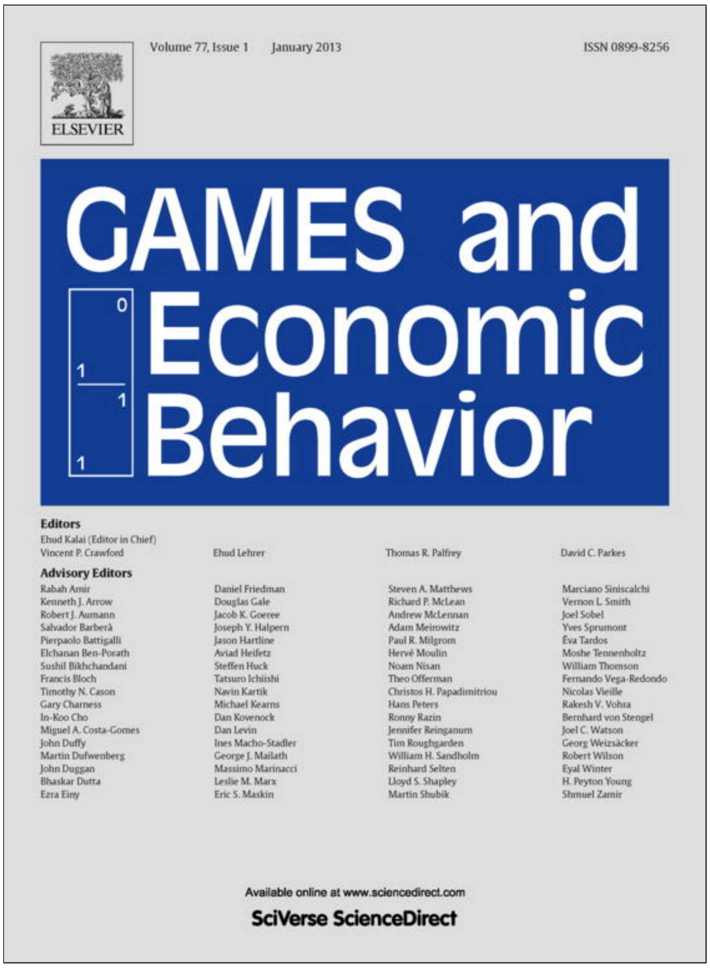
ELSEVIER
Volume 77, Issue 1 January 2013
ISSN 0899-8256
GAMES and Economic Behavior
Editors
Ehud Kalai (Editor in Chief)
Vincent P. Crawford
Ehud Lehrer
Thomas R. Palfrey
David C. Parkes
Advisory Editors
Rabah Amir
Kenneth J. Arrow
Robert J. Aumann
Salvador Barberà
Pierpaolo Battigalli
Elchanan Ben-Porath
Sushil Bikhchandani
Francis Bloch
Timothy N. Cason
Gary Charness
In-Koo Cho
Miguel A. Costa-Gomes
John Duffy
Martin Dufwenberg
John Duggan
Bhaskar Dutta
Ezra Einy
Daniel Friedman
Douglas Gale
Jacob K. Goeree
Joseph Y. Halpern
Jason Hartline
Aviad Heifetz
Steffen Huck
Tatsuro Ichiishi
Navin Kartik
Michael Kearns
Dan Kovenock
Dan Levin
Ines Macho-Stadler
George J. Mailath
Massimo Marinacci
Leslie M. Marx
Eric S. Maskin
Steven A. Matthews
Richard P. McLean
Andrew McLennan
Adam Meirowitz
Paul R. Milgrom
Hervé Moulin
Noam Nisan
Theo Offerman
Christos H. Papadimitriou
Hans Peters
Ronny Razin
Jennifer Reinganum
Tim Roughgarden
William H. Sandholm
Reinhard Selten
Lloyd S. Shapley
Martin Shubik
Marciano Siniscalchi
Vernon L. Smith
Joel Sobel
Yves Sprumont
Éva Tardos
Moshe Tennenholtz
William Thomson
Fernando Vega-Redondo
Nicolas Vieille
Rakesh V. Vohra
Bernhard von Stengel
Joel C. Watson
Georg Weizsäcker
Robert Wilson
Eyal Winter
H. Peyton Young
Shmuel Zamir
Available online at www.sciencedirect.com
SciVerse ScienceDirect

Games and Economic Behavior 77 (2013) 377–404

Contents lists available at SciVerse ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Multi-task incentive contract and performance measurement with multidimensional types

Dawen Meng [a], Guoqiang Tian [b,*]

[a] Institute for Advanced Research, Shanghai University of Finance and Economics; Key Laboratory of Mathematical Economics (SHUFE), Ministry of Education, Shanghai 200433, China
[b] Department of Economics, Texas A&M University, College Station, TX 77843, USA

A R T I C L E I N F O

Article history:
Received 30 August 2011
Available online 2 November 2012

JEL classification:
D82
D86

Keywords:
Multi-task incentive model
Multidimensional mechanism design
Low-powered incentive
Missing incentive

A B S T R A C T

This paper provides a new explanation for the dominance of the low-powered incentive contract over the high-powered incentive contract using a hybrid model of moral hazard and adverse selection. We first show that unobservable risk aversion or cost leads to low-powered incentives. We then consider the case where both risk aversion and cost of the agent are unobservable to the principal. This multidimensional mechanism design problem is solved under two assumptions with regard to the structures of performance measurement system and wage contract. It is shown that if the deterministic and stochastic components of performance measures vary proportionally, the principal is inclined to provide a low-powered incentive contract. Moreover, it is shown that if the base wage depends on a quadratic function rather than the direction of the performance wage vector, no incentive is provided for most of the performance measures in an orthogonal performance measurement system.

1. Introduction

The central topic of moral hazard problem is to provide incentive contracts to motivate the agent's effort. Higher incentive pay will induce the agent to work harder and consequently bring higher surplus to the principal. However, the arrangements employers typically reach with their employees in reality look quite different from the incentive contracts derived by economic theorists. Low-powered incentives are very common in practice, especially within organizations. Many firms prefer to pay fixed compensation and offer continued employment to all but clearly unsatisfactory employees. Good examples are the government agencies and public firms, which are generally blamed for poor performance because their managers and workers lack high-powered incentives. Based on the standard transaction–cost and principal–agent economics, several theories have been provided to explain why low-powered incentives are employed even if objective performance measures are available and agents are highly responsive to incentive pay.

Williamson (1985) argues that weak incentive arises from opportunism and incompleteness of contracts. He shows that the use of high-powered incentives would raise undesirable side problems such as exploitation, inefficient asset utilization and accounting manipulations. For example, if supplying a single large customer would require a firm to make a large investment in an asset that cannot be used readily for other purposes, the supplier may reasonably fear exploitation by the customer: once the investment is made, the customer could force a lower price on the supplier. The problem is not simply that one party to the transaction may end up being treated unfairly. The bigger problem is that as people will anticipate this possibility, the transaction may not take place at all. Even if the manufacturer intends to keep his commitment, a transaction

* Corresponding author.
E-mail addresses: devinmeng@yahoo.com.cn (D. Meng), tian@econmail.tamu.edu (G. Tian).

0899-8256/$ – see front matter
http://dx.doi.org/10.1016/j.geb.2012.10.012

beneficial to both sides may be aborted because the supplier cannot trust him. One possible solution to this problem is to write a court enforceable contract specifying how each party must behave under a number of different contingencies. Unfortunately, as Williamson points out, contracts are not always effective in preventing opportunism in that due to limits to human information-processing abilities, it is often impossible to anticipate all possible contingencies, let alone specify them in a contract. This leaves scope for opportunism, so the supplier and manufacturer have to replace the high-powered market transaction with low-powered incentive inside firms.

Holmstrom and Milgrom (1991) show that the power of incentives on some tasks relies on the principal's ability to monitor other aspects of the agent's performance. The agents may shift their effort from some activities where their individual contributions are poorly measured to the better-measured and well-compensated activities. For this reason, high-powered incentive may be dysfunctional in multi-task environment.

The conclusion of Holmstrom and Milgrom (1991) relies largely on the assumptions that agent is risk averse and the tasks are substitute. On the contrary, Baker (1992) shows that low-powered incentive might arise even with risk neutral agent when the performance measure used and the principal's true objective are weakly correlated. That means if the performance measure does not respond to the agent's actions in the same way that the principal's objective responds to these actions, the firm will reduce the intensity of the incentive contract.

Aside from piece rates or commissions, another way that firms use to compensate agent is by relative performance evaluation such as awarding promotions to a member of a work group who performs best. In fact, in some political organizations such as government agencies, the agents are rewarded mainly on relative performance measures rather than on their individual output. One function of relative performance evaluation is allowing the principal to use flatter incentives.[1] In this sense, the literature justifying relative performance evaluations also gives partial explanations on the arising of low-powered incentive.

Lazear and Rosen (1981) show in a standard single moral hazard framework that the promotion-based incentive scheme can achieve the same results as other incentive schemes can. They argue that the dominance of the promotion-based incentive scheme over the piece-rate linear scheme and the standard bonus scheme arises from the fact that obtaining ordinal measures generally requires less resources than obtaining cardinal measures. Green and Stokey (1983) and Nalebuff and Stiglitz (1983) show also in the single moral hazard framework that the relative performance evaluation incentive scheme may dominate the absolute performance evaluation scheme when the agents are risk averse and there are shocks that are common to all the agents. Obviously, the promotion-based incentive scheme, by filtering out common randomness, can reduce the risk that would otherwise be imposed on the agents and requires compensation. Therefore, the relative performance evaluation improves the principal's efficiency.

Another class of literature closely related to the present paper is in the area of multidimensional mechanism design. The multidimensional mechanism design problem arises when the agent possesses multiple characteristics. Its implementability is much more complicated than that in the unidimensional mechanism design problem because of the lack of a natural order on types.

The studies on this topic can be traced back to Laffont et al. (1987), Rochet (1987), and Wilson (1993), etc.[2] The most notable publications in this field include Armstrong (1996), Rochet and Choné (1998) and Basov (2001) among many others. Armstrong (1996) formulates this problem in a multiproduct nonlinear pricing setting. In this seminal paper, he develops an integration along rays technique and characterizes the pricing contract for the case with cost-based tariff. Rochet and Choné (1998) analyze a general multidimensional screening model. They show that, in general, the monopolist will use mechanisms in which there is bunching, i.e., different consumer-types will be treated equally. They develop a methodology – sweeping technique, for dealing with bunching in multiple dimensions. Basov (2001) takes advantage of control theoretic tools and develops a "generalized Hamiltonian approach" for solving the multidimensional mechanism design problem.

In this paper, we provide a new explanation for the dominance of low-powered over high-powered incentives. Our contributions are two-fold. First, in contrast to most of the existing literature dealing with the power-of-incentive issue in the framework of pure moral hazard, our analysis is made in a hybrid model of both moral hazard and adverse selection.[3] The standard moral hazard model concerns only the trade-off between insurance and incentives. In these environments, the compensation based on certain "risky" performance measure serves the dual functions of increasing both profits and risk. A tension between these two functions arises when the agent is risk averse. Higher pay induces the agent to exert a higher

[1] Rank order tournaments is a simple and widely used form of relative performance evaluation. This classical form of relative performance compensation has the particularity of using only an ordinal ranking of performance. By awarding high and low prizes based on relative performance, a principal can elicit a higher effort level than with a scheme that involves the same wage bill but equal wages. An agent's performance is increasing in the spread between the winner and loser prize, ceteris paribus, rather than the absolute payment levels. It therefore allows the principal to elicit a higher effort level using lower performance pay.

[2] We thank one of the anonymous referees for reminding us of these papers.

[3] Laffont and Tirole (1986), Picard (1987), Guesnerie et al. (1989), Melumad and Reichelstein (1989), Caillaud et al. (1992), Basov and Bardsley (2005), among many others, discuss the welfare losses issue in the joint presence of moral hazard and adverse selection. Their main conclusion is that under risk neutrality, if the noise in production technology is independent of the agent's type, and the effort is unidimensional, hidden action does not create any additional welfare loss, comparatively to the models with observable actions. Basov and Danilkina (2010) show that hidden action will generically lead to a welfare loss even if both parties are risk neutral and production technology is independent of the agent's type, provided that both effort and private information are multidimensional. Our model, however, is trying to answer a different question: with risk-averse agent, whether hidden information leads to a flatter wage contract comparatively to the model with observable information? We are thankful to a referee for bringing our attention to these papers.

level of effort and thus increases the principal's profit. On the other hand, high wage also exposes the agent to unwanted risk, which requires an extra risk premium as compensation. Consequently, when choosing contract, the principal trades off the benefits of more effort against higher wage costs. Most of the existing studies assume that only the agent's actions are unobservable. In contrast, our paper assumes that the agent is privately informed about both his actions and types. The principal therefore is faced with an additional tradeoff – the tradeoff between efficiency and rent extraction. We show that the incentive contract is inclined to be flatter in the hybrid model than in the pure moral hazard model. Furthermore, it is worth noting that if the agent owns multidimensional private information and the base wage depends only on the Σ-norm of the performance wage, the efficient linear compensation rule contains no incentive component for most of the performance measures.

Secondly, we develop a "delegating" method for the complex multidimensional mechanism design problem. The intuition behind this method is a tradeoff between authority and complexity. As a centralized way of resource allocation, an incentive mechanism vests all the decision-making authority in the principal, but it needs to process information transmitted by the agent. The multidimensional information increases the principal's information processing cost and complexity of writing a contract. The more authority the principal owns, the more information he has to process. Therefore, in order to save information processing cost and avoid complexity of writing a multidimensional contract, the principal may choose to delegate part of his authority to the agent. Two extreme cases of this tradeoff in practice are a decentralized market economy which distributes all the decision-making authority to individual agents so that their communications requirements can be minimal (see Hurwicz, 1972, 1979, 1986; Mount and Reiter, 1974; Walker, 1977; Osana, 1978; Tian, 1994, 2004, 2006, among others for detailed discussion); and socialist economy in which a central planner has almost all the authority but a great amount of information has to be processed. In this paper, part of the principal's authority is delegated to the agent under the assumption that the fixed component of compensation bases only on a quadratic form (Σ-norm) of the vector of incentive compensation coefficients. This assumption deprives some of the principal's degrees of freedom but decreases drastically the amount of information required. The multidimensional mechanism design problem is therefore relatively easily solved.

The remainder of the paper is organized as follows. The basic multi-task principal–agent model is specified in Section 2, along with a characterization of the pure moral hazard contract. The results with unobservable risk aversion are examined in Section 3. The results with unobservable cost are discussed in Section 4. Section 5 considers the optimal incentive contract in a general environment where risk aversion and cost are both unobservable. Finally, in Section 6, some concluding remarks are given.

2. Basic model

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:

$$V(e) = \beta' e + \eta, \tag{1}$$

where the n-dimensional vector β characterizes the marginal effect of the agent's effort e on $V(e)$, and η is a noise term with zero mean. The agent chooses a vector of efforts $e = (e_1, \dots, e_n)' \in \mathbb{R}^n_+$ at quadratic personal cost $\frac{e'Ce}{2}$, 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:

$$P_i(e) = b_i' e + \varepsilon_i, \quad i = 1, \dots, m, \tag{2}$$

where $b_i \in \mathbb{R}^n$ captures the marginal effect of the agent's effort e on the performance measure $P_i(e)$; $B = (b_1, \dots, b_m)'$ 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 $\epsilon = (\varepsilon_1, \dots, \varepsilon_m)'$ is an $m \times 1$ vector of normally distributed variables with mean zero and variance–covariance matrix Σ.

Definition 1 *(Orthogonality).* A performance system is said to be orthogonal if and only if $b_i' C^{-1} b_j = 0$ and $Cov(\varepsilon_i, \varepsilon_j) = 0$, for $i \neq j$, that is, $B'C^{-1}B$ and Σ are both diagonal matrices.

Definition 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:

$$\rho^c_{ij} = \frac{b_i' C^{-1} b_j}{\sigma_{ij}}. \tag{3}$$

In single-task agency relationships, the signal-to-noise ratio of performance measure $P_i(e) = \omega_i e + \varepsilon_i$ is given by $\omega_i^2/var(\varepsilon_i)$ (see Kim and Suh, 1991). Schnedler (2008) generalizes the signal-to-noise ratio to account for multidimensional effort. In his definition, the signal-to-noise ratio of performance measure $P_i(e) = b_i'e + \sigma_i$, denoted by γ_i, is

$$\gamma_i \equiv \frac{(\nabla P_i(e))'(\nabla P_i(e))}{var(\varepsilon_i)} = \frac{b_i'b_i}{\sigma_i^2}.$$

Performance measures with higher signal-to-noise ratios provide more precise information about the agent's effort choice than those with lower signal-to-noise ratios. Clearly, the signal-to-noise ratio does not account for the agent's task-specific abilities and interaction among tasks. Furthermore, there is no concept measuring to what degree two performance measures are aligned with each other in the existing literature. The definition of *cost-adjusted correlation* given above, however, captures these ideas. 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, and it is clear that $\rho_{ii} = \gamma_i$.

Definition 3 *(Cost-adjusted congruence).* The cost-adjusted congruence of a performance measure $P_i = b_i'e + \varepsilon_i$ is defined as

$$\Gamma_i = \frac{b_i'C^{-1}\beta}{\sqrt{b_i'C^{-1}b_i}\sqrt{\beta'C^{-1}\beta}}. \tag{4}$$

Baker (2002) defines the congruence of a performance measure as cosine of the angle between the vector of payoff sensitivities and the vector of performance measure sensitivities: $\Upsilon_i \equiv \cos(\widehat{b_i, \beta})$. Also, his definition does not consider the agent's task-specific abilities and interaction among tasks. So we adopt a modified measure of congruity given in (4). If C is a scalar matrix, then $\Gamma_i = \Upsilon_i$. In this paper, a performance measure with nonzero cost-adjusted congruence is said to be *congruent*; a performance measure with unit cost-adjusted congruence is said to be *perfectly congruent*. We assume in this paper 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:

$$W(e) = w_0 + w'P(e), \tag{5}$$

where $P(e) = (P_1(e), \ldots, P_m(e))'$, w_0 denotes the base wage, and $w = (w_1, \ldots, w_m)'$ 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

$$CE_a = w_0 + w'Be - \frac{1}{2}e'Ce - \frac{r}{2}w'\Sigma w. \tag{6}$$

The principal's problem is to design a contract (w_0, w) that maximizes her expected profit Π_p while ensuring the agent's participation and eliciting his optimal effort.

The optimization problem of the principal is thus formulated as:

$$\begin{cases} \max\limits_{\{w_0,\, w,\, e\}} \ \beta'e - w_0 - w'Be \\ \quad \text{s.t:} \quad IR\colon\ w_0 + w'Be - \frac{1}{2}e'Ce - \frac{r}{2}w'\Sigma w \geqslant 0 \\ \qquad\qquad IC\colon\ e \in \arg\max\limits_{\tilde{e}} \left[w_0 + w'B\tilde{e} - \frac{1}{2}\tilde{e}'C\tilde{e} - \frac{r}{2}w'\Sigma w \right]. \end{cases}$$

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:

$$\max_{w \in \mathbb{R}^n} \left[\beta'C^{-1}B'w - \frac{1}{2}w'\left(BC^{-1}B' + r\Sigma\right)w \right].$$

The optimal wage contract and effort to be elicited are therefore:

$$w^p = \left[BC^{-1}B' + r\Sigma\right]^{-1} BC^{-1}\beta, \tag{7}$$

$$w_0^p = \frac{rw^{p'}\Sigma w^p - w^{p'}BC^{-1}B'w^p}{2}, \tag{8}$$

$$e^p = C^{-1}B'w^p. \tag{9}$$

The resulting surplus of the principal is[4]

$$\Pi^p = \frac{1}{2}\beta' C^{-1} B' \left[B C^{-1} B' + r\Sigma \right]^{-1} B C^{-1} \beta. \tag{10}$$

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 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, which is in line with many previously known studies such as those of Feltham and Xie (1994), Baker (2002) and Thiele (2010).

3. The optimal contract with unobservable 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 private information of the agent, and its distribution function $F(r)$ and density function $f(r)$ supported on $[\underline{r}, \bar{r}]$ are common knowledge to all parties. This assumption is different to most of the previous studies in which risk aversion is regarded as a publicly observed variable. 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.

The timing of this hybrid model is as follows. At date 0, nature determines r, only the agent knows it. At date 1 the principal offers the agent an employment contract $(w_0(\hat{r}), w(\hat{r}))$ based upon the agent's report $\hat{r}$. If this contract guarantees at least the same expected utility as his reservation wage (which is normalized to zero), the agent accepts the contract and reports $\hat{r}$. At date 2, after accepting his contract and reporting his type $\hat{r}$, the agent implements an effort vector e. At date 3, the agent's performance measure $P(e)$ as well as his contribution to firm value, $V(e)$, are realized. At date 4, the payment is made.

A contract $\{w_0(\hat{r}), w(\hat{r})\}$ is said to be implementable if the following incentive compatibility condition is satisfied:

$$w_0(r) + \frac{1}{2} w(r)' \left[B C^{-1} B' - r\Sigma \right] w(r) \geqslant w_0(\hat{r}) + \frac{1}{2} w(\hat{r})' \left[B C^{-1} B' - r\Sigma \right] w(\hat{r}). \tag{11}$$

Let $U(r, \hat{r}) \equiv w_0(\hat{r}) + \frac{1}{2} w(\hat{r})' [B C^{-1} B' - r\Sigma] w(\hat{r})$,[5] and $U(r) \equiv U(r, r)$, then the implementability condition of $\{U(r), w(r)\}$ is stated equivalently as:

$$\exists w_0 : [\underline{r}, \bar{r}] \to \mathbb{R}_+,\ \forall (r, \hat{r}) \in [\underline{r}, \bar{r}]^2, \quad U(r) = \max_{\hat{r}} \left\{ w_0(\hat{r}) + \frac{1}{2} w(\hat{r})' \left[B C^{-1} B' - r\Sigma \right] w(\hat{r}) \right\}. \tag{12}$$

The "Taxation Principle" (cf. Guesnerie, 1981; Hammond, 1979 and also Rochet, 1985) states that (12) is equivalent to the following very similar condition

$$\exists w_0 : \mathbb{R}^m \to \mathbb{R}_+,\ \forall r \in [\underline{r}, \bar{r}], \quad U(r) = \max_{w} \left\{ w_0(w) + \frac{1}{2} w' \left[B C^{-1} B' - r\Sigma \right] w \right\}. \tag{13}$$

It is possible to show that $U(\cdot)$ is continuous, convex[6] (thus almost everywhere differentiable), and satisfies the envelop condition:

$$U'(r) = -\frac{1}{2} w' \Sigma w. \tag{14}$$

Conversely, if (14) holds and $U(r)$ is convex, then

$$U(r) \geqslant U(\hat{r}) + (r - \hat{r}) U'(\hat{r}) = U(\hat{r}) - \frac{1}{2}(r - \hat{r}) w'(\hat{r}) \Sigma w(\hat{r}),$$

which implies the incentive compatibility condition $U(r) \geqslant U(r, \hat{r})$. Formally, we have

Lemma 1. *The surplus function $U(r)$ and performance wage function $w(r)$ are implementable if and only if:*

(1) *envelop condition (14) is satisfied;*
(2) *$U(r)$ is convex in r.*

---

4 Superscript "p" denotes "pure moral hazard".

5 Substituting $e^* = C^{-1} B' w$ into expression (6) yields $U = w_0 + \frac{1}{2} w' (B C^{-1} B' - r\Sigma) w$.

6 One way to define the convex functions is through representing them as maximum of the 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 $a = -\frac{1}{2} w' \Sigma w$, $b = w_0(w) + \frac{1}{2} w' B C^{-1} B' w$, and thus $U(r) = \max_{(a,b) \in \mathbb{R}_- \times \mathbb{R}_+} (ar + b)$ is a convex function in r.

Substituting $U(r)$ into the principal's expected payoff, we get

$$\begin{aligned}\Pi &= \int_{\underline{r}}^{\bar{r}} \left[\beta' e^* - w_0(r) - w(r)' B e^*\right] f(r)\, dr \\ &= \int_{\underline{r}}^{\bar{r}} \left\{\beta' C^{-1} B' w(r) - \frac{1}{2} w(r)' \left[B C^{-1} B' + r\Sigma\right] w(r) - U(r)\right\} f(r)\, dr.\end{aligned}$$

The principal's optimization problem is therefore:

$$\max_{U(r),\, w(r)} \Pi, \quad \text{s.t.:} \quad U(r) \geqslant 0, \quad U'(r) = -\frac{1}{2} w(r)' \Sigma w(r), \quad U(r) \text{ is convex.} \tag{15}$$

The following proposition summarizes the solution of the principal's problem:

Proposition 1. *If $\Phi(r)$ is nondecreasing,*[7] *then the optimal wage contract is given by*

$$w^h(r) = \left[B C^{-1} B' + \Phi(r)\Sigma\right]^{-1} B C^{-1} \beta, \tag{16}$$

$$w_0^h(r) = \frac{1}{2} \int_r^{\bar{r}} w^h(\tilde{r})' \Sigma w^h(\tilde{r})\, d\tilde{r} - \frac{1}{2} w^h(r)' \left[B C^{-1} B' - r\Sigma\right] w^h(r), \tag{17}$$

where $\Phi(r) \equiv r + \frac{F(r)}{f(r)}$.[8]

Proof. See Appendix A. □

The following conditions prove to be sufficient for the emergence of low-powered incentives:

Condition 3.1. Σ is diagonal.

Condition 3.2. Matrix $B C^{-1} B'$ is diagonal.

Condition 3.3. Matrices $B C^{-1} B'$ and Σ commute: $B C^{-1} B' \Sigma = \Sigma B C^{-1} B'$.[9]

Condition 3.4. The following inequality holds:

$$2r\lambda_m^2 + \rho > 0, \tag{18}$$

where

$$\begin{aligned}\rho &= \max\left\{\min_{i=1,m} \lambda_i \mu_m \frac{(\sqrt{k_\lambda}+1)^2 - k_\mu(\sqrt{k_\lambda}-1)^2}{2\sqrt{k_\lambda}},\ \min_{i=1,m} \mu_i \lambda_m \frac{(\sqrt{k_\mu}+1)^2 - k_\lambda(\sqrt{k_\mu}-1)^2}{2\sqrt{k_\mu}}\right\} \\ &= \begin{cases} \lambda_m \mu_m \frac{(\sqrt{k_\lambda}+1)^2 - k_\mu(\sqrt{k_\lambda}-1)^2}{2\sqrt{k_\lambda}} & \text{if } \sqrt{k_\mu} \leqslant \frac{\sqrt{k_\lambda}+1}{\sqrt{k_\lambda}-1},\ k_\mu \geqslant k_\lambda, \\ \lambda_m \mu_m \frac{(\sqrt{k_\mu}+1)^2 - k_\lambda(\sqrt{k_\mu}-1)^2}{2\sqrt{k_\mu}} & \text{if } \sqrt{k_\mu} \leqslant \frac{\sqrt{k_\lambda}+1}{\sqrt{k_\lambda}-1},\ k_\mu < k_\lambda, \\ \lambda_1 \mu_m \frac{(\sqrt{k_\lambda}+1)^2 - k_\mu(\sqrt{k_\lambda}-1)^2}{2\sqrt{k_\lambda}} & \text{if } \sqrt{k_\mu} > \frac{\sqrt{k_\lambda}+1}{\sqrt{k_\lambda}-1},\ k_\mu \geqslant k_\lambda, \\ \lambda_m \mu_1 \frac{(\sqrt{k_\mu}+1)^2 - k_\lambda(\sqrt{k_\mu}-1)^2}{2\sqrt{k_\mu}} & \text{if } \sqrt{k_\mu} > \frac{\sqrt{k_\lambda}+1}{\sqrt{k_\lambda}-1},\ k_\lambda < k_\lambda. \end{cases}\end{aligned}$$

λ_i, μ_i are the i-th eigenvalues of Σ, $B C^{-1} B'$ respectively in a descending enumeration. $k_\lambda = \frac{\lambda_1}{\lambda_m}$ and $k_\mu = \frac{\mu_1}{\mu_m}$ denote the spectral condition number of Σ and $B C^{-1} B'$ respectively.

Condition 3.5. There exists a positive number λ such that $B C^{-1} B' = \lambda \Sigma$.

[7] This condition is weaker than and could be implied by the monotone hazard rate property: $\frac{d}{dr}[\frac{F(r)}{f(r)}] \geqslant 0$.

[8] Superscript "h" denotes "hybrid model of moral hazard and adverse selection".

[9] That is, $B C^{-1} B' \Sigma$ is symmetric.

Condition 3.1 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 3.2 states that $b_i' C^{-1} b_j = 0$ 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 3.4 holds when agent is sufficiently risk averse or when either matrix $BC^{-1}B'$ or Σ is well-conditioned.[10] Several important special cases are:

- the performance measures system is orthogonal. In this case Conditions 3.1 to 3.3 are all satisfied;
- Σ is a scalar matrix, in which case Conditions 3.1, 3.3 and 3.4 are satisfied;
- $BC^{-1}B'$ is a scalar matrix, in which case Conditions 3.2, 3.3 and 3.4 are satisfied.

Condition 3.5 emphasizes that the covariance matrix Σ 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 1.

1. *Given any one of Conditions 3.1 to 3.4, there exists an $i \in \{1 \cdots m\}$, such that $|w_i^h(r)| < |w_i^p(r)|$ for all $r \in (\underline{r}, \bar{r}]$.*
2. *If both Condition 3.1 and Condition 3.2 are satisfied, then $|w_i^h(r)| < |w_i^p(r)|$ for all $r \in (\underline{r}, \bar{r}]$ and all $i \in \{1 \cdots m\}$.*
3. *Let ω_i, $i \in \mathcal{K} \equiv \{1, 2, \ldots, k\}$ denote k distinct generalized eigenvalues of $BC^{-1}B'$ relative to Σ, $\mathcal{V}_i \equiv \mathcal{N}(BC^{-1}B' - \omega_i \Sigma)$ be the eigenspace corresponding to ω_i, and $\mathcal{V}_i^{\perp}$ be its orthogonal complement. Suppose that $BC^{-1}\beta \notin \bigcup_{i \in \mathcal{K}} \mathcal{V}_i^{\perp}$, then there exists a positive number $k \in (0, 1)$ such that $w^h = k w^p$ if and only if Condition 3.5 is met.*

Proof. See Appendix A. □

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 3.1 to 3.4, wage vector w is shortened in different quadratic-form norms compared with the pure moral hazard case. Under Condition 3.5, 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 $rw'\Sigma w$. Consequently, the efficiency-rent tradeoff alters only the overall intensity of wage vector, not its relative allocation among performance measures. The opposite of this result is true under the premise $BC^{-1}\beta \notin \bigcup_{i \in \mathcal{K}} \mathcal{V}_i^{\perp}$. To see this, consider the following example with diagonal matrices $BC^{-1}B'$ and Σ:

$$BC^{-1}B' = \begin{bmatrix} b_1' C^{-1} b_1 & & & \\ & b_2' C^{-1} b_2 & & \\ & & \ddots & \\ & & & b_m' C^{-1} b_m \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_m^2 \end{bmatrix}.$$

The i-th generalized eigenvalue of $BC^{-1}B'$ relative to Σ is $\omega_i = \rho_{ii} \equiv \frac{b_i' C^{-1} b_i}{\sigma_i^2}$, the associated normalized generalized eigenvector e_i has its i-th entry one and every other entry zero. $\mathcal{V}_i^{\perp} = \{\mathbf{v} = (v_1, v_2, \ldots, v_m) \in \mathbb{R}^m \mid v_i = 0\}$. $BC^{-1}\beta \notin \bigcup_{i \in \mathcal{K}} \mathcal{V}_i^{\perp}$ requires that $b_i' C^{-1} \beta \neq 0$ for all $i \in \{1, 2, \ldots, m\}$. That means any performance measure is congruent, i.e., $\Gamma_i \neq 0$ (though not necessarily be perfectly congruent). It follows from $w^h = k w^p$ that

$$\frac{b_i' C^{-1} \beta}{b_i' C^{-1} b_i + r\sigma_i^2} = k \frac{b_i' C^{-1} \beta}{b_i' C^{-1} b_i + \Phi(r)\sigma_i^2}, \quad \forall i.$$

Consequently, we get $BC^{-1}B' = \frac{\Phi(r) - kr}{k-1} \Sigma$. In summary, for an orthogonal system with congruent measures, wage vectors w^p and w^h have the same direction if and only if all the performance measures share the same cost-adjusted signal-to-noise ratio.

It may be remarked that adverse selection is followed by moral hazard in our hybrid model. As such, it is reasonable for us to choose pure moral hazard situation as the benchmark and a starting point, and thus it enables us to analyze this model via backward induction. Also, note that the order of analysis matters. If we start from a pure adverse selection model, we may not get a result similar to Theorem 1. To see this, we make the following simple calculation. The principal's optimization problem under pure adverse selection (e is observable, r is unobservable) is

[10] Matrices with condition numbers near 1 are said to be well-conditioned, while matrices with high condition numbers are said to be ill-conditioned.

$$\max_{e,\, w_0(r),\, w(r)} \int_{\underline{r}}^{\bar{r}} \big(\beta' e - w_0 - w' B e\big) f(r)\, dr$$

$$\text{s.t.:}\quad IR:\ w_0(r) + w'(r)Be - \frac{1}{2}e'Ce - \frac{r}{2}w'(r)\Sigma w(r) \geqslant 0,$$

$$IC:\ r = \arg\max_{\tilde{r}} \left[w_0(\tilde{r}) + w'(\tilde{r})Be - \frac{1}{2}e'Ce - \frac{r}{2}w'(\tilde{r})\Sigma w(\tilde{r}) \right].$$

This problem could be simplified to

$$\max_{e,\, w(r)} \int_{\underline{r}}^{\bar{r}} \left[\beta' e - \frac{1}{2}e'Ce - \frac{1}{2}\left(r + \frac{F(r)}{f(r)}\right) w'(r)\Sigma w(r) \right] f(r)\, dr.$$

The optimal wage contract under pure adverse selection entails: $e^* = C^{-1}\beta$, $w^*(r) = 0$. Comparing it with (16) yields $|w_i^*| < |w_i^h|$ for all i. Therefore, the power of incentives in hybrid model is not lower than that in pure adverse selection model.[11]

4. The optimal contract with unobservable cost

In this section we assume that the cost parameter is private information of the agent. To avoid the complicated multidimensional mechanism design issue momentarily, 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 $[\underline{\delta}, \bar{\delta}]$, according to a cumulative distribution function $G(\delta)$ and density $g(\delta)$. The timeline of this problem is analogous to that in Section 3 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:

$$w_0(\delta) + \frac{1}{2}w(\delta)'\big[\delta BB' - r\Sigma\big]w(\delta) \geqslant w_0(\hat{\delta}) + \frac{1}{2}w(\hat{\delta})'\big[\delta BB' - r\Sigma\big]w(\hat{\delta}), \quad \forall (\delta, \hat{\delta}) \in [\underline{\delta}, \bar{\delta}]^2. \tag{19}$$

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

$$\exists w_0 : [\underline{\delta}, \bar{\delta}] \to \mathbb{R}_+,\ \forall (\delta, \hat{\delta}) \in [\underline{\delta}, \bar{\delta}]^2, \quad U(\delta) = \max_{\hat{\delta}} \left\{ w_0(\hat{\delta}) + \frac{1}{2}w(\hat{\delta})'\big[\delta BB' - r\Sigma\big]w(\hat{\delta}) \right\} \tag{20}$$

or equivalently,

$$\exists w_0 : \mathbb{R} \to \mathbb{R}_+,\ \forall \delta \in [\underline{\delta}, \bar{\delta}], \quad U(\delta) = \max_{w \in \mathbb{R}^m} \left\{ w_0(w) + \frac{1}{2}w'\big[\delta BB' - r\Sigma\big]w \right\}. \tag{21}$$

$U(\delta)$ is necessarily continuous, increasing and convex in δ[12] and satisfies the envelop condition:

$$U'(\delta) = \frac{1}{2}w'BB'w. \tag{22}$$

Conversely, similar to the case with unobservable risk aversion, the convexity of $U(\delta)$ and envelop condition (22) implies

$$U(\delta) \geqslant U(\hat{\delta}) + (\delta - \hat{\delta})U'(\hat{\delta}) = U(\hat{\delta}) + \frac{1}{2}(\delta - \hat{\delta})w'BB'w = U(\delta, \hat{\delta}),$$

which in turn implies the implementability of contract. We summarize the above discussion in the following lemma:

Lemma 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 δ.*

The second-best δ-contingent contract solves the following optimization problem:

$$\begin{cases} \displaystyle\max_{w(\delta),\, U(\delta)} \int_{\underline{\delta}}^{\bar{\delta}} \left\{ \delta\beta' B' w(\delta) - \frac{1}{2}w(\delta)'\big[\delta BB' + r\Sigma\big]w(\delta) - U(\delta) \right\} g(\delta)\, d\delta \\[2ex] \text{s.t.:}\quad U(\delta) \geqslant 0, \quad U'(\delta) = \dfrac{w'BB'w}{2}, \quad U(\delta) \text{ is convex.} \end{cases}$$

---

[11] We are thankful to a referee for bringing this point to our attention.

[12] 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 δ.

Proposition 2. *With unobservable cost, if $\delta H(\delta)$ is nonincreasing,*[13] *then the optimal wage is given by*

$$w^h(\delta)=\left(H(\delta)BB'+\frac{r\Sigma}{\delta}\right)^{-1}B\beta, \tag{23}$$

$$w_0^h(\delta)=\frac{1}{2}\int\limits_{\underline{\delta}}^{\delta}w^h(\tilde{\delta})'BB'w^h(\tilde{\delta})\,d\tilde{\delta}-\frac{1}{2}w^h(\delta)'\big[\delta BB'-r\Sigma\big]w^h(\delta), \tag{24}$$

where $H(\delta)\equiv 1+\frac{1-G(\delta)}{\delta g(\delta)}$.

Proof. See Appendix A. □

The following conditions justify the adoption of low-powered incentives in the case with unobservable cost parameter:

Condition 4.1. BB' is a diagonal matrix.

Condition 4.2. Matrices BB' and Σ commute: $BB'\Sigma=\Sigma BB'$.[14]

Condition 4.3. The following inequality holds:

$$2\nu_m^2+\frac{r}{\delta}\eta>0. \tag{25}$$

$$\eta=\max\left\{\min_{i=1,m}\lambda_i\nu_m\frac{(\sqrt{k_\lambda}+1)^2-k_\nu(\sqrt{k_\lambda}-1)^2}{2\sqrt{k_\lambda}},\ \min_{i=1,m}\nu_i\lambda_m\frac{(\sqrt{k_\nu}+1)^2-k_\lambda(\sqrt{k_\nu}-1)^2}{2\sqrt{k_\nu}}\right\}$$

$$=\begin{cases}\lambda_m\nu_m\frac{(\sqrt{k_\lambda}+1)^2-k_\nu(\sqrt{k_\lambda}-1)^2}{2\sqrt{k_\lambda}} & \text{if } \sqrt{k_\nu}\leqslant\frac{\sqrt{k_\lambda}+1}{\sqrt{k_\lambda}-1},\ k_\nu\geqslant k_\lambda,\\ \lambda_m\nu_m\frac{(\sqrt{k_\nu}+1)^2-k_\lambda(\sqrt{k_\nu}-1)^2}{2\sqrt{k_\nu}} & \text{if } \sqrt{k_\nu}\leqslant\frac{\sqrt{k_\lambda}+1}{\sqrt{k_\lambda}-1},\ k_\nu<k_\lambda,\\ \lambda_1\nu_m\frac{(\sqrt{k_\lambda}+1)^2-k_\nu(\sqrt{k_\lambda}-1)^2}{2\sqrt{k_\lambda}} & \text{if } \sqrt{k_\nu}>\frac{\sqrt{k_\lambda}+1}{\sqrt{k_\lambda}-1},\ k_\nu\geqslant k_\lambda,\\ \lambda_m\nu_1\frac{(\sqrt{k_\nu}+1)^2-k_\lambda(\sqrt{k_\nu}-1)^2}{2\sqrt{k_\nu}} & \text{if } \sqrt{k_\nu}>\frac{\sqrt{k_\lambda}+1}{\sqrt{k_\lambda}-1},\ k_\lambda<k_\lambda,\end{cases}$$

represents the lower bound of eigenvalues of Jordan product $BB'\Sigma+\Sigma BB'$; λ_i, ν_i are the i-th eigenvalues of Σ and BB' respectively in a descending enumeration. $k_\lambda=\frac{\lambda_1}{\lambda_m}$ and $k_\nu=\frac{\nu_1}{\nu_m}$ denote the spectral condition number of Σ and BB' respectively.

Condition 4.4. There exists a positive number k such that $BB'=k\Sigma$.

In the special case where performance measures system is orthogonal, Conditions 3.1, 4.1 and 4.2 are satisfied. If Σ (resp. BB') is a scalar matrix, then Conditions 3.1 (resp. 4.1) and 4.3 are both satisfied. Besides, Condition 4.3 could hold even for nondiagonal BB' and Σ, provided either of them is well-conditioned or $\frac{r}{\delta}$ is sufficiently small.

Theorem 2.

1. *Given any of Conditions 3.1, 4.1, 4.2, 4.3, there exists at least one $i\in\{1,\dots,m\}$, such that $|w_i^h(\delta)|<|w_i^p(\delta)|$ for all $\delta\in[\underline{\delta},\bar{\delta})$;*
2. *If both Conditions 3.1 and 4.1 are satisfied, namely, the performance measure system is orthogonal, then $|w_i^h(\delta)|<|w_i^p(\delta)|$ for all $\delta\in[\underline{\delta},\bar{\delta})$ and all i;*
3. *Let τ_i, $i\in\mathcal{L}\equiv\{1,2,\dots,l\}$ denote l distinct generalized eigenvalues of BB' relative to Σ, $\mathcal{U}_i\equiv\mathcal{N}(BB'-\tau_i\Sigma)$ be the eigenspace corresponding to τ_i, $\mathcal{U}_i^{\perp}$ be its orthogonal complement. Suppose that $B\beta\notin\bigcup_{i\in\mathcal{L}}\mathcal{U}_i^{\perp}$, then there exists a positive number $s\in(0,1)$ such that $w^h=sw^p$ if and only if Condition 4.4 is met.*

Proof. The proof of this theorem is similar to that in Theorem 1 and it is omitted here. □

When the agent possesses private information on his own cost, a more efficient agent (the agent with higher δ) 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 2 gives various conditions ensuring low-powered

[13] This assumption is a bit stronger than the usual monotone inverse hazard rate condition. It holds for any nondecreasing $g(\cdot)$.

[14] Again, it is true if $BB'\Sigma$ is symmetric.

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 Σ is well-conditioned (k_ν or k_λ is close to one), or the agent is nearly risk neutral (r is very small), or the agent is highly efficient (δ 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).

5. The optimal contract with both unobservable cost and risk aversion

In this section we assume that both efficiency parameter δ and risk aversion r are unobservable to the principal, their joint PDF is $f(\delta, r)$ supported on a convex region $\mathcal{D}$. Timing of the model is analogous to that in Section 3 and Section 4 except that the agent is now required to report both $\hat{r}$ and $\hat{\delta}$. The multidimensional mechanism design model differs markedly from and is significantly more complex than its one-dimensional counterpart, essentially because different types of agents cannot be unambiguously ordered. For lack of methodology in the most general sense, various assumptions and methods have been used to solve the multidimensional mechanism design models in the existing literature. In a nonlinear pricing setup, Armstrong (1996) adopts an integration along rays procedure solving the relaxed problem of the principal, but the envelop condition could be satisfied by the pointwise maximizer only by accident, let alone the convexity condition. In order for the contract to be implementable, he makes separable assumptions on the indirect utility and density functions. Rochet and Choné (1998) develop a general technique for dealing with the multidimensional screening problem, but it is workable only in the case where the dimensionality of type space is the same as the number of the principal's available instruments. The generalized Hamiltonian approach developed by Basov (2001) circumvents this difficulty but it obtains the optimal contract from a system of partial differential equations, which usually has no analytic solution. Therefore, one often has to rely upon the numerical techniques except for some very special function form.

In the following, we will treat the choosing of optimal performance wage as a multidimensional mechanism design problem. In order to get an explicit analytic solution, we impose restrictions on the set of implementable allocations by assuming that the performance evaluation system is such that $BB' = k\Sigma$ or the base wage is based on the Σ-norm of performance wage vector.

5.1. The performance measurement system with constant correlation

If correlation between any pair of performance measures is constant, i.e., $b_i'b_j/\sigma_{ij} \equiv k$, $\forall i, j$, then $BB' = k\Sigma$. This means the deterministic part of a performance measurement system covaries with its stochastic counterpart. With this assumption, the agent's surplus could be represented as a function of a scalar $\theta_1 \equiv k\delta - r$,

$$U(\delta, r) = \max_w \left[w_0(w) + \frac{1}{2}\theta_1 w'\Sigma w \right] \equiv u(\theta_1). \tag{26}$$

Then, as in the previous sections, we get the convexity and envelop conditions ($u(\theta_1)$ is convex in θ_1 and $u'(\theta_1) = \frac{1}{2}w'\Sigma w$), which in turn implies the implementability of contract. We also define $\theta_2 \equiv k\delta + r$. Then the original type vector (δ, r) is transformed linearly to (θ_1, θ_2). Notice that θ_1 is the only variable affecting the agent's choice, and the principal may elicit it using an incentive compatible contract when it is unknown. θ_2 has no informative value to the agent since it is irrelevant to the agent's decision-making. As such, the principal lacks instrument to elicit it. That means the principal cannot trust the information about θ_2 reported by the agent even if he is telling the truth. Therefore, the wage contract may depend on θ_1 whether it is observable or not. However, it may depend on θ_2 only when it is observable.

Let Θ denote the support of the transformed types (θ_1, θ_2), which is defined by $\underline{\theta}_1 \leqslant \theta_1 \leqslant \bar{\theta}_1$, $\underline{\theta}_2(\theta_1) \leqslant \theta_2 \leqslant \bar{\theta}_2(\theta_1)$ (or $\underline{\theta}_2 \leqslant \theta_2 \leqslant \bar{\theta}_2$, $\underline{\theta}_1(\theta_2) \leqslant \theta_1 \leqslant \bar{\theta}_1(\theta_2)$). Let $\varphi(\theta_1, \theta_2) = f(\frac{\theta_1+\theta_2}{2k}, \frac{\theta_2-\theta_1}{2})|\mathbf{J}(\theta_1, \theta_2)| = f(\frac{\theta_1+\theta_2}{2k}, \frac{\theta_2-\theta_1}{2})\frac{1}{2k}$ denote the joint density of (θ_1, θ_2), where $\mathbf{J}(\theta_1, \theta_2) \equiv \det(\frac{\partial(\delta,r)}{\partial(\theta_1,\theta_2)}) = \frac{1}{2k}$ is the Jacobian determinant of the coordinate transformation. $\varphi_1(\theta_1) \equiv \int_{\underline{\theta}_2(\theta_1)}^{\bar{\theta}_2(\theta_1)} \varphi(\theta_1, \theta_2)\, d\theta_2$ and $\Phi_1(\theta_1) \equiv \int_{\underline{\theta}_1}^{\theta_1} \varphi_1(\theta_1)\, d\theta_1$ represent the marginal PDF and marginal CDF of θ_1. $\varphi_{i|j}(\theta_i|\theta_j)$ and $\Phi_{i|j}(\theta_i|\theta_j) = \int_{\underline{\theta}_i(\theta_j)}^{\theta_i} \varphi_{i|j}(\theta_i|\theta_j)\, d\theta_i$ are respectively the conditional PDF and CDF of θ_i given θ_j. Denote by $\mathbb{H}_1(\theta_1) \equiv \frac{1-\Phi_1(\theta_1)}{\varphi_1(\theta_1)}$ the inverse hazard rate of θ_1, and $\mathbb{H}_{1|2}(\theta_1|\theta_2) = \frac{1-\Phi_{1|2}(\theta_1|\theta_2)}{\varphi_{1|2}(\theta_1|\theta_1)}$ the conditional inverse hazard rate of θ_1 given θ_2. Our main result in this subsection makes use of the following assumptions:

Assumption 1. The inverse hazard rate $\mathbb{H}_1(\theta_1)$ is nonincreasing in θ_1.

Assumption 2. The conditional inverse hazard rate $\mathbb{H}_{1|2}(\theta_1|\theta_2)$ is nonincreasing in θ_1 for all $\theta_2 \in \Theta_2$.

Assumption 3. $k \leqslant \frac{\sigma_r}{\sigma_\delta}$, σ_δ and σ_r are respectively standard deviations of r and δ.

Assumption 4. The PDF of θ_1 at left endpoint $\underline{\theta}_1$ is such that $\underline{\theta}_1\varphi_1(\underline{\theta}_1) \geqslant 1$.

Assumptions 1 and 2 are the familiar monotone inverse hazard rate conditions on the marginal and conditional distributions. Condition 3 implies $Cov(\theta_1, \theta_2) \leqslant 0$. The optimality of wage requires it to be nonincreasing in θ_2, while the implementability constraint requires wage to be increasing in θ_1 (note that $u(\theta_1)$ is convex in θ_1 and $u'(\theta_1) = \frac{1}{2} w' \Sigma w$), so we need to assume these two parameters to be negatively correlated. This condition holds if k is small or σ_r is sufficiently large relative to σ_δ. Assumption 4 ensures that the virtual valuation $\theta_1 - \mathbb{H}(\theta_1)$ is nonnegative everywhere, this assumption is commonly made in the mechanism design and auction literature.

It turns out that there are four cases (labeled as I to IV) of interest depending on the observabilities of θ_1 and θ_2:

	θ_2 observable	θ_2 unobservable
θ_1 observable	case I	case II
θ_1 unobservable	case III	case IV

Notice that in cases I and III, the wage contract could be written as $w_0^i(\theta_1, \theta_2)$, $w^i(\theta_1, \theta_2)$, $i \in \{I, III\}$. While in cases II and IV, they can be written as $w_0^i(\theta_1)$, $w^i(\theta_1)$, $i \in \{II, IV\}$.

The following theorem shows that unobservability of θ_1 may lead to low-powered incentives, but it is not necessarily the case with unobservable θ_2:

Theorem 3. *Suppose that a performance system $\mathcal{P} = (B, \Sigma)$ has constant correlation, i.e., $b_i' b_j / \sigma_{ij} \equiv k$, $\forall i, j$, and that Assumptions 1 to 4 are satisfied. Then we have*

1. *the strength of incentives in the cases with unobservable θ_1 is lower than that with observable θ_1: $|w_i^{III}(\theta_1, \theta_2)| < |w_i^I(\theta_1, \theta_2)|$, $|w_i^{IV}(\theta_1)| < |w_i^{II}(\theta_1)|$ for all $i \in \{1, \dots, m\}$, $\theta_2 \in \Theta_2$ and $\theta_1 \in [\underline{\theta}_1, \bar{\theta}_1)$; all wage vectors point in the same direction; the comparison between cases with observable and unobservable θ_2 is ambiguous;*
2. *the principal's expected surpluses in cases with observable and unobservable θ_1 are ordered as:*

$$\Pi^{III} \leqslant \Pi^I, \Pi^{IV} \leqslant \Pi^{II}. \tag{27}$$

The efficiency parameter δ and the risk aversion r affect the agent's payoff in different ways. δ affects his effort provision ($e^* = \delta B' w$) and thus term $(w_0 + w' B e^* - \frac{1}{2\delta} e^{*\prime} e^*)$ in his profit, while r affects his risk premium $(\frac{r}{2} w' \Sigma w)$. Misreporting these two parameters helps the agent get information rents with two degrees of freedom. However, if the correlation between any pair of performance measures is constant, all the information relevant to the agent's decision-making is contained in a scalar θ_1, and the agent is in fact deprived of one of his degrees of freedom. Therefore, the multidimensional mechanism design problem simplifies to the traditional one-dimensional problem. The low-powered incentives are resulted from traditional rent extraction-efficiency trade-off. Note that θ_2 is an irrelevant parameter to the agent's decision-making, it affects the principal's payoff by adding uncertainty rather than altering information rents. So the wage contract in cases I and II (cases III and IV as well) cannot be unambiguously ordered.

It is worth noting that if there is a single performance measure ($m = 1$), the condition $BB' = k\Sigma$ is clearly satisfied, and the mechanism design problem of principal is in fact unidimensional. Therefore, the multidimensional mechanism design problem may arise only in the joint presence of multidimensional types and multiple performance measures.

5.2. The Σ-norm based base wage: $w_0 = w_0(w' \Sigma w)$

In this subsection, we assume that the base wage lies only on the Σ-norm of performance wage vector w, that is, $w_0 = w_0(w' \Sigma w)$, where $w_0(\cdot)$ is a function of a scalar variable. That is to say, the employer determines the base wage solely on the Σ-norm of wage vector w rather than on its direction. There are two reasons for us to make this assumption.

First, this wage structure is very common in reality. It ties the employee's base wage (w_0) with the risk imposed on him (measured by $w' \Sigma w$). In managerial practice, employers usually set positions with varying degrees of importance and risk inside firms. The position with high pressure and high risk is typically accompanied by high compensation. Agency theory asserts that business risk will be positively related to base pay in managerial compensation contracts. The agent's risk-averse behavior induced by higher strength of incentive compensation might be mitigated by increasing his wealth through higher base pay levels. This idea, which may goes back to Bernoulli (1954), asserts that people's reactions to risk are inversely related to their level of wealth, greater wealth makes losses relatively less painful. Greater base pay increases the agent's wealth thereby offsetting some of the potential losses associated with both business risk and high-powered incentive compensation. Indeed, classic agency theory proposes that insurance will come in the form of higher base pay (Baiman, 1990). This premise is echoed in internal labor markets theory which asserts that people will require higher base pay levels in exchange for reduced stability in their employment and future income stream (Osterman, 1992).

Secondly, this assumption is made for tractability. If both δ and r are unknown to the principal, he has to process multidimensional information reported by the agent and prevent the agent from misreporting either one of them. This is a very difficult task (this point will be shown later). To reduce the information required by a mechanism and thus make the model tractable, one may impose this restriction on the structure of wage contract.

The contract menu (U, w) is called Σ-implementable if it belongs to

$$\mathcal{M}^{\Sigma} = \left\{ \begin{array}{l} (U, w) \in \mathbb{R}_+ \times \mathbb{R}^m \mid \exists w_0 : \mathbb{R}_+ \to \mathbb{R}_+, \text{ such that} \\ U(\delta, r) = \max\limits_{\tilde{w} \in \mathbb{R}^m} \left[w_0\big(\tilde{w}'\Sigma\tilde{w}\big) + \dfrac{1}{2}\tilde{w}'\big(\delta BB' - r\Sigma\big)\tilde{w} \right] \text{and} \\ w(\delta, r) = \arg\max\limits_{\tilde{w} \in \mathbb{R}^m} \left[w_0\big(\tilde{w}'\Sigma\tilde{w}\big) + \dfrac{1}{2}\tilde{w}'\big(\delta BB' - r\Sigma\big)\tilde{w} \right] \end{array} \right\}. \tag{28}$$

Let

$$\mathcal{M} = \left\{ \begin{array}{l} (U, w) \in \mathbb{R}_+ \times \mathbb{R}^m \mid \exists w_0 : \mathbb{R}^m \to \mathbb{R}_+, \text{ such that} \\ U(\delta, r) = \max\limits_{\tilde{w} \in \mathbb{R}^m} \left[w_0(\tilde{w}) + \dfrac{1}{2}\tilde{w}'\big(\delta BB' - r\Sigma\big)\tilde{w} \right] \text{and} \\ w(\delta, r) = \arg\max\limits_{\tilde{w} \in \mathbb{R}^m} \left[w_0(\tilde{w}) + \dfrac{1}{2}\tilde{w}'\big(\delta BB' - r\Sigma\big)\tilde{w} \right] \end{array} \right\} \tag{29}$$

be the set of implementable allocations.[15] It is obvious that a Σ-implementable mechanism is implementable but it is not true vice versa: $\mathcal{M}^{\Sigma} \subsetneqq \mathcal{M}$. When $(U, w) \in \mathcal{M}^{\Sigma}$, the agent's information rent accrued is

$$\begin{aligned} U(\delta, r) &= \max_{w \in \mathbb{R}^m} \left\{ w_0\big(w'\Sigma w\big) + \frac{1}{2} w'\big[\delta BB' - r\Sigma\big] w \right\} \\ &= \max_x \max_{w:\, w'\Sigma w = x^2} \left\{ w_0\big(w'\Sigma w\big) + \frac{1}{2} w'\big[\delta BB' - r\Sigma\big] w \right\} \\ &= \max_x \left[w_0\big(x^2\big) + \frac{1}{2}\left(\delta \max_{w'\Sigma w = x^2} \frac{w'BB'w}{w'\Sigma w} - r \right) x^2 \right] \\ &= \max_x \left[w_0\big(x^2\big) + \frac{1}{2}\vartheta_1 x^2 \right] \\ &\equiv u(\vartheta_1), \end{aligned} \tag{30}$$

where $\vartheta_1 = \delta\lambda_1 - r$,

$$\lambda_1 = \max_{w'\Sigma w = x^2} \frac{w'BB'w}{w'\Sigma w} = \lambda_1\big(\Sigma^{-1/2}BB'\Sigma^{-1/2}\big) = \lambda_1\big(BB'\Sigma^{-1}\big)$$

is the first (largest) eigenvalue of matrix $BB'\Sigma^{-1}$. The corresponding set of optimal wages for the agent is

$$\mathcal{W}(x) = \big\{ w \in \mathbb{R}^m \bigm| \Sigma^{1/2} w \in \mathcal{N}\big(\Sigma^{-1/2}BB'\Sigma^{-1/2} - \lambda_1 I\big),\ w'\Sigma w = x^2 \big\},$$

where $\mathcal{N}(\Sigma^{-1/2}BB'\Sigma^{-1/2} - \lambda_1 I)$ denotes the eigenspace of matrix $\Sigma^{-1/2}BB'\Sigma^{-1/2}$ corresponding to λ_1. (See Lemma A.5 in Appendix A for detailed discussion.) As discussed in previous sections, (30) implies the envelop condition $u'(\vartheta_1) = \frac{1}{2}x^2$ and the convexity of $u(\vartheta_1)$ in ϑ_1, which is conversely sufficient for the implementability of contract. We further define $\vartheta_2 = \delta\lambda_1 + r$, then the principal's optimization problem is reformulated as:

$$\begin{aligned} &\max_x \iint\limits_{\mathcal{D}_\vartheta} \left[\frac{\vartheta_1 + \vartheta_2}{2\lambda_1} \max_{w \in \mathcal{W}(x)} w'B\beta - \frac{1}{2}\vartheta_2 x^2 - u(\vartheta_1) \right] \psi(\vartheta_1, \vartheta_2)\, d\vartheta_1\, d\vartheta_2 \\ &\quad \text{s.t.:}\quad u'(\vartheta_1) = \frac{1}{2}x^2, \quad u(\cdot) \text{ is a convex function}, u(\vartheta_1) \geqslant 0, \end{aligned} \tag{31}$$

here $\psi(\vartheta_1, \vartheta_2) \equiv f(\frac{\vartheta_1+\vartheta_2}{2\lambda_1}, \frac{\vartheta_2-\vartheta_1}{2})\mathbf{J}(\vartheta_1, \vartheta_2) = f(\frac{\vartheta_1+\vartheta_2}{2\lambda_1}, \frac{\vartheta_2-\vartheta_1}{2})\frac{1}{2\lambda_1}$ is the joint density function supported on a convex region $\mathcal{D}_\vartheta$. $\mathcal{D}_\vartheta$ is defined by $\underline{\vartheta}_1 \leqslant \vartheta_1 \leqslant \overline{\vartheta}_1$ and $\underline{\vartheta}_2(\vartheta_1) \leqslant \vartheta_2 \leqslant \overline{\vartheta}_2(\vartheta_1)$.

For expositional convenience we define the following notations

$$\mu(\vartheta_1) \equiv \int_{\underline{\vartheta}_2(\vartheta_1)}^{\overline{\vartheta}_2(\vartheta_1)} (\vartheta_1 + \vartheta_2)\psi(\vartheta_1, \vartheta_2)\, d\vartheta_2,$$

$$\varrho(\vartheta_1) \equiv \int_{\underline{\vartheta}_2(\vartheta_1)}^{\overline{\vartheta}_2(\vartheta_1)} \vartheta_2 \psi(\vartheta_1, \vartheta_2)\, d\vartheta_2,$$

15 Here we abuse notations and still use $w_0(\cdot)$ to represent the base wage function.

$$\psi_1(\vartheta_1) \equiv \int_{\underline{\vartheta}_2(\vartheta_1)}^{\overline{\vartheta}_2(\vartheta_1)} \psi(\vartheta_1, \vartheta_2)\, d\vartheta_2,$$

$$\Psi_1(\vartheta_1) \equiv \int_{\underline{\vartheta}_1}^{\vartheta_1} \psi_1(s)\, ds,$$

$$\mathcal{H}(\vartheta_1) \equiv \frac{1-\Psi_1(\vartheta_1)}{\psi_1(\vartheta_1)}.$$

Analogous to the previous subsection, we need the following assumptions:

Assumption 5. $\mathcal{H}(\vartheta_1)$ is decreasing in ϑ_1.

Assumption 6. $\lambda_1 \leqslant \frac{\sigma_r}{\sigma_\delta}$, σ_r and σ_δ are respectively standard deviations of r and δ.

Assumption 7. The PDF at left endpoint $\underline{\vartheta}_1$ is such that: $\underline{\vartheta}_1 \psi_1(\underline{\vartheta}_1) \geqslant 1$.

Assumption 5 is the traditional monotone hazard rate property of parameter ϑ_1; Assumption 6 hold if σ_r is sufficiently large relative to σ_δ; Assumption 7 ensures that the virtual valuation $\vartheta_1 - \mathcal{H}_1(\vartheta_1)$ is nonnegative over the whole interval $[\underline{\vartheta}_1, \overline{\vartheta}_1]$. Under these assumptions, we then get the main result in this subsection:

Theorem 4. *Suppose that Assumptions 5 to 7 are satisfied, then the Σ-implementable contract entails*

$$w^*(\vartheta_1) = \frac{1}{2\lambda_1} \frac{\vartheta_1 + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1)}{\mathcal{H}(\vartheta_1) + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1)} \Sigma^{-1/2} Q_k Q_k' \Sigma^{-1/2} B\beta, \tag{32}$$

$$u^*(\vartheta_1) = \frac{\beta' B' \Sigma^{-1/2} Q_k Q_k' \Sigma^{-1/2} B\beta}{8\lambda_1^2} \int_{\underline{\vartheta}_1}^{\vartheta_1} \left(\frac{\vartheta_1 + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1)}{\mathcal{H}(\vartheta_1) + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1)} \right)^2 d\vartheta_1. \tag{33}$$

The surplus of the principal is:

$$\Pi^* = \frac{1}{8\lambda_1^2} \mathbb{E}_{\vartheta_1} \left[\frac{(\vartheta_1 + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1))^2}{\mathcal{H}(\vartheta_1) + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1)} \right] \beta' B' \Sigma^{-1/2} Q_k Q_k' \Sigma^{-1/2} B\beta. \tag{34}$$

$Q_k Q_k'$ is the spectral projector of matrix $\Sigma^{-1/2} BB' \Sigma^{-1/2}$ corresponding to the first eigenvalue λ_1.

Proof. See Appendix A. □

In the original model, the contract $\{w_0(\delta, r), w(\delta, r)\}$ is implementable if for all $(\delta, r, \hat{\delta}, \hat{r}) \in \mathcal{D}^2$, the following incentive compatibility condition is satisfied:

$$w_0(\delta, r) + \frac{1}{2} w(\delta, r)' \big(\delta BB' - r\Sigma\big) w(\delta, r) \geqslant w_0(\hat{\delta}, \hat{r}) + \frac{1}{2} w(\hat{\delta}, \hat{r})' \big(\delta BB' - r\Sigma\big) w(\hat{\delta}, \hat{r}). \tag{35}$$

Let

$$U(\delta, r) \equiv w_0(\delta, r) + \frac{1}{2} w(\delta, r)' \big(\delta BB' - r\Sigma\big) w(\delta, r)$$

and

$$U(\hat{\delta}, \hat{r}; \delta, r) \equiv w_0(\hat{\delta}, \hat{r}) + \frac{1}{2} w(\hat{\delta}, \hat{r})' \big(\delta BB' - r\Sigma\big) w(\hat{\delta}, \hat{r}).$$

Then $\{U(\delta, r), w(\delta, r)\}$ is implementable if

$$U(\delta, r) = \max_{(\hat{\delta}, \hat{r}) \in \mathcal{D}} \left\{ w_0(\hat{\delta}, \hat{r}) + \frac{1}{2} w(\hat{\delta}, \hat{r})' \big[\delta BB' - r\Sigma\big] w(\hat{\delta}, \hat{r}) \right\}. \tag{36}$$

Applying “taxation principle”, it could be equivalently represented as:

$$U(\delta, r) = \max_{w \in \mathbb{R}^m} \left\{ w_0(w) + \frac{1}{2} w' \big[\delta BB' - r\Sigma\big] w \right\}.$$

It implies that (i) the envelop conditions $\frac{\partial U}{\partial \delta} = \frac{1}{2} w' B B' w$, $\frac{\partial U}{\partial r} = -\frac{1}{2} w' \Sigma' w$ hold; (ii) $U(\delta, r)$ is convex in (δ, r). Conversely, given the envelop and convexity conditions, we have the following incentive compatibility condition:

$$\begin{aligned} U(\delta, r) &\geqslant U(\hat{\delta}, \hat{r}) + (\delta - \hat{\delta}) \frac{\partial U}{\partial \delta} + (r - \hat{r}) \frac{\partial U}{\partial r} \\ &= U(\hat{\delta}, \hat{r}) + \frac{1}{2} (\delta - \hat{\delta}) w(\hat{\delta}, \hat{r})' B B' w(\hat{\delta}, \hat{r}) - \frac{1}{2} \big(r - r'\big) w(\hat{\delta}, \hat{r})' \Sigma w(\hat{\delta}, \hat{r}) \\ &= U(\hat{\delta}, \hat{r}, \delta, r). \end{aligned}$$

The first inequality follows from the convexity condition, and the second is obtained by using the envelop condition. $U(\delta, r)$ and $w(\delta, r)$ are therefore implementable. Thus the principal's optimization problem is

$$\max_{U, w} \iint_{\mathcal{D}} \left[\delta w' B \beta - \frac{1}{2} w' \big(\delta B B' + r \Sigma \big) w - U(\delta, r) \right] d\delta \, dr$$

$$\text{s.t.:} \quad \frac{\partial U}{\partial \delta} = \frac{1}{2} w' B B' w, \quad \frac{\partial U}{\partial r} = -\frac{1}{2} w' \Sigma' w, \quad U(\delta, r) \geqslant 0, \ U(\delta, r) \text{ is convex.}$$

Ignoring momentarily the convexity condition, the principal's relaxed problem could be regarded as an optimal control problem with multiple controls and double-fold integrals. The generalized Hamiltonian approach offered by Basov (2001) is applicable to this problem. His method however ensures the existence of solution to the relaxed problem rather than offers a feasible way for getting it. One often has to rely upon the numerical techniques to get solution from a system of partial differential equations. A more serious drawback of his approach is that the solution to the relaxed problem usually cannot solve the complete problem because the convexity condition could only be satisfied by accident. In fact the envelop and convexity conditions require that the vector field $(\frac{1}{2} w' B B' w, -\frac{1}{2} w' \Sigma' w)$ has a convex potential function. This puts severe restrictions on the set of implementable wages and makes the multidimensional problem much more complex than its unidimensional counterpart because the latter requires only that $\frac{1}{2} w' B B' w$ or $-\frac{1}{2} w' \Sigma w$ has a convex antiderivative function.

In order to get an explicit solution to the complete problem, we therefore sacrifice some of the principal's degrees of freedom by restricting our attention in the set of Σ-implementable allocations $\mathcal{M}^{\Sigma}$. We decompose the information contained in vector w into two aspects: its Σ-norm ($\sqrt{w' \Sigma w} = x$) and its direction. Meanwhile, the type vector (δ, r) is transformed linearly to $(\vartheta_1, \vartheta_2)$. Notice that, the Σ-norm of wage vector depends only on ϑ_1, while its direction is at free disposal of the agent and depends on neither ϑ_1 nor ϑ_2. Our Σ-norm-based assumption on the base wage w_0 limits greatly the power of the principal since he now has only the discretion to choose Σ-norm of wage vector contingent on the agent's report $\hat{\vartheta}_1$. The agent, on the contrary, is vested the right of choosing the direction of wage vector. Therefore, this procedure is in fact a process of delegating part of the principal's authority to the agent. Under the assumptions we made, the multidimensional mechanism problem is solved with the same amount of computational work as in the one-dimensional screening problem after performing integration with respect to the irrelevant variable ϑ_2.

In a special case of orthogonal performance measurement system, we have the following corollary:

Corollary 1. *For an orthogonal performance measurement system, there is no incentive in the performance measures with non-largest signal–noise ratio.*

Proof. See Appendix A. □

As mentioned above, the wage vector is determined by two aspects: its overall intensity (Σ-norm) and relative allocation among performance measures (direction). In our dimensionality-reducing procedure, the authority of choosing relative allocation is delegated to the agent. Then for an orthogonal system in which performance measures are totally independent to each other, the agent inclines to allocate the overall intensity to the measures with larger sensitivity (measured by $\|b_i\|^2$) and smaller randomness (measured by σ_i^2). Therefore he will put the overall intensity of incentives on the measures with the largest signal–noise ratio $\|b_i\|^2/\sigma_i^2$, and the measures with non-largest signal–noise ratios will be assigned zero incentive. The ideas that underlie this analysis have many applications. For example, in partially decentralized political system, if the central government gives grants to local governments to award themselves, but continues to control their total budgets, this will certainly weaken incentives of local governments to allocate budgetary resources on relatively nonsensitive and high-risk measures. It is commonly observed that in many developing countries under decentralization, local governments are usually enthusiastic about improving their economic performance, but they lack motivations to do the works which are either highly risky (say purely theoretical research) or cannot bring them notable achievements in the short run (say environmental protection). Our prediction fits well with these phenomena.

Holmstrom and Milgrom (1991) show that missing incentive clauses are commonly observed in practice, even when good, objective output measures are available and agents are highly responsive to incentive pay. In their model, there exist multiple performance measures with varying degrees of accuracy (the tasks and performance measures are one-to-one corresponding to each other, that is, $B = I$), and the tasks are substitute to each other. In this setup, employees will

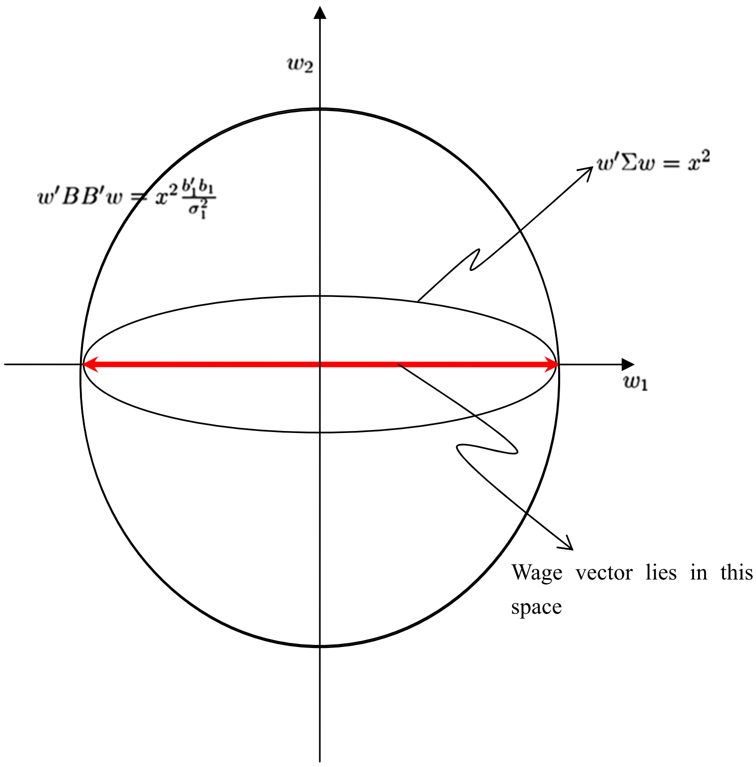

Fig. 1. Orthogonal system with $\frac{b_1'b_1}{\sigma_1^2} > \frac{b_2'b_2}{\sigma_1^2}$.

concentrate their attention (effort) on improving the performance measure tied to high compensation, to the exclusion of hard-to-measure or even non-observable but important tasks. Therefore an optimal incentive contract can be to pay a fixed wage independent of measured performance. Corollary 1 here offers a different explanation to the missing incentive phenomenon. Notice that in this corollary, we assume that the performance measures are orthogonal to each other, which is quite different to the substitute condition required by Holmstrom and Milgrom's paper.

Our result is illustrated in Figs. 1 and 2 for the case with $m = 2$; Figs. 3 and 4 for the case with $m = 3$. When Σ is diagonal and the Σ-norm is fixed, w lies on an ellipse (resp. ellipsoid if $m = 3$) whose axes coincide with the Cartesian axes. The agent will choose a vector w to maximize $w'BB'w$. If BB' is also diagonal, the optimal w must lie on the Cartesian axes or Cartesian planes.

The following corollary provides a comparison of surpluses obtained using two performance measurement systems with the same largest signal–noise ratio:

Corollary 2. *If two orthogonal performance measurement systems $\mathcal{P}_1 \equiv (B_1, \Sigma_{11})$ and $\mathcal{P}_2 \equiv (B_2, \Sigma_{22})$ are such that matrices $\Sigma_{11}^{-1/2}B_1B_1'\Sigma_{11}^{-1/2}$ and $\Sigma_{22}^{-1/2}B_2B_2'\Sigma_{22}^{-1/2}$ have the same first eigenvalues λ_1 and the multiplicities of λ_1 in these two matrices are, respectively, k_1 and k_2, then $\Pi^*(\mathcal{P}_1) \geqslant \Pi^*(\mathcal{P}_2)$ if and only if the sum of squares of congruences of the first k_1 performance measures in $\mathcal{P}_1$ is larger than that of the first k_2 performance measures in $\mathcal{P}_2$, i.e., $\sum_{i=1}^{k_1}\Upsilon_{i1}^2 \geqslant \sum_{i=1}^{k_2}\Upsilon_{i2}^2$.*

Proof. See Appendix A. $\square$

We next discuss the value of additional performance measures to an existing set. Let $\mathcal{P}_1 = (B_1, \Sigma_{11})$ represent a performance measurement system that reports m_1 measures and let

$$\mathcal{P} \equiv (B, \Sigma) = \left[\begin{pmatrix} B_1 \\ B_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right]$$

represent a system that reports additional m_2 measures $\mathcal{P}_2 = (B_2, \Sigma_{22})$. $\mathcal{P}_1$ and $\mathcal{P}_2$ are supposed to be orthogonal to each other. That is to say, $\Sigma_{12} = \mathbf{0}$, $\Sigma_{21} = \mathbf{0}$, $B_1'B_2 = \mathbf{0}$, $B_2'B_1 = \mathbf{0}$. Denote the set of eigenvalues of $\Sigma_{11}^{-1/2}B_1B_1'\Sigma_{11}^{-1/2}$ and $\Sigma_{22}^{-1/2}B_2B_2'\Sigma_{22}^{-1/2}$, respectively, by λ_i, $i = 1, \ldots, m_1$, and μ_j, $j = 1, \ldots, m_2$. The first eigenvalues $\lambda_1 = \max_{1\leqslant i\leqslant m_1}\lambda_i$ and $\mu_1 = \max_{1\leqslant j\leqslant m_2}\mu_j$ have multiplicities k_1 and k_2, respectively. The following corollary provides a specification of the incremental expected value of the additional performance measures provided by $\mathcal{P}$:

Corollary 3. *If $\lambda_1 > \mu_1$, then $\Pi^*(\mathcal{P}) = \Pi^*(\mathcal{P}_1)$; if $\lambda_1 = \mu_1$, then $\Pi^*(\mathcal{P}) = \Pi^*(\mathcal{P}_1) + \Pi^*(\mathcal{P}_2) > \Pi^*(\mathcal{P}_1)$; if $\lambda_1 < \mu_1$, then $\Pi^*(\mathcal{P}) = \Pi^*(\mathcal{P}_2)$.*

Proof. See Appendix A. $\square$

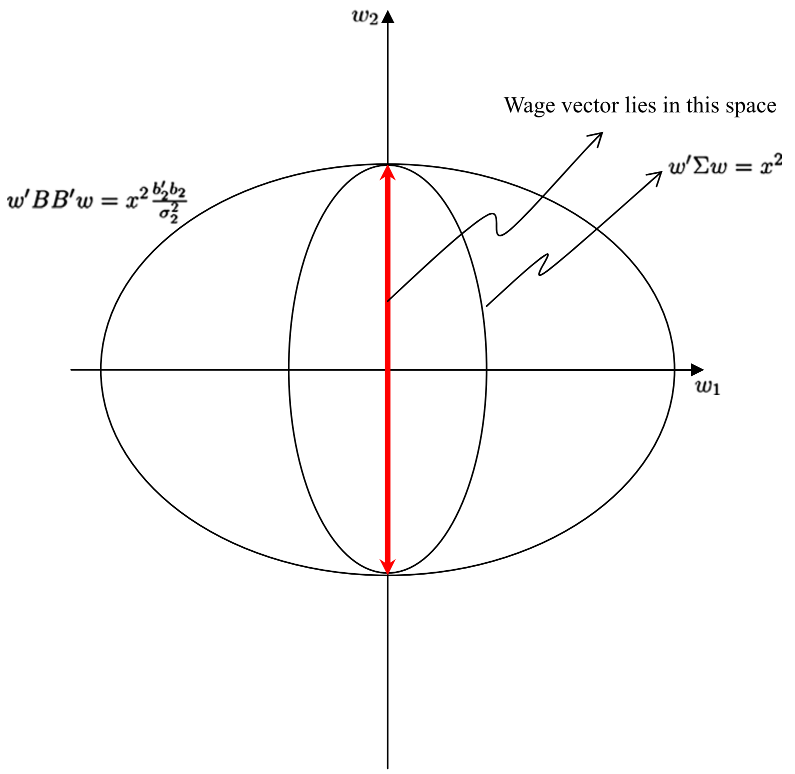

Fig. 2. Orthogonal system with $\frac{b_1'b_1}{\sigma_1^2} < \frac{b_2'b_2}{\sigma_1^2}$.

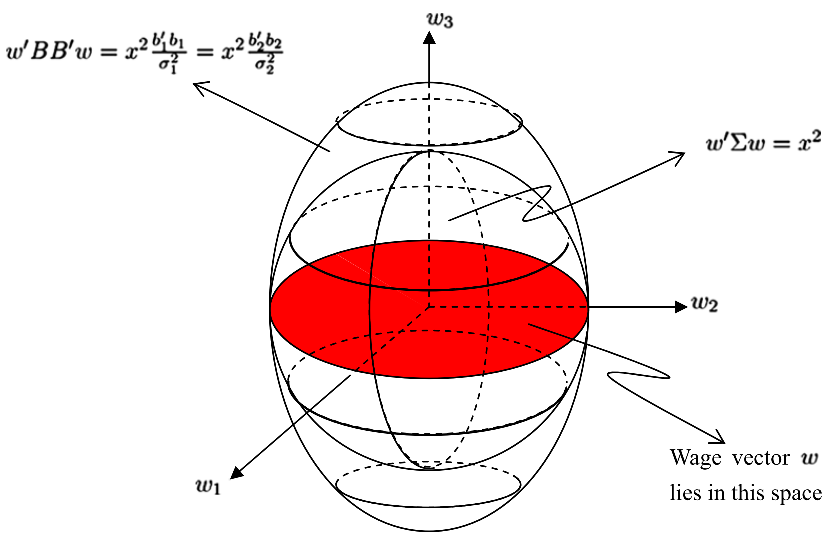

Fig. 3. Orthogonal system with $\frac{b_1'b_1}{\sigma_1^2} = \frac{b_2'b_2}{\sigma_1^2} > \frac{b_3'b_3}{\sigma_3^2}$.

In the environment of complete information (with observable costs and risk aversion), Feltham and Xie (1994) show that the incremental value of additional performance measures is always nonnegative because the principal can always assign zero incentive to the additional measures. In this case, the principal's surplus obtained using the original performance system $\mathcal{P}_1 = (B_1, \Sigma_{11})$ is

$$\Pi^p(\mathcal{P}_1) = \frac{\delta}{2}\beta' B_1' \left(B_1 B_1' + \frac{r}{\delta}\Sigma_{11} \right)^{-1} B_1 \beta;$$

the surplus obtained using the augmented performance measurement system

$$\mathcal{P} = (\mathcal{P}_1, \mathcal{P}_2) = \left[\begin{pmatrix} B_1 \\ B_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right]$$

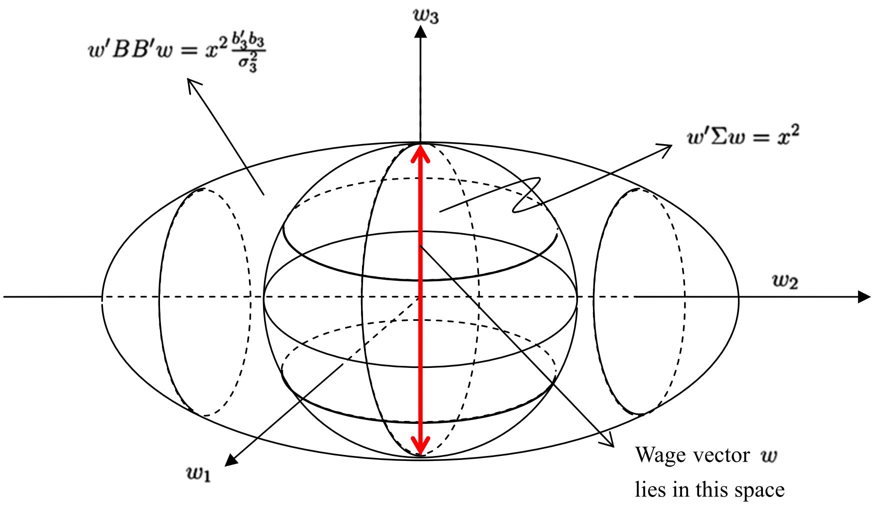

Fig. 4. Orthogonal system with $\frac{b_1' b_1}{\sigma_1^2} = \frac{b_2' b_2}{\sigma_1^2} < \frac{b_3' b_3}{\sigma_3^2}$.

is

$$\Pi^p(\mathcal{P}) = \frac{\delta}{2}\beta' B'\left(BB' + \frac{r}{\delta}\Sigma\right)^{-1} B\beta.$$

The incremental value of additional performance measures is thus:

$$\Delta\Pi = \Pi^p(\mathcal{P}) - \Pi^p(\mathcal{P}_1) = \frac{\delta}{2}\beta'(D - D_1)\beta,$$

where

$$D_1 = B_1'\left(B_1 B_1' + \frac{r}{\delta}\Sigma\right)^{-1} B_1,$$

$$\begin{aligned}
D &= B'\left(BB' + \frac{r}{\delta}\Sigma\right)^{-1} B \\
&= [B_1', B_2']\begin{bmatrix} B_1B_1' + \frac{r}{\delta}\Sigma_{11} & B_1B_2' + \frac{r}{\delta}\Sigma_{12} \\ B_2B_1' + \frac{r}{\delta}\Sigma_{21} & B_2B_2' + \frac{r}{\delta}\Sigma_{22}\end{bmatrix}^{-1}\begin{bmatrix} B_1 \\ B_2\end{bmatrix} \\
&= [B_1', B_2']\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22}\end{bmatrix}^{-1}\begin{bmatrix} B_1 \\ B_2\end{bmatrix} \\
&= [B_1', B_2']\begin{bmatrix} H_{11}^{-1} + H_{11}^{-1}H_{12}H_{22\cdot 1}^{-1}H_{21}H_{11}^{-1} & -H_{11}^{-1}H_{12}H_{22\cdot 1}^{-1} \\ -H_{22\cdot 1}^{-1}H_{21}H_{11}^{-1} & H_{22\cdot 1}^{-1}\end{bmatrix}\begin{bmatrix} B_1 \\ B_2\end{bmatrix} \\
&= B_1'H_{11}^{-1}B_1 + B_1'H_{11}^{-1}H_{12}H_{22\cdot 1}^{-1}H_{21}H_{11}B_1 - B_1'H_{11}H_{12}H_{22\cdot 1}^{-1}B_2 - B_2'H_{22\cdot 1}^{-1}H_{21}H_{11}^{-1}B_1 + B_2'H_{22\cdot 1}^{-1}B_2,
\end{aligned}$$

$$H_{ij} = B_iB_j' + \frac{r}{\delta}\Sigma_{ij},$$

$$H_{22\cdot 1} = H_{22} - H_{21}H_{11}^{-1}H_{12}.$$

It follows that $D - D_1 = B_1'H_{11}^{-1}H_{12}H_{22\cdot 1}^{-1}H_{21}H_{11}B_1 - B_1'H_{11}H_{12}H_{22\cdot 1}^{-1}B_2 - B_2'H_{22\cdot 1}^{-1}H_{21}H_{11}^{-1}B_1 + B_2'H_{22\cdot 1}^{-1}B_2 = [B_1'H_{11}^{-1}H_{12} - B_2']H_{22\cdot 1}^{-1}[H_{21}H_{11}^{-1}B_1 - B_2]$ is a semi-positive definite matrix. It in turn implies that $\Delta\Pi = \frac{\delta}{2}\beta'(D - D_1)\beta \geqslant 0$. As a special case, if $\mathcal{P}_1$ is orthogonal to $\mathcal{P}_2$, then $D - D_1 = B_2'H_{22}^{-1}B_2$, therefore, $\Delta\Pi = \frac{\delta}{2}\beta' B_2(B_2B_2' + \frac{r}{\delta}\Sigma_{22})^{-1}B_2\beta = \Pi^p(\mathcal{P}_2) \geqslant 0$.

The incremental value then is zero if and only if the measures provided by the original performance measurement system are a sufficient statistic for the measures provided by the augmented system, with respect to the agent's effort. (If there exists a constant matrix Ξ, such that $B_2 = \Xi B_1$, $\Sigma_{21} = \Xi\Sigma_{11}$, see Feltham and Xie, 1994 for detailed discussion.) According to this result, adding a performance measurement system which is orthogonal to the original one will increase the surplus for sure. Our result, on the contrary, states that the incremental value is zero if $\lambda_1 > \mu_1$, positive if $\lambda_1 = \mu_1$, and ambiguous if $\lambda_1 < \mu_1$. These new results come from the assumption that w_0 is based on the Σ-norm of w. Under this assumption, the performance measures associated with non-largest eigenvalues will be given zero weights, and the incremental value is therefore determined by the first eigenvalues of the original and new performance measurement systems.

6. Conclusion

In this paper, we explain the phenomenon of low-powered incentives from a new perspective. We consider the case where the agent possesses private information about his own risk aversion and the cost of efforts. Besides the rents motivating the agent's efforts, the principal has to give up some additional information rents to the agent in order to elicit his truthtelling. She has to consider two tradeoffs when choosing the optimal incentive contract. One is the tradeoff between insurance and incentives; the other is the tradeoff between efficiency and rent extraction. The former is the fundamental issue in moral hazard problem; while the latter lies in the core of adverse selection problem. We show that adverse selection usually leads to a flatter incentive contract comparatively to pure moral hazard situation. We further discuss this issue in the framework of multidimensionally hidden information. We first assume that the deterministic and stochastic components of a performance measurement system is proportional to each other. Under this assumption, the agent's private information relevant to his decision-making is contained in a single scalar variable. The power of incentive is lower than that in the case where this scalar variable is observable. Furthermore, we assume the base wage depends only on the Σ-norm of performance wage vector, by which the principal reduces the complexity of computation by delegating part of the principal's decision-making authority to the agent. In this setup, we find that most performance measures are assigned to zero incentives. This provides a new explanation for the missing of incentive clauses in contracts.

Acknowledgments

We would like to thank the co-editor and three anonymous referees for valuable comments and suggestions that significantly improved the exposition of the paper. We are also grateful to Linyue for her excellent research assistance. The first author is grateful to Veikko Thiele, participants at the 4th World Congress of the Game Theory Society in Istanbul, seminar participants at FUDAN, ECUST, DBUFE and JXUFE for helpful comments. Financial supports from the National Natural Science Foundation of China (NSFC-70773073), 211 Leading Academic Discipline Program at Shanghai University of Finance and Economics (Phase 4), Shanghai Leading Academic Discipline Project (B801), the Humanities and Social Sciences Project at the Ministry of Education of China (09YJC790184), and Innovation Program of Shanghai Municipal Education Commission (10YS164) are also gratefully acknowledged.

Appendix A. Proofs and lemmas

A.1. Proof of Proposition 1

Using the envelop condition $U'(r) = -\frac{1}{2}w'\Sigma w$, the participation constraint $U(r) \geqslant 0$ simplifies to $U(\bar{r}) \geqslant 0$. 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

$$U(r) = \int_{r}^{\bar{r}} \frac{1}{2} w(\tilde{r})'\Sigma w(\tilde{r})\, dr. \tag{A.1}$$

The principal's objective function becomes

$$\Pi = \int_{\underline{r}}^{\bar{r}} \left\{ \beta' C^{-1} B' w(r) - \frac{1}{2} w(r)' \left[BC^{-1}B' + r\Sigma \right] w(r) - \int_{r}^{\bar{r}} \frac{1}{2} w(\tilde{r})'\Sigma w(\tilde{r})\, dr \right\} f(r)\, dr$$

which, by an integration of parts, gives

$$\int_{\underline{r}}^{\bar{r}} \left\{ \beta' C^{-1} B' w(r) - \frac{1}{2} w(r)' \left[BC^{-1}B' + \left(r + \frac{F(r)}{f(r)} \right) \Sigma \right] w(r) \right\} f(r)\, dr.$$

Maximizing pointwise the above expression, we get

$$w^h(r) = \left[BC^{-1}B' + \Phi(r)\Sigma \right]^{-1} BC^{-1}\beta$$

and

$$w_0^h(r) = \frac{1}{2} \int_{r}^{\bar{r}} w^h(\tilde{r})'\Sigma w^h(\tilde{r})\, d\tilde{r} - \frac{1}{2} w^h(r)' \left[BC^{-1}B' - r\Sigma \right] w^h(r).$$

The only work left is to verify the convexity of $U(r)$. Notice that

$$U''(r) = -\left(D_r w^h \right)' \Sigma w^h = \Phi'(r) w^h(r)' \Sigma \left[BC^{-1}B' + \Phi(r)\Sigma \right]^{-1} \Sigma w^h(r).$$

The second equality comes from the fact that the derivative of w^h with respect to r is[16]

$$D_r w^h = -\left[BC^{-1}B' + \Phi(r)\Sigma\right]^{-1}\Phi'(r)\Sigma\left[BC^{-1}B' + \Phi(r)\Sigma\right]^{-1}BC^{-1}\beta$$
$$= -\Phi'(r)\left[BC^{-1}B' + \Phi(r)\Sigma\right]^{-1}\Sigma w^h.$$

It is clear that $U''(r) \geqslant 0$ because $\Phi'(r) \geqslant 0$ and the matrix $\Sigma[BC^{-1}B' + \Phi(r)\Sigma]^{-1}\Sigma$ is positive definite. The proof is completed.

A.2. Proof of Theorem 1

1. (a) Applying Lemma A.1 (see Appendix A.9. Lemmas below), there exist a nonsingular matrix U such that

$$U'BC^{-1}B'U = \Lambda, \qquad U'\Sigma U = I.$$

$\Lambda \equiv diag\{\omega_1, \ldots, \omega_m\}$ is a diagonal matrix composed of the generalized eigenvalues. We therefore have

$$\left(w^p\right)'\Sigma w^p = \beta' C^{-1}B'U(\Lambda + rI)^{-2}U'BC^{-1}\beta,$$
$$\left(w^h\right)'\Sigma w^h = \beta' C^{-1}B'U\left[\Lambda + \Phi(r)I\right]^{-2}U'BC^{-1}\beta.$$

It is obvious that $(w^h)'\Sigma w^h < (w^p)'\Sigma w^p$, $\forall r \in (\underline{r}, \bar{r}]$. If Σ is diagonal, then there exists at least one $i \in \{1 \cdots m\}$, such that $|w_i^h| < |w_i^p|$ for all $r \in (\underline{r}, \bar{r}]$.

(b) Also, we have

$$\left(w^p\right)'BC^{-1}B'w^p = \beta' C^{-1}B'V\left(r\Lambda^{-1} + I\right)^{-2}V'BC^{-2}\beta,$$
$$\left(w^h\right)'BC^{-1}B'w^h = \beta' C^{-1}B'V\left[\Phi(r)\Lambda^{-1} + I\right]^{-2}V'BC^{-1}\beta,$$

where $V = U\Lambda^{-1/2}$. Consequently, we have $(w^h)'BC^{-1}B'w^h < (w^p)'BC^{-1}B'w^p$, $\forall r \in (\underline{r}, \bar{r}]$. If $BC^{-1}B'$ is diagonal, then there exists at least one $i \in \{1 \cdots m\}$, such that $|w_i^h| < |w_i^p|$ for all $r \in (\underline{r}, \bar{r}]$.

(c) If $BC^{-1}B'\Sigma = \Sigma BC^{-1}B'$, then $BC^{-1}B'$ and Σ can be simultaneously diagonalized by an orthogonal matrix; that is, there exists an $m \times m$ orthogonal matrix P such that $P'BC^{-1}B'P = D_1$ and $P'\Sigma P = D_2$ for some diagonal matrices D_1 and D_2. It follows that

$$\left(w^p\right)'w^p = \beta' C^{-1}B'P(D_1 + rD_2)^{-2}P'BC^{-1}\beta,$$
$$\left(w^h\right)'w^h = \beta' C^{-1}B'P\left[D_1 + \Phi(r)D_2\right]^{-2}P'BC^{-1}\beta.$$

We conclude that the power of incentive will be reduced for at least one performance measure.

(d) It follows from the expression of w^p that

$$\left(w^p\right)'w^p = \beta C^{-1}B'\left[BC^{-1}B' + r\Sigma\right]^{-2}BC^{-1}\beta. \tag{A.2}$$

Differentiating (A.2) with respect to r yields

$$\frac{\partial w'w}{\partial r} = -\beta' C^{-1}B'\left[BC^{-1}B' + r\Sigma\right]^{-2}\left[\Sigma BC^{-1}B' + BC^{-1}B'\Sigma + 2r\Sigma^2\right]\left[BC^{-1}B' + r\Sigma\right]^{-2}BC^{-1}\beta. \tag{A.3}$$

Making use of Lemma A.2 and Lemma A.3, we have

$$\lambda_m\left(\Sigma BC^{-1}B' + BC^{-1}B'\Sigma + 2r\Sigma^2\right) \geqslant \lambda_m\left(\Sigma BC^{-1}B' + BC^{-1}B'\Sigma\right) + 2r\lambda_m\left(\Sigma^2\right) \geqslant \rho + 2r\lambda_m^2, \tag{A.4}$$

where

$$\rho = \max\left\{\min_{i=1,m} \lambda_i\mu_m \frac{(\sqrt{k_\lambda}+1)^2 - k_\mu(\sqrt{k_\lambda}-1)^2}{2\sqrt{k_\lambda}}, \min_{i=1,m} \mu_i\lambda_m \frac{(\sqrt{k_\mu}+1)^2 - k_\lambda(\sqrt{k_\mu}-1)^2}{2\sqrt{k_\mu}}\right\}$$
$$= \begin{cases} \lambda_m\mu_m \frac{(\sqrt{k_\lambda}+1)^2 - k_\mu(\sqrt{k_\lambda}-1)^2}{2\sqrt{k_\lambda}} & \text{if } \sqrt{k_\mu} \leqslant \frac{\sqrt{k_\lambda}+1}{\sqrt{k_\lambda}-1},\ k_\mu \geqslant k_\lambda, \\ \lambda_m\mu_m \frac{(\sqrt{k_\mu}+1)^2 - k_\lambda(\sqrt{k_\mu}-1)^2}{2\sqrt{k_\mu}} & \text{if } \sqrt{k_\mu} \leqslant \frac{\sqrt{k_\lambda}+1}{\sqrt{k_\lambda}-1},\ k_\mu < k_\lambda, \\ \lambda_1\mu_m \frac{(\sqrt{k_\lambda}+1)^2 - k_\mu(\sqrt{k_\lambda}-1)^2}{2\sqrt{k_\lambda}} & \text{if } \sqrt{k_\mu} > \frac{\sqrt{k_\lambda}+1}{\sqrt{k_\lambda}-1},\ k_\mu \geqslant k_\lambda, \\ \lambda_m\mu_1 \frac{(\sqrt{k_\mu}+1)^2 - k_\lambda(\sqrt{k_\mu}-1)^2}{2\sqrt{k_\mu}} & \text{if } \sqrt{k_\mu} > \frac{\sqrt{k_\lambda}+1}{\sqrt{k_\lambda}-1},\ k_\lambda < k_\lambda. \end{cases}$$

[16] Let A be a nonsingular, $m \times m$ matrix whose elements are functions of the scalar parameter α, then

$$\frac{\partial A^{-1}}{\partial \alpha} = -A^{-1}\frac{\partial A}{\partial \alpha}A^{-1}.$$

If $\rho + 2r\lambda_m^2 > 0$ holds, then matrix $\Sigma BC^{-1}B' + BC^{-1}B'\Sigma + 2r\Sigma^2$ is positive definite, and consequently $\frac{\partial w'w}{\partial r} < 0$. Therefore $\|w^p\| > \|w^h\|$. Then we get the result that there exists at least one i such that $|w_i^p(r)| < |w_i^h(r)|$, $\forall r \in (\underline{r}, \bar{r}]$.

2. If both Conditions 3.1 and 3.2 are satisfied, then

$$w_i^p(r) = \frac{b_i' C^{-1}\beta}{b_i' C^{-1} b_i + r\sigma_i^2},$$
$$w_i^h(r) = \frac{b_i' C^{-1}\beta}{b_i' C^{-1} b_i + \Phi(r)\sigma_i^2}.$$

It is obvious that $|w_i^p(r)| < |w_i^h(r)|$ for all i and all $r \in (\underline{r}, \bar{r}]$.

3. If $BC^{-1}B' = \lambda\Sigma$, it is obvious that $w^h = \frac{\lambda+r}{\lambda+\Phi(r)} w^p$. Now we need only to prove this conclusion from the opposite direction. As shown above, $BC^{-1}B'$ and Σ can be simultaneously diagonalized by nonsingular matrix U, then we have

$$w^p(r) = U(\Lambda + r)^{-1} U' BC^{-1}\beta,$$
$$w^h(r) = U\big(\Lambda + \Phi(r)\big)^{-1} U' BC^{-1}\beta.$$

If $w^h(r) = kw^p(r)$ then $\frac{ku_i' BC^{-1}\beta}{r+\omega_i} = \frac{u_i' BC^{-1}\beta}{\Phi(r)+\omega_i}$, $\forall i$, where u_i is the i-th column of U. Since $BC^{-1}\beta \notin \bigcup_{i\in\mathcal{K}} \mathcal{V}_i^{\perp}$, $u_i' BC^{-1}\beta \neq 0$ for all i. Then we get $\omega_i = \lambda \equiv \frac{k\Phi(r)-r}{1-k}$. It follows that $BC^{-1}B' = \lambda(XX')^{-1} = \lambda\Sigma$.

A.3. Proof of Proposition 2

Using integration by parts, we get

$$\int_{\underline{\delta}}^{\bar{\delta}} U(\delta) g(\delta) = \int_{\underline{\delta}}^{\bar{\delta}} \left[\frac{1-G(\delta)}{g(\delta)}\right] \frac{w'BB'w}{2}\, dG(\delta).$$

Substituting it into the expression of the principal's expected surplus and optimizing it with respect to w, we get the second-best performance wage $w^h(\delta)$, and $w_0^h(\delta)$ is also easily obtained. We now check the convexity of $U(\delta)$. The first order derivative of $w^h(\delta)$ is

$$\begin{aligned}
D_\delta w^h(\delta) &= -\left[H(\delta)BB' + \frac{r}{\delta}\Sigma\right]^{-1}\left[H'(\delta)BB' - \frac{r\Sigma}{\delta^2}\right]\left[H(\delta)BB' + \frac{r\Sigma}{\delta}\right]^{-1} B\beta \\
&= -\left[H(\delta)BB' + \frac{r\Sigma}{\delta}\right]^{-1}\left[H'(\delta)BB' - \frac{r\Sigma}{\delta^2}\right] w^h(\delta) \\
&= -\left[H(\delta)BB' + \frac{r\Sigma}{\delta}\right]^{-1}\left\{-\frac{H(\delta)}{\delta}BB' - \frac{r\Sigma}{\delta^2} + \left[\frac{H(\delta)}{\delta} + H'(\delta)\right]BB'\right\} w^h(\delta) \\
&= \frac{1}{\delta}\left\{\big(BB'\big)^{-1} - \left[H(\delta)BB' + \frac{r\Sigma}{\delta}\right]^{-1}\big[H(\delta) + \delta H'(\delta)\big]\right\} BB' w^h(\delta).
\end{aligned}$$

It can be verified that the matrix $\frac{1}{\delta}\{(BB')^{-1} - [H(\delta)BB' + \frac{r\Sigma}{\delta}]^{-1}[H(\delta) + \delta H'(\delta)]\}$ is positive definite since $\delta + \frac{1-G(\delta)}{g(\delta)} = \delta H(\delta)$ is decreasing. Therefore

$$\begin{aligned}
U''(\delta) &= D_\delta w^h(\delta) BB' w^h(\delta) \\
&= \frac{1}{\delta} w^h(\delta)' BB' \left\{\big(BB'\big)^{-1} - \left[H(\delta)BB' + \frac{r\Sigma}{\delta}\right]^{-1}\big[H(\delta) + \delta H'(\delta)\big]\right\} BB' w^h(\delta) \geqslant 0,
\end{aligned}$$

which implies the convexity of $U(\delta)$.

A.4. Proof of Theorem 3

- Case IV: The principal's objective in case IV is rewritten as:

$$\Pi^{IV} = \iint_{\mathcal{D}} \left[\delta w' B\beta - \frac{1}{2} w'\big(\delta BB' + r\Sigma\big)w - U(\delta, r)\right] f(\delta, r)\, d\delta\, dr$$
$$= \int_{\underline{\theta}_1}^{\bar{\theta}_1} \left[\frac{w' B\beta}{2k} g(\theta_1) - \frac{1}{2} w'\Sigma w h(\theta_1) - u(\theta_1)\varphi_1(\theta)\right] d\theta_1, \tag{A.5}$$

where $g(\theta_1) \equiv \int (\theta_1 + \theta_2)\varphi(\theta_1, \theta_2)\, d\theta_2$, $h(\theta_1) \equiv \int \theta_2 \varphi(\theta_1, \theta_2)\, d\theta_2$.[17] As a consequence, the principal's optimal contract design problem simplifies to a unidimensional mechanism design problem:

$$\max_{w(\cdot),\, u(\cdot)} \Pi^{IV}, \quad \text{s.t.:} \quad u'(\theta_1) = \frac{1}{2} w'\Sigma w, \quad u(\theta_1) \text{ is a convex function},\ u(\theta_1) \geqslant 0. \tag{A.6}$$

Using the integration by parts technique, the principal's objective can be expressed as:

$$\Pi^{IV} = \int_{\underline{\theta}_1}^{\bar{\theta}_1} \left\{\frac{1}{2k} g(\theta_1) w' B\beta - \frac{w'\Sigma w}{2}\big[h(\theta_1) + 1 - \Phi_1(\theta_1)\big]\right\} d\theta_1. \tag{A.7}$$

We ignore momentarily the convexity condition and simply maximize this expression pointwise with respect to w to get:

$$w^{IV}(\theta_1) = \frac{1}{2k} \frac{g(\theta_1)}{h(\theta_1) + 1 - \Phi_1(\theta_1)} \Sigma^{-1} B\beta = \rho(\theta_1)\Sigma^{-1} B\beta, \tag{A.8}$$

where

$$\rho(\theta_1) \equiv \frac{1}{2k} \frac{g(\theta_1)}{h(\theta_1) + 1 - \Phi_1(\theta_1)} = \frac{1}{2k} \frac{\theta_1 + \mathbb{E}_{\theta_2}(\theta_2|\theta_1)}{\mathbb{H}_1(\theta_1) + \mathbb{E}_{\theta_2}(\theta_2|\theta_1)}.$$ [18]

The only task left is to check the convexity of function $u(\theta_1)$. Since

$$u''(\theta_1) = (D_{\theta_1} w)'\Sigma w = \rho'(\theta_1)\rho(\theta_1)\beta' B'\Sigma^{-1} B\beta,$$

$u(\cdot)$ is convex if and only if $\rho(\cdot)$ is nondecreasing. It holds provided that: (i) $\mathbb{H}_1(\theta_1)$ is nonincreasing; (ii) $\eta(\theta_1) \equiv \mathbb{E}_{\theta_2}(\theta_2|\theta_1)$ is nonincreasing; and (iii) $\theta_1 - \mathbb{H}_1(\theta_1) \geqslant 0$ for all $\theta_1 \in [\underline{\theta}_1, \bar{\theta}_1]$. Condition (i) is the familiar monotone hazard rate property, while condition (ii) is equivalent to $Cov(\theta_1, \theta_2) < 0$ (see Lemma A.4). Note that $Cov(\theta_1, \theta_2) = k^2\sigma_\delta^2 - \sigma_r^2$, it holds if and only if Assumption 3 is satisfied. Under monotone hazard rate condition, condition (iii) is satisfied if $\underline{\theta}_1 \varphi_1(\underline{\theta}_1) \geqslant 1$. Substituting (A.8) into (A.7), we get the principal's expected profit.

$$\Pi^{IV} = \frac{1}{8k^2}\mathbb{E}_{\theta_1}\left[\frac{(\theta_1 + \mathbb{E}_{\theta_2}(\theta_2|\theta_1))^2}{\mathbb{H}_1(\theta_1) + \mathbb{E}_{\theta_2}(\theta_2|\theta_1)}\right]\beta' B'\Sigma^{-1} B\beta. \tag{A.9}$$

- Case II: If θ_1 is observable and θ_2 is unknown to the principal, i.e., in case II, we only need to consider the participation constraint $u(\theta_1) \geqslant 0$ in (A.6), then the wage contract and the corresponding surplus are

$$w^{II}(\theta_1) = \frac{1}{2k} \frac{\theta_1 + \mathbb{E}_{\theta_2}(\theta_2|\theta_1)}{\mathbb{E}_{\theta_2}(\theta_2|\theta_1)} \Sigma^{-1} B\beta, \tag{A.10}$$

$$\Pi^{II} = \frac{1}{8k^2}\mathbb{E}_{\theta_1}\left[\frac{(\theta_1 + \mathbb{E}_{\theta_2}(\theta_2|\theta_1))^2}{\mathbb{E}_{\theta_2}(\theta_2|\theta_1)}\right]\beta' B'\Sigma^{-1} B\beta. \tag{A.11}$$

- Case III: In case III, the principal's objective function is

$$\Pi^{III} = \iint_{\Theta} \left[\frac{\theta_1 + \theta_2}{2k} w' B\beta - \frac{1}{2} w'\Sigma w \theta_2 - u(\theta_1, \theta_2)\right] \varphi(\theta_1, \theta_2)\, d\theta_1\, d\theta_2.$$

Notice that in this case w depends on both θ_1 and θ_2, so it cannot be reduced to one-fold integral as in (A.5). The principal's optimization problem could thus be rewritten as

$$\max_{w} \Pi^{III}, \quad \text{s.t.:} \quad \frac{\partial u}{\partial \theta_1} = \frac{1}{2} w'\Sigma w, \quad u(\theta_1, \theta_2) \text{ is convex in } \theta_1,\ u(\theta_1, \theta_2) \geqslant 0. \tag{A.12}$$

---

[17] To simplify notation, we drop the limits of integrals.

[18] $\mathbb{E}_{\theta_i}(\cdot)$ is the expectation operator with respect to θ_i.

The envelop condition implies $u(\theta_1,\theta_2)=\int_{\underline{\theta}_1(\theta_2)}^{\theta_1}\frac{1}{2}w'\Sigma w\,d\theta_1+s(\theta_2)$. Substituting it into the principal's objective function and using the integration by parts technique yields

$$\Pi^{III}=\iint_{\Theta}\left\{\frac{\theta_1+\theta_2}{2k}w'B\beta-\frac{1}{2}w'\Sigma w\big[\theta_2+\mathbb{H}_{1|2}(\theta_1|\theta_2)\big]\right\}\varphi(\theta_1,\theta_2)\,d\theta_1\,d\theta_2+\iint_{\Theta}s(\theta_2)\varphi_2(\theta_2)\,d\theta_1\,d\theta_2. \tag{A.13}$$

It is optimal to choose a function $s(\theta_2)$ such that $\iint_{\Theta}s(\theta_1)\varphi_2(\theta_2)\,d\theta_1\,d\theta_2=0$. With Assumption 4 ensuring the convexity condition, we get the optimal wage contract and surplus as follows:

$$w^{III}(\theta_1,\theta_2)=\frac{1}{2k}\frac{\theta_1+\theta_2}{\mathbb{H}_{1|2}(\theta_1|\theta_2)+\theta_2}\Sigma^{-1}B\beta, \tag{A.14}$$

$$\Pi^{III}=\frac{1}{8k^2}\mathbb{E}_{\theta}\left[\frac{(\theta_1+\theta_2)^2}{\mathbb{H}_{1|2}(\theta_1|\theta_2)+\theta_2}\right]\beta'B'\Sigma^{-1}B\beta. \tag{A.15}$$

- Case I: If θ_1 and θ_2 are both observable, we need only to consider IR constraint in (A.12), the wage contract and surplus are therefore

$$w^{I}(\theta_1,\theta_2)=\frac{1}{2k}\frac{\theta_1+\theta_2}{\theta_2}\Sigma^{-1}B\beta, \tag{A.16}$$

$$\Pi^{I}=\frac{1}{8k^2}\mathbb{E}_{\theta}\left[\frac{(\theta_1+\theta_2)^2}{\theta_2}\right]\beta'B'\Sigma^{-1}B\beta. \tag{A.17}$$

It is obvious that $|w_i^{IV}|\leqslant|w_i^{II}|$, $|w_i^{III}|\leqslant|w_i^{I}|$, $\forall i$ and $\Pi^{III}\leqslant\Pi^{I}$, $\Pi^{IV}\leqslant\Pi^{II}$, but the comparison between cases I and II (cases III and IV as well) is ambiguous.

A.5. Proof of Theorem 4

Letting $y=\Sigma^{1/2}w$, the embedded program $\max_{w\in\mathcal{W}(x)}w'B\beta$ can be expressed as

$$\max_{y}y'\Sigma^{-1/2}B\beta,\quad \text{s.t.:}\quad y'y=x^2,\quad y\in\mathcal{N}\big(\Sigma^{-1/2}BB'\Sigma^{-1/2}-\lambda_1 I\big).$$

Applying Lemma A.6, we get the solution and the corresponding maximized value of this program

$$y^*=x\frac{Q_kQ_k'\Sigma^{-1/2}B\beta}{\sqrt{\beta'B'\Sigma^{-1/2}Q_kQ_k'\Sigma^{-1/2}B\beta}},\qquad \Pi^*=x\sqrt{\beta'B'\Sigma^{-1/2}Q_kQ_k'\Sigma^{-1/2}B\beta}.$$

Following from the spectral representation theorem in linear algebra, Q_kQ_k' is unique although Q_k is usually not. (See Lemma A.7 for detailed discussion.)

The maxima to the original program $\max_{w\in\mathcal{W}(x)}w'B\beta$ is therefore

$$w^*=x\frac{\Sigma^{-1/2}Q_kQ_k'\Sigma^{-1/2}B\beta}{\sqrt{\beta'B'\Sigma^{-1/2}Q_kQ_k'\Sigma^{-1/2}B\beta}}. \tag{A.18}$$

Substituting this expression into (31), we can rewrite the optimization problem of the principal as

$$\max_{u,x}\iint_{\mathcal{D}_{\vartheta}}\left[\frac{\vartheta_1+\vartheta_2}{2\lambda_1}\sqrt{\beta'B'\Sigma^{-1/2}Q_kQ_k'\Sigma^{-1/2}B\beta}\,x-\frac{1}{2}\vartheta_2x^2-u(\vartheta_1)\right]\psi(\vartheta_1,\vartheta_2)\,d\vartheta_1\,d\vartheta_2$$
$$\text{s.t.:}\quad u'(\vartheta_1)=\frac{1}{2}x^2,\quad u(\vartheta_1)\text{ is a convex function},\ u(\vartheta_1)\geqslant 0. \tag{A.19}$$

Integrating with respect to ϑ_2, the above optimization can be simplified to a standard one-dimensional screening problem:

$$\max_{u,x}\int_{\underline{\vartheta}_1}^{\overline{\vartheta}_1}\left[x\frac{\mu(\vartheta_1)}{2\lambda_1}\sqrt{\beta'B'\Sigma^{-1/2}Q_kQ_k'\Sigma^{-1/2}B\beta}-\frac{x^2}{2}\varrho(\vartheta_1)-u(\vartheta_1)\psi_1(\vartheta_1)\right]d\vartheta_1$$
$$\text{s.t.:}\quad u'(\vartheta_1)=\frac{x^2}{2},\quad u(\vartheta_1)\text{ is a convex function},\ u(\vartheta_1)\geqslant 0. \tag{A.20}$$

Ignoring for a while the convexity condition and applying the standard technique, we obtain the solution to the relaxed problem:

$$x^*(\vartheta_1) = \frac{\sqrt{\beta' B' \Sigma^{-1/2} Q_k Q_k' \Sigma^{-1/2} B \beta}}{2\lambda_1} \frac{\mu(\vartheta_1)}{\varrho(\vartheta_1) + [1 - \Psi_1(\vartheta_1)]}$$
$$= \frac{\sqrt{\beta' B' \Sigma^{-1/2} Q_k Q_k' \Sigma^{-1/2} B \beta}}{2\lambda_1} \frac{\vartheta_1 + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1)}{\mathcal{H}(\vartheta_1) + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1)}. \tag{A.21}$$

It can be verified that under Assumptions 5 to 7, $x(\vartheta_1)$ is increasing in ϑ_1. In turn, this implies the convexity of $u(\vartheta_1)$. Substituting (A.21) into (A.18) we get the optimal wage

$$w^*(\vartheta_1) = \frac{1}{2\lambda_1} \frac{\vartheta_1 + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1)}{\mathcal{H}(\vartheta_1) + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1)} \Sigma^{-1/2} Q_k Q_k' \Sigma^{-1/2} B \beta. \tag{A.22}$$

The information rent accrued to the agent and surplus of the principal are also easily obtained:

$$u^*(\vartheta_1) = \frac{\beta' B' \Sigma^{-1/2} Q_k Q_k' \Sigma^{-1/2} B \beta}{8\lambda_1^2} \int_{\underline{\vartheta}_1}^{\vartheta_1} \left(\frac{\vartheta_1 + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1)}{\mathcal{H}(\vartheta_1) + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1)} \right)^2 d\vartheta_1, \tag{A.23}$$

$$\Pi^* = \frac{1}{8\lambda_1^2} \mathbb{E}_{\vartheta_1} \left[\frac{(\vartheta_1 + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1))^2}{\mathcal{H}(\vartheta_1) + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1)} \right] \beta' B' \Sigma^{-1/2} Q_k Q_k' \Sigma^{-1/2} B \beta. \tag{A.24}$$

A.6. Proof of Corollary 1

Let $\mathcal{P} = (B, \Sigma)$ be an orthogonal performance measurement system with $\Sigma = diag\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2\}$, $BB' = diag\{\|b_1\|^2, \ldots, \|b_m\|^2\}$, then $\Sigma^{-\frac{1}{2}} BB' \Sigma^{-\frac{1}{2}} = diag\{\lambda_1, \ldots, \lambda_m\}$, $\lambda_i = \|b_i\|^2/\sigma_i^2$, $i = 1, \ldots, m$, are eigenvalues in descending order. The largest element λ_1 has multiplicity k, that is, $\lambda_1 = \lambda_2 = \cdots = \lambda_k > \lambda_{k+1} \geqslant \cdots \geqslant \lambda_m$. Let $\mathbf{p} = (p_1, \ldots, p_m)' \in \mathbb{R}^m$ be the normalized eigenvector associated with λ_1. Then

$$\mathbf{p}' \Sigma^{-1/2} BB' \Sigma^{-1/2} \mathbf{p} = \lambda_1.$$

It follows that

$$\lambda_1 \sum_{j=1}^{k} p_j^2 + \sum_{j=k+1}^{p} \lambda_j p_j^2 = \lambda_1 \sum_{j=1}^{p} p_j^2.$$

Then we obtain

$$p_j = 0, \quad \forall j = k+1, \ldots, m.$$

Therefore we write

$$Q_k = \begin{bmatrix} \tilde{Q}_k \\ \mathbf{0} \end{bmatrix},$$

where $\tilde{Q}_k$ is a $k \times k$ orthogonal matrix. Substituting it into (A.22), we get

$$w^*(\vartheta_1) = \frac{1}{2\lambda_1} \frac{\vartheta_1 + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1)}{\mathcal{H}(\vartheta_1) + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1)} \begin{bmatrix} b_1' \beta / \sigma_1^2 \\ \vdots \\ b_k' \beta / \sigma_k^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{A.25}$$

The optimal wages paid for the performance measures associated with the non-largest eigenvalues are zero: $w_i^*(\vartheta_1) = 0$, for all $i = k+1, \ldots, m$.

A.7. Proof of Corollary 2

Suppose that $\mathcal{P}_1$ and $\mathcal{P}_2$ are orthogonal systems with the same first eigenvalues $\lambda_1 = \lambda_1(\Sigma_{11}^{-1/2} B_1 B_1' \Sigma_{11}^{-1/2}) = \lambda_1(\Sigma_{22}^{-1/2} B_2 B_2' \Sigma_{22}^{-1/2})$, and the multiplicities of λ_1 are respectively k_1 and k_2. Then the surplus obtained using system $\mathcal{P}_1$ is

$$
\begin{aligned}
\Pi^*(\mathcal{P}_1) &= \kappa(\lambda_1)\beta' B_1' \Sigma_{11}^{-1/2} Q_1 Q_1' \Sigma_{11}^{-1/2} B_1 \beta \\
&= \kappa(\lambda_1)\beta' B_1' \begin{bmatrix} \Lambda_{k_1} & \mathbf{0} \\ \mathbf{0} & \Lambda_{m_1-k_1} \end{bmatrix} \begin{bmatrix} I_{k_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Lambda_{k_1} & \mathbf{0} \\ \mathbf{0} & \Lambda_{m_1-k_1} \end{bmatrix} B_1 \beta \\
&= \kappa(\lambda_1)\beta' \sum_{i=1}^{k_1} \frac{b_i^1 b_i^{1'}}{\sigma_i^2} \beta \\
&= \kappa(\lambda_1) \sum_{i=1}^{k_1} \frac{\|b_i^1\|^2 \|\beta\|^2 \cos^2(\widehat{b_i^1, \beta})}{\sigma_i^2} \\
&= \kappa(\lambda_1)\lambda_1 \|\beta\|^2 \sum_{i=1}^{k_1} \cos^2\bigl(\widehat{b_i^1, \beta}\bigr) \\
&= \kappa(\lambda_1)\lambda_1 \|\beta\|^2 \sum_{i=1}^{k_1} \Upsilon_{i1}^2,
\end{aligned}
$$

where

$$
\kappa(\lambda_1) = \frac{1}{8\lambda_1^2} \mathbb{E}_{\vartheta_1}\left[\frac{(\vartheta_1 + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1))^2}{\mathcal{H}(\vartheta_1) + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1)}\right].^{19}
$$

$\Lambda_{k_1} = diag\{\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_{k_1}}\}$, $\Lambda_{m_1-k_1} = diag\{\frac{1}{\sigma_{k_1+1}}, \ldots, \frac{1}{\sigma_{m_1}}\}$, b_i^1, $i = 1, \ldots, m_1$, are the columns of B_1', $\Upsilon_{i1} = \cos(\widehat{b_i^1, \beta})$. Similarly, the surplus obtained using system $\mathcal{P}_2$ is:

$$
\Pi^*(\mathcal{P}_2) = \kappa(\lambda_1)\lambda_1 \|\beta\|^2 \sum_{i=1}^{k_2} \cos^2\bigl(\widehat{b_i^2, \beta}\bigr) = \kappa(\lambda_1)\lambda_1 \|\beta\|^2 \sum_{i=1}^{k_2} \Upsilon_{i2}^2, \tag{A.26}
$$

where b_i^2, $i = 1, \ldots, m_2$, are columns of matrix B_2', $\Upsilon_{i2} = \cos(\widehat{b_i^2, \beta})$. It follows that $\Pi^*(\mathcal{P}_1) \geqslant \Pi^*(\mathcal{P}_2)$ if and only if $\sum_{i=1}^{k_1} \Upsilon_{i1}^2 \geqslant \sum_{i=1}^{k_1} \Upsilon_{i2}^2$.

A.8. Proof of Corollary 3

The matrix

$$
\Sigma^{-1/2} B B' \Sigma^{-1/2} = \begin{bmatrix} \Sigma_{11}^{-1/2} B_1 B_1' \Sigma_{11}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22}^{-1/2} B_2 B_2' \Sigma_{22}^{-1/2} \end{bmatrix}
$$

has $m_1 + m_2$ eigenvalues $\lambda_1, \ldots, \lambda_{m_1}, \mu_1, \ldots, \mu_{m_2}$.

1. If $\lambda_1 > \mu_1$, then the first eigenvalue of $\Sigma^{-1/2} B B' \Sigma^{-1/2}$ is λ_1, its multiplicity is still k_1. If $\mathbf{q} \in \mathbb{R}^{m_1}$ is an eigenvector of $\Sigma_{11}^{-1/2} B_1 B_1' \Sigma_{11}^{-1/2}$ associated with λ_1, then $\hat{\mathbf{q}} = (\mathbf{q}, \mathbf{0})' \in \mathbb{R}^{m_1+m_2}$ is clearly the eigenvector of $\Sigma^{-1/2} B B' \Sigma^{-1/2}$ associated with λ_1. Subsequently, if $Q_1 Q_1'$ is the spectral projector of $\Sigma_{11}^{-1/2} B_1 B_1' \Sigma_{11}^{-1/2}$ associated with λ_1, then

$$
\hat{Q}_1 \hat{Q}_1' = \begin{pmatrix} Q_1 \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} Q_1' & \mathbf{0} \end{pmatrix} = \begin{pmatrix} Q_1 Q_1' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}
$$

is the spectral projector of $\Sigma^{-1/2} B B' \Sigma^{-1/2}$ associated with λ_1. Then the expected revenue of the principal with augmented performance system $\mathcal{P}$ is

$$
\begin{aligned}
\Pi^*(\mathcal{P}) &= \frac{1}{8\lambda_1^2} \mathbb{E}_{\vartheta_1}\left[\frac{(\vartheta_1 + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1))^2}{\mathcal{H}(\vartheta_1) + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1)}\right] \beta' B' \Sigma^{-1/2} \hat{Q}_1 \hat{Q}_1' \Sigma^{-1/2} B \beta \\
&= \frac{1}{8\lambda_1^2} \mathbb{E}_{\vartheta_1}\left[\frac{(\vartheta_1 + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1))^2}{\mathcal{H}(\vartheta_1) + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1)}\right] \beta' B_1' \Sigma_{11}^{-1/2} Q_1 Q_1' \Sigma_{11}^{-1/2} B_1 \beta \\
&= \Pi^*(\mathcal{P}_1).
\end{aligned} \tag{A.27}
$$

---

[19] Note that the term $\mathbb{E}_{\vartheta_1}[\frac{(\vartheta_1+\mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1))^2}{\mathcal{H}(\vartheta_1)+\mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1)}]$ also depends on λ_1.

2. If $\lambda_1 = \mu_1$, then the first eigenvalue of $\Sigma^{-1/2}BB'\Sigma^{-1/2}$ is λ_1, but its multiplicity is now $k_1 + k_2$. Let Q_1 (Q_2) represent an $m_1 \times k_1$ ($m_1 \times k_1$) matrix whose columns are orthonormal eigenvectors of $\Sigma_{11}^{-1/2}B_1B_1'\Sigma_{11}^{-1/2}$ ($\Sigma_{22}^{-1/2}B_2B_2'\Sigma_{22}^{-1/2}$) associated with λ_1 (μ_1). Then the columns of matrix

$$\hat{Q} = \begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix} \in \mathbb{R}^{m_1+m_2}_{k_1+k_2}$$

form an orthonormal basis for eigenspace $\mathcal{N}(\Sigma^{-1/2}BB'\Sigma^{-1/2} - \lambda_1 I)$

$$\begin{aligned}
\Pi^*(\mathcal{P}) &= \frac{1}{8\lambda_1^2}\mathbb{E}_{\vartheta_1}\left[\frac{(\vartheta_1 + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1))^2}{\mathcal{H}(\vartheta_1) + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1)}\right]\beta' B' \Sigma^{-1/2}\hat{Q}\hat{Q}'\Sigma^{-1/2}B\beta \\
&= \frac{1}{8\lambda_1^2}\mathbb{E}_{\vartheta_1}\left[\frac{(\vartheta_1 + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1))^2}{\mathcal{H}(\vartheta_1) + \mathbb{E}_{\vartheta_2}(\vartheta_2|\vartheta_1)}\right]\beta'\big(B_1'\Sigma_{11}^{-1/2}Q_1Q_1'\Sigma_{11}^{-1/2}B_1 + B_2'\Sigma_{22}^{-1/2}Q_2Q_2'\Sigma_{22}^{-1/2}B_2\big)\beta \\
&= \Pi^*(\mathcal{P}_1) + \Pi^*(\mathcal{P}_2) \\
&> \Pi^*(\mathcal{P}_1).
\end{aligned}$$

3. The case $\lambda_1 < \mu_1$ is similar to $\lambda_1 > \mu_1$. Then the proof is thus omitted.

A.9. Lemmas

Lemma A.1. *Two real symmetric, positive-definite matrices can be simultaneously diagonalized.*

Proof. Let A and B be two real symmetric positive-definite matrices, $\{\tau_i\}$ be the set of generalized eigenvalues of A relative to B, that is, for some nonzero v_i, $Av_i = \tau_i B v_i$. B has a Cholesky decomposition, $B = T'T$, where T is an upper triangular matrix with positive diagonal elements. We can therefore get

$$\big(T'\big)^{-1}AT^{-1}u_i = \tau_i u_i,$$

where $u_i = Tv_i$. Note that since A is symmetric, $(T')^{-1}AT^{-1}$ is symmetric, and since τ_i is an eigenvalue of this matrix, it is real. Its associated eigenvector (with respect to $(T')^{-1}AT^{-1}$) is likewise real, so is the generalized eigenvector v_i. Since $(T')^{-1}AT^{-1}$ is symmetric, the ordinary eigenvectors can be chosen to be orthogonal. This implies that the generalized eigenvectors of A relative to B can be chosen to be B-orthogonal. Let $V = (v_1, \ldots, v_n)$ be a matrix having as columns the generalized B-normalized and B-orthogonal eigenvectors and $\Lambda = diag\{\tau_1, \ldots, \tau_n\}$ be a diagonal matrix consisting of generalized eigenvalues, then we have the final result: $V'AV = \Lambda$, $V'BV = I$.[20] □

Lemma A.2. *Let A and B be two $m \times m$ real symmetric matrices, then*

$$\lambda_m(A + B) \geqslant \lambda_m(A) + \lambda_m(B).$$[21]

Proof.

$$\lambda_m(A + B) = \min_{x \neq 0}\frac{x'(A + B)x}{x'x} \geqslant \min_{x \neq 0}\frac{x'Ax}{x'x} + \min_{x \neq 0}\frac{x'Bx}{x'x} = \lambda_m(A) + \lambda_m(B). \qquad \square$$

Lemma A.3. *Let A and B stand for $m \times m$ real symmetric positive definite matrices and C is their Jordan product: $C = AB + BA$. a_i, b_i, c_i denote the eigenvalues of A, B, and C in a descending enumeration respectively. Then we have*

$$c_m \geqslant \tilde{C}, \tag{A.28}$$

where

$$\begin{aligned}
\tilde{C} &= \max\left\{\min_{i=1,m} a_i b_m \frac{(\sqrt{k_a} + 1)^2 - k_b(\sqrt{k_a} - 1)^2}{2\sqrt{k_a}},\ \min_{i=1,m} b_i a_m \frac{(\sqrt{k_b} + 1)^2 - k_a(\sqrt{k_b} - 1)^2}{2\sqrt{k_b}}\right\} \\
&= \begin{cases}
a_m b_m \frac{(\sqrt{k_a}+1)^2 - k_b(\sqrt{k_a}-1)^2}{2\sqrt{k_a}} & \textit{if } \sqrt{k_b} \leqslant \frac{\sqrt{k_a}+1}{\sqrt{k_a}-1},\ k_b \geqslant k_a, \\
a_m b_m \frac{(\sqrt{k_b}+1)^2 - k_b(\sqrt{k_b}-1)^2}{2\sqrt{k_b}} & \textit{if } \sqrt{k_b} \leqslant \frac{\sqrt{k_a}+1}{\sqrt{k_a}-1},\ k_b < k_a, \\
a_1 b_m \frac{(\sqrt{k_a}+1)^2 - k_b(\sqrt{k_a}-1)^2}{2\sqrt{k_a}} & \textit{if } \sqrt{k_b} > \frac{\sqrt{k_a}+1}{\sqrt{k_a}-1},\ k_b \geqslant k_a, \\
a_m b_1 \frac{(\sqrt{k_b}+1)^2 - k_b(\sqrt{k_b}-1)^2}{2\sqrt{k_b}} & \textit{if } \sqrt{k_b} > \frac{\sqrt{k_a}+1}{\sqrt{k_a}-1},\ k_b < k_a,
\end{cases}
\end{aligned} \tag{A.29}$$

[20] Note that this is not necessarily an orthogonal diagonalization, and then this result does not extend to the case of three or more matrices.
[21] $\lambda_m(\cdot)$ stands for the least eigenvalue of a matrix.

where a_i, b_i and ρ_i are the i-th eigenvalues of A, B and their Jordan product C respectively in a descending enumeration. $k_a = \frac{a_1}{a_m}$ and $k_b = \frac{b_1}{b_m}$ denote the spectral condition numbers of A and B, respectively.

Proof. See Alikakos and Bates (1984) for detailed discussion. □

Lemma A.4. *$E(Y|X)$ is nonincreasing in X if and only if $Cov(X,Y) \leqslant 0$.*

Proof. Since $Cov(X,Y) = E(X)E(Y) - E(XY) = E(X)E[E(Y|X)] - E[E(XY|X)] = E(X)E[E(Y|X)] - E[XE(Y|X)] = Cov[X, E(Y|X)]$, $Cov(X,Y) \leqslant 0$ if and only if $E(Y|X)$ is a nonincreasing function of X.[22] □

Lemma A.5. *Let A, B be $m \times m$ symmetric matrices and $B > 0$. Then*

$$\max_{x \neq 0} \frac{x'Ax}{x'Bx} = \lambda_1\big(B^{-1/2}AB^{-1/2}\big)$$

and the optimal x satisfies: $B^{-1/2}x \in \mathcal{N}(B^{-1/2}AB^{-1/2} - \lambda_1 I)$.

Proof. Let $\frac{B^{1/2}x}{\sqrt{x'Bx}} = y$. Then

$$\max_{x \neq 0} \frac{x'Ax}{x'Bx} = \max_{\|y\|=1} y'B^{-1/2}AB^{-1/2}y.$$

Since $B^{-1/2}AB^{-1/2}$ is a symmetric matrix, there exists an orthogonal matrix P such that $P'B^{-1/2}AB^{-1/2}P = diag\{\lambda_1, \ldots, \lambda_m\}$. $\lambda_1, \ldots, \lambda_m$ are eigenvalues of $B^{-1/2}AB^{-1/2}$ in descending order; λ_1 has multiplicity k. Let $P'y = z$, then

$$\max_{\|y\|=1} y'B^{-1/2}AB^{-1/2}y = \max_{\|z\|=1} z' diag\{\lambda_1, \ldots, \lambda_m\} z = \max_{1 \leqslant j \leqslant m} \lambda_j.$$

The optimal solution to this problem is $z = (z_1, \ldots, z_k, 0, \ldots, 0)'$ with $\sum_{j=1}^k z_j^2 = 1$. We get $y = Pz = \sum_{i=1}^k z_i p_i$. p_i, $i = 1, \ldots, k$, are eigenvectors associated with $\lambda_1 = \cdots = \lambda_k$. Therefore $y \in \mathcal{N}(B^{-1/2}AB^{-1/2} - \lambda_1 I)$, which in turn implies that $B^{-1/2}x \in \mathcal{N}(B^{-1/2}AB^{-1/2} - \lambda_1 I)$. □

Lemma A.6. *The solution x^* and maximized value Π^* to program*

$$\max_x: \alpha'x, \quad s.t.: \quad \|x\| = a, \quad x \in \mathcal{N}(A - \lambda I) \tag{A.30}$$

are:

(i) *If $\alpha' Q_k Q_k' \alpha \neq 0$*

$$x^* = a \frac{Q_k Q_k' \alpha}{\sqrt{\alpha' Q_k Q_k' \alpha}},$$

$$\Pi^* = a\sqrt{\alpha' Q_k Q_k' \alpha}.$$

(ii) *If $\alpha' Q_k Q_k' \alpha = 0$*

x^* is an arbitrary element in $\mathcal{N}(A - \lambda I)$ with norm a,

$$\Pi^* = 0$$

where λ is an eigenvalue of symmetric matrix A with multiplicity k, $\mathcal{N}(A - \lambda I)$ represents the eigenspace of A associated with λ, $Q_k = (q_1, q_2, \ldots, q_k)$ are a set of orthonormal eigenvectors of A corresponding to λ.

Proof. Since A is a real symmetric matrix, there exists an orthogonal matrix $Q = (q_1, \ldots, q_n) = (Q_k, Q_{-k})$ such that

$$Q'AQ = diag\{\lambda, \ldots, \lambda, \lambda_{k+1}, \ldots, \lambda_n\}.$$

$Q_k = (q_1, q_2, \ldots, q_k)$ are a set of orthonormal eigenvectors associated with λ, $Q_{-k} = (q_{k+1}, \ldots, q_n)$ is the set of remaining orthonormal eigenvectors. Applying the spectral decomposition theorem in matrix algebra (see Lemma A.7), the spectral projector matrix $Q_k Q_k'$ is unique although Q_k is in general not unique:

$$(A - \lambda I)x = 0 \quad \Longleftrightarrow \quad Q\, diag\{0, \ldots, 0, \lambda - \lambda_{k+1}, \ldots, \lambda - \lambda_n\} Q' x = 0.$$

22 Here we use a result in probability theory: $Cov(\varphi_1(X), \varphi_2(X)) \leqslant 0$ iff $\varphi_1'(X)\varphi_2'(X) \leqslant 0$.

Letting $Q'x = y$, we get

$$diag\{0, \ldots, 0, \lambda - \lambda_{k+1}, \ldots, \lambda - \lambda_n\} y = 0,$$

which implies $y_i = 0$, $i = k+1, \ldots, n$. Then the program (A.30) can be rewritten as

$$\max_y : \alpha' Q y, \quad \text{s.t.:} \quad \|y\| = a, \quad y = (y_1, \ldots, y_k, 0, \ldots, 0).$$

- If $\alpha' Q_k Q_k' \alpha = 0$, then $Q_k' \alpha = 0$. Therefore, for any vector y with entries $y_{k+1} = \cdots = k_n = 0$, we have

$$\alpha' Q y = \left(0, \ldots, 0, \alpha' q_{k+1}, \alpha' q_n\right) \begin{bmatrix} y_1 \\ \vdots \\ y_k \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0.$$

  The solution to the program (A.30) is therefore an arbitrary vector in $\mathcal{N}(A - \lambda I)$ with norm a.
- If $\alpha' Q_k Q_k' \alpha \neq 0$, it is optimal to choose

$$y^* = \frac{a}{\sqrt{\alpha' Q_k Q_k' \alpha}} (Q_k, \mathbf{0})' \alpha,$$

  the corresponding optimal value is

$$\Pi^* = a\sqrt{\alpha' Q_k Q_k' \alpha}.$$

  The maxima for the original program is therefore

$$x^* = Q y^* = a \frac{Q_k Q_k' \alpha}{\sqrt{\alpha' Q_k Q_k' \alpha}}. \qquad \square$$

Lemma A.7 *(Uniqueness of spectral representation). A represents an $n \times n$ symmetric matrix, Q represents an $n \times n$ orthogonal matrix, $D = diag\{d_1, \ldots, d_n\}$ is an $n \times n$ diagonal matrix such that $Q'AQ = D$. (Note that every real symmetric matrix is orthogonally diagonalizable.) The i-th columns of Q are $\mathbf{q}_i$, $i = 1, \ldots, n$, respectively. $\lambda_1, \ldots, \lambda_k$ represent the distinct eigenvalues of A, $\nu_1, \ldots, \nu_k$ represent the multiplicities of $\lambda_1, \ldots, \lambda_k$, respectively. For $j = 1, \ldots, k$, $S_j = \{i\colon d_i = \lambda_j\}$ represents the set comprising the ν_j values of i such that $d_i = \lambda_j$. Then A can be expressed uniquely (aside from the ordering of the terms) as*

$$A = \sum_{j=1}^{k} \lambda_j E_j, \tag{A.31}$$

where (for $j = 1, \ldots, k$) $E_j = \sum_{i \in S_j} \mathbf{q}_i \mathbf{q}_i'$, $\mathbf{q}_i$, $i \in S_j$ are eigenvectors associated with λ_j.

Proof. Suppose that P is an $n \times n$ orthogonal matrix and $D^* = \{d_i\}$ an $n \times n$ diagonal matrix such that $P'AP = D^*$ (where P and D^* are possibly different from Q and D). Further, denote the first, n-th columns of P by $p_1, \ldots, p_n$, respectively, and (for $j = 1, \ldots, k$) let $S_j = \{i\colon d_i^* = \lambda_j\}$. Then, analogous to the decomposition $A = \sum_{j=1}^{k} \lambda_j E_j$, we have the decomposition

$$A = \sum_{j=1}^{k} \lambda_j F_j,$$

where (for $j = 1, \ldots, k$) $F_j = \sum_{j \in S_j^*} p_j p_j'$. Now, for $j = 1, \ldots, k$, let $Q_j = (\mathbf{q}_{i_1}, \ldots, \mathbf{q}_{i_{\nu_j}})$ and $P_j = (\mathbf{p}_{i_1^*}, \ldots, \mathbf{p}_{i_{\nu_j}^*})$, where $i_1, \ldots, i_{\nu_j}$ and $i_1^*, \ldots, i_{\nu_j}^*$ are the elements of S_j and S_j^*, respectively. Then, $\mathcal{C}(P_j) = \mathcal{N}(A - \lambda_j I) = \mathcal{C}(Q_j)$ (the symbol $\mathcal{C}(A)$ denotes the column space of a matrix A), so that $P_j = Q_j L_j$ for some $\nu_j \times \nu_j$ matrix L_j. Moreover, since clearly $Q_j' Q_j = I_{\nu_j}$ and $P_j' P_j = I_{\nu_j}$,

$$L_j' L_j = L_j' Q_j' Q_j L_j = P_j' P_j = I,$$

implying that L_j is an orthogonal matrix. Thus,

$$F_j = P_j P_j' = Q_j L_j L_j' Q_j' = Q_j I Q_j' = Q_j Q_j' = E_j.$$

We conclude that the decomposition $A = \sum_{j=1}^{k} \lambda_j F_j$ is identical to the decomposition $A = \sum_{j=1}^{k} \lambda_j E_j$, and hence that the decomposition $A = \sum_{j=1}^{k} \lambda_j E_j$ is unique (aside from the ordering of terms). $\square$

References

Alikakos, N., Bates, P., 1984. Estimates for the eigenvalues of the Jordan product of hermitian matrices. Linear Algebra Appl. 57, 41–56.

Armstrong, M., 1996. Multiproduct nonlinear pricing. Econometrica 64, 51–75.

Baiman, S., 1990. Agency research in managerial accounting, a second look. Account. Organ. Soc. 15, 341–371.

Baker, G., 1992. Incentive contracts and performance measurement. J. Polit. Economy 100, 598–614.

Baker, G.P., 2002. Distortion and risk in optimal incentive contracts. J. Human Res. 37, 728–751.

Basov, S., 2001. Hamiltonian approach to multi-dimensional screening. J. Math. Econ. 36, 77–94.

Basov, S., Bardsley, P., 2005. A general model of coexisting hidden action and hidden information. Working paper No. 958, Department of Economics, University of Melbourne. http://www.economics.unimelb.edu.au/downloads/wpapers-05/958.pdf.

Basov, S., Danilkina, S., 2010. Multitasking, multidimensional screening, and moral hazard with risk neutral agents. Econ. Rec. 86 (Special Issue), 80–86.

Bernoulli, D., 1954. Exposition of a new theory of the measurement of risk. Econometrica 22, 23–36; (1758).

Caillaud, B., Guesnerie, R., Rey, P., 1992. Noisy observation in adverse selection models. Rev. Econ. Stud. 59, 595–615.

Feltham, G.A., Xie, J., 1994. Performance measure congruity and diversity in multi-task principal–agent relations. Acc. Rev. 69, 429–453.

Green, J., Stokey, N., 1983. A comparison of tournaments and contracts. J. Polit. Economy 91, 349–364.

Guesnerie, R., 1981. On taxation and incentives, further reflections on the limits of redistribution. Discussion paper No. 89, Sonderforschungsbereich 21, University of Bonn.

Guesnerie, R., Picard, P., Rey, P., 1989. Adverse selection and moral hazard with risk neutral agents. Europ. Econ. Rev. 33, 807–823.

Hammond, P., 1979. Straightforward individual incentive compatibility in large economies. Rev. Econ. Stud. 46, 263–282.

Holmstrom, B., Milgrom, P., 1991. Multitask principal–agent analyses. J. Law, Econ., Organ. 7, 24–52.

Hurwicz, L., 1972. On informationally decentralized systems. In: Radner, R., McGuire, C.B. (Eds.), Decision and Organization, Volume in Honor of J. Marschak. North-Holland, pp. 297–336.

Hurwicz, L., 1979. On the dimension requirements of informationally decentralized Pareto-satisfactory processes. In: Arrow, K., Hurwicz, L. (Eds.), Studies in Resource Allocation Processes. Cambridge University Press.

Hurwicz, L., 1986. On informational decentralization and efficiency in resource allocation mechanism. In: Reiter, S. (Ed.), Studies in Mathematical Economics. Mathematical Association of America.

Laffont, J., Tirole, J., 1986. Using cost observation to regulate firms. J. Polit. Economy 94, 614–641.

Laffont, J.J., Maskin, E., Rochet, J.C., 1987. Optimal nonlinear pricing with two-dimensional characteristics. In: Groves, T., Radner, R., Reiter, S. (Eds.), Information, Incentives, and Economic Mechanisms. University of Minnesota Press, Minneapolis, pp. 256–266.

Lazear, E., Rosen, S., 1981. Rank-order tournaments as optimum labor contracts. J. Polit. Economy 89, 841–864.

Melumad, N., Reichelstein, N., 1989. Value of communication in agencies. J. Econ. Theory 47, 334–368.

Mount, K., Reiter, S., 1974. Informational size of message spaces. J. Econ. Theory 8, 161–191.

Nalebuff, B., Stiglitz, J., 1983. Prices and incentives, towards a general theory of compensation and competition. Bell J. Econ. 14, 21–43.

Osana, H., 1978. On the informational size of message spaces for resource allocation process. J. Econ. Theory 17, 66–78.

Osterman, P., 1992. Internal labor markets in a changing environment, models and evidence. In: Lewin, D., Mitchell, O.S., Sherer, P.D. (Eds.), Research Frontiers in Industrial Relations and Human Resources. Industrial Relations Research Association, Madison, WI, pp. 273–308.

Picard, P., 1987. On the design of incentive schemes under moral hazard and adverse selection. J. Public Econ. 33, 305–332.

Rochet, J.-C., 1985. The taxation principle and multi-time Hamilton–Jacobi equations. J. Math. Econ. 14, 113–128.

Rochet, J.-C., 1987. A necessary and sufficient condition for rationalizability in a quasi-linear context. J. Math. Econ. 16, 190–200.

Rochet, J.-C., Choné, P., 1998. Ironing, sweeping, and multidimensional screening. Econometrica 66, 783–826.

Thiele, V., 2010. Task-specific abilities in multi-task principal–agent relationships. Lab. Econ. 174, 690–698.

Tian, G., 1994. On informational efficiency and incentive aspects of generalized ratio equilibria. J. Math. Econ. 23, 323–337.

Tian, G., 2004. A unique informationally efficient allocation mechanism in economies with consumption externalities. Int. Econ. Rev. 45, 79–111.

Tian, G., 2006. The unique informational efficiency of the competitive mechanism in economies with production. Soc. Choice Welfare 26, 155–182.

Walker, M., 1977. On the informational size of message spaces. J. Econ. Theory 15, 366–375.

Williamson, O., 1985. The Economic Institutions of Capitalism, Firms, Markets, Relational Contracting. Free Press, New York.

Wilson, R., 1993. Nonlinear Pricing. Oxford University Press, New York.