We call $\mu$ is Pareto-efficient, if there is no other matching $\nu$ that Pareto dominates $\mu$.

Suppose that initially each player owns a house and the initial ownership structure is denoted as $\mu^0$ with $\mu^0(i) = h_i \in H$, which is a matching as well.

In the following, we will discuss the notion of stable matching for housing
Definition 22.3.1 A matching \( \mu \) is a stable matching if \( \mu \) is in the (strong) core of house markets, that is, there exists no coalition \( T \subseteq N \) and any other matching \( \nu \) such that:

\[
\begin{align*}
\nu(i) & \in \{h_j\}_{j \in T}, \forall i \in T; \\
\nu(i) & \succeq_i \mu(i), \forall i \in T; \\
\nu(j) & \succ_j \mu(j) \text{ for some } j \in T.
\end{align*}
\]

That is, a stable matching means that there exists no coalition that can make its members better off. If a matching does not maximize members’ welfare, then some other coalition may form to reach a new matching to improve upon the welfare. When this happens, we call matching \( (T, \nu) \) weakly dominates \( \mu \).

The matching \( \nu \) defined on \( T \) can also be regarded as being weakly Pareto dominant over \( \mu \). Let \( \nu^E \) be an extended matching of \( \nu \) for the set of all players, and define it as:

\[
\begin{align*}
\nu^E(i) &= \nu(i), \forall i \in T, \\
\nu^E(i) &= \mu(i), \forall i \in N \setminus T.
\end{align*}
\]

The above conditions (22.36), (22.37) and (22.38) imply that:

\[
\begin{align*}
\nu^E(i) & \succeq_i \mu(i), \forall i \in N, \\
\nu^E(j) & \succ_j \mu(j) \text{ for some } j \in N.
\end{align*}
\]

Thus, if matching \( \mu \) is in the core of housing markets, it is equivalent to say that there is no other matching \( \nu^E \) that weakly Pareto dominates \( \mu \).

Next we will discuss how to find a core allocation. Shapley and Scarf (1974) introduced the Top Trading Cycles Algorithm based on David Gale’s idea.

We will first introduce the notion of top cycle before moving to the Top Trading Cycle.
Definition 22.3.2 (Top Cycle) We call \( \{i_1, \ldots, i_K\} \) a top cycle of length \( K \) if for any \( k < K \), both \( h_{i_{k+1}} = p_{i_k}(1) \) and \( h_{i_1} = p_{i_K}(1) \) hold. If for \( j \), there is \( h_j = p_j(1) \), then \( j \) is a top cycle itself. We call \( j \) a self-cycle.

In the above definition, the best (most preferred) item for player \( i_k \) is \( h_{i_{k+1}} = p_{i_k}(1) \).

Lemma 22.3.1 Within a finite population, if each agent has only one good, then there must be a top cycle.

Proof. Firstly, consider \( N = 2 \). If \( i \in \{1, 2\} \) and \( h_i = p_i(1) \), obviously, \( \{i\} \) is a top cycle itself. Otherwise, \( h_1 = p_2(1) \) and \( h_2 = p_1(1) \) must hold and so, \( \{1, 2\} \) is a self-cycle.

Next consider \( N = 3 \). If \( i \in \{1, 2, 3\} \) and \( h_i = p_i(1) \), obviously, \( \{i\} \) is a self-cycle. Otherwise, for player 1, we have \( h_2 = p_1(1) \) or \( h_3 = p_1(1) \). If \( h_2 = p_1(1) \), consider the preference list of player 2. If \( h_1 = p_2(1) \), then \( \{1, 2\} \) is a top cycle; if \( h_3 = p_2(1) \), consider the preference list of player 3. If \( h_2 = p_3(1) \), then \( \{2, 3\} \) is a top cycle; if \( h_1 = p_3(1) \), then \( \{1, 2, 3\} \) is a top cycle. When \( h_3 = p_1(1) \), we will obtain similar results. Therefore, when \( N = 3 \), a top cycle exists as well.

Mathematical induction proves that within a finite population, if each agent owns only one good, then there must be a top cycle. Of course, we can relax the assumption of each player having only one good and let players have different numbers of items.

Next we introduce Gale’s Top Trading Cycle Algorithm:

The top trading cycle is a centralized allocation mechanism, where each player reports his/her (strict) preference.

In step 1, each player points to his/her most favorite good. Since players and goods are both finite, there exists at least one top cycle (including self-cycle) and top cycles do not intersect. Let \( H_{T^1} \) be the set of the goods that players \( T^1 \) in the top cycles formed in step 1 and these players make mutual exchange in the top cycles. Players in the self-cycle stick to their own good. Each player in \( T^1 \) obtains the good he/she prefers most in this step.
In step $t$, remove all the players and goods in the top trading cycles of previous steps. With the remaining players, $N \setminus \bigcup_{s \leq t} T^s$ and goods, rank players’ preference for the goods and search for other top cycles among the remaining goods. If there are remaining players and goods, go to step $t + 1$.

However, the algorithm must terminate when all players and goods have participated in top trading cycles (of different steps) until step $t^*$. Players in their own top trading cycles exchange their goods.

We will use the following example to show how the top trading cycles works.

**Example 22.3.1 (House Exchange)** Consider a group consisting of four players. The house of player $i$ is denoted as $h_i$. Each player’s preference list for houses is given as follows (only houses more preferred than the player’s initial house are listed):

\[
\begin{align*}
p_1 &= h_2, \\
p_2 &= h_1, \\
p_3 &= h_1, h_2, h_4, \\
p_4 &= h_3, h_2.
\end{align*}
\]

$T^1 = \{1, 2\}$ is a top cycle in step 1.

In step 2, players $\{3, 4\}$ and their houses are left. The preferences list of player 3 and player 4 for the remaining houses are:

\[
\begin{align*}
p_3 &= h_4, \\
p_4 &= h_3.
\end{align*}
\]

Now we have $T^2 = \{3, 4\}$. All players have participated in the top cycles in step 1 and step 2. So the final matching $\mu$ of the top trading cycle is:

\[
\mu = \begin{pmatrix} 1 & 2 & 3 & 4 \\ h_2 & h_1 & h_4 & h_3 \end{pmatrix}.
\]

Next we will discuss the properties of the top trading cycle.

Roth and Postlewaite (1977) proved that the matching outcomes produced by Gale’s top trading cycle is the unique core matching.
Theorem 22.3.1 (Roth-Postlewaite, 1977) The matching outcome by Gale’s top trading cycle is the unique core allocation of housing market.

Proof. We only need to show that the matching outcome by Gale’s top trading cycle weakly dominates any other matching outcome. Suppose $\mu^{TTC}$ is a matching outcome by the top trading cycle and $\nu$ is any other matching outcome. In the top trading cycle, there exists a series of cycles, $T^1, \ldots, T^*$, which are obtained from step 1 to the last step $t^*$. If there exists a player $i \in T^1$, $\mu^{TTC}(i) \neq \nu(i)$, as players in $T^1$ all obtain their most preferred goods in matching $\mu$, $(T^1, \mu^{TTC})$ weakly dominates $\nu$. If $\mu^{TTC}(i) \neq \nu(i), \forall i \in T^1$ and there exists $j \in T^2$ such that $\mu^{TTC}(j) \neq \nu(j)$, then $(T^1 \cup T^2, \mu^{TTC})$ weakly dominates $\nu$. Similarly, if $\mu^{TTC}(i) \neq \nu(i), \forall i \in \bigcup_{s \leq t^*} T^s$, then there exists $j \in T^*$ such that $\mu^{TTC}(j) \neq \nu(j)$, thus $(H = \bigcup_{s \leq t^*} T^s, \mu^{TTC})$ weakly dominates $\nu$. Therefore, any matching outcome of $\nu \neq \mu^{TTC}$ is weakly dominated by $\mu^{TTC}$.

Next we will discuss the incentive property of the top trading cycle. Suppose each player reports his/her preference for goods and then is matched through the top trading cycle. Roth (1992) showed that the top trading cycles mechanism is strategy-proof, which is different from two-sided matching.

Theorem 22.3.2 (Roth, 1982) As a direct revelation mechanism, the matching outcome in core (that fully characterizes by Gale’s Top Trading Cycle Algorithm) is strategy-proof, that is, it is a dominant strategy to state his or her true preferences.

Since the complete proof of the above theorem is relatively complex (see Roth, 1982), we will just describe the idea behind the theorem.

Suppose the preferences of other players are reported as $p_{-i}$ and the true preference of player $i$ is $p_i$. $p'_i \neq p_i$ is any preference manipulation. We need to prove that: $\mu^{TTC}(p_i, p_{-i}) \succeq i \mu^{TTC}(p'_i, p_{-i})$, where $\mu^{TTC}$ is a matching obtained by the top trading cycle.

With other players’ preferences reported as $p_{-i}$, when player $i$ reports his/her true preference $p_i$, in the top trading cycle $\mu^{TTC}(p_i, p_{-i})$, $T^1, \ldots, T^*$ are the
top cycles from step 1 until step \( t^* \) and \( T^t = \{A^t_1, \ldots, A^t_m\} \), which means that \( A^t_j \subseteq N \) is the \( j \)th chain of the top cycle in step \( t \). Suppose player \( i \) participates in the top cycle in step \( t \), then with \( \mu^{TTC}(p_i, p_{-i}) \), player \( i \)'s payoff is
\[
\max_{h \in RH_k} u_i(h),
\]
in which \( RH_t = H \setminus \bigcup_{s < t} H_{T^s} \) denotes the remaining houses in step \( t \). When player \( i \) misreports his/her preference as \( p'_i \neq p_i \), it can be shown that because of the preference manipulation, there exists no \( A^s_j, s < k \) such that \( \{i\} \cup A^s_j \) forms a top cycle chain.

We can use mathematical induction to show that the conclusion is correct. When \( s = 1 \), each player in top cycle chain \( A^1_j \) obtains his/her most favourite house in \( H \) while \( k \in A^1_j \) prefers \( i \)'s house \( h_i \). Thus, \( \{i\} \cup A^1_j \) can not be a top cycle chain. When \( s = 2 \), each player in \( A^2_j \) obtains his/her most favourite house in \( RH_2 \) while \( k \in A^1_j \) prefers \( i \)'s house \( h_i \). Thus, \( \{i\} \cup A^2_j \) can not be a top cycle chain. Similarly, when \( s < k \), we cannot find any \( A^s_j, s < k \) such that when player \( i \) reports preference \( p'_i \neq p_i \), \( \{i\} \cup A^s_j \) forms a top cycle chain. In this way, when player \( i \) reports \( p'_i \neq p_i \), the set of houses \( H \setminus RH_k \) is not available to \( i \). However, if player \( i \) reports its true preference \( p_i \), he or she can be assigned his/her most favourite house in \( RH_k \). Therefore, it is a weakly dominant strategy for the player to state his/her true preferences \( p_i \).

Previously we have proved that the top trading cycle is the unique core allocation, which means that it is Pareto-efficient and individually rational. Moreover, the top trading cycle is strategy-proof. Next we will show that if a mechanism \( \nu \) is Pareto-efficient, individually rational and strategy-proof, then it must be a top trading cycle.

**Theorem 22.3.3 (Ma, 1994)** A matching mechanism is Pareto-efficient, individually rational and strategy-proof if and only if it is a top trading cycles mechanism.

**Proof.** Suppose \( \nu(.) \neq \mu^{TTC}(.) \). Then there exists at least a preference \( p \) such that \( \nu(p) \neq \mu^{TTC}(p) \). Define \( J(\mu, \mu', p) = \{j \in N : \mu_j(p) \succ_j \mu'_j(p)\} \). The proof contains the following four steps.

1032
Step 1: Let $\mu$ and $\mu'$ be two matching mechanisms that are Pareto-efficient. If $\mu \neq \mu'$, then $J(\mu, \mu', p) \neq \emptyset$. If $J(\mu, \mu', p) \neq \emptyset$, then $\mu_j'(p) \succeq_j \mu_j(p), \forall j \in N$. Under strict preference, $\mu \neq \mu'$ implies that there must exist some $k \in N$ such that $\mu'_k(p) >_k \mu_k(p)$, which contradicts the fact that $\mu$ is Pareto-efficient.

Step 2: There exists $j \in J(\mu^{TTC}(p), \nu(p), p)$ such that $\mu_j^{TTC}(p) >_j \nu_j(p) >_j h_j$.

In step 1, we have shown that there exists some $j \in J(\mu^{TTC}, \nu, p)$. Suppose there exists no $j \in J(\mu^{TTC}, \nu, p)$ such that $\mu_j^{TTC}(p) >_j \nu_j(p) >_j h_j$, then it implies that $\nu_j(p) = h_j, \forall j \in J(\mu^{TTC}, \nu, p)$. Let $S = N \setminus \{J(\mu^{TTC}, \nu, p) \cup J(\nu, \mu^{TTC}, p)\}$ and $T = S \cup J(\nu, \mu^{TTC}, p)$. Then $(T, \nu(p))$ weakly dominates $\mu^{TTC}$, which contradicts the fact that $\mu^{TTC}$ is a core.

Let $T_p = \{i \in N : \exists h \in H, \mu_i^{TTC}(p) >_i h \succ_i h_i\}$ and define the preference ordering $p' = (p'_1, \ldots, p'_n)$ such that:

$$p'_i = \begin{cases} 
\mu_i^{TTC}(p), & \text{if } i \in T_p \\
p_i, & \text{if } i \notin T_p.
\end{cases} \quad (22.39)$$

Obviously:

$$\mu^{TTC}(p) = \mu^{TTC}(p') = \mu^{TTC}(p'_{-T}, p_T), \forall T \subseteq N. \quad (22.40)$$

Step 3: $\mu^{TTC}(p') = \nu(p')$. Suppose $\mu^{TTC}(p') \neq \nu(p')$. Then, by Step 2, there exists $j \in J(\mu^{TTC}(p'), \nu(p'), p)$ such that

$$\mu_j^{TTC}(p')p'_j \nu_j(p')p'_j h_j. \quad (22.41)$$

However, considering the preference ordering $p'$ in (22.39) and (22.40), we know that $j \in N$, $h_j$ ranking behind $\mu_j^{TTC}(p')$ under $p'_j$ or $h_j = \mu_j^{TTC}(p')$ all contradict to (22.41).

Step 4: $\mu^{TTC}(p'_{-T}, p_T) = \nu(p'_{-T}, p_T), \forall T \subseteq N, T = N$ implies that $\mu^{TTC}(p) = \nu(p)$. We can use mathematical induction to verify the conclusion. When $|T| = 0$, it can be deduced from step 3 directly. Suppose when $|T| = k$, the conclusion holds.

When $|T| = k + 1$, there exists $\mu^{TTC}(p'_{-T}, p_T) \neq \nu(p'_{-T}, p_T)$. Let $q = (p'_{-T}, p_T)$. From step 2, there exists $j \in J(\mu^{TTC}(q), \nu(q), q)$ such that:

$$\mu_j^{TTC}(q)q_j \nu_j(q)q_j h_j. \quad (22.42)$$

1033
When \( j \in N \setminus T \), with equations (22.40) and (22.42), we have:
\[
\mu_j^{TTC}(p)p_j'\nu_j(q)p_j'h_j, \quad (22.43)
\]
which contradicts with the definition of \( p' \).

When \( j \in T \), from (22.40), we obtain:
\[
\mu_j^{TTC}(q) = \mu^{TTC}(q_{-j}, p'_j). \quad (22.44)
\]

From the assumption of mathematical induction, there is:
\[
\mu_j^{TTC}(q_{-j}, p'_j) = \nu_j(q_{-j}, p'_j). \quad (22.45)
\]

From (22.42), (22.44) and (22.45), we know that:
\[
\nu_j(q_{-j}, p'_j)p_j\nu_j(q). \quad (22.46)
\]

Substituting \( q = (p'_{-T}, p_T) \) into (22.47), we will have:
\[
\nu_j(p'_{-T}, p_T\setminus\{j\}, p'_j)p_j\nu_j(p'_{-T}, p_T), \quad (22.47)
\]
which contradicts with \( \nu \) being strategy-proof. ■

The above theorem reveals that in one-sided matching, the top trading cycles mechanism has good properties: it is the only mechanism that is individually rational, Pareto-efficient and strategy-proof.

### 22.3.2 House Allocation Problem

The house allocation problem is also a one-sided matching between individuals and indivisible goods. Different from that in house market, houses in the house allocation problem are collectively owned. So, one problem is how to properly allocate goods to individuals. First of all, we need to determine the standard of efficiency. Pareto efficiency is often used. The model here is the same as that in house market, except that it gives no initial ownership of the houses to individuals. The house allocation problem was first studied in Hylland and Zeckhauser (1977).
One simple and most common method for allocation in real world is to pick a particular order for participants in \( N \) to select and then let them successively choose an item in that order. This mechanism is called serial dictatorship mechanism.

**Definition 22.3.3 (Serial Dictatorship Mechanism)** A choice order is determined by function \( f : N \rightarrow N \), where \( f(k), k \in N \) denotes that the choice order function of the \( k \)th player is \( f(k) \). A mechanism is called *serial dictatorship*, if, ex ante, there is a choice order determined by function \( f \), according to which, \( f(1) = i \) means player \( i \) is the first to choose his/her favourite good in \( H \). Then in sequence, \( f(k) = j \) means player \( j \) is the \( k \)-the to chooses his/her favourite good among the remaining goods.

The following example illustrates how the serial dictatorship mechanism works.

**Example 22.3.2** Suppose \( H = \{a, b, c\} \), \( N = \{1, 2, 3\} \), and the choice order function is \( f(1) = 2, f(2) = 3, f(3) = 1 \). The preference list of players for houses are:

\[
\begin{align*}
p_1 &= a, b, c; \\
p_2 &= a, c, b; \\
p_3 &= b, c, a.
\end{align*}
\]

The matching outcome obtained by the serial dictatorship mechanism is:

\[
\mu^f(p) = \begin{pmatrix}
1 & 2 & 3 \\
1 & 2 & 3
\end{pmatrix}.
\]

Obviously, the mechanism is Pareto-efficient and strategy-proof.

A mechanism is *core from assigned endowments* if property rights are ex ante assigned according to some rule and then exchanges are made through the top
trading cycles mechanism. It is also a way of allocation. We know from the last subsection that the matching outcome obtained by such mechanism has good properties, that is, it is the unique core allocation and strategy-proof. Abdulkadiroglu and Sonmez (1998) showed that the above two mechanisms are equivalent to some extent.

First, we need to introduce some notations. \( N = \{1, \ldots, n\} \) is the set of players and \( H = \{h_1, \ldots, h_n\} \) is the set of indivisible goods. Let \( \mu : N \to H \) be a matching. The set of all matchings is denoted by \( \mathcal{M} \), and the set of all Pareto-optimal matchings is denoted by \( \mathcal{E}^p \subset \mathcal{M} \).

Let \( \omega \) be the initial allocation of property rights, which is a matching as well. If \( |N| = |H| = n \), then there exist \( |\mathcal{M}| = n! \) different ways of matchings. Let \( p_i \) be the preference ordering (\( \succ_i \) is the corresponding preference) of player \( i \) for houses in set \( H \). For housing market \( (N, H, P, \omega) \), there is the unique core, denoted by \( \mu^\omega \), which is the outcome obtained by the top trading cycle. Consider all possible initial property rights allocations and denote the corresponding cores by \( \mathcal{E}^{TTC} = \{\mu^\omega : \omega \in \mathcal{M}\} \).

Let \( f : N \to N \) be the choice order function and denote all possible choice order functions as \( \mathcal{F} \). Obviously \( |\mathcal{F}| = n! \). Let the matching outcomes obtained by the serial dictatorship mechanism in the choice order \( f \) be \( \mu^f \) and the set of all \( \mu^f \) is denoted by \( \mathcal{E}^{SD} = \{\mu^f : f \in \mathcal{F}\} \).

Next we will discuss the matching outcomes of core mechanism (TTC) with given property rights and of the serial dictatorship mechanism, respectively.

Given the (strict) preference ordering \( p \) and initial property endowment \( \omega \) for all players, in the top trading cycle, let \( T = \{T^1, \ldots, T^{t^*}\} \) denote the set of all top cycles which is terminated at step \( t^* \). \( T^t \subseteq N \) is the set of top cycles in step \( t \) and \( T \) can be regarded as a division to the set of players \( N \). Let \( k_t = |T^t| \), then \( n = \sum_{t=1}^{t^*} k_t \). For player \( i \), let \( C_i(\mathring{H}) \), \( \mathring{H} \subseteq H \), denote the optimal choice made by player \( i \) from house subset \( \mathring{H} \), in which \( C_i(\mathring{H}) = \{h : h \succ_i h', \forall h' \neq h, h' \in \mathring{H}\} \).

In the top trading cycle, let \( RH_t \) denote the subset of houses remaining at the beginning of step \( t \leq t^* \). Obviously, \( RH_1 = H \). When \( i \in T^1 \), \( \mu^\omega_i = C_i(RH_1) \),
the houses chosen in step 1 and those removed from the housing market can be denoted as \( H_1 = \{ \omega_i : i \in T^1 \} \). At the beginning of step 2, \( RH_2 = RH_1 \setminus H^1 \) is the set of remaining houses. When \( i \in T^2 \), \( \mu^\omega_i = C_i(RH_2) \), the houses chosen in step 2 and those removed from the housing market can be denoted as \( H_2 = \{ \omega_i : i \in T^2 \} \). Similarly, at the beginning of step \( t \leq t^* \), the remaining houses is in set \( RH_t = RH_{t-1} \setminus H_{t-1} = H \setminus \bigcup_{1 \leq s < t} \{ H_s \} \). When \( i \in T^t \), we have \( \mu^\omega_i = C_i(RH_t) \).

In the serial dictatorship mechanism, the choice order of players is \( f(1), \ldots, f(n) \). Thus, player \( i = f(1) \) is the first to choose good \( \mu^{f(1)}_i = C_{f(1)}(RH^sd_1) \) and player \( f(k) \leq n \), who is the \( k \)th to choose, will obtain good \( \mu^{f(k)}_i = C_{f(k)}(RH^sd_k) \), in which \( RH^sd_k = H \setminus \bigcup_{1 \leq s < k} \{ \mu^{f(s)}_i \} \).

It is easy to see that given any \( \mu^* \), if the choice order function \( f \) satisfies: \( \{ f(1), \ldots, f(k_1) \} = T^1 \), \( \{ f(k_1 + 1), \ldots, f(k_1 + k_2) \} = T^2 \), \ldots, \( \{ f(k_1 + \ldots + k_{t-1} + 1), \ldots, f(k_1 + \ldots + k_{t-1} + k_t) \} = T^t \), then \( \mu^f = \mu^\omega \). In the same time, for any choice order function \( f \) and the matching \( \mu^f \) obtained by the serial dictatorship mechanism, when the initial allocation of property rights \( \omega \) satisfies: \( \omega_{f(k)} = \mu^f_{f(k)} \), it is easy to show that \( \mu^\omega = \mu^f \).

Besides, we know that both the core from assigned property rights and the serial dictatorship mechanism are Pareto-efficient. For any Pareto-efficient matching \( \mu \in \mathcal{E}^P \), when \( \omega = \mu \), we will have \( \mu^\omega_i \succeq \mu(i), \forall i \in N \) since the core from assigned property rights is individually rational. Since each player’s preference is strict, \( \mu^\omega = \mu = \omega \) must hold. Otherwise, it will contradict with the fact of \( \mu \) being Pareto-efficient.

Abdulkadiroglu and Sonmez (1998) proved the following conclusion.

**Theorem 22.3.4** For any Pareto-efficient matching \( \nu(,) \), there exist a serial dictatorship matching and a top trading cycles mechanism from assigned property rights that both can achieve \( \nu(,) \) and are Pareto-efficient, that is, \( \mathcal{E}^P = \mathcal{E}^{TTC} = \mathcal{E}^{SD} \).

From Theorem 22.3.4, we can have the following corollary.

**Corollary 22.3.1** A matching is Pareto-efficient if and only if it is a matching
obtained by a serial dictatorship mechanism. A matching is Pareto-efficient if and only if it is a core matching from assigned property rights (or a matching of the top trading cycle).

For any Pareto-efficient matching \( \eta \), let \( \mathcal{M}_\eta = \{ \omega : \mu^\omega = \eta \} \) denote the set of initial property rights which enable a core matching to be a Pareto-efficient matching \( \eta \); similarly, let \( \mathcal{F}_\eta = \{ f : \mu^f = \eta \} \) be the set of choice order functions which enable a serial dictatorship matching to be a Pareto-efficient matching \( \eta \). Abdulkadiroglu and Sonmez (1998) further proved a stronger conclusion, that is \( |\mathcal{M}_\eta| = |\mathcal{F}_\eta|, \forall \eta \in \mathcal{E}^p \). Since its proof is relatively complex, interested readers can refer to their paper for more details. In addition, Long and Tian (2014) showed that Abdulkadiroglu and Sonmez’s Theorem 22.3.4 still holds under weak preferences.

In the above, we discussed the relationship between a matching mechanism with a given choice order and a matching mechanism with assigned initial property rights. However, disputes may arise due to the human intervention of assigning a choice order or initial property rights. For example, the choice order may raise disputes since someone may obtain the priority to choose by using his/her power, which will certainly cause social unfairness. In the real world, one way to deal with the human intervention is to arrange the choice order and the initial property rights in random order, for example, by using a random draw.

If the choice order and the property rights assignment are both based on probability distribution, the former is called random serial dictatorship mechanism or random priority mechanism and the latter is called core from random endowment. Consider the uniform distribution, that is, the probabilities for all possible choice order functions are the same, \( \text{prob}(f) = \frac{1}{n!}, \forall f \in \mathcal{F} \); or the probabilities for all initial property rights are the same, which is \( \text{prob}(\omega) = \frac{1}{n!}, \forall \omega \in \mathcal{M} \). Under the uniform distribution, the matching of random serial dictatorship \( \psi^{r.sd} \) is:

\[
\psi^{r.sd} = \sum_{f \in \mathcal{F}} \frac{1}{n!} \mu^f.
\]

Under the uniform distribution, the core matching from random endowments \( \psi^{c.\text{cre}} \)
\[ \psi_{\text{cre}} = \sum_{\omega \in \mathcal{M}} \frac{1}{n!} \mu_\omega. \]  

By the conclusion \(|\mathcal{M}^\eta| = |\mathcal{F}^\eta|\), \(\forall \eta \in \mathcal{E}^p\) and Theorem 22.3.4, we can have the following corollary.

**Corollary 22.3.2** Under the uniform distribution, a matching obtained by a random serial dictatorship \(\psi_{\text{rsd}}\) and a core matching from random endowments \(\psi_{\text{cre}}\) are the same.

### 22.3.3 House Allocation with Existing Tenants

In this subsection, we discuss the problems of house allocation and housing market together. For many goods, when allocated, there exists mixed ownerships of property, which involves both private property rights and public property rights. Also, we take house allocation as example. In a society, one institution is in charge of house allocation; some individuals already have houses to live in while some new players are waiting for allocation. For instance, think of dormitory allocation where some students are freshmen, and have no house yet, while others are existing tenants. Such kind of problem is called the problem of tenants’ house allocation. Discussion in this subsection is mainly referred to Abdulkadiroglu and Sonmez (1999): house allocation with existing tenants.

We first present some notations and definitions. Let \(\langle N_E, N_U, H_E, H_V, p \rangle\) denote the problem of tenants’ house allocation, in which \(N_E\) denotes the set of tenants, \(N_U\) the set of new players, \(H_E = \{h_i\}_{i \in N_E}\) the set of houses occupied already, \(H_V\) the set of houses to be allocated, and \(h_0 \in H_V\) means no house. If \(|H_E| + |H_V| < |N_E| + |N_U|\), then inevitably some players \(i \in N = N_E \cup N_U\) cannot get a house and \(h_i = h_0\). \(H = H_E \cup H_V\) is the set of all houses.

A matching \(\mu : N \rightarrow H\) is: for \(i \neq j\) with \(i, j \in N\), we have \(\mu(i), \mu(j) \in H\) such that either \(\mu(i) = \mu(j) = h_0\) or \(\mu(i) \neq \mu(j)\). In other words, it is possible that more than one new players are not assigned to any house. Let \(p(i)\) be
(strict) preference ordering of player $i$ and his preference be $\succ_i$ that satisfies $h \succ_i h_0, \forall h \neq h_0, h \in H, i \in N$.

Matching $\mu$ is Pareto efficient, if there does not exist any other $\nu$ such that $\nu(i) \succeq_i \mu(i), \forall i \in N$ and $\nu(j) \succ_j \mu(j)$ for some $j \in N$. Matching $\mu$ is individually rational if there does not exist any $i \in N_E$ such that $h_i \succ_i \mu(i)$. Stochastic matching is a probability distribution of matching. A common tenant matching mechanism is called the random serial dictatorship with squatting rights, which is described as follows:

(1) Each tenant determines whether to participate in a (random) allocation of houses or not. If not, he can keep the house that he has owned, otherwise his house is added to the set which will be allocated.

(2) Determine players’ order of choosing according to a given probability distribution that can be uniform or make some group to be more advantaged.

(3) Once the order of choice is determined, houses will be allocated by serial dictatorship mechanism.

Although this mechanism has wide applications in real world, it is not always ex post individually rational or ex post Pareto efficient. The following example reveals this point.

**Example 22.3.3** Consider a matching of three individuals and three houses. The set of tenants is $N_E = \{1\}$ and the set of new players is $N_U = \{2, 3\}$; the houses occupied already is $H_E = \{h_1\}$ and the house unoccupied is $H_v = \{h_2, h_3\}$. Tenant 1 decides whether to participate in allocation. If not, he still keeps his house $h_1$; otherwise, the houses that can be allocated are $H = \{h_1, h_2, h_3\}$. Houses are allocated by (uniform) serial dictatorship mechanism. Suppose preference
order of players are as followed:

\[ p(1) = h_2, h_1, h_3; \]
\[ p(2) = h_1, h_2, h_3; \]
\[ p(3) = h_2, h_1, h_3. \]

Since tenant 1 will decide whether to participate in the allocation or not before the allocation, we need to compare his expected utility when he participates and that when he does not. Suppose the preference of tenant 1 satisfies with fundamental axiom of expected utility and his utility to the three houses is:

\[ u_1(h_1) = 3, \quad u_1(h_2) = 4, \quad u_1(h_3) = 1. \]

If tenant 1 does not participate in the allocation, his expected utility is 3. Otherwise, if he participates, there are six preference ordering functions with the same probability. Below, we will solve the matching outcome of each of these six sequences according to serial dictatorship mechanism.

\[ f^1 = (i_1, i_2, i_3), \mu^{f^1}(1) = h_2, \mu^{f^1}(2) = h_1, \mu^{f^1}(3) = h_3; \]
\[ f^2 = (i_1, i_3, i_2), \mu^{f^2}(1) = h_2, \mu^{f^2}(2) = h_3, \mu^{f^2}(3) = h_1; \]
\[ f^3 = (i_2, i_1, i_3), \mu^{f^3}(1) = h_2, \mu^{f^3}(2) = h_1, \mu^{f^3}(3) = h_3; \]
\[ f^4 = (i_2, i_3, i_1), \mu^{f^4}(1) = h_3, \mu^{f^4}(2) = h_1, \mu^{f^4}(3) = h_2; \]
\[ f^5 = (i_3, i_1, i_2), \mu^{f^5}(1) = h_3, \mu^{f^5}(2) = h_3, \mu^{f^5}(3) = h_2; \]
\[ f^6 = (i_3, i_2, i_1), \mu^{f^6}(1) = h_3, \mu^{f^6}(2) = h_1, \mu^{f^6}(3) = h_2. \]

Thus, when participating in allocation, the expected utility of tenant 1 is:

\[ \frac{1}{2}u_1(h_2) + \frac{1}{6}u_1(h_1) + \frac{1}{3}u_1(h_3) = \frac{17}{6} < 3. \]

Then we know that tenant 1 will not participate and the matching outcome of random serial dictatorship mechanisms is:

\[ f'^1 = (i_2, i_3), \mu^{f'^1}(1) = h_2, \mu^{f'^1}(2) = h_1, \mu^{f'^1}(3) = h_3, \]
\[ f'^2 = (i_3, i_2), \mu^{f'^2}(1) = h_1, \mu^{f'^2}(2) = h_3, \mu^{f'^2}(3) = h_2. \]
Under random serial dictatorship mechanism, the probabilities of $\mu^{r_1}, \mu^{r_2}$ are both $\frac{1}{2}$. It is easy to see that $\mu^{r_1}$ is Pareto dominated by $\nu$, in which

$$\nu(1) = h_2, \nu(2) = h_1, \nu(1) = h_3.$$  

Thus, random serial dictatorship with squatting rights is not an ex post Pareto-efficient matching mechanism.

The reason that the above mechanism is not ex post Pareto efficient is because a tenant will be worried about the loss of his benefit and thus does not participate in allocation. Because of the withdrawal of participants, some mutually beneficial matching can not be realized. In order to solve this problem, some guarantee mechanism should be introduced. When there is only one tenant, the following random serial dictatorship mechanism with order adjustment in below can solve this problem.

The random serial dictatorship mechanism with order adjustment can be described as follow:

Step 1, all houses, including the tenant’s houses, are used for allocation;

Step 2, determine the order function through random selection according to a given probability distribution;

Step 3, the player, who is the first in order, points to his most favourite house. Then the second player points to his most favourite one among the remaining houses. This process will be carried out in turn until some individual points to the house of the tenant. Then continue to step 4;

Step 4, if the tenant has already been matched to some house before his house is pointed, then keep the choice order and allocate houses by serial dictatorship mechanism; otherwise, if the tenant has not been matched to any house, then adjust the order by putting the
tenant in front of the individual who points to his house. At the same time, keep others’ order unchanged.

The following example reveals that the random serial dictatorship mechanism with order adjustment satisfies ex post individual rationality and Pareto efficiency simultaneously.

**Example 22.3.4 (Example 22.3.3 continued:)** Preferences of players and the probability distribution of choice order are the same with that in Example 22.3.3. It is easy to see that in Example 22.3.3 there no order adjustment in preference ordering function \( f^1, f^2, f^5 \) and the matching outcome is the same with that in Example 22.3.3, that is:

\[
\begin{align*}
  f^1 &= (i_1, i_2, i_3), \quad \mu_f^1(1) = h_2, \quad \mu_f^1(2) = h_1, \quad \mu_f^1(3) = h_3; \\
  f^2 &= (i_1, i_3, i_2), \quad \mu_f^2(1) = h_2, \quad \mu_f^2(2) = h_3, \quad \mu_f^2(3) = h_1; \\
  f^5 &= (i_3, i_1, i_2), \quad \mu_f^5(1) = h_1, \quad \mu_f^5(2) = h_3, \quad \mu_f^5(3) = h_2.
\end{align*}
\]

In \( f^3 \) and \( f^4 \), player 2, the first to choose, points to the house of tenant 1 who has not been matched to any houses. Thus the choice order needs to be changed. Let \( f^3A \) and \( f^4A \) be the corresponding adjustment orders, respectively:

\[
\begin{align*}
  f^3 &= (i_2, i_1, i_3), \quad f^3A = (i_1, i_2, i_3), \quad \mu_{f^3A}(1) = h_2, \quad \mu_{f^3A}(2) = h_1, \quad \mu_{f^3A}(3) = h_3; \\
  f^4 &= (i_2, i_3, i_1), \quad f^4A = (i_1, i_2, i_3), \quad \mu_{f^4A}(1) = h_2, \quad \mu_{f^4A}(2) = h_1, \quad \mu_{f^4A}(3) = h_3.
\end{align*}
\]

In \( f^6 \), player 2, the second to choose, points to the house of tenant 1 who has not been matched to any houses. Thus the choice order needs to be changed. Let \( f^6A \) be the corresponding adjustment order

\[
\begin{align*}
  f^6 &= (i_3, i_2, i_1), \quad f^6A = (i_3, i_1, i_2), \quad \mu_{f^6A}(1) = h_1, \quad \mu_{f^6A}(2) = h_3, \quad \mu_{f^6A}(3) = h_2.
\end{align*}
\]

The random serial dictatorship mechanism with order adjustment results in the following matching

\[
\begin{align*}
  \mu^1 &= \begin{pmatrix}
    1 & 2 & 3 \\
    h_2 & h_1 & h_3
  \end{pmatrix}, \quad \mu^2 = \begin{pmatrix}
    1 & 2 & 3 \\
    h_2 & h_3 & h_1
  \end{pmatrix}, \quad \mu^3 = \begin{pmatrix}
    1 & 2 & 3 \\
    h_1 & h_3 & h_2
  \end{pmatrix}.
\end{align*}
\]
Obviously, this mechanism satisfies ex post individual rationality and Pareto efficiency at the same time.

If the number of tenants is more than 1, the above mechanism does not work any more. When there are two tenants pointing to each other’s house, the above order adjustment will not apply.

Not all guarantee mechanisms for tenants can realize ex post individual rationality and Pareto efficiency. The matching mechanism below is called the random serial dictatorship with waiting list, which is widely used in certain fields, such as allocation of dormitories in university. It can be described as follow:

At the beginning $RH_1 = H_V$ is the set of all houses that can be allocated. For new individuals $i \in N_U$, the set of acceptable houses is $RH_1 = H_V$; for tenants $i \in N_E$, the set of acceptable houses is the houses which can be chosen and better than the one he has already owned.

In period 1, for individual $f(1)$ who has the highest priority (the first in order), if there is at least one acceptable house, then he is matched to his most favourite one; his matching house is excluded from the set of allocation houses. If $f(1) \in N_E$ is matched, then the house $h_{f(1)}$ he used to own is put into the set which can be allocated. In sequence, by the order of priority, each individual chooses the most favourite one among his acceptable houses. If a tenant is matched to a house successfully, then the house he used to own is put into the set which can be allocated. And matching gets into next period.

In period $t$, the houses that can be allocated are those that can be allocated at the end of previous period. The individual who has the highest priority of all individuals and has not been matched yet in previous steps can choose the most favourite one among his acceptable houses, and then the house that is matched to him is excluded from the set of houses to be allocated. If he is matched to a house successfully and he is a tenant, then the house he used to own is put into the set which can be allocated. In sequence, by the priority, the other individuals who have not been matched continue to choose according the rule and the house set to be allocated adjusts accordingly.
At the end of the step, if the set of houses to be allocated is not empty and some individuals still are not matched to houses, then matching process gets into next period, otherwise, ends. For individual $i \in N_U$, if he does not been matched, then he will be matched to $h_0$; for individual $i \in N_E$, he will be matched to $h_i$.

However, the following example reveals that random serial dictatorship with waiting list is not ex post Pareto efficient.

**Example 22.3.5** Suppose $N_E = \{1, 2, 3\}, N_U = \emptyset, H_E = \{h_1, h_2, h_3\}, H_V = \{h_4\}$. The preference of individuals is:

$$p(1) = h_2, h_3, h_1, h_4;$$
$$p(2) = h_3, h_1, h_2, h_4;$$
$$p(3) = h_1, h_4, h_3, h_2.$$

Suppose choice order function is $f(i) = i, i = 1, 2, 3$. The matching process of random serial dictatorship with waiting list can be described as:

**Period 1**, houses that can be allocated are $RH_1 = \{h_4\}$ and only individual 3 is matched, $\mu(3) = h_4$, and then $h_3$ is put into the set of houses that can be allocated.

**Period 2**, houses that can be allocated are $RH_2 = \{h_3\}$ and only individual 1 is matched, $\mu(1) = h_3$, and then $h_1$ is put into the set of houses that can be allocated.

**Period 3**, houses that can be allocated are $RH_3 = \{h_1\}$ and only individual 2 is matched, $\mu(2) = h_1$, and then $h_2$ is put into the set of houses that can be allocated. Since all individuals have been matched, the matching ends.

The outcome of random serial dictatorship with waiting list is:

$$\mu = \begin{pmatrix}
1 & 2 & 3 \\
3 & h_3 & h_4 \\
h_1 & h_4 \\
\end{pmatrix}. $$
This matching outcome can guarantee that benefit of each tenant will not be damaged, that is, it satisfies ex post individual rationality; however, it is not Pareto efficient since it is Pareto dominated by the following matching:

$$
\mu' = \begin{pmatrix}
1 & 2 & 3 \\
h_2 & h_3 & h_1
\end{pmatrix}.
$$

Abdulkadiroglu and Sonmez (1999) combined serial dictatorship with the top trading cycle and proposed the generalized top trading cycles mechanism to solve the efficiency problem in tenants’ house allocation. Given a preference ordering function $f$ and the profile of preference orders $p$ as reported by each person, the following steps describe how the generalized top trading cycles mechanism works:

Period 1, $RH_1 = H_V$ is the set of houses that can be allocated at the beginning. Each one points to his most favourite house. Each tenant has the highest priority to the house that he has already occupied and the individual $f(1)$, who is the first to choose, has the highest priority to all houses that will be allocated. Since the numbers of individuals and houses are both finite, there must exist a top cycle chain $A^1 = \{i_1, \ldots, i_k\}$ such that for individual $i_j, j < k$, the highest priority of the most favourite house is $i_{j+1}$ and for $i_k$, the highest priority of the most favourite house is $i_1$. At this time, each individual in top cycle chain is matched to his most favourite house. Then these individuals and houses which has been allocated are excluded from the matching in next period. If a tenant is in top cycle, then after the matching in this period, the house he occupied before will be put into the set that can be allocated next period. The set at the beginning of period 1 excluding the houses that have already been allocated and the set of added houses together are called the set of houses that can be allocated at the end of period 1. This set of houses together with the individuals who are not in period 1’s top cycle chain will enter the matching in the next period.
Period $t$, the set of houses $RH_t$ which can be allocated is the set at the end of the previous period. In this period, each remaining individual points to his favourite among remaining houses. Each tenant has the highest priority to the house that he has already occupied. The individual, who is the first to choose among remaining individuals, has the highest priority to all allocated houses in this period. Similarly, there exists a top cycle chain in this period. Each individual in the chain is matched to the house that he has pointed to. They and the houses that have been matched are excluded from the matching in next period. If a tenant is in the chain, then the house he has occupied will be put into the set that can be allocated at the end of this period. When all individuals are matched or the set of houses that can be allocated is empty or all tenants are matched, the matching process ends, otherwise it gets into next period. If the set of houses that can be allocated is empty but some new individuals are not matched, they are matched to $h_0$ at the end.

Let us start with an example to understand the generalized top trading cycle mechanism.

**Example 22.3.6** Suppose $N_E = \{1, 2, 3, 4\}, N_U = \{5\}, H_E = \{h_1, h_2, h_3, h_4\}, H_V = \{h_5, h_6, h_7\}$. Suppose $i \in N_E$ owns house $h_i$, and the order function $f$ satisfies $f(i) = i, i \in N$. The preference ordering of participants is:

- $p(1) = h_2, h_6, h_5, h_1, h_4, h_3, h_7$;
- $p(2) = h_7, h_1, h_6, h_5, h_4, h_3, h_2$;
- $p(3) = h_2, h_1, h_4, h_7, h_3, h_6, h_5$;
- $p(4) = h_2, h_4, h_3, h_6, h_1, h_7, h_5$;
- $p(5) = h_4, h_3, h_7, h_1, h_2, h_5, h_6$.

**Sept 1**: The set of available houses for step 1 is $RH_1 = H_V = \{h_5, h_6, h_7\}$. The top cycles are 1 and 2, and the matchings for them are $\mu(1) = h_2$ and $\mu(2) = h_7$;
Step 2: The set of available houses for step 2 is \(RH_2 = \{h_1, h_5, h_6\}\). The top cycles are 3 and 4, and the matchings for them are \(\mu(3) = h_1\) and \(\mu(4) = h_4\).

Step 3: The set of available houses for step 3 is \(RH_3 = \{h_3, h_5, h_6\}\). The top cycle is 5, and the matching is \(\mu(5) = h_3\). The matching ends and the matching outcomes are:

\[
\mu = \begin{pmatrix}
1 & 2 & 3 & 4 & 5 \\
\ h_2 & h_7 & h_1 & h_4 & h_3
\end{pmatrix}.
\]

Obviously, the matching outcome satisfies individually rationality and Pareto efficiency. The generalized top trading cycles mechanism ensures that each tenant is motivated to take part in. It is because when and only when tenant \(i\) is in a certain period of a top cycle, he will give the house he occupied for allocation. In a top cycle, the house he will be matched to surely will not be worse than the one he occupied.

When all participants are tenants, the above generalized top trading cycles mechanism reduces to the top trading cycles mechanism. When none of the participants are tenants, the above mechanism reduces to the serial dictatorship mechanism. When there is only one tenant, it reduces to the random serial dictatorship with order adjustment.

Now we discuss the characteristics of the generalized top trading cycles mechanism. Like the top trading cycles mechanism, it is Pareto efficient and strategy-proof.

Given any preference profile and ordering, a tenant will not be matched to a house worse than the one he has occupied before because the condition for a tenant to give up his house is for him to enter a certain top cycle. Besides, every tenant is matched to a house through top trading cycle. As we discussed before, matches for tenants in the generalized top trading cycles mechanism are Pareto efficient, and meanwhile, every tenant reporting true preference is dominant strategy. And for the new coming participant, his matching process combines the serial dictatorship with the top trading cycles mechanism. Only when he becomes the participant of the highest priority in the allocation of available houses at some stage, he can
participate in the top trading cycles and get a match.

To improve his interest, he needs to have an earlier order of choice, which inevitably undermines the interest of those who have been matched before. Therefore, for the participants of any new coming group, the matching outcome of the generalized top trading cycles mechanism is Pareto efficient. In addition, for the newcomer, his preference for reporting does not affect his eligibility to participate in the top cycle. Once he becomes the participant of the highest priority in a set of houses available for allocation, truthful reporting of his preference is the dominant strategy in the top trading cycle. We then have the following theorem.

\textbf{Theorem 22.3.5} \textit{For any ordering }$f$\textit{, the generalized top trading cycles mechanism is individually rational, Pareto efficient, and strategy-proof.}

From Theorem 22.3.5, we have the following corollary.

\textbf{Corollary 22.3.3} \textit{For any random ordering functions, the generalized top trading cycles mechanism is ex-post individually rational, ex-post Pareto efficient, and strategy-proof.}

Abdulkadiroglu and Sonmez (1999) proposed the following \textit{you request my house-I get your turn} (or in short YRMH-IGYT) mechanism, which can lead to the outcome of the generalized top trading cycles mechanism. The YRMH-IGYT mechanism is as follows:

For any given ordering $f$, assign the favorite house (or the one he points to) to the first agent $f(1)$, then assign the favorite house to the second agent $f(2)$ among the remaining houses, and so on, until someone demands the house occupied by a tenant. If at that point the tenant whose house is demanded is already matched to a house, then the procedure goes on. Otherwise, the order of the tenant is placed in front of all the rest and then proceed. If at any point a cycle forms, it is formed by tenants exclusively, and each of them demands the house of the tenant next in the cycle.

A cycle $A$ is an ordered list of tenants $\{i_1, i_2, \ldots, i_k\} \subseteq N_E$ (tenant $i_1$ demands the house of tenant $i_2$, $i_2$ demands the house of $i_3$, $\ldots$, $i_k$ demands the house of
In this case, every participant in this cycle is matched to the house he points to, so all the houses that have been matched are excluded from the allocation.

Note that if there is only one tenant, YRMH-IGYT mechanism reduces to the random serial dictatorship with order adjustment.

**Theorem 22.3.6** For a given ordering $f$, the YRMH-IGYT mechanism yields the same outcome as the generalized top trading cycles mechanism.

**Proof.** In any phase, when the remaining participant is $j$ and the remaining house is $G$, YRMH-IGYT mechanism runs to the next stage in the following two possible ways.

Case 1: There is a sequence of participants $\{i_1, i_2, \ldots, i_k\} \subseteq N_E$ ($k = 1$ is possible), where $i_1$ has the highest priority to choose among $J$ houses and demands the house of $i_2$, $i_2$ demands the house of $i_3$, ..., $i_k$ demands an available house $h \in G$. At this point, $i_k$ is matched to house $h$, $i_{k-1}$ is matched to house $h_{i_k}$, ..., and finally $i_1$ is matched to the house of $i_2$. When there are still a set of participants $J$ and a set of houses $G$ remaining in the generalized top cycle, $i_1, i_2, \ldots, i_k$ also form a top trading cycle, and the matching outcome is the same with the generalized top trading cycles mechanism.

Case 2: There is a cycle of agents $\{i_1, i_2, \ldots, i_k\} \subseteq N_E$ ($k = 1$ is possible). When agent $i_1$ is matched to the house of $i_2$, agent $i_2$ is matched to the house of $i_3$, ..., agent $i_k$ is matched to house of $i_1$, the matching outcome is also the same with the generalized top trading cycles mechanism.

Hence the YRMH-IGYT mechanism locates a cycle and implements the associated matches for any sets of remaining participants and houses. This observation is the same with the generalized top trading cycles mechanism.

In the literature, Papai (2000) studied hierarchical exchange rule, which is a generalized framework, including the serial dictatorship mechanism, the top trading cycles mechanism and the YRMH-IGYT mechanism. There are some very good review articles on one-sided matching, among which Abdulkadiroglu and Sonmez (2013) and Sonmez and Unver (2010) provided rather comprehensive
discussions and reviews of the theoretical problems and applications of one-sided matching. On the matter of house allocation, because determined allocation lacks fairness, random allocation gradually becomes a common tool. Bigomolnaia and Moulin (2001) proposed the probabilistic serial mechanism, which, together with the random serial dictatorship mechanism, is the main mechanism in this field. Currently, random allocation is one of main directions of research on one-sided matching. Besides efficiency, fairness, and incentive, when there is external options for participants (that is to say, the participants keep unmatched to any houses), the number of items to be allocated in the mechanism is another concern for society. See Bogomolnaia and Moulin (2015) and Huang and Tian (2017) for relevant discussions.

22.4 Application

Next, we investigate two specific matching problems of great practical significance by applying the two-sided matching and one-sided matching theory: school admission matching and organ transplant matching. We find that despite the differences in the issues discussed, the logic behind the operational mechanism embodies a high degree of internal consistency.

22.4.1 School Choice

Education is an important foundation for social progress and development, and school education is an important way to promote human capital for individuals. Because resources including educational resources (especially high-quality education resources) are scarce, the efficient allocation of educational resources becomes an extremely important problem. At the same time, educational allocation is closely related to social justice. In this way, allocation of educational resources involves not only efficiency, but also fairness, so that it cannot depend solely on the price mechanism but also requires the input of public resources and the guidance of education policy.
In the previous two-sided matching, we found in college admission, the school and the students choose different targets according to their preferences, and when there is no Pareto improving matching between schools and students, the matching becomes stable. The Gale-shapley deferred acceptance algorithm gives such a stable matching mechanism. Furthermore, the matching obtained by the deferred acceptance algorithm has some good properties to the proposing side such as incentive compatibility in the preference disclosure and also it is the optimal to the proposing side among all possible stable matchings.

However, there are many different forms of manifestation of admission problems in different enrollment stages and under different institutional backgrounds. For instance, in university enrollment of many countries, most admissions are determined according to certain indicators, such as grades of national college entrance examination in China. In the compulsory education stage, taking China as an example, students’ enrollment depends on factors such as “hukou” (location of household registration) and the distance between residence and school. At the same time, in the education that aims to improve the quality of citizens and enhance the human capital, students are the most fundamental concern of the education policy, and the school to a large degree is only a carrier of the educational process. Therefore, student enrollment is more inclined to show the characteristics of one-sided matching. In the literature, this problem is called School Choice problem, and the social welfare regards the interest of students as the objective function. Below we have a rigorous model to characterize the admissions problem.

Suppose $N = \{1, 2, \ldots, n\}$ is a collection of students, $S = \{s_1, \ldots, s_m\}$ is a collection of schools, $q = (q_s)_{s \in S}$ is the school enrollment quota, and $\succ_i, i \in N$ is the student’s (strict) preference for school. For school $s$, there is a priority rule for students in enrollment, which is called priority order, $\succ_s$. Similarly, for college admission, the achievement of students determines their priority, and for enrollment in the period of compulsory education, factors like the hukou, residence, and so on determines the student’s priority. The college’s priority for
students can be understood in the two-sided matching framework as a school’s preference (but for school admission, schools do not act as a strategic decision-maker to admit students, so it may be assumed that schools cannot manipulate their preferences). In the framework of one-sided matching, students’ priority can be understood as a sort of order function. However, when different schools (such as ordinary colleges and art colleges) may have different considerations for priority, this is a local priority.

A matching in school choosing $\mu : N \cup S \rightarrow 2^{N \cup S}$ needs to satisfy the following conditions:

- $\mu(i) \subseteq S$, $|\mu(i)| \leq 1$, $\forall i \in N$;
- $\mu(s) \subseteq N$, $|\mu(s)| \leq q_s$, $\forall s \in S$;
- $s \in \mu(i)$ iff $i \in \mu(s)$, $\forall i \in N$, $s \in S$,

where $|\mu(i)| = 0$ means $\mu(i) = \emptyset$, indicating that the participant $i$ is not admitted by any school.

**Definition 22.4.1** Let $\mu : N \cup S \rightarrow 2^{N \cup S}$ be a matching.

- It is individually rational if for any $i \in N$, $\mu(i) \succ_i \emptyset$;
- It is non-wasteful if whenever $s \succ_i \mu(i)$, $\forall i \in N$, $s \in S$, then $|\mu^{-1}(s)| = q_s$;
- It is fair or eliminates justified envy if there doesn’t exist $i \succ_{\mu(i)} j$ that satisfies $\mu(j) \succ_j \mu(i)$;
- It is Pareto efficient if there doesn’t exist $\nu$ such that $\nu(i) \succeq_{i, \mu(i)} \forall i \in N$, and there exists at least one $j \in N$ that satisfies $\nu(j) \succ_j \mu(j)$;
- It is stable if $\mu$ is individually rational, non-wasteful, and eliminates justified envy;
- A stable matching $\mu$ is student-optimal if there exists no other stable matchings $\nu$ that Pareto dominates $\mu$. 

1053
As for the features of matchings for school choice, we mainly focus on efficiency, stability and incentive compatibility (in the revelation of preferences). In reality, many school-choosing mechanisms do not have the feature of incentive compatibility, such as the Boston mechanism studied by Abdulkadiroglu and Sonmez (2003). Let us first describe the Boston mechanism:

In each school, a priority rule (such as home location, siblings if they attended the school, etc.) is given beforehand.

Each student reports his preference to the school, such as the first choice of school, second choice of school, etc.

Round 1: The school considers only the students who choose it as their first choice of school: among all the students whose first choices are $s$, according to the priority $\succ_s$, the highest $q_s$ students are enrolled. If the quota of enrollments is not completed, the admission proceeds to the next round. If it is completed, the admission is over.

Round $k$: Among the remaining students, the schools which are not yet been fully enrolled will consider the students who choose them as $k$-th choice of school and accept those of the highest priority according to the priority of $\succ_s$. The number of admitted students does not exceed the remaining quota. If in this round, the quota of enrollment is completed, the admission will end, otherwise it will enter the next round. Until all the schools have no vacancy, or all the students are admitted, the admission ends.

The following example reveals that the Boston mechanism does not satisfy incentive compatibility (in the revelation of preferences).

**Example 22.4.1** Suppose that there are three students, three schools, $N = \{1, 2, 3\}, S = \{s_1, s_2, s_3\}, q_s = 1, \forall i \in$, students’ preferences and school priorities
According to the Boston matching mechanism, the matching outcomes are:

\[
\mu^B = \begin{pmatrix}
  s_1 & s_2 & s_3 \\
  3 & 1 & 2
\end{pmatrix}.
\]

First of all, this matching mechanism does not satisfy the elimination of justified envy, because \( s_2 = \mu(1) \succ_2 \mu(2) = s_3 \), and \( 2 \succ s_2 1 \). Secondly, it also does not satisfy incentive compatibility, because if participant 2 reports a preference: \( p_2' = s_2, s_1, s_3 \), the matching outcome is:

\[
\mu'^B = \begin{pmatrix}
  s_1 & s_2 & s_3 \\
  3 & 2 & 1
\end{pmatrix}.
\]

Abdulkadiroglu and Sonmez (2003) proposed two matching mechanisms to solve the problem of incentive compatibility in preference revelation. One is the student-optimal deferred acceptance algorithm \( \mu^{SOSM} \), and the other is the top trading cycles mechanism \( \mu^{TTC} \). The definition of \( \mu^{TTC} \) is below:

**Definition 22.4.2 (Top Trading Cycle Mechanism for School Choice)** The top trading cycles algorithm for school choice is given by the following steps:

1. Step 1: Assign one counter to each school to show how many remaining seats it still has. The initial setting of the counter will be the admission quota of the school. Each student points to his favorite school according to his preference as reported. Each school points to the student who has the highest priority in this school. Because the numbers of students and schools are finite, there exists at least one cycle. In addition, each school will be in one cycle at most, so will each student. In the cycle, each student will be assigned to a
seat in the school that he points to, and then the student will be removed from the algorithm. In each cycle, the count as shown on the counter of each school will be decreased by 1, and the school will be removed when its counter’s count becomes 0. The count of other schools’ counters stays unchanged.

Step k: Each remaining student points to his favorite school among the remaining schools. Each remaining school points to the student who has the highest priority among all remaining students. There exists at least one cycle. Each student in the cycle will be assigned to a seat in the school that he points to and then be removed. In each cycle, the count as shown by the counter of each school will be decreased by 1, and the school will be removed when its counter’s count becomes 0. The count of other schools’ counters stays unchanged.

When each of all the students is assigned to one seat, the algorithm ends. The number of steps of the algorithm will not exceed the cardinal number of the set of students.

Based on the analysis of the previous two sections, we can draw the following conclusions: Firstly, the matching $\mu^{SOSM}$ is stable and is the best matching among all stable matchings for students, and the matching $\mu^{SOSM}$ satisfies incentive compatibility in the preference disclosure.

Secondly, the matching $\mu^{TTC}$ is Pareto efficient and satisfies incentive compatibility in preference revelation.

Based on Example 22.4.1, we continue to discuss these two mechanisms.

**Example 22.4.2** Consider the matching of students and schools, as described in the Example 22.4.1. It is easy to find the deferred acceptance algorithm matching $\mu^{SOSM}$ and the top trading cycles matching $\mu^{TTC}$.

$$
\mu^{SOSM} = \begin{pmatrix}
s_1 & s_2 & s_3 \\
1 & 2 & 3
\end{pmatrix};
\mu^{TTC} = \begin{pmatrix}
s_1 & s_2 & s_3 \\
2 & 1 & 3
\end{pmatrix}.
$$
Obviously, \( \mu^{SOSM} \) is not Pareto efficient, because it is (Pareto) dominant by \( \mu^{TTC} \). At the same time, \( \mu^{TTC} \) also does not satisfy the elimination of justified envy as one of the conditions of stability, because \( s_1 = \mu(2) \succ_3 \mu(3) = 3 \), but \( 3 \succ_{s_1} 2 \).

However, in the above example, the matching outcomes of \( \mu^{TTC} \) Pareto dominate \( \mu^{SOSM} \), but \( \mu^{TTC} \) mechanism does not Pareto dominate \( \mu^{SOSM} \). The following example illustrates this point.

**Example 22.4.3** Suppose there are 3 students, 3 schools, \( N = \{1, 2, 3\} \), \( S = \{s_1, s_2, s_3\} \), \( q_s = 1, \forall i \in S \), student preferences and school priorities are described below, with preferences of participant 2 different from those in Example 22.4.1.

\[
\begin{align*}
p_1 &= s_2, s_1, s_3; \quad & p_{s_1} &= 1, 3, 2; \\
p_2 &= s_1, s_3, s_2; \quad & p_{s_2} &= 2, 1, 3; \\
p_3 &= s_1, s_2, s_3; \quad & p_{s_3} &= 2, 1, 3.
\end{align*}
\]

It is easy to find that the deferred acceptance algorithm matching \( \mu^{SOSM} \) and top trading cycles matching \( \mu^{TTC} \) are:

\[
\begin{align*}
\mu^{SOSM} &= \begin{pmatrix} s_1 & s_2 & s_3 \\ 3 & 1 & 2 \end{pmatrix}; \\
\mu^{TTC} &= \begin{pmatrix} s_1 & s_2 & s_3 \\ 2 & 1 & 3 \end{pmatrix}.
\end{align*}
\]

In this example, there is no Pareto dominant relationship between \( \mu^{SOSM} \) and \( \mu^{TTC} \).

We know that both \( \mu^{SOSM} \) and \( \mu^{TTC} \) satisfy incentive compatibility in preference revelation, and \( \mu^{TTC} \) is Pareto efficient. The above example reveals that \( \mu^{TTC} \) does not Pareto dominate \( \mu^{SOSM} \). This leads to a problem: is there an incentive-compatible Pareto-efficient matching mechanism that Pareto dominates \( \mu^{SOSM} \)?

Keston (2010) gave a negative answer, and provided the following theorem.

**Theorem 22.4.1** If the priority order of the school is strict, then there is no incentive-compatible Pareto-efficient matching mechanism that Pareto dominates \( \mu^{SOSM} \).
Keston (2010) put forward sacrificing incentive compatibility for efficiency, and constructed a matching mechanism to improve the allocative efficiency, which is called the efficiency-adjusted deferred accepted Mechanism, abbreviated as the EADAM mechanism.

Let us first introduce this mechanism.

**Definition 22.4.3** We call student \(i\) the disturbance of school \(s\) in the deferred acceptance mechanism (student-optimal), if the \(i\) offers to the school \(s\) under the deferred acceptance mechanism, so that a student \(j\) is rejected by the school \(s\), but the school \(s\) refuses students \(i\) after receiving the order of the student \(k\).

In the above definition, if student \(i\) is the disturbance of school \(s\), his invitation causes the school \(s\) to reject a student, but finally still does not get \(i\) accepted. As such, \(i\)’s behavior harm \(j\) without benefiting himself. If participants can sign a contract beforehand to promise not to disturb each other, that is, the student \(i\) who is originally the disturbance of the school \(s\) promises not to extend the offer to the school \(s\), then this may improve the matching efficiency.

The following is an example how to find out who is a disturbance.

**Example 22.4.4** Consider a matching between two schools and three students. \(S = \{s_1, s_2\}, N = \{1, 2, 3\}, q_s = 1, \forall s \in S\), and students’ preference order and school priority order are as follows:

\[
\begin{align*}
   p_1 &= s_2, s_1; & p_{s_1} &= 1, 2, 3; \\
   p_2 &= s_1; & p_{s_2} &= 3, 1; \\
   p_3 &= s_1, s_2.
\end{align*}
\]

The process of the student-optimal deferred acceptance algorithm is as follows (the box indicates temporary acceptance):
Then the matching is:

$$\mu = \begin{pmatrix} s_1 & s_2 & \emptyset \\ 2 & 3 & 1 \\ 1,3 \\ 1,2 \vdots \\ \vdots & 2 \\ 1 & 3 & 2 \end{pmatrix}.$$ 

We see that student 2 is a disturbance of \( s_1 \). If student 2 agrees to remove \( s_1 \) from his preference list, that is, he does not select \( s_1 \), then the matching process becomes:

$$\mu_E = \begin{pmatrix} s_1 & s_2 & \emptyset \\ 1 & 3 & 2 \end{pmatrix}.$$ 

Thus the matching is:

$$\mu^E = \begin{pmatrix} s_1 & s_2 & \emptyset \\ 3 & 1 & 2 \end{pmatrix}.$$ 

Obviously, \( \mu^E \) is a Pareto improvement of \( \mu \), and it is a Pareto efficient matching.

Here we briefly describe the EADAM mechanism:

Round 0, run the deferred acceptance algorithm (referred to as DA algorithm).

Round \( k \), find disturbance \((i, s)\) who agrees not to disturb, remove \( s \) from the preferences of \( i \) and rerun the DA algorithm.

Once no disturbance appears in the DA algorithm, the outcome of this round of DA algorithm is the EADAM matching outcome.

Let us use the following example to discuss the EADAM mechanism.
Example 22.4.5 Suppose there are six students $N = \{1, \ldots, 6\}$ and five schools $S = \{s_1, s_2, s_3, s_4, s_5\}$. The admission quota is $q_s = 1, \forall s \neq s_5, q_{s_5} = 2$. The priority order of the school is:

\[
\begin{align*}
p_{s_1} &= 2, 1, 5, 6, 4, 3; \\
p_{s_2} &= 3, 6, 4, 1, \ldots; \\
p_{s_3} &= 1, 6, 2, 3, \ldots; \\
p_{s_4} &= 4, 3, 6, \ldots; \\
p_{s_5} &= \ldots,
\end{align*}
\]

The omitted are orderings that are not different or do not affect the matching. The students’ preferences are:

\[
\begin{align*}
p_1 &= s_2, s_1, s_3, \ldots; \\
p_2 &= s_3, s_1, s_5, \ldots; \\
p_3 &= s_3, s_4, s_2, \ldots; \\
p_4 &= s_1, s_2, s_4, \ldots; \\
p_5 &= s_1, s_5, \ldots; \\
p_6 &= s_4, s_1, s_3, s_2, s_5, \ldots.
\end{align*}
\]

Suppose that all students agree not to be the disturbance, the EADAM mechanism operates as follows:

Round 0, run the DA algorithm:
Round 1, identify \((6, s_2)\) in the 9th step in the previous round as the last group of disturbance and corresponding school, remove \(s_2\) from the preference of student 6, rerun DA algorithm and get:

<table>
<thead>
<tr>
<th>Step</th>
<th>(s_1)</th>
<th>(s_2)</th>
<th>(s_3)</th>
<th>(s_4)</th>
<th>(s_5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5,4</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1,5,6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>6</td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2,1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>1</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>3</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3,6</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>5,6</td>
</tr>
</tbody>
</table>

Round 2, identify \((6, s_3)\) in the 6th step in the previous round as the last group of disturbance and corresponding school, remove \(s_3\) from the preference of student 6, rerun DA algorithm and get:

<table>
<thead>
<tr>
<th>Step</th>
<th>(s_1)</th>
<th>(s_2)</th>
<th>(s_3)</th>
<th>(s_4)</th>
<th>(s_5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5,4</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1,5,6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>6</td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2,1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>1</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>5,6</td>
</tr>
</tbody>
</table>

1061
Round 3, identify \((5, s_1)\) in the 3rd step in the previous round as the last group of disturbance and corresponding school, remove \(s_1\) from the preference of student 5, rerun DA algorithm and get:

<table>
<thead>
<tr>
<th>Step</th>
<th>(s_1)</th>
<th>(s_2)</th>
<th>(s_3)</th>
<th>(s_4)</th>
<th>(s_5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5,4</td>
<td>1</td>
<td>2,3</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>4,1</td>
<td></td>
<td></td>
<td>3,6</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>1,5,6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>5,6</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1,4</td>
<td>2</td>
<td>3</td>
<td>5,6</td>
<td></td>
</tr>
</tbody>
</table>

Round 4, identify \((6, s_1)\) in the 5th step in the previous round as the last group of disturbance and corresponding school, remove \(s_1\) from the preference of student 6, rerun DA algorithm and get:

<table>
<thead>
<tr>
<th>Step</th>
<th>(s_1)</th>
<th>(s_2)</th>
<th>(s_3)</th>
<th>(s_4)</th>
<th>(s_5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>1</td>
<td>2,3</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3,6</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>6,4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>4,1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1,6</td>
<td></td>
<td></td>
<td></td>
<td>5,6</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>5,6</td>
</tr>
</tbody>
</table>

Round 5, the algorithm ends because no disturbance appears in the last round.
The final match outcome is:

\[ \mu^E = \left( \begin{array}{ccccc} s_1 & s_2 & s_3 & s_4 & s_5 \\ 4 & 1 & 2 & 3 & \{5, 6\} \end{array} \right), \]

which is Pareto efficient.

Kesten (2010) proved that, as long as the disturbing students agree to remove the unmatched corresponding school from his preference, the EADAM mechanism is a Pareto improvement for \( \mu^{SOSM} \). If all the students who are disturbance agree to remove the unmatched schools from their preferences, then the EADAM mechanism will result in Pareto efficient matching. For students who do not participate in the agreement, the EADAM mechanism can eliminate their justified envy or satisfy fairness for them. At the same time, under the EADAM mechanism, the welfare of students who do not participate in the agreement will not be reduced even if they do participate. The proofs of these results long and readers who are interested can refer to Kesten (2010).

However, the EADAM mechanism does not satisfy incentive compatibility in preference revelation, as shown by the following example.

**Example 22.4.6** Consider a matching with three students \( N = \{1, 2, 3\} \) and three schools \( S = \{s_1, s_2, s_3\} \), \( q_s = 1, \forall s \in S \), students’ preference and school’s limited order are as follows:

\[ p_1 = s_1, s_2, \quad p_{s_1} = 3, 1, 2; \]
\[ p_2 = s_1, s_2, s_3, \quad p_{s_2} = 2, 3; \]
\[ p_3 = s_3, s_1, \quad p_{s_2} = 2, 1. \]

Suppose all students agree not to be disturbance. In Round 0, DA algorithm works as follows when all students are truly disclosing their preferences:

<table>
<thead>
<tr>
<th>Step</th>
<th>( s_1 )</th>
<th>( s_2 )</th>
<th>( s_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
The matching outcome is

$$\mu^E(p_2, p_{-2}) = \begin{pmatrix} s_1 & s_2 & s_3 \\ 1 & 2 & 3, \end{pmatrix}$$

which is Pareto efficient. However, if student 2 manipulates his preference as: $p'_2 = s_1, s_3, s_2$. In Round 0, DA algorithm works as follows:

<table>
<thead>
<tr>
<th>Step</th>
<th>$s_1$</th>
<th>$s_2$</th>
<th>$s_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1, 2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2, 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3, 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3, 1, 2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Round 1, identify (1, $s_1$) in the 3rd step in the previous round as the last group of disturbed students and corresponding school, remove $s_1$ from the preferences of student 1, rerun DA algorithm and get:

<table>
<thead>
<tr>
<th>Step</th>
<th>$s_1$</th>
<th>$s_2$</th>
<th>$s_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2, 1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2, 1</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

The matching outcomes is

$$\mu^E(p'_2, p_{-2}) = \begin{pmatrix} s_1 & s_2 & s_3 \\ 2 & 1 & 3, \end{pmatrix}.$$
22.4.2 Matching in Organ Transplantation

In this subsection, we use the previous matching results to analyze the matching problem in organ transplantation. Organ transplantation is an efficient method in the treatment of many serious diseases. Organ transplantation is not only a medical problem but also a question of economics. Abdulkadiroglu and Sonmez (2013) made a brief description of the American kidney transplant market between 2009-2010. In July 2010, more than 85,000 patients were awaiting kidney transplants. A total of 16,829 kidney transplants were performed in 2009, of which 10,422 used kidneys from remains donors, and 6,387 were from living donors. In the same year, 35,123 new patients were added to the waiting list of remains kidney transplant, while 4,789 of patients died in the process of waiting.

In China, the world’s most populous country, the problem is likely to be even worse.

Although organ transplant is a market of scarce resource, most countries prohibit organ trade, and organ donation is the only possible source. The following is the discussion of kidney transplants. In kidney transplantation, blood-matching and tissue rejection are important factors that affect the success of surgery. Many kidneys donated to specific individuals cannot be used efficiently because of these problems.

Rapaport (1986) put forward the idea of exchange between recipients-donors portfolio, which allows patients to receive a suitable kidney source for transplant. The exchange of donors between recipients is ethically acceptable, which is called direct exchange. In addition, there is a kind of indirect exchange between a recipient who is unsuitable with his donor and a patient in the waiting list of remains kidney donations, in which the former gives the kidney he receives to the latter in exchange for a higher priority in the waiting list.

Roth, Sonmez and Unver (2004) compared these two kinds of exchange forms in the framework of the one-sided matching. If donated kidneys are compared to houses to be allocated or exchanged, the direct exchange between the kidney sources and the house exchange problem are equivalent, and the top-trading cycles
matching mechanism discussed above has good properties. The indirect exchange between a recipient and a patient in the waiting list is similar to that of house allocation with tenants, and for the latter, YRMH-IGYT (you request my house-I get your turn) mechanism has good properties.

However, there are some different features of the allocation in kidney transplantation and the allocation of houses. In house allocation, the collection of houses to be allocated is given beforehand, but in kidney transplant, remains donations or donations not for specific recipients, the type of kidney donated and the time to get the kidney are unknown and uncertain. Thus, YRMH-IGYT can not be directly applied to the indirect exchange. Roth, Sonmez and Unver (2004) proposed a new matching mechanism, called the top trading cycles and chains mechanism, referred to as the TTCC mechanism. Here we mainly discuss this mechanism. We first need to introduce some notations and definitions.

A kidney exchange (static) problem includes the following components: firstly, there are pairs of recipients and donors; the set of these pairs is \{((k_1, t_1), \ldots, (k_n, t_n))\}, the set of recipients is \(T = \{t_1, \ldots, t_n\}\), and the set of donors’ kidneys is noted as \(K = \{k_1, \ldots, k_n\}\). \(K_i \subseteq K\) is a collection of kidney sources compatible with recipients \(t_i\). Because remains kidneys for patients in the waiting list are uncertain, here we do not give direct depiction of the set of waiting patients and future remains kidney donations.

Assume that the recipient \(t_i\) chooses to give the kidney to a patient in the waiting list, he will be given a priority in the waiting list of remains kidney donations; this priority is a certain order, the sequence and the possible (uncertain) kidney source he gets is \(w\). For the recipient \(t_i\)’s preference \(\succ_i\), the corresponding preference ordering \(p_i\) is a sort on \(K_i \cup \{k_i, w\}\). The key sort in this ordering is the preference order between \(k_i\) and \(w\): if \(k_i \succ_i w\), all kidney sources worse than \(k_i\) will not be considered by \(t_i\), and if \(w \succ_i k_i\), all kidney sources worse than \(w\) will not be considered by \(t_i\).

A matching \(\mu\) in the kidney exchange problem is to match each recipient with a compatible donor, or a kidney donated specific to him, or an order in the waiting
list of remains kidney donation, namely, \( \mu(t_i) \in K_i \cup \{k_i, w\} \). We call matching \( \mu \) is \textit{Pareto efficient}, if there is no other matching \( \nu \) such that \( \nu(i) \succeq i \mu(i), \forall i \in T \), and \( \nu(j) \succ_j \mu(j) \) for some \( j \in T \); matching mechanism \( \mu(.) \) is \textit{Pareto efficient}, if for any preference order profile \( p \), \( \mu(p) \) is a Pareto efficient matching; \( \mu(.) \) is \textit{strategy-proof}, or \textit{incentive-compatible} in preference revelation, if for any true preference order profile of \( p \), there is no \( p_i' \neq p_i \) such that \( \mu(p_i', p_{-i}) \succ_i \mu(p_i, p_{-i}). \)

Now we define a cycle in the kidney exchange problem. A cycle is an ordered sequence consisting of kidney sources and patients, namely, \((k'_1, t'_1, \ldots, k'_m, t'_m)\), and each element in the sequence points to the next element, where \( t'_m \) points to \( k'_1 \). The so-called patient pointing to the kidney source refers to the patient’s favorite kidney source in the selectable set, and the kidney source pointing to the patient means that the kidney is donated specifically to the patient. Apparently any two cycles do not intersect.

Then we define chain in the kidney exchange problem. A chain is an ordered sequence consisting of kidney sources and patients, namely \((k'_1, t'_1, \ldots, k'_m, t'_m)\), in which each element points to the next element, \( t'_m \) points to \( w \), \( t'_1 \)'s kidney source \( k'_1 \) is exchanged to a patient in the waiting list, \((k'_m, t'_m)\) is the head of the chain, and \((k'_1, t'_1)\) is the tail of the chain. Unlike cycles, a recipient and his kidney source may appear in multiple chains, as Example 22.4.7 below reveals. Because of the possibility of intersection in chains, the chain selection mechanism needs to be determined in the TTCC mechanism. The chain selection mechanism is discussed after introducing the TTCC mechanism.

If recipients and donors are limited and different, there is either a cycle or a chain. This is because, suppose there is no chain, it means that each patient \( t \) points to the kidney of other patient. Because the number of patients is finite, similar to the TTC mechanism where there must be a top cycle chain, there will always be a chain.

The top trading cycles and chains mechanism are described below:

\begin{itemize}
\item \textbf{Step 1:} Assume that all initial kidney sources, i.e., \( K \), are available for selection, and each recipient is an active participant. At each
stage:

(a) Each of the remaining active recipients points to his favorite kidney source among the remaining unmatched kidneys, or to $w$ in the waiting list, if the latter is more favored by the patient;

(b) Each remaining passive donated patient continues to point to matches available to him;

(c) Each remaining kidney source points to its recipient.

Step 2: There will be at least one cycle or a $w$ chain or both at each stage before it ends:

(a) If there is no cycle, continue to step 3; if there is a cycle, the matching in the cycle is performed, that is, patients in each cycle all get the kidney source they point to, while in the matching of the next stage, the patients in these chains and the kidney source matched to them are removed.

(b) The remaining recipients point to their favorite kidney sources among the remaining kidneys, or to $w$ in the waiting list, if the latter is more favored by the patients. If there is no chain, continue to step 3. Otherwise, carry on the matching in the chain, and remove the patients and matching kidneys in the chains. Repeat this step until there is no more cycles.

Step 3: If there is no combination pairs of recipients and donors, the algorithm ends. Otherwise, there is at least one chain. Select one of the chains according to the selection mechanism, execute the matching in the chain, that is, except the combination of the
chain head, each patient gets matched to the kidney source he points to, while the patient in the head of the chain points to $w$ in the waiting list. The chains selection rule decides which chain to choose, and also decide:

(a) Whether to remove the patient and kidney source in the selection chain;

(b) If the chosen chain is retained, patients in the chain become passive participants.

Step 4: Each time the $w$ chain is selected, there will be a new cycle and chain; then repeat the 2nd and 3rd steps until there is no active participant, and the algorithm ends. Here are the selection rules for the participants who are not in the head of the chain to be matched to their pointed kidney sources and for the participants in the head of the chain to be matched to some possible chains:

Rule A: Select the longest $w$ chain, remove it after matching, and if there are more than one longest chains, choose one of them in some way;

Rule B: Select the longest $w$ chain, keep it after matching, and if there are more than one longest chains, choose one of them in some way;

Rule C: First give a priority order to the recipient-donor combination, and then select the $w$ chain with the highest priority to be removed after matching;

Rule D: First give a priority order to the recipient-donor combination, and then select the $w$ chain with the highest priority to be retained after the match;

Later we will discuss the good property of rule D. An example is given below to show the operation of the TTCC mechanism (under rule D).
Example 22.4.7 Consider a kidney problem with there are 12 pairs of recipient-donor combinations, \( \{(k_1,t_1), \ldots, (k_{12},t_{12})\} \). \( w \) is the order on the waiting list of remains donations that are gotten through exchange, and the recipients are sorted in ascending order of subscript numbers, that is, \( t_1 \) ranks the highest, \( t_2 \) the second, and so on. Preferences of recipients are as follows:

\[
\begin{align*}
p_1 &= k_9, k_{10}, k_1; \\
p_2 &= k_{11}, k_3, k_5, k_6, k_2; \\
p_3 &= k_2, k_4, k_5, k_6, k_7, k_8, w; \\
p_4 &= k_5, k_9, k_1, k_8, k_{10}, k_3, w; \\
p_5 &= k_3, k_7, k_{11}, k_4, k_5; \\
p_6 &= k_3, k_5, k_8, k_6; \\
p_7 &= k_6, k_1, k_3, k_9, k_{10}, k_1, w; \\
p_8 &= k_6, k_4, k_{11}, k_2, k_3, k_8; \\
p_9 &= k_3, k_{11}, w; \\
p_{10} &= k_{11}, k_1, k_4, k_5, k_6, k_7, w; \\
p_{11} &= k_3, k_6, k_5, k_{11}; \\
p_{12} &= k_{11}, k_3, k_9, k_8, k_{10}, k_{12}.
\end{align*}
\]

Round 1: There exists a unique cycle, i.e. \( C_1 = (k_2 \rightarrow t_2 \rightarrow k_{11} \rightarrow t_{11} \rightarrow k_3 \rightarrow t_3) \). Perform cycle matching \( \mu(t_2) = k_{11}, \mu(t_{11}) = k_3, \mu(t_3) = k_2 \), and in the next round, remove the participants in cycle \( C_1 \) and the matched kidney sources.

Round 2: After \( C_1 \) is removed, a new cycle forms, namely, \( C_2 = (k_5 \rightarrow t_5 \rightarrow k_7 \rightarrow t_7 \rightarrow k_6 \rightarrow t_6) \). Perform cycle matching \( \mu(t_5) = k_7, \mu(t_7) = k_6, \mu(t_6) = k_5 \), and remove the participants in cycle \( C_2 \) and the matching kidney sources in the next round.

Round 3: The removal of \( C_2 \) does not generate a new cycle but produces two longest chains at this time: \( w_1 = (k_8 \rightarrow t_8 \rightarrow k_4 \rightarrow t_4 \rightarrow k_9 \rightarrow t_9) \) and \( w_2 = (k_{10} \rightarrow t_{10} \rightarrow k_1 \rightarrow t_1 \rightarrow k_9 \rightarrow t_9) \). Since there exists a recipient \( t_1 \) who has the highest ranking in the chain \( w_2 \), select \( w_2 \) and perform \( w_2 \) matching, that is \( \mu(t_{10}) = k_1, \mu(t_1) = k_9, \mu(t_9) = w \), and then keep chain \( w_2 \). At the same time, chain \( w_2 \) become passive participants.

Round 4: Fix chain \( w_2 \) and a new cycle forms, that is, \( C_3 = (k_4 \rightarrow t_4 \rightarrow k_8 \rightarrow t_8) \). Perform a cycle matching \( \mu(t_4) = k_8, \mu(t_8) = k_4 \), and remove participants in the cycle \( C_3 \) and matched kidney sources in the next round.

Round 5: Remove \( C_3 \) and no new cycle forms. At this time, the combination pair of \( (k_{12},t_{12}) \) is linked to the end of chain \( w_2 \), which generates a longer chain, \( w_3 = (k_{12} \rightarrow t_{12} \rightarrow k_{10} \rightarrow t_{10} \rightarrow k_1 \rightarrow t_1 \rightarrow k_9 \rightarrow t_9) \). Perform supplementary
match in the new chain, that is, $\mu(t_{12}) = k_{10}$, where all recipients are matched. The TTCC matching outcomes are:

$$
\mu^{TTCC} = \begin{pmatrix}
    t_1 & t_2 & t_3 & t_4 & t_5 & t_6 & t_7 & t_8 & t_9 & t_{10} & t_{11} & t_{12} \\
    k_9 & k_{11} & k_2 & k_8 & k_7 & k_5 & k_6 & k_4 & w & k_1 & k_3 & k_{10}
\end{pmatrix}.
$$

In the TTCC mechanism, the reason for choosing the longest chain is to let more recipients take part in the exchange and improve the efficiency of the matching. Roth, Sonmez and Unver (2004) further proved that the TTCC mechanism under the chain selection Rule D is a Pareto-efficient mechanism, and also a strategy-proof mechanism. The proofs are referred to the original paper. If we interpret the queueing order $w$ as empty houses to be allocated in the tenants’ house allocation model, Krishna and Wang (2007) further proved that the TTCC mechanism under the chain selection Rule D is equivalent to the YRMH-IGYT mechanism. The analysis, design, and improvement of organ transplantation from the perspective of economics are growing; Sonmez and Unver (2013) provided a more comprehensive overview on this respect.
Reference

Books:


Articles:


34. S. Papai (2000), Strategyproof assignment by hierarchical exchange, Econometrica, 68, 1403-1434.


