Voting over Selfishly Optimal Income Tax Schedules with Tax-Driven Migrations*

Darong Dai† Guoqiang Tian‡

This Version: July, 2018

Abstract

This paper studies majority voting over selfishly optimal nonlinear income tax schedules proposed by a continuum of workers who can migrate between two competing jurisdictions. Both skill and migration cost are the private information of each worker. Assuming quasilinear-in-consumption preferences, the tax schedule proposed by the median skill type is the Condorcet winner that redistributes resources from the rich and poor toward the middle. While it features negative marginal tax rates for low skills, it features positive marginal tax rates for high skills with small migration elasticities. We establish the skill-dependent threshold of migration elasticity for all types. If their migration elasticities are higher than their respective thresholds, then migration induces lower marginal tax rates than does autarky; otherwise, migration induces higher marginal tax rates for the jurisdiction facing net labor inflow in low skills while net labor outflow in high skills.

Keywords: Migration; Redistributive taxation; Nonlinear income tax; Majority voting; Median voter; Tax competition.

JEL classification codes: D72; D82; H21; J61.

1 Introduction

As barriers to labor mobility have been lowered and education and language skills have improved, governments are facing the challenge that the base of labor income tax is becoming more mobile. This is especially true for highly skilled workers. For example, Kleven et al. (2013), Kleven et al. (2014) and Akcigit et al. (2016) estimate large migration elasticities with respect to income tax rate for these types of workers.

*We thank the helpful comments and suggestions from Yonghong An, Pedro Bento, Laurent Bouton, Klaus Desmet, Andrew T. Foerster, Andrew Glover, Dennis W. Jansen, Quan Li, Jason Lindo, Liqun Liu, Chen-Yu Pan, Tatevik Sekhposyan, Sang-Chul Suh, John Weymark, Kei-Mu Yi, Yuzhe Zhang, Sarah Zubairy and the participants of Macro Research Group and Macro Student Conference (October 2017) at Texas A&M, the Midwest Economic Theory Conference at SMU (November 2017), the 14th Meeting of the Society for Social Choice and Welfare at Seoul (June 2018), and the 2018 Asian Meeting of the Econometric Society at Seoul. The usual disclaimer certainly applies.

†Institute for Advanced Research (IAR), Shanghai University of Finance and Economics, Shanghai 200433, P. R. China. E-mail: dai.darong@mail.shufe.edu.cn

‡Department of Economics, Texas A&M University, College Station, TX 77843, USA. E-mail: gtian@tamu.edu

To analyze how the possibility of geographic mobility affects the design of redistributive taxation, the literature (see, e.g., Mirrlees, 1982; Simula and Trannoy, 2010, 2012; Piketty and Saez, 2013; Lehmann et al., 2014) that builds on the seminal work of Mirrlees (1971) focuses on the normative perspective. Little attention has been paid to the positive perspective. We hence address this question: how would the schedule of redistributive taxation look like when workers can vote both in the ballot box and with their feet?[1] In particular, answering this question allows us to reexamine the conventional wisdom claiming that geographic mobility limits the ability of government to redistribute incomes via a tax-transfer system (see Stigler, 1957).

To our knowledge, the answer is not yet well established. Indeed, the literature either assumes away asymmetric information (e.g., Cremer and Pestieau, 1998; Hindriks, 2001), restricts attention to flat tax (e.g., Hindriks, 2001) and special connections between skills and migration costs in a two-type setting that rules out countervailing incentives (e.g., Hamilton and Pestieau, 2005), or focuses on probabilistic voting in a representative democracy (e.g., Brett, 2016). These simplifications make it possible to obtain sharp predictions, whereas reasonable doubts about the generality and robustness of their predictions may arise.

Intuitively, the combination of labor mobility and majority voting results in a complex interaction whereby the taxation policies chosen by competing jurisdictions determine whom they attract and whom they attract determine their choices of taxation policies.[2] Our goal is to answer the above question with taking into account this interesting interaction and without resorting to those simplifications used by the literature.

To this end, we consider an economy consisting of two jurisdictions, between which workers born in each one can move by paying certain amount of migration cost. In each jurisdiction, workers differ in both skill level that measures their labor productivity and migration cost that measures their foot-voting capability. While the ex ante distribution of skill levels and the conditional distribution of migration costs are common knowledge, the values of any worker's labor productivity and migration cost are only known to herself. As usual, taxation is based on the residence principle[3], and we focus on the tax competition induced by the mobile tax base between these two jurisdictions. Taking as given the income taxes implemented in both jurisdictions, workers make individual decisions along two margins: optimal labor supply on the intensive margin and optimal residence choice on the extensive margin. In particular, by allowing for location choice, both the reservation utilities of workers and the ex post skill distribution are endogenously determined. This feature, on one hand, enables us to design more realistic income tax schedules. On the other hand, it makes the qualitative characterization much more challenging.

Following Röell (2012), Bohn and Stuart (2013), and Brett and Weymark (2016, 2017, 2017b), we are

[1] Morelli et al. (2012) also consider the case with mobility and voting, but income taxation is optimally determined in their model and only the constitutional choice is determined by majority voting. They still focus on the normative approach in taxation design.

[2] A real example featured by such a combination is Switzerland that consists of 26 cantons. Each canton determines its own income tax rate via direct democracy, and also there are lots of inter-canton migrations.

[3] In practice, almost all countries use the residence principle. The only exceptions are the US and Israel, where the citizens pay domestic income tax based on their global income.

interested in selfishly optimal income tax schedules. Each worker can be viewed as a citizen candidate who can propose an income tax schedule that maximizes the utility of her own type. Then, the pairwise majority rule is used to select the tax schedule that is going to be implemented in equilibrium.[4] Importantly, each worker proposes an income tax schedule as if she were representing the government. Following the mechanism design approach, each proposer must take into account individual responses along both intensive and extensive margins and design incentive-compatible allocations satisfying the government budget constraint.

We rely on the first-order approach in the text and leave the complete solution to the tax design problem to Appendix B due to its complexity. By assuming quasilinear-in-consumption preferences, we show that the tax schedule proposed by the median skill type is a Condorcet winner in the majority rule equilibrium, which provides support for the empirical finding of Corneo and Neher (2015). They show by using survey data that most democracies implement the preferred redistribution of the median voter and also the probability to serve the median voter increases with the quality of democracy. Moreover, the tax rates implemented in the experiment designed by Agranov and Palfrey (2015) closely track the preferences of the median skill worker, and the cross-national empirical evidence of Gründler and Köllner (2017) emphasizes the political channel as well as the middle class in determining the extent of redistribution.

The current tax schedule exhibits the following characteristics.

First, it coincides with the maximax tax schedule for types below the median skill level and coincides with the maximin tax schedule for types above the median skill level. Thus governments under direct democracy and majority rule tend to redistribute from both the poor and the rich toward the middle class. This prediction[5] extends the Director's Law (see Stigler, 1970) to the circumstance with both labor mobility and inter-jurisdictional tax competition.

Second, marginal tax rates are negative for low incomes, but for high incomes there is an endogenously determined and skill-dependent threshold of migration elasticity such that they are negative only when migration elasticity is above this threshold; otherwise, they are nonnegative. Provided that income taxation in the United States is based on citizenship other than residence principle, the migration elasticities of high incomes may not be that large, this prediction hence explains in some sense why effective marginal tax rates in the United States are negative for low incomes and positive for high incomes (see Congressional Budget Office, 2012).

Third, it creates three potential discontinuities, one at the skill level of the proposer and the other two at the endpoints of the ex post skill distribution, in the resulting income schedule. In the voting equilibrium, there always exists a downward discontinuity at the median skill level. Importantly, we identify the indirect utility level of the least skilled as the determinant factor of the type of the other two discontinuities. If it lies between a negative threshold and zero, then both of them are downward.

[4] In a population with the majority consisting of "poor" individuals, Höchtl et al. (2012) experimentally find that redistribution outcomes look as if all voters were exclusively motivated by self-interest. We hence argue that it is somewhat reasonable to focus attention on selfishly optimal income taxes in the current political economy.

[5] It also theoretically supports the empirical finding of Jacobs et al. (2017) that all Dutch political parties give a higher political weight to middle incomes than to the poor and the rich.

If it lies above zero, then one is upward at the lowest skill level and the other is downward at the highest skill level. If, however, it lies below the negative threshold, then one is downward at the lowest skill level and the other is upward at the highest skill level. We hence need to build more than one bridges when ironing the tax schedule to satisfy the monotonicity constraint placed on incomes. As such, truth-telling allocation calls for bunching not only in the middle-income class, but also either in the low-income class, or in the high-income class, or in both.

Fourth, by allowing for inter-jurisdictional migrations that endogenize the ex post skill distribution and median skill level, the resulting level of redistribution deviates from that under autarky. In a comparison with the autarky equilibrium of Brett and Weymark (2017), we obtain the following results. If migration elasticities are large for both low and high skills, then migrations induce lower marginal tax rates and a lower level of redistribution than does autarky. If, however, their migration elasticities are small, then the jurisdiction facing net labor inflow in low skills while net labor outflow in high skills imposes higher tax rates on both low and high skills. So, high skills pay more taxes, low skills receive less transfers, and the median skill receives more transfers than in autarky. Also, if the median skill level is different between migration equilibrium and autarky equilibrium, then these predictions hold for all but those between these two median skill levels. The direction of change depends on identifying which median voter to be the richer (or poorer).

To complement the theoretical analysis, the model is simulated with empirical parameter estimates based on the U.S. data. Since the literature suggests that tax-driven migrations be more likely to occur in the population of high skills (e.g., Doquier and Marfouk, 2006) and the upper part of empirical income distribution be well fitted by Pareto distribution (e.g., Saez, 2001), our numerical experiments focus on top-income workers. Following Atkinson et al. (2011) and Diamond and Saez (2011), we set the ex post Pareto index to be 1.5. For the ex ante Pareto index in autarky (without migrations), we allow it to have varying values around 1.5. The ex post and ex ante degrees of income inequality are not necessarily the same. To do counterfactual simulations, we manipulate the value of another parameter, which is an approximation of the elasticity of utility with respect to pre-tax income, in our tax formula such that it generates the realistic top marginal tax rate of 42.5% (see Diamond and Saez, 2011). Our calculation reveals that there can be large differences of marginal tax rate between our prediction and that in autarky. For example, for a labor supply elasticity of 0.25 (e.g., Saez et al., 2012), a migration elasticity of 0.25 (e.g., Lehmann et al., 2014) and an ex ante Pareto index of 1.5, the autarky equilibrium generates a top tax rate 34.4% higher than 42.5%. Everything else equal, if the labor supply elasticity increases to the highest possible estimate 0.4 (e.g., Saez et al., 2012), the autarky equilibrium still generates a top tax rate 27.5% higher than 42.5%. Naturally, these numbers of differences should only be interpreted within the current economic context and be taken as providing a rough assessment of the quantitative significance of migrations in terms of reducing equilibrium distortions imposed on top-income workers.

The rest of the paper is organized as follows. Section 2 discusses the relation of our paper to the literature. Section 3 sets up the model of the economy. Section 4 derives and characterizes selfishly optimal nonlinear income tax schedules. Section 5 establishes the voting equilibrium. Section 6 qualitatively and quantitatively identifies the effect of migrations on the equilibrium marginal tax rates. Section 7 concludes. Proofs are relegated to Appendix A.

2 Literature Review

The current study relates to two strands of literature and we shall discuss one by one.

First, it relates to the literature studying selfishly optimal nonlinear taxation determined by the majority rule,[6] such as Röell (2012), Bohn and Stuart (2013), and Brett and Weymark (2016, 2017, 2017b). While Röell (2012) and Brett and Weymark (2017b) consider the case with discrete skill levels, the others use a continuum version of the Röell model. They all but Brett and Weymark (2017) impose a minimum-utility constraint, but they all demonstrate that the schedule proposed by the median skill type is a Condorcet winner. By sacrificing the minimum-utility constraint and assuming quasilinear-in-consumption preferences, Brett and Weymark (2017) realize a complete characterization of selfishly optimal nonlinear tax schedules without calling for quite technical analysis.

By adding the location choice for workers, both the reservation utility of the standard participation constraint and the ex post skill distribution are endogenously determined. So, our extension makes the underlying tax design issue more realistic and indeed modifies the predictions obtained by Brett and Weymark (2017) in three main ways.

Firstly, there tend to be more than one discontinuities in the tax schedule, and hence the ironing procedure used in deriving the complete solution is much more involved than theirs. Secondly, marginal tax rates for high incomes are not definitely positive as suggested by Brett and Weymark, and they are rather negative for large migration elasticities. And thirdly, migrations reduce marginal tax rates under a reasonable range of migration elasticities, and also the level of redistribution is likely to be either lower or higher than suggested by Brett and Weymark, partially depending on whether the ex post median skill level lies at the right or at the left of the ex ante median skill level.

Second, it relates to the literature analyzing how the change of skill distribution affects equilibrium tax rates as well as the level of redistribution, such as Leite-Monteiro (1997), Hamilton and Pestieau (2005), Brett and Weymark (2011), and Lehmann et al. (2014). Regardless of whether they use discrete or continuous skill distributions, they all follow the normative approach and focus on optimal income taxation. This paper follows the positive approach and stresses the fact that tax schedules are, directly or indirectly, chosen by self-interested voters in democracies. In this sense, it could be regarded as a complementary work to the literature.

[6] Voting over selfishly optimal tax schedules has also been studied by Meltzer and Richard (1981) and Hindriks and De Donder (2003), but the former study focuses on linear taxes and the latter study focuses on quadratic tax schedules.

3 The Model

We consider an economy consisting of two jurisdictions, called A and B and not necessarily symmetric. Both adopt direct democracy in determining its redistributive taxation schedule. The measure of workers in A is normalized to 1, while that of B is denoted by n_-, for $0 < n_- \leq 1$. In what follows, we will focus on A because similar assumptions hold for B. To save on notation, whenever needed, we will use the subscript "$_-$" to indicate variables associated to B. Each worker is characterized by three characteristics: her native jurisdiction A or B, her skill (or labor productivity) $w \in [\underline{w}, \overline{w}]$ with $0 < \underline{w} < \overline{w}$, and the migration cost $m \in \mathbb{R}^+$ she supports if deciding to relocate. In particular, if she faces an infinitely large migration cost, then she is immobile. Following Lehmann et al. (2014), we do not make any restriction on the correlation between skills and migration costs.[7]

The ex ante skill density function, $f(w) = F'(w) > 0$, is assumed to be differentiable for all $w \in [\underline{w}, \overline{w}]$. For each skill w, $g(m|w)$ denotes the conditional density of the migration cost and $G(m|w) = \int_0^m g(x|w)dx$ the conditional cumulative distribution function. The joint density of (m, w) is thus $g(m|w)f(w)$, and $G(m|w)f(w)$ is the mass of workers of skill w with migration costs lower than m.

Following Mirrlees (1971) and Lehmann et al. (2014), governments cannot observe workers' type (w, m) and can only condition transfers on earnings y via an income tax function, $T(\cdot)$. As usual, taxes are levied according to the residence principle. So, migration threat actually induces tax competition between these two jurisdictions, and we are in line with Lehmann et al. (2014) to consider the tax competition wherein each government takes the income tax policy of the opponent as given.

A worker with skill level w produces w units of a consumption good per unit of labor time in a perfectly competitive labor market and earns a before-tax income of

$$y = wl, \tag{1}$$

in which $l \geq 0$ denotes labor supply. A worker has nonnegative consumption c that is also her after-tax income, namely

$$c = y - T(y). \tag{2}$$

Preferences over consumption and labor supply are represented by the quasilinear-in-consumption utility function[8]

$$\tilde{u}(c, l; m) = c - h(l) - \mathbb{I} \cdot m,$$

which is common to all workers with $\mathbb{I}$ being equal to 1 if she decides to migrate and to 0 otherwise. The disutility function h is increasing, strictly convex and three-times continuously differentiable, and also satisfies the usual normalization $h(0) = h'(0) = 0$. The government can observe a worker's before-

[7] The simpler assumption is that migration costs and skill levels are independently distributed, as adopted by Bierbrauer et al. (2013) and Blumkin et al. (2015). It, nevertheless, seems to be inconsistent with the empirical finding of Doquier and Marfouk (2006) that migration rates increase in skill.

[8] This assumption not only simplifies the theoretical derivation but also seems to be empirically reasonable by eliminating the income effect on taxable income (e.g., Gruber and Saez, 2002).

and after-tax incomes, but not her labor supply. Using (1), the utility function in terms of observable variables is written as

$$u(c, y; w, m) = c - h\left(\frac{y}{w}\right) - \mathbb{I} \cdot m. \tag{3}$$

It is easy to verify that the standard single-crossing property is satisfied.

So, the individual choice of each worker is along two margins: the intensive margin on optimal labor supply and the extensive margin on optimal residence choice.

3.1 Intensive Margin

If a worker decides to stay in jurisdiction A, then she maximizes (3) subject to $\mathbb{I} = 0$ and (2), yielding the first-order condition:

$$T'(y(w)) = 1 - \frac{1}{w} h'\left(\frac{y(w)}{w}\right),$$

whenever T is differentiable. If it is not differentiable at some incomes, then the marginal tax rate is not well-defined. To avoid unnecessary technical issues, we follow Brett and Weymark (2017) and directly define the function of marginal tax rate as

$$\tau(w) \equiv 1 - \frac{1}{w} h'\left(\frac{y(w)}{w}\right), \quad \forall w \in [\underline{w}, \overline{w}]. \tag{4}$$

That is, marginal tax rate is equal to one minus the marginal rate of substitution between consumption and income.

We then define the resulting indirect utility as

$$U(w) \equiv c(w) - h\left(\frac{y(w)}{w}\right), \quad \forall w \in [\underline{w}, \overline{w}]. \tag{5}$$

Incentive compatibility requires that

$$U(w) = \max_{w' \in [\underline{w}, \overline{w}]} c(w') - h\left(\frac{y(w')}{w}\right), \quad \forall w \in [\underline{w}, \overline{w}].$$

The necessary condition is thus

$$U'(w) = h'\left(\frac{y(w)}{w}\right) \frac{y(w)}{w^2}, \quad \forall w \in [\underline{w}, \overline{w}], \tag{6}$$

which gives the first-order incentive compatibility (FOIC) condition. Sufficiency is guaranteed by the second-order incentive compatibility (SOIC) condition

$$y'(w) \geq 0, \quad \forall w \in [\underline{w}, \overline{w}]. \tag{7}$$

If constraint (7) does not bind, then the first-order approach is appropriate.

3.2 Migration Decision

For a worker of type (w, m) born in jurisdiction A, she will migrate to jurisdiction B if and only if $m < U_{-}(w) - U(w)$. As in Lehmann et al. (2014), after combining the migration decisions made by workers born in both jurisdictions, the density of residents of skill w in jurisdiction A can be written as:

$$\phi(\Delta(w); w) \equiv \begin{cases} f(w) + G_{-}(\Delta(w)|w) f_{-}(w) n_{-} & \text{for } \Delta(w) \geq 0, \\ (1 - G(-\Delta(w)|w)) f(w) & \text{for } \Delta(w) \leq 0 \end{cases} \tag{8}$$

with $\Delta(w) \equiv U(w) - U_{-}(w)$. To ensure that $\phi(\cdot; w)$ is differentiable, we impose the technical restriction that $g(0|w)f(w) = g_{-}(0|w)f_{-}(w)n_{-}$, which is verified when these two jurisdictions are symmetric or when there is a fixed cost of migration, namely $g(0|w) = g_{-}(0|w) = 0$. We can then, as in Lehmann et al. (2014), define the elasticity of migration as

$$\theta(\Delta(w); w) \equiv \frac{\partial \phi(\Delta(w); w)}{\partial \Delta(w)} \frac{c(w)}{\phi(\Delta(w); w)} \tag{9}$$

To save on notations, we let $\tilde{f}(w) \equiv \phi(\Delta(w); w)$ and $\tilde{\theta}(w) \equiv \theta(\Delta(w); w)$.

4 Selfishly Optimal Nonlinear Income Tax Schedules

To focus on redistributive taxation, the government budget constraint can be written as

$$\int_{\underline{w}}^{\overline{w}} [y(w) - c(w)] \tilde{f}(w) dw \geq 0, \tag{10}$$

where we have used (2). Provided the quasi-linearity in consumption, (10) must be binding. In particular, the individual participation constraint has been incorporated into this fiscal budget constraint through the term of ex post skill density $\tilde{f}$.

As in Brett and Weymark (2016, 2017), each worker can propose an income tax schedule satisfying incentive compatibility constraints (6)-(7) and the government budget constraint (10), and then pairwise majority rule is used to determine which of these schedules shall be implemented. That is, each worker can be seen as a citizen candidate who may be elected as the representative agent of the government.

By applying the Taxation Principle[9] (see Hammond, 1979; Guesnerie, 1995) that allows us to restrict attention to simple direct mechanisms[10], for a worker of type $k \in [\underline{w}, \overline{w}]$, proposing an income tax schedule is equivalent to proposing an allocation schedule $\{c(w), y(w)\}_{w \in [\underline{w}, \overline{w}]}$ which solves the maximization problem

$$\max_{\{c(w), y(w)\}_{w \in [\underline{w}, \overline{w}]}} U(k) \quad \text{subject to } (5), (6), (7) \text{ and } (10), \tag{11}$$

[9]It states that there is an equivalence between admissible allocations and allocations that are decentralizable via an income tax system.

[10]If individual skills are drawn independently, Bierbrauer (2011) proves that the optimal sophisticated mechanism with strategic interdependence is a simple mechanism as long as individuals exhibit decreasing risk aversion.

taking as given the allocation schedule in the opponent jurisdiction. By (11), the resulting allocation schedule is indeed *selfishly optimal* for the proposer.

We leave the complete solution to Appendix B, and here we rely on the first-order approach that considers a simpler while still useful case in which the SOIC condition (7) is ignored. Formally, problem (11) is relaxed as

$$\max_{\{c(w),y(w)\}_{w\in[\underline{w},\overline{w}]}} U(k) \quad \text{subject to } (5), (6) \text{ and } (10). \tag{12}$$

Solving problem (12) leads to the following theorem.

Theorem 4.1 *The selfishly optimal schedule of pre-tax incomes proposed by any worker of type $k \in (\underline{w}, \overline{w})$ is given by*

$$y(w) = \begin{cases} \underline{y}^M(w) & \text{for } w = \underline{w}, \\ y^M(w) & \text{for } w \in (\underline{w}, k), \\ y^R(w) & \text{for } w \in (k, \overline{w}), \\ \overline{y}^R(w) & \text{for } w = \overline{w}. \end{cases} \tag{13}$$

Proof. See Appendix A. ■

Since we focus on selfishly optimal income tax schedules, a proposer of type k wishes to redistribute incomes from all other types toward her own type. To this end, for types greater than her own, she optimally proposes the maximin income schedule, denoted by $y^R(\cdot)$, whereas for types smaller than her own, she optimally proposes the maximax income schedule, denoted by $y^M(\cdot)$.

By applying the formula (4) to Lemmas 7.2 and 7.3 stated in Appendix A and using Theorem 4.1, we summarize the resulting prediction as the second theorem.

Theorem 4.2 *The selfishly optimal income tax schedule proposed by any worker of type $k \in (\underline{w}, \overline{w})$ is given by*

$$\tau(w) = \begin{cases} \underline{\tau}^M(w) & \text{for } w = \underline{w}, \\ \tau^M(w) & \text{for } w \in (\underline{w}, k), \\ \tau^R(w) & \text{for } w \in (k, \overline{w}), \\ \overline{\tau}^R(w) & \text{for } w = \overline{w} \end{cases} \tag{14}$$

in which these marginal tax rates are explicitly given as

$$\underline{\tau}^M(\underline{w}) = -\frac{\partial \tilde{f}(\underline{w})}{\partial y(\underline{w})} \frac{1}{\tilde{f}(\underline{w})} \left[y(\underline{w}) - h\left(\frac{y(\underline{w})}{\underline{w}}\right) + \frac{y(\underline{w})}{\underline{w}} h'\left(\frac{y(\underline{w})}{\underline{w}}\right) - U(\underline{w}) \right], \tag{15}$$

$$\begin{aligned} \tau^M(w) = {} & \frac{\Gamma(w,\overline{w}) - \Gamma(\underline{w},\overline{w})}{w\tilde{f}(w)} \left[\frac{1}{w} h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^2} h''\left(\frac{y(w)}{w}\right) \right] \\ & - \frac{\partial \tilde{f}(w)}{\partial y(w)} \frac{1}{\tilde{f}(w)} \left[y(w) - h\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w} h'\left(\frac{y(w)}{w}\right) \right], \end{aligned} \tag{16}$$

$$\tau^R(w) = \frac{\Gamma(w,\overline{w})}{w\tilde{f}(w)}\left[\frac{1}{w}h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^2}h''\left(\frac{y(w)}{w}\right)\right] - \frac{\partial \tilde{f}(w)}{\partial y(w)}\frac{1}{\tilde{f}(w)}\left[y(w) - h\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w}h'\left(\frac{y(w)}{w}\right)\right] \tag{17}$$

and

$$\overline{\tau}^R(\overline{w}) = -\frac{\partial \tilde{f}(\overline{w})}{\partial y(\overline{w})}\frac{1}{\tilde{f}(\overline{w})}\left[y(\overline{w}) - h\left(\frac{y(\overline{w})}{\overline{w}}\right) - U(\underline{w})\right], \tag{18}$$

with $\partial \tilde{f}(w)/\partial y(w)$ determined by equations (39) and (40) stated in Appendix A for any $w \in [\underline{w}, \overline{w}]$.

These marginal tax rates depart from those obtained by Brett and Weymark (2017) in two important aspects. First, in addition to the discontinuity between the maximax tax schedule and the maximin tax schedule, we show that there may exist another two discontinuities: one at the lowest type within the maximax schedule and the other at the highest type within the maximin schedule. Second, as the ex post skill distribution is endogenously determined as a function of income and consumption, the migration decision along the extensive margin imposes non-trivial effects on these tax rates.

By using Theorem 4.2, we obtain the following two propositions.

Proposition 4.1 *Regarding the sign of these marginal tax rates given by (15)-(16), we have the following predictions.*

(i) For workers of type $\underline{w}$, we have:

(i-a) *There is a threshold, written as $\widehat{T}^M(y(\underline{w})) \equiv -\frac{y(\underline{w})}{\underline{w}}h'\left(\frac{y(\underline{w})}{\underline{w}}\right) = -\underline{w}U'(\underline{w}) < 0$, of tax liability such that*

$$\underline{\tau}^M(\underline{w}) \begin{cases} < 0 & \text{for } T^M(y(\underline{w})) > \widehat{T}^M(y(\underline{w})), \\ = 0 & \text{for } T^M(y(\underline{w})) = \widehat{T}^M(y(\underline{w})), \\ > 0 & \text{for } T^M(y(\underline{w})) < \widehat{T}^M(y(\underline{w})); \end{cases} \tag{19}$$

(i-b) *Suppose $\tilde{\theta}(\underline{w}) \cdot MU_y(\underline{w}) \cdot Q_{y,c}(\underline{w}) \neq 1$, in which MU_y and $Q_{y,c}$ denote, respectively, the marginal utility of pre-tax income and the ratio of pre-tax income to after-tax income.*

- *If $0 < \tilde{\theta}(\underline{w}) \cdot MU_y(\underline{w}) \cdot Q_{y,c}(\underline{w}) < 1$, then*

$$\underline{\tau}^M(\underline{w}) \begin{cases} < 0 & \text{for } ATR^M(\underline{w}) > -1, \\ = 0 & \text{for } ATR^M(\underline{w}) = -1, \\ > 0 & \text{for } ATR^M(\underline{w}) < -1; \end{cases} \tag{20}$$

- *If* $\tilde{\theta}(\underline{w}) \cdot MU_y(\underline{w}) \cdot Q_{y,c}(\underline{w}) > 1$, *then*

$$\underline{\tau}^M(\underline{w}) \begin{cases} < 0 & \text{for } ATR^M(\underline{w}) < -1, \\ = 0 & \text{for } ATR^M(\underline{w}) = -1, \\ > 0 & \text{for } ATR^M(\underline{w}) > -1; \end{cases} \tag{21}$$

in which ATR denotes average tax rate.

(ii) For workers of type $w \in (\underline{w}, k)$, *we have* $\tau^M(w) < 0$.

Proof. See Appendix A. ∎

The sign of these maximax marginal tax rates departs from that of Brett and Weymark (2017) as follows. Instead of showing that the marginal tax rate is always zero for the lowest skilled, we show that it is zero only when their tax liability is equal to a negative critical value or when their average tax rate is equal to -1. That is, it is more likely to be either strictly positive or strictly negative. In particular, if the elasticity of migration for the lowest skilled is small, it is negative when their average tax rate is relatively large and is positive when their average tax rate is relatively small, as shown in (20).

Proposition 4.2 *Regarding the sign of these marginal tax rates given by (17)-(18), we have the following predictions.*

(i) For workers of type $w \in (k, \overline{w})$, *there is a threshold of the elasticity of migration, written as* $\tilde{\theta}_\star(w) \equiv \frac{\Gamma(w,\overline{w})}{w\tilde{f}(w)} \left[\frac{c(w)}{wl(w)+l(w)h'(l(w))-h(l(w))} \right] > 0$, *such that*

$$\tau^R(w) \begin{cases} < 0 & \text{for } \tilde{\theta}(w) > \tilde{\theta}_\star(w), \\ = 0 & \text{for } \tilde{\theta}(w) = \tilde{\theta}_\star(w), \\ > 0 & \text{for } \tilde{\theta}(w) < \tilde{\theta}_\star(w). \end{cases} \tag{22}$$

(ii) For workers of type $\overline{w}$, *we have*

$$\tau^R(\overline{w}) \begin{cases} < 0 & \text{for } T^R(y(\overline{w})) > U(\underline{w}) - U(\overline{w}), \\ = 0 & \text{for } T^R(y(\overline{w})) = U(\underline{w}) - U(\overline{w}), \\ > 0 & \text{for } T^R(y(\overline{w})) < U(\underline{w}) - U(\overline{w}), \end{cases} \tag{23}$$

in which $U(\underline{w}) - U(\overline{w}) < 0$.

Proof. See Appendix A. ∎

The sign of these maximin marginal tax rates departs from that of Brett and Weymark (2017) in two ways. First, instead of showing that the marginal tax rate is always zero for the highest skilled, we show that it is zero only when their tax liability is equal to a negative critical value defined as the utility difference between the lowest skilled and the highest skilled. Second, instead of showing that

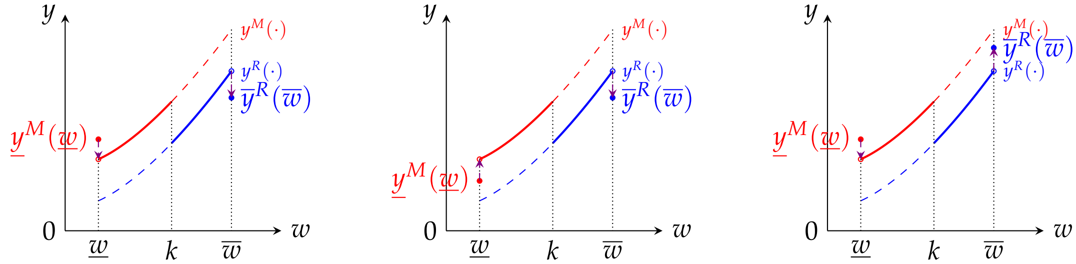

Figure 1: A type-k proposer: $-\overline{w}U'(\overline{w}) < U(\underline{w}) < 0$, $U(\underline{w}) > 0$ and $U(\underline{w}) < -\overline{w}U'(\overline{w})$.

the marginal tax rate is always positive for workers of types higher than the type k of the proposer, it is positive only when their elasticity of migration is smaller than an endogenously determined critical value.

Intuition: Here we explain intuitively why selfishly optimal tax schedules feature negative marginal tax rates for low skills while positive ones for high skills. Everything else being equal, for a selfish proposer motivated to extract resources from the remaining types, the best way is to tax more on higher skills than on lower skills. The reason is that higher skills have higher wage rates implying that they have bigger opportunity costs of leisure, whereas lower skills have lower wage rates implying that they have smaller opportunity costs of leisure. The proposer faces the *tradeoff* between maximizing resources extracted from other types and maximizing their incentives to produce resources available for extraction. In addition, since the proposer cannot observe the types of other workers and also the other workers have migration freedom, she cannot tax higher skills that much, otherwise either tax base shrinks or higher skills are induced to mimic lower types in which she actually needs to pay them information rent to guarantee truth-telling. Since higher skills are allocated with incomes and consumptions no smaller than those to lower skills, lower types have incentives to mimic higher types and hence the proposer needs to pay them certain information rent to prevent them from mimicking. That is, the selfish proposer can just achieve the second-best allocation scheme. This is why lower skills face negative marginal tax rates.

By using Theorems 4.1 and 4.2, the following proposition is obtained.

Proposition 4.3 *For any proposer of type $k \in (\underline{w}, \overline{w})$, the proposed income schedule given by (13) may have the following discontinuities.*

(i) At $w = \underline{w}$, we have:

- *If $U(\underline{w}) > 0$, then there is an upward discontinuity;*
- *If $U(\underline{w}) = 0$, then there is no discontinuity;*
- *If $U(\underline{w}) < 0$, then there is a downward discontinuity.*

(ii) At $w = k$, there is a downward discontinuity.

(iii) At $w = \overline{w}$, we have:

- *If* $U(\underline{w}) > -\overline{w}U'(\overline{w})$, *then there is a downward discontinuity;*
- *If* $U(\underline{w}) = -\overline{w}U'(\overline{w})$, *then there is no discontinuity;*
- *If* $U(\underline{w}) < -\overline{w}U'(\overline{w})$, *then there is an upward discontinuity.*

Proof. See Appendix A. ■

These discontinuities are graphically shown in Figure 1.

In an economy with an exogenous skill distribution, the tax schedule of Brett and Weymark (2016, 2017) only exhibits the downward discontinuity at the type of the proposer, as shown in part (ii). In the current economy with an endogenous skill distribution, we show that there may be another two downward discontinuities at the lowest type and the highest type. Importantly, we identify the utility of the lowest skilled as the key variable determining whether or not these two discontinuities are indeed downward. In order to iron this tax schedule, as shown in Appendix B, it is useful to know when the SOIC condition is violated at the endpoints of the ex post skill distribution.

5 The Voting Equilibrium

Following one common practice in literature, majority rule is used to select the income tax schedule that shall be implemented. Each worker is assumed to have one vote. As argued by Roberts (1977), if political parties in a democratic system choose the income tax schedule to maximize the likelihood of being elected then it is somewhat reasonable to view the chosen tax schedule as being determined, albeit indirectly, by a *pairwise majority voting* process. In each round, workers vote over two arbitrarily-selected alternatives. The one that survives all rounds is hence the winner.

To distinguish allocation schedules by the types of the proposers who propose them, we let $(c(w,k)$, $y(w,k))$ denote the selfishly optimal allocation assigned to a worker of type w by a proposer of type k. The utility obtained by a worker with skill level w under the schedule proposed by type k is hence

$$U(w,k) = c(w,k) - h\left(\frac{y(w,k)}{w}\right). \tag{24}$$

Theorem 5.1 *The selfishly optimal income tax schedule for the median skill type is a Condorcet winner when pairwise majority voting is restricted to the income tax schedules that are selfishly optimal for some skill type.*

Proof. See Appendix A. ■

We hence establish the existence of a Condorcet winner in the current political economy. In particular, as there is a continuum of tax schedules in our problem, the single-crossing condition used by Gans and Smart (1996) is not sufficient to prove the existence of a Condorcet winner. Indeed, here we need to first establish the single-peakedness of preferences and then appeal to Black's (1948) Median Voter Theorem.

6 Identifying the Effect of Migrations on Equilibrium Tax Schedule

6.1 Qualitative Characterization

In what follows, we let w_m denote the median skill level of the ex ante distribution $F(w)$, and let $\tilde{w}_m$ denote that of the ex post distribution $\Gamma(\underline{w}, w) = \int_{\underline{w}}^{w} \tilde{f}(t)dt$. To identify the effects of migrations placed on marginal tax rates, we shall compare our marginal tax rates to those of Brett and Weymark (2017) derived in autarky. As is clear soon, migrations affect marginal tax rates through endogenizing the skill distribution that is a key part of the tax formula and is the determinant of the median skill level. So, migrations affect both the distortion level and the redistribution scale.

We summarize our main findings as two propositions.

Proposition 6.1 *Suppose*

$$\Theta^M(w) < \tilde{\theta}(w) < \Theta^{MR}(w) \;\; \textit{for} \;\; \forall w \in (\underline{w}, w_m]$$

and

$$\Theta^R(w) < \tilde{\theta}(w) < \Theta^{MR}(w) \;\; \textit{for} \;\; \forall w \in (w_m, \overline{w})$$

with these endogenously determined positive bounds $\Theta^M(w)$, $\Theta^R(w)$ and $\Theta^{MR}(w)$ given in Appendix A, then we have the following predictions.

(i) If $\tilde{w}_m = w_m$, then workers of type $w \in (\underline{w}, \overline{w})$ face lower tax rates than in autarky.

(ii) If $\tilde{w}_m < w_m$, then result (i) holds for workers of type $w \in (\underline{w}, \tilde{w}_m] \cup (w_m, \overline{w})$, whereas workers of type $w \in (\tilde{w}_m, w_m]$ face higher tax rates than in autarky.

(iii) If $\tilde{w}_m > w_m$, then result (i) still holds. In particular, workers of type $w \in (w_m, \tilde{w}_m]$ face even lower tax rates than when $\tilde{w}_m = w_m$.

Proof. See Appendix A and Figure 2.[11] ■

Intuition: If migration elasticities of high skills are reasonably large, then their migration threat is strong enough to motivate the government to impose lower tax rates on them relative to the autarky in which migrations and tax competition are forbidden. Since high types face lower tax rates and also low types have reasonably large migration elasticities, the incentives for low types to mimic high types become stronger, which implies that the government needs to transfer more to them to prevent them from mimicking. This analysis explains why migrations under these conditions generally induce lower tax rates than suggested by the autarky equilibrium for both low and high skills. In addition, for types belonging to $(\tilde{w}_m, w_m]$ in case (ii) of Proposition 6.1, they face higher tax rates than in autarky because the median voter under migrations is poorer. They belong to the low-income class in autarky while belonging to the high-income class under migrations, so their status changes from receiving transfers to paying taxes.

[11] In these figures, $\hat{y}^R(\cdot)$ and $\hat{y}^M(\cdot)$ denote the maximin and maximax income schedules derived by Brett and Weymark (2017).

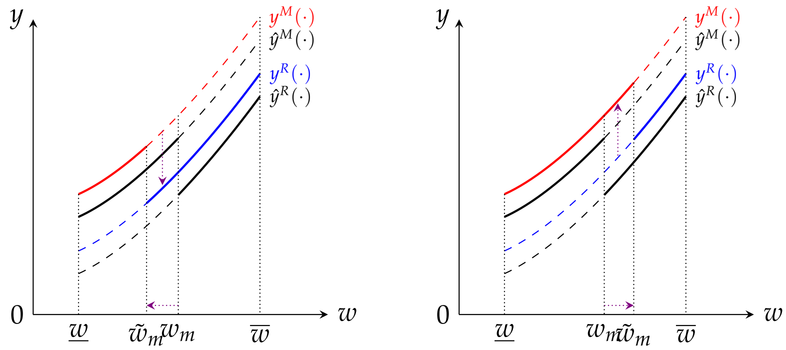

Figure 2: Graphic Illustration of Proposition 6.1.

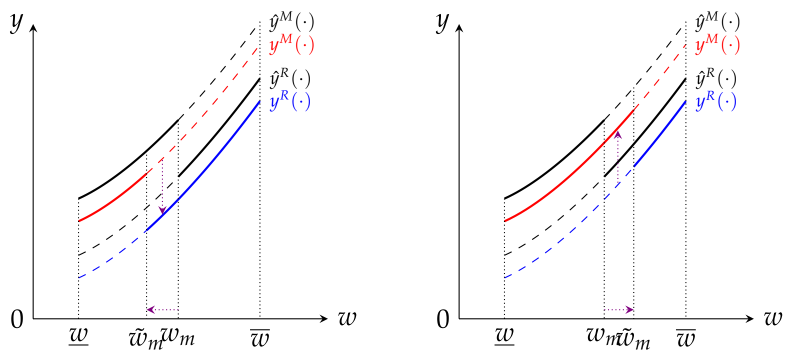

Figure 3: Graphic Illustration of Proposition 6.2.

Proposition 6.2 *Suppose*

$$\begin{cases} \tilde{\theta}(w) < \Theta^M(w) \text{ and } f(w)\Gamma(\underline{w}, w) < \tilde{f}(w)F(w) & \text{for } \forall w \in (\underline{w}, w_m], \\ \tilde{\theta}(w) < \Theta^R(w) \text{ and } f(w)[\Gamma(\underline{w}, \overline{w}) - \Gamma(\underline{w}, w)] > \tilde{f}(w)[1 - F(w)] & \text{for } \forall w \in (w_m, \overline{w}), \end{cases} \tag{25}$$

then we have the following predictions.

(i) If $\tilde{w}_m = w_m$, then workers of type $w \in (\underline{w}, \overline{w})$ face higher tax rates than in autarky.

(ii) If $\tilde{w}_m < w_m$, then result (i) still holds. In particular, workers of type $w \in (\tilde{w}_m, w_m]$ face even higher tax rates than when $\tilde{w}_m = w_m$.

(iii) If $\tilde{w}_m > w_m$, then result (i) holds for workers of type $w \in (\underline{w}, w_m] \cup (\tilde{w}_m, \overline{w})$, whereas workers of type $w \in (w_m, \tilde{w}_m]$ face lower tax rates than in autarky.

Proof. See Appendix A and Figure 3. ■

Intuition: Assumption (25) provides three conditions: (a) migration elasticity is small for both low and high skills; (b) the jurisdiction under consideration faces net labor inflow in low skills but (c) net labor outflow in high skills. Condition (c) means that tax base shrinks relative to the autarky, so marginal tax rates imposed on remaining high skills must be higher than those in autarky, which is somehow guaranteed by condition (a). As now high skills face higher tax rates, low skills have weaker incentives to mimic them, which implies that, everything else equal, the information rents paid to low skills can be smaller. Also, for the same amount of transfers, condition (b) implies that each low-skill worker receives less than they may receive in autarky. As such, both high and low skills face higher marginal tax rates than in autarky. For workers between ex ante and ex post median skill levels, the reasoning used in analyzing Proposition 6.1 still applies.

We have identified conditions determining the relation between $\tilde{w}_m$ and w_m in Appendix C. Essentially, these conditions rely on the following four indexes: (1) whether the ex post measure of workers of all skill levels is greater than, equal to, or smaller than the ex ante one; (2) whether the net labor inflow of skill levels below the ex ante median skill level is positive or not; (3) whether the net labor inflow of skill levels above the ex ante median skill level is positive or not; and (4) the relative magnitude of these two net labor inflows.

6.2 Quantitative Simulation

This subsection proposes numerical experiments[12] of the difference between the current top tax rate and that derived in autarky. We use parameter values estimated based on the U.S. data. This exercise allows us to quantitatively measure the effect of migrations on equilibrium top tax rates. In terms of the equilibrium level of distortion, it also highlights the joint consideration of majority voting and migration in designing redistributive taxation.

It has been argued by Saez (2001) and Atkinson et al. (2011) that Pareto distribution fits the empirical income distribution at high income levels reasonably well. Assuming un-truncated distributions, namely $\overline{w} = \infty$, with ex ante Pareto index $a > 1$ and ex post Pareto index $\tilde{a} > 1$, we can thus obtain

$$\frac{1 - F(w)}{wf(w)} = \frac{1}{a} \quad \text{and} \quad \frac{\Gamma(w, \overline{w})}{w\tilde{f}(w)} = \frac{1}{\tilde{a}}, \tag{26}$$

which measure, respectively, the ex ante and the ex post degrees of income inequality. By using U.S. tax return micro data for 2005, Diamond and Saez (2011) find that the empirical Pareto coefficient is approximately a constant around 1.5 for adjusted gross incomes higher than $200,000, we hence set $\tilde{a} = 1.5$. Following Diamond (1998), we assume that $h(l)$ takes the isoelastic form

$$h(l) = l^{1+\frac{1}{\varepsilon}} / \left(1 + \frac{1}{\varepsilon}\right) \tag{27}$$

for constant elasticity $\varepsilon > 0$. Saez et al. (2012) survey the recent literature and conclude that the best available estimates range from 0.12 to 0.4 in the U.S. As usual, we take a central value of $\varepsilon = 0.25$ for the benchmark exercise. In addition, we, without loss of generality, normalize the upper bound of l to be 1 and focus on interior values, namely $l \in (0, 1)$. Following Lehmann et al. (2014), we set the elasticity of migration for top-income workers to be $\tilde{\theta}(w) = 0.25$.

[12] In fact, noting that the real-world redistributive taxation policy is determined by a political process rather than by a benevolent social planner, numerical experiments based on the current political-economy approach might be more reliable than those based on the normative approach.

By applying (26) and (27), we have the formula of the difference between marginal tax rates as

$$\delta_{\text{MTR}} \equiv \hat{\tau}^R(w) - \tau^R(w) = \frac{\frac{1}{a}\left(1+\frac{1}{\varepsilon}\right)}{1+\frac{1}{a}\left(1+\frac{1}{\varepsilon}\right)} - \frac{\frac{1}{\tilde{a}}\left(1+\frac{1}{\varepsilon}\right)}{1+\frac{1}{\tilde{a}}\left(1+\frac{1}{\varepsilon}\right)} + \frac{\tilde{\theta}(w)\xi(w)}{1+\frac{1}{\tilde{a}}\left(1+\frac{1}{\varepsilon}\right)}\left[1+\frac{l(w)^{\frac{1}{\varepsilon}}}{w(1+\varepsilon)}\right]$$

in which

$$\xi(w) \equiv \frac{\partial U(w)}{\partial y(w)}\frac{y(w)}{c(w)}$$

can be interpreted as an approximation of the elasticity of utility with respect to pre-tax income under quasilinear-in-consumption preferences. To make our tax formula generate the current 42.5% top U.S. marginal tax rate (see Diamond and Saez, 2011), we consider two cases:

- Case (a): $\tilde{a} = 1.5$, $\varepsilon = 0.25$, $\tilde{\theta} = 0.25$, $\xi = 5.96 \Rightarrow \tau^R = 42.5\%$.
- Case (b): $\tilde{a} = 1.5$, $\varepsilon = 0.40$, $\tilde{\theta} = 0.25$, $\xi = 3.64 \Rightarrow \tau^R = 42.5\%$.

We give our findings under varying values of a in Table 1. As is obvious, there can be large differences of marginal tax rates between the current prediction and that in autarky.

Table 1: δ_{MTR} (%) under Cases (a) and (b)

a	1.35	1.4	1.5	1.7	2.0	3.0
$\varepsilon = 0.25$	36.2	35.6	34.4	32.1	28.9	20.0
$\varepsilon = 0.40$	29.6	28.9	27.5	24.8	21.1	11.4

7 Conclusion

In this paper we have examined the feature of redistributive taxation when voters are geographically mobile at the expense of some unobserved migration cost. Without loss of generality, we consider two jurisdictions that can be interpreted as two local states of the United States or two European countries. We have established the voting equilibrium under the majority rule and have fully characterized the income tax schedule that would enact the wishes of median voters. The resulting redistributive policy highlights a complex interaction between voting-with-hand and voting-with-feet, which hence makes the level of distortion and redistribution tend to deviate from that occurs in autarky. We provide the sufficient conditions associated to the elasticity and level of migrations to clearly identify the relative level of redistribution between migration and autarky.

Given what we have established, the goal of this paper is achieved. For future research, our model can be modified or extended in at least three directions. First, as both quasilinear-in-consumption and quasilinear-in-labor preferences are used in taxation literature, a parallel analysis can be conducted under the second type of preferences, and potentially different implications for redistributive taxation is worthwhile studying. Second, instead of being completely motivated by self-interest, we may expect voters of certain skills to exhibit other-regarding or pro-social preferences. Incorporating this concern into our model could make the resulting tax schedule a better approximation to reality. And third, by imposing specific distribution functions of skill and migration cost as well as specific correlations between these two unobservable variables, one could investigate possible sorting types in the voting equilibrium.

Without resorting to a benevolent social planner and with each voter being self-interested, we still obtain an equilibrium tax schedule that achieves a desirable balance between equity and efficiency. We argue that private information and foot-voting possibility are the key. The self-interested median voter faces the tradeoff between maximizing resources extracted from other types and maximizing resources available for extraction. For low skills, the median voter (or government) transfers a positive amount of resources, mainly as information rents, to them to induce truth-telling; for high skills, especially those with high foot-voting capabilities, the median voter will not tax them as in the scenario wherein they cannot exit in order to avoid brain drain and restore a desirable tax base for redistribution. As a result, even the median voter benefits the most from this tax schedule,[13] both equity and efficiency concerns are taken into account seriously under such institutional arrangement. The emergence of the so-called tyranny of middle class is (partially) avoided. In sum, it is the *majority-rule democracy (or voice freedom)* plus the *relevant private information* and *migration (exit) freedom* of heterogeneous voters that lead to the desirable balance between equity and efficiency among a population of completely self-interested voters.

[13]This is actually desirable when the middle class consists of the major part of a society.

References

[1] Agranov, M. and T. R. Palfrey [2015]: "Equilibrium Tax Rates and Income Redistribution: A Laboratory Study." *Journal of Public Economics*, 130: 45-58.

[2] Akcigit, U., S. Baslandze and S. Stantcheva [2016]: "Taxation and the International Mobility of Inventors." *American Economic Review*, 106(10): 2930-2981.

[3] Atkinson, A. B., Piketty, T. and E. Saez [2011]: "Top Incomes in the Long Run of History." *Journal of Economic Literature*, 49(1): 3-71.

[4] Bierbrauer, F. J., C. Brett and J. A. Weymark [2013]: "Strategic Nonlinear Income Tax Competition with Perfect Labor Mobility." *Games and Economic Behavior*, 82: 292-311.

[5] Bierbrauer, F. J. [2011]: "On the Optimality of Optimal Income Taxation." *Journal of Economic Theory*, 146(5): 2105-2116.

[6] Black, D. [1948]: "On the Rationale of Group Decision-making." *Journal of Political Economy*, 56(1): 23-34.

[7] Blumkin, T., E. Sadka and Y. Shem-Tov [2015]: "International Tax Competition: Zero Tax Rate at the Top Re-Established." *International Tax and Public Finance*, 22(5): 760-776.

[8] Bohn, H. and C. Stuart [2013]: "Revenue Extraction by Median Voters." Unpublished Manuscript. Department of Economics, University of California, Santa Barbara.

[9] Brett, C. [2016]: "Probabilistic Voting over Nonlinear Income Taxes with International Migration." Unpublished Manuscript. Department of Economics, Mount Allison University.

[10] Brett, C. and J. A. Weymark [2011]: "How Optimal Nonlinear Income Taxes Change when the Distribution of the Population Changes." *Journal of Public Economics*, 95(11-12): 1239-1247.

[11] Brett, C. and J. A. Weymark [2016]: "Voting over Selfishly Optimal Nonlinear Income Tax Schedules with a Minimum-Utility Constraint." *Journal of Mathematical Economics*, 67: 18-31.

[12] Brett, C. and J. A. Weymark [2017]: "Voting over Selfishly Optimal Nonlinear Income Tax Schedules." *Games and Economic Behavior*, 101: 172-188.

[13] Brett, C. and J. A. Weymark [2017b]: "Majority Rule and Selfishly Optimal Nonlinear Income Tax Schedules with Discrete Skill Levels." Vanderbilt University Department of Economics Working Papers, VUECON-17-00015.

[14] Congressional Budget Office [2012]: "Effective Marginal Tax Rates for Low- and Moderate-Income Workers." Publication No.4149, Congressional Budget Office, Congress of the United States.

[15] Corneo, G. and F. Neher [2015]: "Democratic Redistribution and Rule of the Majority." *European Journal of Political Economy*, 40(Part A): 96-109.

[16] Cremer, H. and P. Pestieau [1998]: "Social Insurance, Majority Voting and Labor Mobility." *Journal of Public Economics*, 68(3): 397-420.

[17] Hindriks, J. and P. De Donder [2003]: "The Politics of Progressive Income Taxation with Incentive Effects." *Journal of Public Economics*, 87(11): 2491-2505.

[18] Diamond, P. [1998]: "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates." *American Economic Review*, 88(1): 83-95.

[19] Diamond, P. and E. Saez [2011]: "The Case for a Progressive Tax: From Basic Research to Policy Recommendations." *Journal of Economic Perspectives*, 25(4): 165-90.

[20] Doquier, F. and A. Marfouk [2006]: "International Migration by Educational Attainment (1990-2000)" in C. Ozden and M. Schiff (eds.) *International Migration, Remittances and Development*, Chapter 5, New York: Palgrave Macmillan.

[21] Gans, J. S. and M. Smart [1996]: "Majority Voting with Single-Crossing Preferences." *Journal of Public Economics*, 59(2): 219-237.

[22] Gruber, J. and E. Saez [2002]: "The Elasticity of Taxable Income: Evidence and Implications." *Journal of Public Economics*, 84(1): 1-32.

[23] Gründler, K. and S. Köllner [2017]: "Determinants of Governmental Redistribution: Income Distribution, Development Levels, and the Role of Perceptions." *Journal of Comparative Economics*, 45(4): 930-962.

[24] Guesnerie, R. [1995]: "*A Contribution to the Pure Theory of Taxation*." Cambridge University Press, Cambridge.

[25] Hamilton, J. and P. Pestieau [2005]: "Optimal Income Taxation and the Ability Distribution: Implications for Migration Equilibria." *International Tax and Public Finance*, 12(1): 29-45.

[26] Hammond, P. J. [1979]: "Straightforward Individual Incentive Compatibility in Large Economies." *Review of Economic Studies*, 46(2): 263-282.

[27] Hindriks, J. [2001]: "Mobility and Redistributive Politics." *Journal of Public Economic Theory*, 3(1): 95-120.

[28] Höchtl, W., R. Sausgruber and J.-R. Tyran [2012]: "Inequality Aversion and Voting on Redistribution." *European Economic Review*, 56(7): 1406-1421.

[29] Kleven, H. J., C. Landais and E. Saez [2013]: "Taxation and International Migration of Superstars: Evidence from the European Football Market." *American Economic Review*, 103(5): 1892-1924.

[30] Kleven, H. J., C. Landais, E. Saez and E. Schultz [2014]: "Migration and Wage Effects of Taxing Top Earners: Evidence from the Foreigners' Tax Scheme in Denmark." *Quarterly Journal of Economics*, 129(1): 333-378.

[31] Lehmann, E., L. Simula and A. Trannoy [2014]: "Tax Me If You Can! Optimal Nonlinear Income Tax between Competing Governments." *Quarterly Journal of Economics*, 129(4): 1995-2030.

[32] Leite-Monteiro, M. [1997]: "Redistributive Policy with Labour Mobility across Countries." *Journal of Public Economics*, 65(2): 229-244.

[33] Meltzer, A. H. and S. F. Richard [1981]: "A Rational Theory of the Size of Government." *Journal of Political Economy*, 89(5): 914-927.

[34] Mirrlees, J. A. [1971]: "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies*, 38(2): 175-208.

[35] Mirrlees, J. A. [1982]: "Migration and Optimal Income Taxes." *Journal of Public Economics*, 18(3): 319-341.

[36] Morelli, M., H. Yang and L. Ye [2012]: "Competitive Nonlinear Taxation and Constitutional Choice." *American Economic Journal: Microeconomics*, 4(1): 142-175.

[37] Piketty, T. and E. Saez [2013]: "Optimal Labor Income Taxation," in Chapter 7 of *Handbook of Public Economics*, 5: 391-474.

[38] Roberts, K. W. S. [1977]: "Voting over Income Tax Schedules." *Journal of Public Economics*, 8(3): 329-340.

[39] Röell, A. A. [2012]: "Voting over Nonlinear Income Tax Schedules." Unpublised Manuscript. School of International and Public Affairs, Columbia University.

[40] Saez, E. [2001]: "Using Elasticities to Derive Optimal Income Tax Rates." *Review of Economic Studies*, 68(1): 205-229.

[41] Saez, E., J. B. Slemrod and S. H. Giertz [2012]: "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review." *Journal of Economic Literature*, 50(1): 3-50.

[42] Simula, L. and A. Trannoy [2010]: "Optimal Income Tax under the Threat of Migration by Top-Income Earners." *Journal of Public Economics*, 94(1-2): 163-173.

[43] Simula, L. and A. Trannoy [2012]: "Shall we Keep the Highly Skilled at Home? The Optimal Income Tax Perspective." *Social Choice and Welfare*, 39(4): 751-782.

[44] Stigler, G. J. [1957]: "The Tenable Range of Functions of Local Government." In: *Federal Expenditure Policy for Economic Growth and Stability: Papers Submitted by Panelists Appearing Before the Subcommittee on Fiscal Policy*. Joint Economic Committee, Congress of the U.S. Washington.

[45] Stigler, G. J. [1970]: "Director's Law of Public Income Redistribution." *The Journal of Law and Economics*, 13(1): 1-10.

[46] Topkis, D. M. [1978]: "Minimizing a Submodular Function on a Lattice." *Operations Research*, 26(2): 305-321.

[47] Jacobs, B., E. L. W. Jongen and F. T. Zoutman [2017]: "Revealed Social Preferences of Dutch Political Parties." *Journal of Public Economics*, 156(C): 81-100.

Appendix A: Proofs

Proof of Theorem 4.1. We shall complete the proof in 4 steps.

Step 1. In order to solve problem (12), we first give the following lemma.

Lemma 7.1 *The optimal schedule of before-tax incomes $y(\cdot)$ for type k's problem (12) is obtained by solving the following unconstrained maximization problem*

$$\begin{aligned} \max_{y(\cdot)} \; & \frac{1}{\Gamma(\underline{w},\overline{w})} \int_{\underline{w}}^{\overline{w}} \left\{ \left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \tilde{f}(w) - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \Gamma(w,\overline{w}) \right\} dw \\ & + \int_{\underline{w}}^{k} \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) dw \end{aligned} \tag{28}$$

with $\Gamma(w,\overline{w}) \equiv \int_w^{\overline{w}} \tilde{f}(t)dt \geq 0.$

Proof. By using (6), we have

$$U(w) = U(\underline{w}) + \int_{\underline{w}}^{w} \frac{y(t)}{t^2} h'\left(\frac{y(t)}{t}\right) dt. \tag{29}$$

Integrating over the ex post support of the skill distribution yields

$$\int_{\underline{w}}^{\overline{w}} U(w)\tilde{f}(w)dw = U(\underline{w}) \int_{\underline{w}}^{\overline{w}} \tilde{f}(w)dw + \int_{\underline{w}}^{\overline{w}} \left[\int_{\underline{w}}^{w} \frac{y(t)}{t^2} h'\left(\frac{y(t)}{t}\right) dt \right] \tilde{f}(w)dw. \tag{30}$$

Reversing the order of integration in (30) gives rise to

$$\int_{\underline{w}}^{\overline{w}} U(w)\tilde{f}(w)dw = U(\underline{w}) \int_{\underline{w}}^{\overline{w}} \tilde{f}(w)dw + \int_{\underline{w}}^{\overline{w}} \frac{y(t)}{t^2} h'\left(\frac{y(t)}{t}\right) \left[\int_{t}^{\overline{w}} \tilde{f}(w)dw \right] dt. \tag{31}$$

Also, it follows from (5) that

$$\int_{\underline{w}}^{\overline{w}} U(w)\tilde{f}(w)dw = \int_{\underline{w}}^{\overline{w}} c(w)\tilde{f}(w)dw - \int_{\underline{w}}^{\overline{w}} h\left(\frac{y(w)}{w}\right) \tilde{f}(w)dw. \tag{32}$$

Applying the equality form of (10) to (32) shows that

$$\int_{\underline{w}}^{\overline{w}} U(w)\tilde{f}(w)dw = \int_{\underline{w}}^{\overline{w}} y(w)\tilde{f}(w)dw - \int_{\underline{w}}^{\overline{w}} h\left(\frac{y(w)}{w}\right) \tilde{f}(w)dw. \tag{33}$$

Combining (31) and (33) leads us to

$$\begin{aligned} U(\underline{w}) \int_{\underline{w}}^{\overline{w}} \tilde{f}(w)dw = & \int_{\underline{w}}^{\overline{w}} y(w)\tilde{f}(w)dw \\ & - \int_{\underline{w}}^{\overline{w}} h\left(\frac{y(w)}{w}\right) \tilde{f}(w)dw - \int_{\underline{w}}^{\overline{w}} \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \left[\int_{w}^{\overline{w}} \tilde{f}(t)dt \right] dw. \end{aligned} \tag{34}$$

Define

$$\Gamma(w,\overline{w}) \equiv \int_{w}^{\overline{w}} \tilde{f}(t)dt, \tag{35}$$

then we can rewrite (34) as

$$U(\underline{w}) = \frac{1}{\Gamma(\underline{w},\overline{w})} \int_{\underline{w}}^{\overline{w}} \left\{ \left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \tilde{f}(w) - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \Gamma(w,\overline{w}) \right\} dw. \tag{36}$$

Substituting (36) into (29) and setting $w = k$, then the maximand in (28) is established.

Step 2. We prove the following lemma.

Lemma 7.2 *By setting $k = \underline{w}$ in problem (28), the maximin income schedule, denoted by*

$$\left\{ \underline{y}^R(\underline{w}), \left\{ y^R(w) \right\}_{w\in(\underline{w},\overline{w})}, \overline{y}^R(\overline{w}) \right\},$$

is obtained as follows.

(i) $\{y^R(w)\}_{w\in(\underline{w},\overline{w})}$ and $\overline{y}^R(\overline{w})$ are solutions to equations

$$\begin{aligned} &\left[1 - \frac{1}{w} h'\left(\frac{y(w)}{w}\right)\right] \tilde{f}(w) + \left[y(w) - h\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w} h'\left(\frac{y(w)}{w}\right)\right] \frac{\partial \tilde{f}(w)}{\partial y(w)} \\ &\qquad = \left[\frac{1}{w^2} h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^3} h''\left(\frac{y(w)}{w}\right)\right] \Gamma(w,\overline{w}) \end{aligned} \tag{37}$$

and

$$\left[1 - \frac{1}{\overline{w}} h'\left(\frac{y(\overline{w})}{\overline{w}}\right)\right] \tilde{f}(\overline{w}) + \left[y(\overline{w}) - h\left(\frac{y(\overline{w})}{\overline{w}}\right) - U(\underline{w})\right] \frac{\partial \tilde{f}(\overline{w})}{\partial y(\overline{w})} = 0, \tag{38}$$

respectively, in which

$$\frac{\partial \tilde{f}(w)}{\partial y(w)} = \begin{cases} g_-(\Delta(w)|w) f_-(w) n_- \frac{\partial U(w)}{\partial y(w)} & \text{for } \Delta(w) \geq 0, \\ g(-\Delta(w)|w) f(w) \frac{\partial U(w)}{\partial y(w)} & \text{for } \Delta(w) \leq 0 \end{cases} \tag{39}$$

with

$$\frac{\partial U(w)}{\partial y(w)} = \frac{1}{w} h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^2} h''\left(\frac{y(w)}{w}\right) \tag{40}$$

for $\forall w \in (\underline{w},\overline{w}]$.

(ii) Given the established $\{y^R(w)\}_{w\in(\underline{w},\overline{w})}$ and $\overline{y}^R(\overline{w})$, $\underline{y}^R(\underline{w})$ is obtained by solving the balanced government budget constraint.

Proof. By setting $k = \underline{w}$, the maximand of problem (28) is hence given by (36). It is straightforward that the corresponding maximization problem can be solved point-wise, and we just need to consider two cases.

Case I: $w \in (\underline{w},\overline{w})$. To solve the first-order condition $\partial U(\underline{w})/\partial y(w) = 0$ using (36), it is sufficient to solve the equation

$$\begin{aligned} &\left[1 - \frac{1}{w} h'\left(\frac{y(w)}{w}\right)\right] \tilde{f}(w) + \left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \frac{\partial \tilde{f}(w)}{\partial y(w)} - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \frac{\partial \Gamma(w,\overline{w})}{\partial y(w)} \\ &\qquad = \left[\frac{1}{w^2} h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^3} h''\left(\frac{y(w)}{w}\right)\right] \Gamma(w,\overline{w}). \end{aligned} \tag{41}$$

By applying the formula of integration by parts to (35), we have

$$\Gamma(w,\overline{w}) = \tilde{f}(\overline{w})\overline{w} - \tilde{f}(w)w - \int_{w}^{\overline{w}} \tilde{f}'(t)tdt,$$

by which we obtain

$$\frac{\partial\Gamma(w,\overline{w})}{\partial y(w)} = -w\frac{\partial\tilde{f}(w)}{\partial y(w)} \text{ and } \frac{\partial\Gamma(w,\overline{w})}{\partial y(\overline{w})} = \overline{w}\frac{\partial\tilde{f}(\overline{w})}{\partial y(\overline{w})}. \tag{42}$$

Applying (42) to (41) and rearranging the algebra, the desired (37) is hence established. By setting $k = w$ in the maximand of problem (28), then for $\forall w \in (\underline{w},\overline{w})$ (40) is immediate by applying the formula of integration by parts. Also, by using (8), (39) is immediate.

Case II: $w = \overline{w}$. By using (36) and the formula of integration by parts again, we have the following first-order condition:

$$\begin{aligned}\frac{\partial U(\underline{w})}{\partial y(\overline{w})} =& -\frac{U(\underline{w})}{\Gamma(\underline{w},\overline{w})}\frac{\partial\Gamma(\underline{w},\overline{w})}{\partial y(\overline{w})} \\ &+ \frac{\overline{w}}{\Gamma(\underline{w},\overline{w})}\left\{\left[1-\frac{1}{\overline{w}}h'\left(\frac{y(\overline{w})}{\overline{w}}\right)\right]\tilde{f}(\overline{w}) + \left[y(\overline{w}) - h\left(\frac{y(\overline{w})}{\overline{w}}\right)\right]\frac{\partial\tilde{f}(\overline{w})}{\partial y(\overline{w})}\right\} = 0.\end{aligned} \tag{43}$$

By applying (42) to (43) and rearranging the algebra, equation (38) is thus obtained. In addition, by setting $k = \overline{w}$ in the maximand of problem (28) and making use of the formula of integration by parts, we obtain

$$\begin{aligned}\frac{\partial U(\overline{w})}{\partial y(\overline{w})} =& -\frac{U(\underline{w})}{\Gamma(\underline{w},\overline{w})}\frac{\partial\Gamma(\underline{w},\overline{w})}{\partial y(\overline{w})} \\ &+ \frac{\overline{w}}{\Gamma(\underline{w},\overline{w})}\left\{\left[1-\frac{1}{\overline{w}}h'\left(\frac{y(\overline{w})}{\overline{w}}\right)\right]\tilde{f}(\overline{w}) + \left[y(\overline{w}) - h\left(\frac{y(\overline{w})}{\overline{w}}\right)\right]\frac{\partial\tilde{f}(\overline{w})}{\partial y(\overline{w})}\right\} \\ &+ \overline{w}\left[\frac{1}{\overline{w}^2}h'\left(\frac{y(\overline{w})}{\overline{w}}\right) + \frac{y(\overline{w})}{\overline{w}^3}h''\left(\frac{y(\overline{w})}{\overline{w}}\right)\right].\end{aligned} \tag{44}$$

By applying (43) to (44), then we arrive at that (40) also holds for $w = \overline{w}$.

Step 3. We also need the following lemma.

Lemma 7.3 *By setting $k = \overline{w}$ in problem (28), the maximax income schedule, denoted by*

$$\left\{\underline{y}^M(\underline{w}), \left\{y^M(w)\right\}_{w\in(\underline{w},\overline{w})}, \overline{y}^M(\overline{w})\right\},$$

is obtained as follows.

(i) $\{y^M(w)\}_{w\in(\underline{w},\overline{w})}$ and $\underline{y}^M(\underline{w})$ are solutions to equations

$$\begin{aligned}&\left[1-\frac{1}{w}h'\left(\frac{y(w)}{w}\right)\right]\tilde{f}(w) + \left[y(w) - h\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w}h'\left(\frac{y(w)}{w}\right)\right]\frac{\partial\tilde{f}(w)}{\partial y(w)} \\ &\qquad = \left[\frac{1}{w^2}h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^3}h''\left(\frac{y(w)}{w}\right)\right][\Gamma(w,\overline{w}) - \Gamma(\underline{w},\overline{w})]\end{aligned} \tag{45}$$

and

$$\left[1-\frac{1}{\underline{w}}h'\left(\frac{y(\underline{w})}{\underline{w}}\right)\right]\tilde{f}(\underline{w})+\left[y(\underline{w})-h\left(\frac{y(\underline{w})}{\underline{w}}\right)+\frac{y(\underline{w})}{\underline{w}}h'\left(\frac{y(\underline{w})}{\underline{w}}\right)-U(\underline{w})\right]\frac{\partial\tilde{f}(\underline{w})}{\partial y(\underline{w})}=0, \tag{46}$$

respectively, in which $\partial\tilde{f}(w)/\partial y(w)$ is determined by equations (39) and (40) for $\forall w\in[\underline{w},\overline{w})$.

(ii) Given the established $\{y^M(w)\}_{w\in(\underline{w},\overline{w})}$ and $\underline{y}^M(\underline{w})$, $\overline{y}^M(\overline{w})$ is obtained by solving the balanced government budget constraint.

Proof. By setting $k=\overline{w}$, the maximand of problem (28) can be written as

$$\begin{aligned}U(\overline{w})=&\frac{1}{\Gamma(\underline{w},\overline{w})}\int_{\underline{w}}^{\overline{w}}\left\{\left[y(w)-h\left(\frac{y(w)}{w}\right)\right]\tilde{f}(w)-\frac{y(w)}{w^2}h'\left(\frac{y(w)}{w}\right)\Gamma(w,\overline{w})\right\}dw\\&+\int_{\underline{w}}^{\overline{w}}\frac{y(w)}{w^2}h'\left(\frac{y(w)}{w}\right)dw.\end{aligned} \tag{47}$$

The corresponding maximization problem can be solved point-wise, and we just need to consider two cases.

<u>Case I</u>: $w\in(\underline{w},\overline{w})$. To solve the first-order condition $\partial U(\overline{w})/\partial y(w)=0$ using (47), it is sufficient to solve the equation

$$\begin{aligned}&\left[1-\frac{1}{w}h'\left(\frac{y(w)}{w}\right)\right]\tilde{f}(w)+\left[y(w)-h\left(\frac{y(w)}{w}\right)\right]\frac{\partial\tilde{f}(w)}{\partial y(w)}-\frac{y(w)}{w^2}h'\left(\frac{y(w)}{w}\right)\frac{\partial\Gamma(w,\overline{w})}{\partial y(w)}\\&\qquad=\left[\frac{1}{w^2}h'\left(\frac{y(w)}{w}\right)+\frac{y(w)}{w^3}h''\left(\frac{y(w)}{w}\right)\right][\Gamma(w,\overline{w})-\Gamma(\underline{w},\overline{w})].\end{aligned} \tag{48}$$

By using (42) and rearranging the algebra, (45) follows from (48). Similar to the proof of Lemma 7.2, it is easy to verify that $\{\partial\tilde{f}(w)/\partial y(w)\}_{w\in(\underline{w},\overline{w})}$ is determined by equations (39) and (40).

<u>Case II</u>: $w=\underline{w}$. By using (47) and the formula of integration by parts, we have the first-order condition:

$$\begin{aligned}\frac{\partial U(\overline{w})}{\partial y(\underline{w})}=&-\frac{U(\underline{w})}{\Gamma(\underline{w},\overline{w})}\frac{\partial\Gamma(\underline{w},\overline{w})}{\partial y(\underline{w})}\\&-\frac{\underline{w}}{\Gamma(\underline{w},\overline{w})}\left\{\left[1-\frac{1}{\underline{w}}h'\left(\frac{y(\underline{w})}{\underline{w}}\right)\right]\tilde{f}(\underline{w})+\left[y(\underline{w})-h\left(\frac{y(\underline{w})}{\underline{w}}\right)\right]\frac{\partial\tilde{f}(\underline{w})}{\partial y(\underline{w})}\right\}\\&+\frac{\underline{w}}{\Gamma(\underline{w},\overline{w})}\left\{\left[\frac{1}{\underline{w}^2}h'\left(\frac{y(\underline{w})}{\underline{w}}\right)+\frac{y(\underline{w})}{\underline{w}^3}h''\left(\frac{y(\underline{w})}{\underline{w}}\right)\right]\Gamma(\underline{w},\overline{w})+\frac{y(\underline{w})}{\underline{w}^2}h'\left(\frac{y(\underline{w})}{\underline{w}}\right)\frac{\partial\Gamma(\underline{w},\overline{w})}{\partial y(\underline{w})}\right\}\\&-\underline{w}\left[\frac{1}{\underline{w}^2}h'\left(\frac{y(\underline{w})}{\underline{w}}\right)+\frac{y(\underline{w})}{\underline{w}^3}h''\left(\frac{y(\underline{w})}{\underline{w}}\right)\right]=0.\end{aligned} \tag{49}$$

By applying (42) and rearranging the algebra, (49) can be simplified as being the desired (46). In order to solve for the term $\partial \tilde{f}(\underline{w})/\partial y(\underline{w})$ appearing in (49), we get from (36) that

$$\begin{aligned}\frac{\partial U(\underline{w})}{\partial y(\underline{w})} =& -\frac{U(\underline{w})}{\Gamma(\underline{w},\overline{w})}\frac{\partial \Gamma(\underline{w},\overline{w})}{\partial y(\underline{w})} + \underline{w}\left[\frac{1}{\underline{w}^2}h'\left(\frac{y(\underline{w})}{\underline{w}}\right) + \frac{y(\underline{w})}{\underline{w}^3}h''\left(\frac{y(\underline{w})}{\underline{w}}\right)\right] \\ & - \frac{\underline{w}}{\Gamma(\underline{w},\overline{w})}\left\{\left[1-\frac{1}{\underline{w}}h'\left(\frac{y(\underline{w})}{\underline{w}}\right)\right]\tilde{f}(\underline{w}) + \left[y(\underline{w}) - h\left(\frac{y(\underline{w})}{\underline{w}}\right)\right]\frac{\partial \tilde{f}(\underline{w})}{\partial y(\underline{w})}\right\} \\ & + \frac{\underline{w}}{\Gamma(\underline{w},\overline{w})}\frac{y(\underline{w})}{\underline{w}^2}h'\left(\frac{y(\underline{w})}{\underline{w}}\right)\frac{\partial \Gamma(\underline{w},\overline{w})}{\partial y(\underline{w})}.\end{aligned} \tag{50}$$

Applying (49) to (50) results in the desired equation (40) for $w = \underline{w}$.

Step 4. By using the maximization problem (28) stated in Lemma 7.1, it is easy to show that

$$\frac{\partial U(k)}{\partial y(w)} = \frac{\partial U(\overline{w})}{\partial y(w)} \quad \text{for } \forall w \in [\underline{w}, k)$$

and

$$\frac{\partial U(k)}{\partial y(w)} = \frac{\partial U(\underline{w})}{\partial y(w)} \quad \text{for } \forall w \in (k, \overline{w}],$$

for $\forall k \in (\underline{w}, \overline{w})$. Therefore, the desired income schedule (13) follows from a direct application of Lemmas 7.2 and 7.3. ■

Proof of Proposition 4.1. We shall complete the proof in 3 steps.

Step 1. Making use of (2) and (5) reveals that

$$y(\underline{w}) - h\left(\frac{y(\underline{w})}{\underline{w}}\right) - U(\underline{w}) = T^M(y(\underline{w})). \tag{51}$$

Plugging (51) in (15) produces

$$\underline{\tau}^M(\underline{w}) = \underbrace{-\frac{\partial \tilde{f}(\underline{w})}{\partial y(\underline{w})}\frac{1}{\tilde{f}(\underline{w})}}_{<0}\left[T^M(y(\underline{w})) + \frac{y(\underline{w})}{\underline{w}}h'\left(\frac{y(\underline{w})}{\underline{w}}\right)\right], \tag{52}$$

in which we have used (39), (40) and the strict convexity of h. Also, by applying (6), (52) can be equivalently written as

$$\underline{\tau}^M(\underline{w}) = -\frac{\partial \tilde{f}(\underline{w})}{\partial y(\underline{w})}\frac{1}{\tilde{f}(\underline{w})}\left[T^M(y(\underline{w})) + \underline{w}U'(\underline{w})\right]. \tag{53}$$

So, the required assertion (19) follows from (52) and (53).

Step 2. In fact, we can further rewrite (52) as

$$\begin{aligned}
\underline{\tau}^M(\underline{w}) &= -\frac{\partial \tilde{f}(\underline{w})}{\partial y(\underline{w})}\frac{y(\underline{w})}{\tilde{f}(\underline{w})}\left[\frac{T^M(y(\underline{w}))}{y(\underline{w})}+\frac{1}{\underline{w}}h'\left(\frac{y(\underline{w})}{\underline{w}}\right)\right] \\
&= -\frac{\partial \tilde{f}(\underline{w})}{\partial U(\underline{w})}\frac{\partial U(\underline{w})}{\partial y(\underline{w})}\frac{y(\underline{w})}{\tilde{f}(\underline{w})}\left[\frac{T^M(y(\underline{w}))}{y(\underline{w})}+\frac{1}{\underline{w}}h'\left(\frac{y(\underline{w})}{\underline{w}}\right)\right] \\
&= -\frac{\partial \tilde{f}(\underline{w})}{\partial \Delta(\underline{w})}\frac{c(\underline{w})}{\tilde{f}(\underline{w})}\cdot\frac{\partial U(\underline{w})}{\partial y(\underline{w})}\cdot\frac{y(\underline{w})}{c(\underline{w})}\left[\frac{T^M(y(\underline{w}))}{y(\underline{w})}+\frac{1}{\underline{w}}h'\left(\frac{y(\underline{w})}{\underline{w}}\right)\right] \\
&= -\tilde{\theta}(\underline{w})\cdot \mathrm{MU}_y(\underline{w})\cdot \mathrm{Q}_{y,c}(\underline{w})\left[\mathrm{ATR}^M(\underline{w})+\frac{1}{\underline{w}}h'\left(\frac{y(\underline{w})}{\underline{w}}\right)\right],
\end{aligned} \tag{54}$$

in which we have used the chain rule of calculus, (8) and (9). Applying (4) to (54) reveals that

$$\underline{\tau}^M(\underline{w}) = -\tilde{\theta}(\underline{w})\cdot \mathrm{MU}_y(\underline{w})\cdot \mathrm{Q}_{y,c}(\underline{w})\left[\mathrm{ATR}^M(\underline{w})+1-\underline{\tau}^M(\underline{w})\right],$$

rearranging the algebra of which results in

$$\underline{\tau}^M(\underline{w}) = -\frac{\tilde{\theta}(\underline{w})\cdot \mathrm{MU}_y(\underline{w})\cdot \mathrm{Q}_{y,c}(\underline{w})}{1-\tilde{\theta}(\underline{w})\cdot \mathrm{MU}_y(\underline{w})\cdot \mathrm{Q}_{y,c}(\underline{w})}\left[\mathrm{ATR}^M(\underline{w})+1\right] \tag{55}$$

whenever $\tilde{\theta}(\underline{w})\cdot \mathrm{MU}_y(\underline{w})\cdot \mathrm{Q}_{y,c}(\underline{w}) \neq 1$. As a result, assertions (20) and (21) are obtained by using (55).

Step 3. We get from Lemma 7.1 that

$$\Gamma(w,\overline{w})-\Gamma(\underline{w},\overline{w}) = -\int_{\underline{w}}^{w}\tilde{f}(t)dt = -\Gamma(\underline{w},w) < 0, \quad \forall w > \underline{w}. \tag{56}$$

Also, we can show that

$$\begin{aligned}
& y(w)-h\left(\frac{y(w)}{w}\right)+\frac{y(w)}{w}h'\left(\frac{y(w)}{w}\right) \geq 0 \\
\Leftrightarrow\ & w \geq \frac{h(l(w))}{l(w)}\left[1-\frac{l(w)h'(l(w))}{h(l(w))}\right].
\end{aligned} \tag{57}$$

Let's define $H(l) \equiv lh'(l)-h(l)$. We have $H'(l) = lh''(l) > 0$ for $\forall l > 0$, which implies that $H(l)$ is strictly increasing in l and hence $H(l) > H(0) = 0$. In other words, $lh'(l)/h(l) > 1$, and hence the right hand side of the second inequality of (57) is negative. Given that $\underline{w} > 0$ by assumption, we conclude by using (57) again that

$$y(w)-h\left(\frac{y(w)}{w}\right)+\frac{y(w)}{w}h'\left(\frac{y(w)}{w}\right) > 0. \tag{58}$$

So combining (16), (56), (58), (39), (40) and the strict convexity of h all together gives rise to the desired assertion (ii). ■

Proof of Proposition 4.2. We shall complete the proof in 2 steps.

Step 1. By using the chain rule of calculus, (8) and (9), we have

$$\begin{aligned}
&\frac{\partial \tilde{f}(w)}{\partial y(w)} \frac{1}{\tilde{f}(w)} \left[h\left(\frac{y(w)}{w}\right) - y(w) - \frac{y(w)}{w} h'\left(\frac{y(w)}{w}\right) \right] \\
&= \frac{\partial \tilde{f}(w)}{\partial \Delta(w)} \frac{c(w)}{\tilde{f}(w)} \frac{\partial U(w)}{\partial y(w)} \frac{1}{c(w)} \left[h\left(\frac{y(w)}{w}\right) - y(w) - \frac{y(w)}{w} h'\left(\frac{y(w)}{w}\right) \right] \\
&= \tilde{\theta}(w) \frac{\partial U(w)}{\partial y(w)} \frac{1}{c(w)} \left[h\left(\frac{y(w)}{w}\right) - y(w) - \frac{y(w)}{w} h'\left(\frac{y(w)}{w}\right) \right].
\end{aligned} \tag{59}$$

Then we get from (58), (59) and (17) that

$$\tau^R(w) = \underbrace{\frac{\partial U(w)}{\partial y(w)}}_{>0} \left\{ \underbrace{\frac{\Gamma(w, \overline{w})}{w \tilde{f}(w)}}_{>0} - \underbrace{\frac{\tilde{\theta}(w)}{c(w)} \left[y(w) - h\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w} h'\left(\frac{y(w)}{w}\right) \right]}_{>0} \right\}, \tag{60}$$

by which assertion (22) is immediate.

Step 2. It follows from (6) that $U(\overline{w}) > U(\underline{w})$. Making use (2) and (5) shows that

$$y(\overline{w}) - h\left(\frac{y(\overline{w})}{\overline{w}}\right) - U(\underline{w}) = T^R(y(\overline{w})) + \underbrace{U(\overline{w}) - U(\underline{w})}_{>0},$$

by which the desired assertion (23) follows. ■

Proof of Proposition 4.3. We shall complete the proof in 3 steps.

Step 1. As w approaches $\underline{w}$ from the above, we get from (16) that

$$\tau^M(\underline{w}) = -\frac{\partial \tilde{f}(\underline{w})}{\partial y(\underline{w})} \frac{1}{\tilde{f}(\underline{w})} \left[y(\underline{w}) - h\left(\frac{y(\underline{w})}{\underline{w}}\right) + \frac{y(\underline{w})}{\underline{w}} h'\left(\frac{y(\underline{w})}{\underline{w}}\right) \right], \tag{61}$$

in which we have used the continuity of $\tau^M(w)$ over the interval $(\underline{w}, k)$. Comparing this formula (61) with (15) reveals that

$$\underline{\tau}^M(\underline{w}) - \tau^M(\underline{w}) = \underbrace{\frac{\partial \tilde{f}(\underline{w})}{\partial y(\underline{w})} \frac{1}{\tilde{f}(\underline{w})}}_{>0} \cdot U(\underline{w}).$$

So, for example, if $U(\underline{w}) > 0$, then we have $\underline{\tau}^M(\underline{w}) > \tau^M(\underline{w})$. We have three cases to consider. First, if $\underline{\tau}^M(\underline{w}) \geq 0 > \tau^M(\underline{w})$, then under tax $\tau^M(\underline{w})$ each type-$\underline{w}$ worker has her income distorted upward compared to the full-information solution, whereas her income is either not distorted or is distorted downward compared to the full-information solution under tax $\underline{\tau}^M(\underline{w})$. And the similar observation applies for the case $\underline{\tau}^M(\underline{w}) > 0 \geq \tau^M(\underline{w})$. Second, if $\underline{\tau}^M(\underline{w}) > \tau^M(\underline{w}) > 0$, then each type-$\underline{w}$ worker has her income distorted downward compared to the full-information solution under both taxes but the magnitude of distortion is bigger under tax $\underline{\tau}^M(\underline{w})$. And the similar observation applies for the case $\underline{\tau}^M(\underline{w}) > \tau^M(\underline{w}) \geq 0$. Third, if $0 > \underline{\tau}^M(\underline{w}) > \tau^M(\underline{w})$, then each type-$\underline{w}$ worker has her income distorted upward compared to the full-information solution under both taxes but the magnitude of distortion is bigger under tax $\tau^M(\underline{w})$. And the similar observation applies for the case $0 \geq \underline{\tau}^M(\underline{w}) > \tau^M(\underline{w})$. Thus, no matter which case we consider, we see an upward discontinuity of the income schedule whenever $\underline{\tau}^M(\underline{w}) > \tau^M(\underline{w})$. For the other cases, we can analyze in a quite similar way, and we omit them to economize on the space. The desired assertion in (i) hence follows.

Step 2. It follows from (16) and (17) that

$$\tau^M(w) - \tau^R(w) = -\frac{\Gamma(\underline{w}, \overline{w})}{w\tilde{f}(w)}\left[\frac{1}{w}h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^2}h''\left(\frac{y(w)}{w}\right)\right] < 0,$$

then we can apply the same reasoning used in step 1 to obtain the assertion stated in (ii).

Step 3. As w approaches $\overline{w}$ from the below, we get from (17) that

$$\tau^R(\overline{w}) = -\frac{\partial \tilde{f}(\overline{w})}{\partial y(\overline{w})}\frac{1}{\tilde{f}(\overline{w})}\left[y(\overline{w}) - h\left(\frac{y(\overline{w})}{\overline{w}}\right) + \frac{y(\overline{w})}{\overline{w}}h'\left(\frac{y(\overline{w})}{\overline{w}}\right)\right], \tag{62}$$

in which we have used the continuity of $\tau^R(w)$ over the interval $(k, \overline{w})$. Comparing this formula (62) with (18) reveals that

$$\tau^R(\overline{w}) - \overline{\tau}^R(\overline{w}) = \underbrace{-\frac{\partial \tilde{f}(\overline{w})}{\partial y(\overline{w})}\frac{1}{\tilde{f}(\overline{w})}}_{<0}\left[\frac{y(\overline{w})}{\overline{w}}h'\left(\frac{y(\overline{w})}{\overline{w}}\right) + U(\underline{w})\right],$$

which combined with (6) and the same reasoning used in step 1 leads us to the desired assertion shown in part (iii). ■

Proof of Theorem 5.1. We shall complete the proof in 3 steps.

Step 1. Let's consider two alternative proposers of types k_1 and k_2, for $k_1 < k_2$. Since their proposed income schedules coincide with the maximax schedule for types below their type and coincide with the maximin schedule for types above their type, and also the maximax income schedule lies everywhere above the maximin income schedule, the higher the type of the proposer, the more workers whose types are below the proposer and the more workers who are allocated with the maximax incomes. Precisely, if the proposer changes from type k_1 to type k_2, all workers of types belong to set $[k_1, k_2]$ are strictly better off in terms of pre-tax income while all other workers with the remaining types are neutral to this change. We hence have that $y(w, k_1) \leq y(w, k_2)$ for $\forall w, k_1, k_2 \in [\underline{w}, \overline{w}]$, and $y(w, k_1) < y(w, k_2)$ for $\forall w \in [k_1, k_2]$ (see Figure 4).

In addition, as all proposers face the same government budget and incentive constraints, each proposer must weakly prefer what she obtains with her own schedule to what any other worker proposer for her. Formally,

$$U(w, w) \;\geq\; U(w, k) \quad \text{for } \forall w, k \in [\underline{w}, \overline{w}]. \tag{63}$$

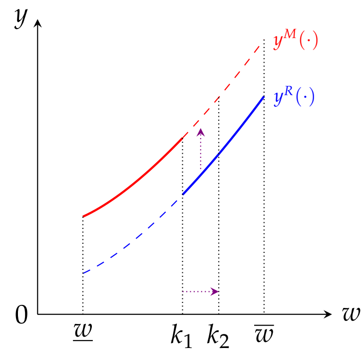

Figure 4: The type of proposer changes from k_1 to k_2.

We next show that a worker of any type w has a weakly single-peaked preference on the set of types. To this end, we need to consider two cases with the proof procedure being directly brought from Brett and Weymark (2017).

Step 2. First, we consider the right hand side of w. That is, let's pick arbitrarily three types w, k_1, k_2 satisfying $w < k_1 < k_2$.

By using (6) and (63), we have

$$\begin{aligned} U(w,k_1) &= U(k_1,k_1) - \int_w^{k_1} h'\left(\frac{y(t,k_1)}{t}\right)\frac{y(t,k_1)}{t^2}dt \\ &\geq U(k_1,k_2) - \int_w^{k_1} h'\left(\frac{y(t,k_1)}{t}\right)\frac{y(t,k_1)}{t^2}dt. \end{aligned} \tag{64}$$

Similarly, we can get by (6) that

$$U(w,k_2) = U(k_1,k_2) - \int_w^{k_1} h'\left(\frac{y(t,k_2)}{t}\right)\frac{y(t,k_2)}{t^2}dt. \tag{65}$$

Solving for $U(k_1,k_2)$ from (65) and inserting it into (64) produces

$$U(w,k_1) - U(w,k_2) \geq \int_w^{k_1}\left[h'\left(\frac{y(t,k_2)}{t}\right)\frac{y(t,k_2)}{t^2} - h'\left(\frac{y(t,k_1)}{t}\right)\frac{y(t,k_1)}{t^2}\right]dt. \tag{66}$$

Since h is strictly increasing and convex, we hence have by using (66) that $U(w,k_1) \geq U(w,k_2)$, which combined with (63) reveals that

$$U(w,w) \geq U(w,k_1) \geq U(w,k_2), \quad \forall w < k_1 < k_2. \tag{67}$$

Step 3. Second, for the case with $w > k_1 > k_2$, we also get by using (6) and (63) that

$$\begin{aligned} U(w,k_1) &= U(k_1,k_1) + \int_{k_1}^{w} h'\left(\frac{y(t,k_1)}{t}\right)\frac{y(t,k_1)}{t^2}dt \\ &\geq U(k_1,k_2) + \int_{k_1}^{w} h'\left(\frac{y(t,k_1)}{t}\right)\frac{y(t,k_1)}{t^2}dt. \end{aligned} \tag{68}$$

Similarly,

$$U(w,k_2) \;=\; U(k_1,k_2) + \int_{k_1}^{w} h'\left(\frac{y(t,k_2)}{t}\right)\frac{y(t,k_2)}{t^2}dt. \tag{69}$$

Making use of (68) and (69) gives rise to

$$U(w,k_1) - U(w,k_2) \;\geq\; \int_{k_1}^{w}\left[h'\left(\frac{y(t,k_1)}{t}\right)\frac{y(t,k_1)}{t^2} - h'\left(\frac{y(t,k_2)}{t}\right)\frac{y(t,k_2)}{t^2}\right]dt. \tag{70}$$

By applying the same reasoning used in step 2 to (70), we arrive at

$$U(w,w) \;\geq\; U(w,k_1) \;\geq\; U(w,k_2), \quad \forall w > k_1 > k_2. \tag{71}$$

Accordingly, (67) combined with (71) reveals that the preference of any given type of worker is indeed (weakly) single-peaked on the set of types. By applying the Black's Median Voter Theorem (see Black, 1948), the desired assertion hence follows. ■

Proof of Propositions 6.1 and 6.2. To prove these two propositions, we just need to prove the following lemma.

Lemma 7.4 *Suppose $\tilde{w}_m = w_m$, then we have the following predictions.*

(i) We have $\tau^R(w) > \hat{\tau}^M(w) > \tau^M(w)$ if $\Theta^M(w) < \tilde{\theta}(w) < \Theta^{MR}(w)$ and $\hat{\tau}^R(w) > \tau^R(w) > \hat{\tau}^M(w)$ if $\Theta^R(w) < \tilde{\theta}(w) < \Theta^{MR}(w)$, in which $\hat{\tau}^M(w)$ and $\hat{\tau}^R(w)$ denote the maximax and maximin marginal tax rates derived by Brett and Weymark (2017) and $\Theta^M(w)$, $\Theta^R(w)$ and $\Theta^{MR}(w)$ are respectively given by (80), (83) and (86).

(ii) If the ex ante and ex post skill distributions satisfy

$$\begin{cases} \frac{f(w)}{F(w)} < \frac{\tilde{f}(w)}{\Gamma(\underline{w},w)} & \text{for } \forall w \in (\underline{w}, w_m], \\ \frac{f(w)}{1-F(w)} > \frac{\tilde{f}(w)}{\Gamma(\underline{w},\overline{w})-\Gamma(\underline{w},w)} & \text{for } \forall w \in (w_m, \overline{w}), \end{cases} \tag{72}$$

then $\hat{\tau}^M(w) < \tau^M(w) < \hat{\tau}^R(w)$ if $\tilde{\theta}(w) < \Theta^M(w)$ and $\tau^M(w) < \hat{\tau}^R(w) < \tau^R(w)$ if $\tilde{\theta}(w) < \Theta^R(w)$.

We shall complete the proof this lemma in 5 steps.

<u>*Step 1.*</u> The tax formulas of Brett and Weymark (2017) are given as follows:

$$\hat{\tau}^R(w) = \frac{1-F(w)}{wf(w)}\left[\frac{1}{w}h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^2}h''\left(\frac{y(w)}{w}\right)\right], \tag{73}$$

$$\hat{\tau}^M(w) = -\frac{F(w)}{wf(w)}\left[\frac{1}{w}h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^2}h''\left(\frac{y(w)}{w}\right)\right]. \tag{74}$$

By using our notation given by (40), (73) and (74) can be rewritten as

$$\hat{\tau}^R(w) = \frac{1-F(w)}{wf(w)}\frac{\partial U(w)}{\partial y(w)}, \tag{75}$$

$$\hat{\tau}^M(w) = -\frac{F(w)}{wf(w)}\frac{\partial U(w)}{\partial y(w)}. \tag{76}$$

Step 2. We can use (56) to rewrite (16) as

$$\begin{aligned}\tau^M(w) \;=\; & \frac{\partial \tilde{f}(w)}{\partial y(w)}\frac{1}{\tilde{f}(w)}\left[h\left(\frac{y(w)}{w}\right) - y(w) - \frac{y(w)}{w}h'\left(\frac{y(w)}{w}\right)\right] \\ & - \frac{\Gamma(\underline{w},w)}{w\tilde{f}(w)}\left[\frac{1}{w}h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^2}h''\left(\frac{y(w)}{w}\right)\right].\end{aligned} \tag{77}$$

Substituting (59) into (77) and using (40) and (58), we can rewrite $\tau^M(w)$ as

$$\tau^M(w) = \underbrace{\frac{\partial U(w)}{\partial y(w)}}_{>0}\left\{\underbrace{\frac{\tilde{\theta}(w)}{c(w)}\left[h\left(\frac{y(w)}{w}\right) - y(w) - \frac{y(w)}{w}h'\left(\frac{y(w)}{w}\right)\right]}_{<0} - \underbrace{\frac{\Gamma(\underline{w},w)}{w\tilde{f}(w)}}_{>0}\right\}. \tag{78}$$

By using (76) and (78), we obtain

$$\begin{aligned}&\hat{\tau}^M(w) - \tau^M(w) \\ &= \underbrace{\frac{\partial U(w)}{\partial y(w)}}_{>0}\left\{\frac{\tilde{\theta}(w)}{c(w)}\left[y(w) - h\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w}h'\left(\frac{y(w)}{w}\right)\right] - \left[\frac{F(w)}{wf(w)} - \frac{\Gamma(\underline{w},w)}{w\tilde{f}(w)}\right]\right\}\end{aligned} \tag{79}$$

for $\forall w \in (\underline{w}, w_m]$. We first define by using (79) that

$$\Theta^M(w) \equiv \frac{c(w)}{wl(w) - h(l(w)) + l(w)h'(l(w))}\left[\frac{F(w)}{wf(w)} - \frac{\Gamma(\underline{w},w)}{w\tilde{f}(w)}\right]. \tag{80}$$

So using (79) and (80) gives rise to

$$\hat{\tau}^M(w)\begin{cases} < \tau^M(w) & \text{if } \tilde{\theta}(w) < \Theta^M(w), \\ = \tau^M(w) & \text{if } \tilde{\theta}(w) = \Theta^M(w), \\ > \tau^M(w) & \text{if } \tilde{\theta}(w) > \Theta^M(w). \end{cases} \tag{81}$$

Since by definition we have $\tilde{\theta}(w) > 0$, hence $\hat{\tau}^M(w) \leq \tau^M(w)$ predicted by (81) additionally requires that

$$\frac{\Gamma(\underline{w},w)}{\tilde{f}(w)} < \frac{F(w)}{f(w)},$$

as desired by (72).

Step 3. By using (75) and (60), we obtain

$$\begin{aligned}&\hat{\tau}^R(w) - \tau^R(w) \\ &= \underbrace{\frac{\partial U(w)}{\partial y(w)}}_{>0}\left\{\frac{\tilde{\theta}(w)}{c(w)}\left[y(w) - h\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w}h'\left(\frac{y(w)}{w}\right)\right] - \left[\frac{\Gamma(w,\overline{w})}{w\tilde{f}(w)} - \frac{1-F(w)}{wf(w)}\right]\right\}\end{aligned} \tag{82}$$

for $\forall w \in (w_m, \overline{w})$. We first define by using (82) that

$$\Theta^R(w) \equiv \frac{c(w)}{wl(w) - h(l(w)) + l(w)h'(l(w))} \left[\frac{\Gamma(w,\overline{w})}{w\tilde{f}(w)} - \frac{1-F(w)}{wf(w)} \right]. \tag{83}$$

Using (82) and (83) gives rise to

$$\hat{\tau}^R(w) \begin{cases} < \tau^R(w) & \text{if } \tilde{\theta}(w) < \Theta^R(w), \\ = \tau^R(w) & \text{if } \tilde{\theta}(w) = \Theta^R(w), \\ > \tau^R(w) & \text{if } \tilde{\theta}(w) > \Theta^R(w). \end{cases} \tag{84}$$

Since by definition we have $\tilde{\theta}(w) > 0$, hence $\hat{\tau}^R(w) \leq \tau^R(w)$ predicted by (84) additionally requires that

$$\frac{\Gamma(w,\overline{w})}{\tilde{f}(w)} > \frac{1-F(w)}{f(w)},$$

as desired by (72).

Step 4. Applying (76) and (60) reveals that

$$\begin{aligned} &\hat{\tau}^M(w) - \tau^R(w) \\ &\quad = \underbrace{-\frac{\partial U(w)}{\partial y(w)}}_{<0} \left\{ \underbrace{\frac{\Gamma(w,\overline{w})}{w\tilde{f}(w)} + \frac{F(w)}{wf(w)}}_{>0} - \frac{\tilde{\theta}(w)}{c(w)} \left[y(w) - h\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w} h'\left(\frac{y(w)}{w}\right) \right] \right\} \end{aligned} \tag{85}$$

for $\forall w \in (w_m, \overline{w})$. We first define by using (85) that

$$\Theta^{MR}(w) \equiv \frac{c(w)}{wl(w) - h(l(w)) + l(w)h'(l(w))} \left[\frac{\Gamma(w,\overline{w})}{w\tilde{f}(w)} + \frac{F(w)}{wf(w)} \right]. \tag{86}$$

Thus, using (85) and (86) gives rise to

$$\hat{\tau}^M(w) \begin{cases} < \tau^R(w) & \text{if } \tilde{\theta}(w) < \Theta^{MR}(w), \\ = \tau^R(w) & \text{if } \tilde{\theta}(w) = \Theta^{MR}(w), \\ > \tau^R(w) & \text{if } \tilde{\theta}(w) > \Theta^{MR}(w). \end{cases} \tag{87}$$

Step 5. Applying (75) and (78) reveals that

$$\begin{aligned} &\hat{\tau}^R(w) - \tau^M(w) \\ &\quad = \underbrace{\frac{\partial U(w)}{\partial y(w)}}_{>0} \left\{ \underbrace{\frac{\Gamma(\underline{w},w)}{w\tilde{f}(w)} + \frac{1-F(w)}{wf(w)}}_{>0} - \underbrace{\frac{\tilde{\theta}(w)}{c(w)} \left[h\left(\frac{y(w)}{w}\right) - y(w) - \frac{y(w)}{w} h'\left(\frac{y(w)}{w}\right) \right]}_{<0} \right\} \end{aligned} \tag{88}$$

for $\forall w \in (\underline{w}, w_m]$. Then, it is straightforward by (88) that $\hat{\tau}^R(w) > \tau^M(w)$. Finally, it is easy to verify by using equations (80), (83) and (86) that $\Theta^M(w) < \Theta^{MR}(w)$ and $\Theta^R(w) < \Theta^{MR}(w)$ for $\forall w \in (\underline{w}, \overline{w})$, as desired in part (i) of Lemma 7.4. ■

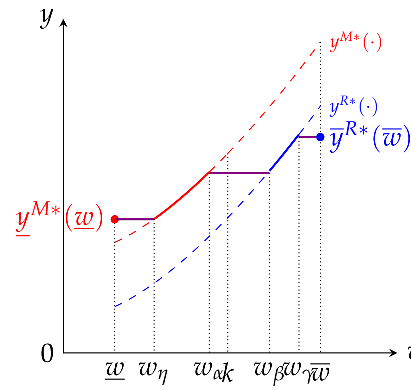

Figure 5: Income schedule with three bridges.

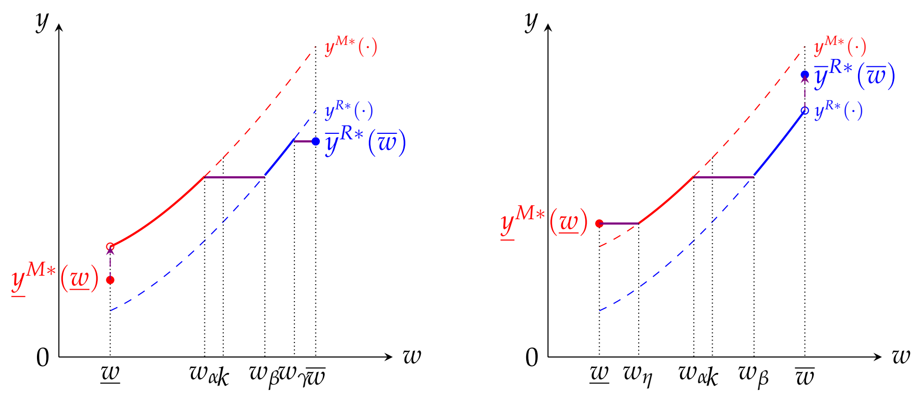

Figure 6: Income schedule with two bridges and an upward discontinuity.

Appendix B: The Complete Solution of Tax Design (Not for Publication!)

Following Brett and Weymark (2016, 2017), if either the maximin or the maximax income schedule obtained using the first-order approach fails to satisfy the SOIC condition (7), then it is necessary to bunch all types in a decreasing part of the schedule with some types who are in an increasing part. This kind of surgery is known as ironing, and any bunching region must be a closed interval.

Correspondingly, we let $\underline{y}^{M*}(\underline{w})$, $y^{M*}(\cdot)$, $y^{R*}(\cdot)$ and $\overline{y}^{M*}(\overline{w})$ denote the optimal maximax and maximin income schedules when the SOIC condition has been taken into account. We now show that it is optimal for the proposer of type k to build some bridges (see Figures 5-6), one of which includes her own type between the maximax and maximin parts of this schedule.

Theorem 7.1 *For any proposer of type $k \in (\underline{w}, \overline{w})$, the selfishly optimal schedule of pre-tax incomes, denoted by $y^*(\cdot)$, is given as follows.*

(i) If $U(\underline{w}) \in (-\overline{w}U'(\overline{w}), 0)$, then

$$y^*(w) = \begin{cases} \underline{y}^{M*}(\underline{w}) & \text{for } w \in [\underline{w}, w_\eta] \text{ if } w_\eta \leq w_\alpha, \\ y^{M*}(w) & \text{for } w \in (w_\eta, w_\alpha), \\ y^{M*}(w_\alpha) & \text{for } w \in [w_\alpha, w_\beta] \text{ if } w_\alpha > \underline{w}, \\ y^{R*}(w_\beta) & \text{for } w \in [w_\alpha, w_\beta] \text{ if } w_\beta < \overline{w}, \\ y^{R*}(w) & \text{for } w \in (w_\beta, w_\gamma), \\ \overline{y}^{R*}(\overline{w}) & \text{for } w \in [w_\gamma, \overline{w}] \text{ if } w_\gamma \geq w_\beta, \end{cases} \tag{89}$$

for some $w_\eta, w_\alpha, w_\beta, w_\gamma \in (\underline{w}, \overline{w})$ with $w_\alpha < w_\beta$ and $k \in [w_\alpha, w_\beta]$.

(ii) If $U(\underline{w}) \geq 0$, then

$$y^*(w) = \begin{cases} \underline{y}^{M*}(\underline{w}) & \text{for } w = \underline{w}, \\ y^{M*}(w) & \text{for } w \in (\underline{w}, w_\alpha), \\ y^{M*}(w_\alpha) & \text{for } w \in [w_\alpha, w_\beta] \text{ if } w_\alpha > \underline{w}, \\ y^{R*}(w_\beta) & \text{for } w \in [w_\alpha, w_\beta] \text{ if } w_\beta < \overline{w}, \\ y^{R*}(w) & \text{for } w \in (w_\beta, w_\gamma), \\ \overline{y}^{R*}(\overline{w}) & \text{for } w \in [w_\gamma, \overline{w}] \text{ if } w_\gamma \geq w_\beta, \end{cases} \tag{90}$$

for some $w_\alpha, w_\beta, w_\gamma \in (\underline{w}, \overline{w})$ with $w_\alpha < w_\beta$ and $k \in [w_\alpha, w_\beta]$.

(iii) If $U(\underline{w}) \leq -\overline{w}U'(\overline{w})$, then

$$y^*(w) = \begin{cases} \underline{y}^{M*}(\underline{w}) & \text{for } w \in [\underline{w}, w_\eta] \text{ if } w_\eta \leq w_\alpha, \\ y^{M*}(w) & \text{for } w \in (w_\eta, w_\alpha), \\ y^{M*}(w_\alpha) & \text{for } w \in [w_\alpha, w_\beta] \text{ if } w_\alpha > \underline{w}, \\ y^{R*}(w_\beta) & \text{for } w \in [w_\alpha, w_\beta] \text{ if } w_\beta < \overline{w}, \\ y^{R*}(w) & \text{for } w \in (w_\beta, \overline{w}), \\ \overline{y}^{R*}(\overline{w}) & \text{for } w = \overline{w}, \end{cases} \tag{91}$$

for some $w_\eta, w_\alpha, w_\beta \in (\underline{w}, \overline{w})$ with $w_\alpha < w_\beta$ and $k \in [w_\alpha, w_\beta]$.

Proof. We shall complete the proof in 5 steps.

Step 1. By Proposition 4.3 (ii), there always exists a downward discontinuity at $w = k$, which hence requires the surgery of ironing in the current complete solution. By Proposition 4.3 (i) and (iii), there would be two additional downward discontinuities at, respectively, $w = \underline{w}$ and $w = \overline{w}$ whenever $-\overline{w}U'(\overline{w}) < U(\underline{w}) < 0$ holds. This case corresponds to part (i) of Theorem 7.1. Similarly, based on Proposition 4.3, part (ii) of Theorem 7.1 considers the case with two discontinuities at $w = k$ and

$w = \overline{w}$, and part (iii) of Theorem 7.1 considers the case with two discontinuities at $w = k$ and $w = \underline{w}$. Importantly, it follows from Proposition 4.3 that at least one of these two endpoints $w = \underline{w}$ and $w = \overline{w}$ exhibits a downward discontinuity in the schedule derived under the first-order approach. Due to the similarity between these cases, here we just prove part (i) of Theorem 7.1 to economize on the space.

Step 2. Let's fix the bridge endpoints $w_\eta, w_\alpha, w_\beta$ and w_γ, and let $y^*(\underline{w}, w_\eta)$, $y^*(w_\alpha, w_\beta)$ and $y^*(w_\gamma, \overline{w})$ denote the optimal incomes on these bridges over income intervals $[\underline{w}, w_\eta]$, $[w_\alpha, w_\beta]$ and $[w_\gamma, \overline{w}]$, respectively. Without loss of generality, we assume that the bridge over $[w_\alpha, w_\beta]$ cannot begin in the interior of a bunching interval of the maximax schedule $y^{M*}(\cdot)$, nor can it end in the interior of a bunching interval of maximin schedule $y^{R*}(\cdot)$.

Step 3. In what follows, let $\mathcal{B}^M$ and $\mathcal{B}^R$ denote the types that are bunched with some other types in the complete solution to the maximax and maximin problems, respectively. Also, whenever w is bunched, we let interval $[w_-, w_+]$ denote the set of types bunched with w.

We now equivalently rewrite the maximand of problem (28) as follows

$$\begin{aligned} U(k) = &\int_{\underline{w}}^{k} \left\{ \left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \frac{\tilde{f}(w)}{\Gamma(\underline{w}, \overline{w})} + \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \left[1 - \frac{\Gamma(w, \overline{w})}{\Gamma(\underline{w}, \overline{w})} \right] \right\} dw \\ &+ \int_{k}^{\overline{w}} \left\{ \left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \frac{\tilde{f}(w)}{\Gamma(\underline{w}, \overline{w})} - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \frac{\Gamma(w, \overline{w})}{\Gamma(\underline{w}, \overline{w})} \right\} dw. \end{aligned} \tag{92}$$

Taking into account the bunching possibility, (92) should be modified as follows:

$$\begin{aligned} U^*(k) = &\int_{\underline{w}}^{k} \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w | w \notin \mathcal{B}^M\}} dw \\ &+ \int_{\underline{w}}^{k} \tilde{\Phi}^{M*}(w, y(w), \Gamma(w_-, w_+), \Gamma(w_-, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w | w \in \mathcal{B}^M\}} dw \\ &+ \int_{k}^{\overline{w}} \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w | w \notin \mathcal{B}^R\}} dw \\ &+ \int_{k}^{\overline{w}} \tilde{\Phi}^{R*}(w, y(w), \Gamma(w_-, w_+), \Gamma(w_+, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w | w \in \mathcal{B}^R\}} dw, \end{aligned} \tag{93}$$

in which

$$\begin{aligned} &\tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \\ \equiv &\left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \frac{\tilde{f}(w)}{\Gamma(\underline{w}, \overline{w})} + \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \left[1 - \frac{\Gamma(w, \overline{w})}{\Gamma(\underline{w}, \overline{w})} \right]; \\ &\tilde{\Phi}^{M*}(w, y(w), \Gamma(w_-, w_+), \Gamma(w_-, \overline{w}), \Gamma(\underline{w}, \overline{w})) \\ \equiv &\left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \frac{\Gamma(w_-, w_+)}{\Gamma(\underline{w}, \overline{w})} + \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \left[1 - \frac{\Gamma(w_-, \overline{w})}{\Gamma(\underline{w}, \overline{w})} \right] \end{aligned} \tag{94}$$

and

$$\begin{aligned} &\tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \\ \equiv &\left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \frac{\tilde{f}(w)}{\Gamma(\underline{w}, \overline{w})} - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \frac{\Gamma(w, \overline{w})}{\Gamma(\underline{w}, \overline{w})}; \\ &\tilde{\Phi}^{R*}(w, y(w), \Gamma(w_-, w_+), \Gamma(w_+, \overline{w}), \Gamma(\underline{w}, \overline{w})) \\ \equiv &\left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \frac{\Gamma(w_-, w_+)}{\Gamma(\underline{w}, \overline{w})} - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \frac{\Gamma(w_+, \overline{w})}{\Gamma(\underline{w}, \overline{w})} \end{aligned} \tag{95}$$

with $\mathbb{I}$ being a standard indicator function.

As ironing does not affect the solution outside a bunching region, no modifications to the integrands in (93) are needed for types that are not bunched. Departing from the first-order approach, if an extra unit of consumption is given to type-w workers, it must be given to all workers who are bunched with them, whose mass is $\Gamma(w_-, w_+)$. Also, if w is bunched, in the maximax case, some of this extra consumption can be reclaimed from workers of lower types than those bunched with w, whose mass is $\Gamma(\underline{w}, \overline{w}) - \Gamma(w_-, \overline{w})$. The corresponding workers in the maximin case are those workers of higher types than those bunched with w, whose mass is $\Gamma(w_+, \overline{w})$.

Step 4. Now, the selfishly optimal income schedule of proposer $k \in (\underline{w}, \overline{w})$ is obtained by solving the following problem:

$$\begin{aligned}
\max_{y(\cdot)} \Bigg\{ & \int_{\underline{w}}^{w_\eta} \tilde{\Phi}^{M*}(w, y(w), \Gamma(w_-, w_+), \Gamma(w_-, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta \in (\underline{w}, w_\alpha)\} \cap \{w_- = \underline{w}\} \cap \{w_+ = w_\eta\}} dw \\
& + \int_{w_\eta}^{w_\alpha} \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta \le w_\alpha\}} dw \\
& + \int_{w_\alpha}^{k} \tilde{\Phi}^{M*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\alpha, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}} dw \\
& + \int_{k}^{w_\beta} \tilde{\Phi}^{R*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\beta < \overline{w}\}} dw \\
& + \int_{w_\beta}^{w_\gamma} \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\gamma \ge w_\beta\}} dw \\
& + \int_{w_\gamma}^{\overline{w}} \tilde{\Phi}^{R*}(w, y(w), \Gamma(w_-, w_+), \Gamma(w_+, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\gamma \in (w_\beta, \overline{w})\} \cap \{w_- = w_\gamma\} \cap \{w_+ = \overline{w}\}} dw \Bigg\},
\end{aligned} \tag{96}$$

subject to

$$y(w) = \begin{cases} y^*(\underline{w}, w_\eta) & \text{for } w \in [\underline{w}, w_\eta], \\ y^*(w_\alpha, w_\beta) & \text{for } w \in [w_\alpha, w_\beta], \\ y^*(w_\gamma, \overline{w}) & \text{for } w \in [w_\gamma, \overline{w}]. \end{cases} \tag{97}$$

In consequence, for the current purpose, problem (96) can be simplified as the following unconstrained maximization problem:

$$\begin{aligned}
\max_{y(\cdot)} \Bigg\{ & \int_{w_\eta}^{w_\alpha} \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta \le w_\alpha\}} dw \\
& + \int_{w_\beta}^{w_\gamma} \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\gamma \ge w_\beta\}} dw \Bigg\}.
\end{aligned}$$

As is obvious, this problem can be solved point-wise, and the solution is given implicitly by the these first-order conditions:

$$\begin{aligned}
&\frac{\partial\tilde{\Phi}^{M*}(w,y(w),\tilde{f}(w),\Gamma(w,\overline{w}),\Gamma(\underline{w},\overline{w}))}{\partial y(w)}\\
&+\frac{\partial\tilde{\Phi}^{M*}(w,y(w),\tilde{f}(w),\Gamma(w,\overline{w}),\Gamma(\underline{w},\overline{w}))}{\partial\tilde{f}(w)}\frac{\partial\tilde{f}(w)}{\partial y(w)}\\
&+\frac{\partial\tilde{\Phi}^{M*}(w,y(w),\tilde{f}(w),\Gamma(w,\overline{w}),\Gamma(\underline{w},\overline{w}))}{\partial\Gamma(w,\overline{w})}\frac{\partial\Gamma(w,\overline{w})}{\partial y(w)}=0 \ \text{ for } \forall w\in(w_\eta,w_\alpha),
\end{aligned} \tag{98}$$

and

$$\begin{aligned}
&\frac{\partial\tilde{\Phi}^{R*}(w,y(w),\tilde{f}(w),\Gamma(w,\overline{w}),\Gamma(\underline{w},\overline{w}))}{\partial y(w)}\\
&+\frac{\partial\tilde{\Phi}^{R*}(w,y(w),\tilde{f}(w),\Gamma(w,\overline{w}),\Gamma(\underline{w},\overline{w}))}{\partial\tilde{f}(w)}\frac{\partial\tilde{f}(w)}{\partial y(w)}\\
&+\frac{\partial\tilde{\Phi}^{R*}(w,y(w),\tilde{f}(w),\Gamma(w,\overline{w}),\Gamma(\underline{w},\overline{w}))}{\partial\Gamma(w,\overline{w})}\frac{\partial\Gamma(w,\overline{w})}{\partial y(w)}=0 \ \text{ for } \forall w\in(w_\beta,w_\gamma).
\end{aligned} \tag{99}$$

As before, we denote the resulting solutions as $y^{M*}(\cdot)$ and $y^{R*}(\cdot)$, respectively.

Step 5. We now show that $y^*(w_\alpha,w_\beta)=y^{M*}(w_\alpha)$ if $w_\alpha>\underline{w}$ and $y^*(w_\alpha,w_\beta)=y^{R*}(w_\beta)$ if $w_\beta<\overline{w}$. Based on the above ironing procedure, we can apply the same reasoning used to prove the Proposition 3 of Brett and Weymark (2017) to show that $y^*(\cdot)$ is continuous on $[\underline{w},\overline{w}]$.

Suppose that there exists a type $k'>k$ for which $y^*(k')$ is not the maximin income, formally $y^*(k')\neq y^{R*}(k')$. The SOIC condition (7) must therefore bind at k', which implies that the slope of $y^*(\cdot)$ is zero at k'. Since $y^*(\cdot)$ is continuous, we obtain that there exists a $w_\beta>k'$ such that $y^*(\cdot)$ is constant on $[k,w_\beta]$ and coincides with the maximin income schedule $y^{R*}(\cdot)$ on $[w_\beta,w_\gamma]$. Similarly, if there exists a type $k'<k$ for which $y^*(k')$ is not the maximax income, formally $y^*(k')\neq y^{M*}(k')$, we can use the same argument to show that there exists a $w_\alpha<k'$ such that $y^*(\cdot)$ is constant on $[w_\alpha,k]$ and coincides with the maximax income schedule $y^{M*}(\cdot)$ on $[w_\eta,w_\alpha]$.

By further setting $y^*(\underline{w},w_\eta)\equiv\underline{y}^{M*}(\underline{w})$ and $y^*(w_\gamma,\overline{w})\equiv\overline{y}^{R*}(\overline{w})$ in (97), then the desired income schedule given by (89) is established. ■

Theorem 7.2 *For any proposer of type $k\in(\underline{w},\overline{w})$, the bridge endpoints are chosen as follows.*

(i) The optimal values of the bridge endpoints w_α and w_β are determined by the first-order condition (102) for $w_\beta<\overline{w}$ and by the first-order condition (118) for $w_\alpha>\underline{w}$.

(ii) The optimal value of the bridge endpoint w_η is the solution to equation $y^{M*}(w)=\underline{y}^{M*}(\underline{w})$, or $w_\eta=(y^{M*})^{-1}(\underline{y}^{M*}(\underline{w}))$.

(iii) The optimal value of the bridge endpoint w_γ is the solution to equation $y^{R*}(w)=\overline{y}^{R*}(\overline{w})$, or $w_\gamma=(y^{R*})^{-1}(\overline{y}^{R*}(\overline{w}))$.

Proof. We shall complete the proof in 4 steps.

Step 1. We first determine the optimal endpoints w_α and w_β of the bridge connecting the maximax income schedule and the maximin income schedule. Our proof employs the procedure developed by

Brett and Weymark (2017).

Suppose $w_\beta < \overline{w}$ holds. By continuity of income schedule $y^*(\cdot)$, we get from Theorem 7.1 that $y^*(w_\beta) = y^{R*}(w_\beta)$. Also, $y^*(w_\beta) = y^*(w_\alpha)$ because income is a constant on the bridge. If we also have $w_\alpha > \underline{w}$, then by continuity again, $y^*(w_\alpha) = y^{M*}(w_\alpha)$. Define

$$\psi(w_\beta) \equiv \begin{cases} (y^{M*})^{-1}(y^{R*}(w_\beta)) & \text{if } w_\alpha > \underline{w}, \\ w_\alpha & \text{if } w_\alpha = \underline{w}. \end{cases} \tag{100}$$

So we can write the proposer k's objective function of choosing w_β as follows:

$$\begin{aligned} \Xi(w_\beta; k) \equiv & \int_{w_\eta}^{\psi(w_\beta)} \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta \le w_\alpha\} \cap \{y(w) = y^{M*}(w)\}} dw \\ & + \int_{\psi(w_\beta)}^{k} \tilde{\Phi}^{M*}(w, y(w), \Gamma(\psi(w_\beta), w_\beta), \Gamma(\psi(w_\beta), \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\} \cap \{y(w) = y^{R*}(w_\beta)\}} dw \\ & + \int_{k}^{w_\beta} \tilde{\Phi}^{R*}(w, y(w), \Gamma(\psi(w_\beta), w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\beta < \overline{w}\} \cap \{y(w) = y^{R*}(w_\beta)\}} dw \\ & + \int_{w_\beta}^{w_\gamma} \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\gamma \ge w_\beta\} \cap \{y(w) = y^{R*}(w)\}} dw. \end{aligned} \tag{101}$$

Thus, the choice of w_β for any worker of type k is the solution to the maximization problem

$$\max_{w_\beta} \ \Xi(w_\beta; k).$$

Using (101), the first-order condition with respect to w_β can be derived as

$$\Psi_1 + \Psi_2(k) + \Psi_3(k) + \Psi_4 = 0, \tag{102}$$

in which

$$\begin{aligned} & \Psi_1 = \\ & \frac{\mathrm{d}\psi(w_\beta)}{\mathrm{d}w_\beta} \Big\{ \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta \le w_\alpha\} \cap \{y(w) = y^{M*}(w)\} \cap \{w = \psi(w_\beta)\}} \\ & - \tilde{\Phi}^{M*}(w, y(w), \Gamma(\psi(w_\beta), w_\beta), \Gamma(\psi(w_\beta), \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\} \cap \{y(w) = y^{R*}(w_\beta)\} \cap \{w = \psi(w_\beta)\}} \Big\}, \end{aligned} \tag{103}$$

$$
\begin{aligned}
&\Psi_2(k) = \\
&\int_{\psi(w_\beta)}^{k} \frac{\partial\tilde{\Phi}^{M*}(w, y^{R*}(w_\beta), \Gamma(\psi(w_\beta), w_\beta), \Gamma(\psi(w_\beta), \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial y^{R*}(w_\beta)} \frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta} \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}} dw \\
&+ \int_{\psi(w_\beta)}^{k} \frac{\partial\tilde{\Phi}^{M*}(w, y^{R*}(w_\beta), \Gamma(\psi(w_\beta), w_\beta), \Gamma(\psi(w_\beta), \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial\Gamma(\psi(w_\beta), w_\beta)} \frac{\partial\Gamma(\psi(w_\beta), w_\beta)}{\partial w_\beta} \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}} dw \\
&+ \int_{\psi(w_\beta)}^{k} \frac{\partial\tilde{\Phi}^{M*}(w, y^{R*}(w_\beta), \Gamma(\psi(w_\beta), w_\beta), \Gamma(\psi(w_\beta), \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial\Gamma(\psi(w_\beta), \overline{w})} \frac{\partial\Gamma(\psi(w_\beta), \overline{w})}{\partial w_\beta} \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}} dw \\
&\equiv \int_{\psi(w_\beta)}^{k} [\Psi_{21}(w) + \Psi_{22}(w) + \Psi_{23}(w)] \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}} dw,
\end{aligned} \tag{104}
$$

$$
\begin{aligned}
&\Psi_3(k) = \\
&\int_{k}^{w_\beta} \frac{\partial\tilde{\Phi}^{R*}(w, y^{R*}(w_\beta), \Gamma(\psi(w_\beta), w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial y^{R*}(w_\beta)} \frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta} \cdot \mathbb{I}_{\{w_\beta < \overline{w}\}} dw \\
&+ \int_{k}^{w_\beta} \frac{\partial\tilde{\Phi}^{R*}(w, y^{R*}(w_\beta), \Gamma(\psi(w_\beta), w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial\Gamma(\psi(w_\beta), w_\beta)} \frac{\partial\Gamma(\psi(w_\beta), w_\beta)}{\partial w_\beta} \cdot \mathbb{I}_{\{w_\beta < \overline{w}\}} dw \\
&+ \int_{k}^{w_\beta} \frac{\partial\tilde{\Phi}^{R*}(w, y^{R*}(w_\beta), \Gamma(\psi(w_\beta), w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial\Gamma(w_\beta, \overline{w})} \frac{\partial\Gamma(w_\beta, \overline{w})}{\partial w_\beta} \cdot \mathbb{I}_{\{w_\beta < \overline{w}\}} dw \\
&\equiv \int_{k}^{w_\beta} [\Psi_{31}(w) + \Psi_{32}(w) + \Psi_{33}(w)] \cdot \mathbb{I}_{\{w_\beta < \overline{w}\}} dw,
\end{aligned} \tag{105}
$$

and

$$
\begin{aligned}
\Psi_4 = {} & \tilde{\Phi}^{R*}(w, y(w), \Gamma(\psi(w_\beta), w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\beta < \overline{w}\} \cap \{y(w) = y^{R*}(w_\beta)\} \cap \{w = w_\beta\}} \\
& - \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\gamma \geq w_\beta\} \cap \{y(w) = y^{R*}(w)\} \cap \{w = w_\beta\}}.
\end{aligned} \tag{106}
$$

Step 2. By using (94) and (95), we can have

$$
\begin{aligned}
&\tilde{\Phi}^{M*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\alpha, \overline{w}), \Gamma(\underline{w}, \overline{w})) \\
&\equiv \left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \frac{\Gamma(w_\alpha, w_\beta)}{\Gamma(\underline{w}, \overline{w})} + \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \left[1 - \frac{\Gamma(w_\alpha, \overline{w})}{\Gamma(\underline{w}, \overline{w})} \right], \\
&\tilde{\Phi}^{R*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w})) \\
&\equiv \left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \frac{\Gamma(w_\alpha, w_\beta)}{\Gamma(\underline{w}, \overline{w})} - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \frac{\Gamma(w_\beta, \overline{w})}{\Gamma(\underline{w}, \overline{w})}
\end{aligned} \tag{107}
$$

for $\forall w \in [w_\alpha, w_\beta]$.

By using (100), (107), (94) and (95), we can rewrite (103) as

$$
\begin{aligned}
\Psi_1 =& \frac{d\psi(w_\beta)}{dw_\beta} \Big\{ \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta \le w_\alpha\} \cap \{y(w) = y^{R*}(w_\beta)\} \cap \{w = \psi(w_\beta)\}} \\
& - \tilde{\Phi}^{M*}(w, y(w), \Gamma(w, w_\beta), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\} \cap \{y(w) = y^{R*}(w_\beta)\} \cap \{w = \psi(w_\beta)\}} \Big\} \\
=& \frac{d\psi(w_\beta)}{dw_\beta} \Big\{ \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \\
& - \tilde{\Phi}^{M*}(w, y(w), \Gamma(w, w_\beta), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \Big\} \cdot \mathbb{I}_{\{w_\eta \le w_\alpha\} \cap \{w_\alpha > \underline{w}\} \cap \{y(w) = y^{R*}(w_\beta)\} \cap \{w = \psi(w_\beta)\}} \\
=& \frac{d\psi(w_\beta)}{dw_\beta} \left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \frac{\tilde{f}(w) - \Gamma(w, w_\beta)}{\Gamma(\underline{w}, \overline{w})} \cdot \mathbb{I}_{\{w_\eta \le w_\alpha\} \cap \{w_\alpha > \underline{w}\} \cap \{y(w) = y^{R*}(w_\beta)\} \cap \{w = \psi(w_\beta)\}} \\
=& \frac{d\psi(w_\beta)}{dw_\beta} \left[y^{R*}(w_\beta) - h\left(\frac{y^{R*}(w_\beta)}{\psi(w_\beta)}\right) \right] \frac{\tilde{f}(\psi(w_\beta)) - \Gamma(\psi(w_\beta), w_\beta)}{\Gamma(\underline{w}, \overline{w})} \cdot \mathbb{I}_{\{w_\eta \le w_\alpha\} \cap \{w_\alpha > \underline{w}\}}.
\end{aligned}
\tag{108}
$$

By using (104) and (107), we have

$$
\begin{aligned}
\Psi_{21}(w) = & \Bigg\{ \left[1 - \frac{1}{w} h'\left(\frac{y^{R*}(w_\beta)}{w}\right) \right] \frac{\Gamma(\psi(w_\beta), w_\beta)}{\Gamma(\underline{w}, \overline{w})} \\
& + \left[\frac{1}{w^2} h'\left(\frac{y^{R*}(w_\beta)}{w}\right) + \frac{y^{R*}(w_\beta)}{w^3} h''\left(\frac{y^{R*}(w_\beta)}{w}\right) \right] \left[1 - \frac{\Gamma(\psi(w_\beta), \overline{w})}{\Gamma(\underline{w}, \overline{w})} \right] \Bigg\} \frac{dy^{R*}(w_\beta)}{dw_\beta},
\end{aligned}
\tag{109}
$$

$$
\begin{aligned}
\Psi_{22}(w) = & \left[y^{R*}(w_\beta) - h\left(\frac{y^{R*}(w_\beta)}{w}\right) \right] \frac{1}{\Gamma(\underline{w}, \overline{w})} \\
& \times \left\{ \tilde{f}(w_\beta) - \tilde{f}(\psi(w_\beta)) \frac{d\psi(w_\beta)}{dw_\beta} + \left[\int_{\psi(w_\beta)}^{w_\beta} \frac{\partial \tilde{f}(w)}{\partial y^{R*}(w_\beta)} \frac{dy^{R*}(w_\beta)}{dw_\beta} dw \right] \right\},
\end{aligned}
\tag{110}
$$

and

$$
\Psi_{23}(w) = \frac{y^{R*}(w_\beta)}{w^2} h'\left(\frac{y^{R*}(w_\beta)}{w}\right) \frac{\tilde{f}(\psi(w_\beta))}{\Gamma(\underline{w}, \overline{w})} \frac{d\psi(w_\beta)}{dw_\beta}.
\tag{111}
$$

Similarly, By using (105) and (107), we have

$$
\begin{aligned}
\Psi_{31}(w) = & \Bigg\{ \left[1 - \frac{1}{w} h'\left(\frac{y^{R*}(w_\beta)}{w}\right) \right] \frac{\Gamma(\psi(w_\beta), w_\beta)}{\Gamma(\underline{w}, \overline{w})} \\
& - \left[\frac{1}{w^2} h'\left(\frac{y^{R*}(w_\beta)}{w}\right) + \frac{y^{R*}(w_\beta)}{w^3} h''\left(\frac{y^{R*}(w_\beta)}{w}\right) \right] \frac{\Gamma(w_\beta, \overline{w})}{\Gamma(\underline{w}, \overline{w})} \Bigg\} \frac{dy^{R*}(w_\beta)}{dw_\beta},
\end{aligned}
\tag{112}
$$

$$
\begin{aligned}
\Psi_{32}(w) = & \left[y^{R*}(w_\beta) - h\left(\frac{y^{R*}(w_\beta)}{w}\right) \right] \frac{1}{\Gamma(\underline{w}, \overline{w})} \\
& \times \left\{ \tilde{f}(w_\beta) - \tilde{f}(\psi(w_\beta)) \frac{d\psi(w_\beta)}{dw_\beta} + \left[\int_{\psi(w_\beta)}^{w_\beta} \frac{\partial \tilde{f}(w)}{\partial y^{R*}(w_\beta)} \frac{dy^{R*}(w_\beta)}{dw_\beta} dw \right] \right\},
\end{aligned}
\tag{113}
$$

and

$$
\Psi_{33}(w) = \frac{y^{R*}(w_\beta)}{w^2} h'\left(\frac{y^{R*}(w_\beta)}{w}\right) \frac{\tilde{f}(w_\beta)}{\Gamma(\underline{w}, \overline{w})}.
\tag{114}
$$

Finally, by using (107), we can rewrite (106) as

$$\begin{aligned}\Psi_4 &= \tilde{\Phi}^{R*}(w, y(w), \Gamma(\psi(w), w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\beta < \overline{w}\} \cap \{y(w) = y^{R*}(w_\beta)\} \cap \{w = w_\beta\}} \\ &\quad - \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\gamma \geq w_\beta\} \cap \{y(w) = y^{R*}(w)\} \cap \{w = w_\beta\}} \\ &= \left[\tilde{\Phi}^{R*}(w, y(w), \Gamma(\psi(w), w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) - \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w}))\right] \\ &\quad \times \mathbb{I}_{\{w_\gamma \geq w_\beta\} \cap \{w_\beta < \overline{w}\} \cap \{y(w) = y^{R*}(w)\} \cap \{w = w_\beta\}} \\ &= \left\{\left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \frac{\Gamma(\psi(w), w) - \tilde{f}(w)}{\Gamma(\underline{w}, \overline{w})}\right\} \mathbb{I}_{\{w_\gamma \geq w_\beta\} \cap \{w_\beta < \overline{w}\} \cap \{y(w) = y^{R*}(w)\} \cap \{w = w_\beta\}}.\end{aligned} \tag{115}$$

Step 3. Suppose $w_\alpha > \underline{w}$ holds. By continuity of income schedule $y^*(\cdot)$, we get from Theorem 7.1 that $y^*(w_\alpha) = y^{M*}(w_\alpha)$. Also, $y^*(w_\beta) = y^*(w_\alpha)$ because income is a constant on the bridge. If we also have $w_\beta < \overline{w}$, then by continuity again, $y^*(w_\beta) = y^{R*}(w_\beta)$. Define

$$\varphi(w_\alpha) \equiv \begin{cases} (y^{R*})^{-1}(y^{M*}(w_\alpha)) & \text{if } w_\beta < \overline{w}, \\ w_\beta & \text{if } w_\beta = \overline{w}. \end{cases} \tag{116}$$

So we can write the proposer k's objective function of choosing w_α as follows:

$$\begin{aligned}\Xi(w_\alpha; k) \equiv &\int_{w_\eta}^{w_\alpha} \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta \leq w_\alpha\} \cap \{y(w) = y^{M*}(w)\}} dw \\ &+ \int_{w_\alpha}^{k} \tilde{\Phi}^{M*}(w, y(w), \Gamma(w_\alpha, \varphi(w_\alpha)), \Gamma(w_\alpha, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\} \cap \{y(w) = y^{M*}(w_\alpha)\}} dw \\ &+ \int_{k}^{\varphi(w_\alpha)} \tilde{\Phi}^{R*}(w, y(w), \Gamma(w_\alpha, \varphi(w_\alpha)), \Gamma(\varphi(w_\alpha), \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\beta < \overline{w}\} \cap \{y(w) = y^{M*}(w_\alpha)\}} dw \\ &+ \int_{\varphi(w_\alpha)}^{w_\gamma} \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\gamma \geq w_\beta\} \cap \{y(w) = y^{R*}(w)\}} dw.\end{aligned} \tag{117}$$

Thus, the choice of w_α for any worker of type k is the solution to the maximization problem

$$\max_{w_\alpha} \; \Xi(w_\alpha; k).$$

Using (117), the first-order condition with respect to w_α can be derived as

$$\Lambda_1 + \Lambda_2(k) + \Lambda_3(k) + \Lambda_4 = 0, \tag{118}$$

in which

$$\begin{aligned}\Lambda_1 &= \left[\tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) - \tilde{\Phi}^{M*}(w, y(w), \Gamma(w, \varphi(w)), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w}))\right] \\ &\quad \times \mathbb{I}_{\{w_\eta \leq w_\alpha\} \cap \{w_\alpha > \underline{w}\} \cap \{y(w) = y^{M*}(w)\} \cap \{w = w_\alpha\}} \\ &= \left\{\left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \frac{\tilde{f}(w) - \Gamma(w, \varphi(w))}{\Gamma(\underline{w}, \overline{w})}\right\} \mathbb{I}_{\{w_\eta \leq w_\alpha\} \cap \{w_\alpha > \underline{w}\} \cap \{y(w) = y^{M*}(w)\} \cap \{w = w_\alpha\}},\end{aligned} \tag{119}$$

$$\Lambda_2(k) = \int_{w_\alpha}^{k} \left[\Lambda_{21}(w) + \Lambda_{22}(w) + \Lambda_{23}(w)\right] \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}} dw, \tag{120}$$

with

$$\begin{aligned} \Lambda_{21}(w) = &\left\{ \left[1 - \frac{1}{w} h'\left(\frac{y^{M*}(w_\alpha)}{w}\right)\right] \frac{\Gamma(w_\alpha, \varphi(w_\alpha))}{\Gamma(\underline{w}, \overline{w})} \right. \\ &\left. + \left[\frac{1}{w^2} h'\left(\frac{y^{M*}(w_\alpha)}{w}\right) + \frac{y^{M*}(w_\alpha)}{w^3} h''\left(\frac{y^{M*}(w_\alpha)}{w}\right)\right] \left[1 - \frac{\Gamma(w_\alpha, \overline{w})}{\Gamma(\underline{w}, \overline{w})}\right] \right\} \frac{\mathrm{d}y^{M*}(w_\alpha)}{\mathrm{d}w_\alpha}, \end{aligned} \tag{121}$$

$$\begin{aligned} \Lambda_{22}(w) = &\left[y^{M*}(w_\alpha) - h\left(\frac{y^{M*}(w_\alpha)}{w}\right)\right] \frac{1}{\Gamma(\underline{w}, \overline{w})} \\ &\times \left\{ \tilde{f}\left(\varphi(w_\alpha)\right) \frac{\mathrm{d}\varphi(w_\alpha)}{\mathrm{d}w_\alpha} - \tilde{f}(w_\alpha) + \left[\int_{w_\alpha}^{\varphi(w_\alpha)} \frac{\partial \tilde{f}(w)}{\partial y^{M*}(w_\alpha)} \frac{\mathrm{d}y^{M*}(w_\alpha)}{\mathrm{d}w_\alpha} dw\right] \right\}, \end{aligned} \tag{122}$$

and

$$\Lambda_{23}(w) = \frac{y^{M*}(w_\alpha)}{w^2} h'\left(\frac{y^{M*}(w_\alpha)}{w}\right) \frac{\tilde{f}\left(w_\alpha\right)}{\Gamma(\underline{w}, \overline{w})}; \tag{123}$$

$$\Lambda_3(k) = \int_{k}^{\varphi(w_\alpha)} \left[\Lambda_{31}(w) + \Lambda_{32}(w) + \Lambda_{33}(w)\right] \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\}} dw, \tag{124}$$

with

$$\begin{aligned} \Lambda_{31}(w) = &\left\{ \left[1 - \frac{1}{w} h'\left(\frac{y^{M*}(w_\alpha)}{w}\right)\right] \frac{\Gamma(w_\alpha, \varphi(w_\alpha))}{\Gamma(\underline{w}, \overline{w})} \right. \\ &\left. - \left[\frac{1}{w^2} h'\left(\frac{y^{M*}(w_\alpha)}{w}\right) + \frac{y^{M*}(w_\alpha)}{w^3} h''\left(\frac{y^{M*}(w_\alpha)}{w}\right)\right] \frac{\Gamma(\varphi(w_\alpha), \overline{w})}{\Gamma(\underline{w}, \overline{w})} \right\} \frac{\mathrm{d}y^{M*}(w_\alpha)}{\mathrm{d}w_\alpha}, \end{aligned} \tag{125}$$

$$\begin{aligned} \Lambda_{32}(w) = &\left[y^{M*}(w_\alpha) - h\left(\frac{y^{M*}(w_\alpha)}{w}\right)\right] \frac{1}{\Gamma(\underline{w}, \overline{w})} \\ &\times \left\{ \tilde{f}\left(\varphi(w_\alpha)\right) \frac{\mathrm{d}\varphi(w_\alpha)}{\mathrm{d}w_\alpha} - \tilde{f}(w_\alpha) + \left[\int_{w_\alpha}^{\varphi(w_\alpha)} \frac{\partial \tilde{f}(w)}{\partial y^{M*}(w_\alpha)} \frac{\mathrm{d}y^{M*}(w_\alpha)}{\mathrm{d}w_\alpha} dw\right] \right\}, \end{aligned} \tag{126}$$

and

$$\Lambda_{33}(w) = \frac{y^{M*}(w_\alpha)}{w^2} h'\left(\frac{y^{M*}(w_\alpha)}{w}\right) \frac{\tilde{f}\left(\varphi(w_\alpha)\right)}{\Gamma(\underline{w}, \overline{w})} \frac{\mathrm{d}\varphi(w_\alpha)}{\mathrm{d}w_\alpha}; \tag{127}$$

and finally

$$\begin{aligned} \Lambda_4 = &\frac{\mathrm{d}\varphi(w_\alpha)}{\mathrm{d}w_\alpha} \times \\ &[\tilde{\Phi}^{R*}(w, y(w), \Gamma(w_\alpha, w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\} \cap \left\{y(w) = y^{M*}(w_\alpha)\right\} \cap \{w = \varphi(w_\alpha)\}} \\ &- \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{w_\gamma \geq w_\beta\right\} \cap \left\{y(w) = y^{M*}(w_\alpha)\right\} \cap \{w = \varphi(w_\alpha)\}}] \\ = &\left[\tilde{\Phi}^{R*}(w, y(w), \Gamma(w_\alpha, w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) - \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w}))\right] \\ &\times \mathbb{I}_{\left\{w_\gamma \geq w_\beta\right\} \cap \left\{w_\beta < \overline{w}\right\} \cap \left\{y(w) = y^{M*}(w_\alpha)\right\} \cap \{w = \varphi(w_\alpha)\}} \frac{\mathrm{d}\varphi(w_\alpha)}{\mathrm{d}w_\alpha} \\ = &\left\{ \left[y^{M*}(w_\alpha) - h\left(\frac{y^{M*}(w_\alpha)}{\varphi(w_\alpha)}\right)\right] \frac{\Gamma\left(w_\alpha, \varphi(w_\alpha)\right) - \tilde{f}(\varphi(w_\alpha))}{\Gamma\left(\underline{w}, \overline{w}\right)} \right\} \frac{\mathrm{d}\varphi(w_\alpha)}{\mathrm{d}w_\alpha} \mathbb{I}_{\left\{w_\gamma \geq w_\beta\right\} \cap \left\{w_\beta < \overline{w}\right\}}. \end{aligned} \tag{128}$$

Step 4. By using (96), we can establish the first-order conditions implicitly solving for $y(\underline{w})$ and $y(\overline{w})$, and we denote the solutions as $\underline{y}^{M*}(\underline{w})$ and $\overline{y}^{R*}(\overline{w})$, respectively. Since for each of these two bridges, one of the endpoints is already fixed, the other endpoint is actually fixed given the established optimal income schedule. That is, w_η is determined by setting $y^{M*}(w_\eta) = \underline{y}^{M*}(\underline{w})$ and w_γ is determined by setting $y^{R*}(w_\gamma) = \overline{y}^{R*}(\overline{w})$. ∎

Proposition 7.1 *If the following condition*

$$\begin{cases} \frac{\mathrm{d}w_\alpha}{\mathrm{d}w_\beta} \geq \frac{\tilde{f}(w_\beta)}{\tilde{f}(w_\alpha)} & \textit{for } w_\beta < \overline{w}, \\ \frac{\tilde{f}(w_\alpha)}{\tilde{f}(w_\beta)} \geq \frac{\mathrm{d}w_\beta}{\mathrm{d}w_\alpha} & \textit{for } w_\alpha > \underline{w} \end{cases} \tag{129}$$

holds, then the bridge endpoints $w_\alpha(k)$ and $w_\beta(k)$ are nondecreasing in k for $\forall k \in [\underline{w}, \overline{w}]$.

Proof. We shall complete the proof in 2 steps.

Step 1. We first consider the case with $w_\beta < \overline{w}$. It follows from (101) and (102) that

$$\frac{\partial^2 \Xi(w_\beta; k)}{\partial w_\beta \partial k} = \frac{\mathrm{d}\Psi_2(k)}{\mathrm{d}k} + \frac{\mathrm{d}\Psi_3(k)}{\mathrm{d}k}. \tag{130}$$

By using equations (109)-(114), (130) can be explicitly expressed as

$$\begin{aligned} & \frac{\mathrm{d}\Psi_2(k)}{\mathrm{d}k} + \frac{\mathrm{d}\Psi_3(k)}{\mathrm{d}k} \\ = \ & [\Psi_{21}(k) + \Psi_{22}(k) + \Psi_{23}(k)] - [\Psi_{31}(k) + \Psi_{32}(k) + \Psi_{33}(k)] \\ = \ & [\Psi_{21}(k) - \Psi_{31}(k)] + \underbrace{[\Psi_{22}(k) - \Psi_{32}(k)]}_{=0} + [\Psi_{23}(k) - \Psi_{33}(k)] \\ = \ & \left[\frac{1}{k^2} h'\left(\frac{y^{R*}(w_\beta)}{k}\right) + \frac{y^{R*}(w_\beta)}{k^3} h''\left(\frac{y^{R*}(w_\beta)}{k}\right)\right] \left[1 - \frac{\Gamma(\psi\left(w_\beta\right), \overline{w})}{\Gamma(\underline{w}, \overline{w})} + \frac{\Gamma(w_\beta, \overline{w})}{\Gamma(\underline{w}, \overline{w})}\right] \frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta} \\ & + \frac{y^{R*}(w_\beta)}{k^2} h'\left(\frac{y^{R*}(w_\beta)}{k}\right) \frac{1}{\Gamma(\underline{w}, \overline{w})} \left[\tilde{f}(\psi(w_\beta)) \frac{\mathrm{d}\psi(w_\beta)}{\mathrm{d}w_\beta} - \tilde{f}(w_\beta)\right]. \end{aligned} \tag{131}$$

Since $\mathrm{d}y^{R*}(w_\beta)/\mathrm{d}w_\beta > 0$ by assumption, we have

$$\left[\frac{1}{k^2} h'\left(\frac{y^{R*}(w_\beta)}{k}\right) + \frac{y^{R*}(w_\beta)}{k^3} h''\left(\frac{y^{R*}(w_\beta)}{k}\right)\right] \left[1 - \frac{\Gamma(\psi\left(w_\beta\right), \overline{w})}{\Gamma(\underline{w}, \overline{w})} + \frac{\Gamma(w_\beta, \overline{w})}{\Gamma(\underline{w}, \overline{w})}\right] \frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta} > 0.$$

As such

$$\frac{\partial^2 \Xi(w_\beta; k)}{\partial w_\beta \partial k} = \frac{\mathrm{d}\Psi_2(k)}{\mathrm{d}k} + \frac{\mathrm{d}\Psi_3(k)}{\mathrm{d}k} > 0 \tag{132}$$

whenever

$$\tilde{f}(\psi(w_\beta)) \frac{\mathrm{d}\psi(w_\beta)}{\mathrm{d}w_\beta} \geq \tilde{f}(w_\beta),$$

as desired. In particular, $\mathrm{d}\psi(w_\beta)/\mathrm{d}w_\beta > 0$ based on the construction of $\psi(\cdot)$ given in the proof of

Theorem 7.2 as well as the monotonicity of the income schedule. In the case of (132), $\Xi(w_\beta;k)$ is a supermodular function, and an application of Topkis Theorem (see Topkis, 1978, Theorem 6.1) implies that $w_\beta(k)$ is nondecreasing in k.

Step 2. We now consider the case with $w_\alpha > \underline{w}$. It follows from (117) and (118) that

$$\frac{\partial^2 \Xi(w_\alpha;k)}{\partial w_\alpha \partial k} = \frac{\mathrm{d}\Lambda_2(k)}{\mathrm{d}k} + \frac{\mathrm{d}\Lambda_3(k)}{\mathrm{d}k}. \tag{133}$$

By using equations (120)-(127), (133) can be explicitly expressed as

$$\begin{aligned}
&\frac{\mathrm{d}\Lambda_2(k)}{\mathrm{d}k} + \frac{\mathrm{d}\Lambda_3(k)}{\mathrm{d}k} \\
=\ & [\Lambda_{21}(k) + \Lambda_{22}(k) + \Lambda_{23}(k)] - [\Lambda_{31}(k) + \Lambda_{32}(k) + \Lambda_{33}(k)] \\
=\ & [\Lambda_{21}(k) - \Lambda_{31}(k)] + \underbrace{[\Lambda_{22}(k) - \Lambda_{32}(k)]}_{=0} + [\Lambda_{23}(k) - \Lambda_{33}(k)] \\
=\ & \left[\frac{1}{k^2} h'\left(\frac{y^{M*}(w_\alpha)}{k}\right) + \frac{y^{M*}(w_\alpha)}{k^3} h''\left(\frac{y^{M*}(w_\alpha)}{k}\right)\right] \left[1 - \frac{\Gamma(w_\alpha, \overline{w})}{\Gamma(\underline{w}, \overline{w})} + \frac{\Gamma(\varphi(w_\alpha), \overline{w})}{\Gamma(\underline{w}, \overline{w})}\right] \frac{\mathrm{d}y^{M*}(w_\alpha)}{\mathrm{d}w_\alpha} \\
& + \frac{y^{M*}(w_\alpha)}{k^2} h'\left(\frac{y^{M*}(w_\alpha)}{k}\right) \frac{1}{\Gamma(\underline{w}, \overline{w})} \left[\tilde{f}(w_\alpha) - \tilde{f}(\varphi(w_\alpha)) \frac{\mathrm{d}\varphi(w_\alpha)}{\mathrm{d}w_\alpha}\right].
\end{aligned} \tag{134}$$

Since $\mathrm{d}y^{M*}(w_\alpha)/\mathrm{d}w_\alpha > 0$ by assumption, we have by (134) that

$$\frac{\partial^2 \Xi(w_\alpha;k)}{\partial w_\alpha \partial k} = \frac{\mathrm{d}\Lambda_2(k)}{\mathrm{d}k} + \frac{\mathrm{d}\Lambda_3(k)}{\mathrm{d}k} > 0 \tag{135}$$

whenever

$$\tilde{f}(w_\alpha) \geq \tilde{f}(\varphi(w_\alpha)) \frac{\mathrm{d}\varphi(w_\alpha)}{\mathrm{d}w_\alpha},$$

as desired in (129). In the case of (135), $\Xi(w_\alpha;k)$ is a supermodular function, and an application of Topkis Theorem (see Topkis, 1978, Theorem 6.1) implies that $w_\alpha(k)$ is nondecreasing in k. ∎

Proposition 7.2 *The bridge endpoint $w_\eta(k)$ is decreasing whereas the bridge endpoint $w_\gamma(k)$ is increasing in the type k of the proposer.*

Proof. We shall complete the proofs in 3 steps.

Step 1. By using (94), we have

$$\begin{aligned}
&\tilde{\Phi}^{M*}(w, y(w), \Gamma(w_-, w_+), \Gamma(w_-, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta \in (\underline{w}, w_\alpha)\} \cap \{w_- = \underline{w}\} \cap \{w_+ = w_\eta\}} \\
= & \left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \frac{\Gamma(\underline{w}, w_\eta)}{\Gamma(\underline{w}, \overline{w})} \cdot \mathbb{I}_{\{w_\eta \in (\underline{w}, w_\alpha)\}}.
\end{aligned} \tag{136}$$

By using (95), we have

$$\begin{aligned}&\tilde{\Phi}^{R*}(w,y(w),\Gamma(w_-,w_+),\Gamma(w_+,\overline{w}),\Gamma(\underline{w},\overline{w}))\cdot\mathbb{I}_{\{w_\gamma\in(w_\beta,\overline{w})\}\cap\{w_-=w_\gamma\}\cap\{w_+=\overline{w}\}}\\=&\left[y(w)-h\left(\frac{y(w)}{w}\right)\right]\frac{\Gamma(w_\gamma,\overline{w})}{\Gamma(\underline{w},\overline{w})}\cdot\mathbb{I}_{\{w_\gamma\in(w_\beta,\overline{w})\}}.\end{aligned}\tag{137}$$

Substituting (136) and (137) into (96) results in

$$\begin{aligned}U^*(k)=&\int_{\underline{w}}^{w_\eta}\left[y(w)-h\left(\frac{y(w)}{w}\right)\right]\frac{\Gamma(\underline{w},w_\eta)}{\Gamma(\underline{w},\overline{w})}\cdot\mathbb{I}_{\{w_\eta\in(\underline{w},w_\alpha)\}}dw\\&+\int_{w_\eta}^{w_\alpha}\tilde{\Phi}^{M*}(w,y(w),\tilde{f}(w),\Gamma(w,\overline{w}),\Gamma(\underline{w},\overline{w}))\cdot\mathbb{I}_{\{w_\eta\le w_\alpha\}}dw\\&+\int_{w_\alpha}^{k}\tilde{\Phi}^{M*}(w,y(w),\Gamma(w_\alpha,w_\beta),\Gamma(w_\alpha,\overline{w}),\Gamma(\underline{w},\overline{w}))\cdot\mathbb{I}_{\{w_\alpha>\underline{w}\}}dw\\&+\int_{k}^{w_\beta}\tilde{\Phi}^{R*}(w,y(w),\Gamma(w_\alpha,w_\beta),\Gamma(w_\beta,\overline{w}),\Gamma(\underline{w},\overline{w}))\cdot\mathbb{I}_{\{w_\beta<\overline{w}\}}dw\\&+\int_{w_\beta}^{w_\gamma}\tilde{\Phi}^{R*}(w,y(w),\tilde{f}(w),\Gamma(w,\overline{w}),\Gamma(\underline{w},\overline{w}))\cdot\mathbb{I}_{\{w_\gamma\ge w_\beta\}}dw\\&+\int_{w_\gamma}^{\overline{w}}\left[y(w)-h\left(\frac{y(w)}{w}\right)\right]\frac{\Gamma(w_\gamma,\overline{w})}{\Gamma(\underline{w},\overline{w})}\cdot\mathbb{I}_{\{w_\gamma\in(w_\beta,\overline{w})\}}dw.\end{aligned}\tag{138}$$

Step 2. With respect to $y(\underline{w})$, we get from (138) and (107) that

$$\begin{aligned}\frac{\partial^2U^*(k)}{\partial y(\underline{w})\partial k}=&\frac{\partial\tilde{\Phi}^{M*}(w,y(w),\Gamma(w_\alpha,w_\beta),\Gamma(w_\alpha,\overline{w}),\Gamma(\underline{w},\overline{w}))}{\partial\Gamma(\underline{w},\overline{w})}\frac{\partial\Gamma(\underline{w},\overline{w})}{\partial y(\underline{w})}\cdot\mathbb{I}_{\{w_\alpha>\underline{w}\}\cap\{w=k\}}\\&-\frac{\partial\tilde{\Phi}^{R*}(w,y(w),\Gamma(w_\alpha,w_\beta),\Gamma(w_\beta,\overline{w}),\Gamma(\underline{w},\overline{w}))}{\partial\Gamma(\underline{w},\overline{w})}\frac{\partial\Gamma(\underline{w},\overline{w})}{\partial y(\underline{w})}\cdot\mathbb{I}_{\{w_\beta<\overline{w}\}\cap\{w=k\}}\\=&\underbrace{\frac{\partial\Gamma(\underline{w},\overline{w})}{\partial y(\underline{w})}}_{<0}\cdot\frac{y(k)}{k^2}h'\left(\frac{y(k)}{k}\right)\frac{1}{\Gamma(\underline{w},\overline{w})^2}\cdot\underbrace{\left[\Gamma(w_\alpha,\overline{w})-\Gamma(w_\beta,\overline{w})\right]}_{>0}\cdot\mathbb{I}_{\{w_\alpha>\underline{w}\}\cap\{w_\beta<\overline{w}\}}\\<&\ 0.\end{aligned}\tag{139}$$

Hence, (139) implies that $U^*(k)$ is a submodular function, and an application of Topkis Theorem (see Topkis, 1978, Theorem 6.1) implies that $y(\underline{w})$ is decreasing in k. Since the other endpoint $w_\eta(k)$ is completely determined by the value of $y(\underline{w})$, we get that $w_\eta(k)$ is decreasing in k.

Step 3. With respect to $y(\overline{w})$, we get from (138) and (107) that

$$\begin{aligned}
\frac{\partial^2 U^*(k)}{\partial y(\overline{w})\partial k} &= \frac{\partial \tilde{\Phi}^{M*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\alpha, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial \Gamma(w_\alpha, \overline{w})} \frac{\partial \Gamma(w_\alpha, \overline{w})}{\partial y(\overline{w})} \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\} \cap \{w = k\}} \\
&\quad + \frac{\partial \tilde{\Phi}^{M*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\alpha, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial \Gamma(\underline{w}, \overline{w})} \frac{\partial \Gamma(\underline{w}, \overline{w})}{\partial y(\underline{w})} \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\} \cap \{w = k\}} \\
&\quad - \frac{\partial \tilde{\Phi}^{R*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial \Gamma(w_\beta, \overline{w})} \frac{\partial \Gamma(w_\beta, \overline{w})}{\partial y(\overline{w})} \cdot \mathbb{I}_{\{w_\beta < \overline{w}\} \cap \{w = k\}} \\
&\quad - \frac{\partial \tilde{\Phi}^{R*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial \Gamma(\underline{w}, \overline{w})} \frac{\partial \Gamma(\underline{w}, \overline{w})}{\partial y(\underline{w})} \cdot \mathbb{I}_{\{w_\beta < \overline{w}\} \cap \{w = k\}} \\
&= \underbrace{\frac{\partial \Gamma(\underline{w}, \overline{w})}{\partial y(\overline{w})}}_{>0} \cdot \frac{y(k)}{k^2} h'\left(\frac{y(k)}{k}\right) \frac{1}{\Gamma(\underline{w}, \overline{w})^2} \cdot \underbrace{\left[\Gamma(w_\alpha, \overline{w}) - \Gamma(w_\beta, \overline{w})\right]}_{>0} \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\} \cap \{w_\beta < \overline{w}\}} \\
&> 0.
\end{aligned} \tag{140}$$

Hence, (140) implies that $U^*(k)$ is a supermodular function, and an application of Topkis Theorem (see Topkis, 1978, Theorem 6.1) implies that $y(\overline{w})$ is increasing in k. Since the other endpoint $w_\gamma(k)$ is completely determined by the value of $y(\overline{w})$, we accordingly have that $w_\gamma(k)$ is increasing in k. ■

Theorem 7.3 *If the following condition*

$$\begin{cases} \frac{\mathrm{d}w_\alpha}{\mathrm{d}w_\beta} \geq \frac{\tilde{f}(w_\beta)}{\tilde{f}(w_\alpha)} & \text{for } w_\beta < \overline{w}, \\ \frac{\tilde{f}(w_\alpha)}{\tilde{f}(w_\beta)} \geq \frac{\mathrm{d}w_\beta}{\mathrm{d}w_\alpha} & \text{for } w_\alpha > \underline{w} \end{cases}$$

holds, then the selfishly optimal income tax schedule over $(\underline{w}, \overline{w}]$ for the median skill type is a Condorcet winner under pairwise majority voting.

Proof. Provided Propositions 7.1 and 7.2, the proof is the same as that of Theorem 5.1. ■

Appendix C: The Relation between Ex Ante and Ex Post Median Skill Levels (Not for Publication!)

After combining the migration decisions, the ex post measure of workers is given by

$$\Gamma(\underline{w}, \overline{w}) = \int_{\underline{w}}^{\overline{w}} \tilde{f}(w) dw \;=\; \int_{\underline{w}}^{w_m} \tilde{f}(w) dw + \int_{w_m}^{\overline{w}} \tilde{f}(w) dw. \tag{141}$$

By using (8), we get the right-hand terms of (141) as

$$\int_{\underline{w}}^{w_m} \tilde{f}(w) dw = \frac{1}{2} \;+\; \underbrace{L^I([\underline{w}, w_m]) - L^O([\underline{w}, w_m])}_{L^{NI}([\underline{w}, w_m]) \;=\; \text{net labor inflow}} \tag{142}$$

and

$$\int_{w_m}^{\overline{w}} \tilde{f}(w)dw = \frac{1}{2} + \underbrace{L^I([w_m,\overline{w}]) - L^O([w_m,\overline{w}])}_{L^{NI}([w_m,\overline{w}]) \,=\, \text{net labor inflow}}, \tag{143}$$

in which the measures of labor inflows are defined as

$$\begin{aligned} L^I([\underline{w},w_m]) &\equiv \int_{\{w\in[\underline{w},w_m]|\Delta(w)\geq 0\}} G_-(\Delta(w)|w)f_-(w)n_-dw, \\ L^I([w_m,\overline{w}]) &\equiv \int_{\{w\in[w_m,\overline{w}]|\Delta(w)\geq 0\}} G_-(\Delta(w)|w)f_-(w)n_-dw, \end{aligned} \tag{144}$$

and the measures of labor outflows are defined as

$$\begin{aligned} L^O([\underline{w},w_m]) &\equiv \int_{\{w\in[\underline{w},w_m]|\Delta(w)\leq 0\}} G(-\Delta(w)|w)f(w)dw, \\ L^O([w_m,\overline{w}]) &\equiv \int_{\{w\in[w_m,\overline{w}]|\Delta(w)\leq 0\}} G(-\Delta(w)|w)f(w)dw. \end{aligned} \tag{145}$$

By using (141)-(145), we can identify the relation of the ex post median skill level $\tilde{w}_m$ with the ex ante median skill level w_m and summarize the results as three propositions.

Proposition 7.3 *Suppose* $\Gamma(\underline{w},\overline{w}) = 1$. *We have: (a) If* $L^{NI}([\underline{w},w_m]) = L^{NI}([w_m,\overline{w}]) = 0$*, then* $\tilde{w}_m = w_m$*; (b) If* $L^{NI}([\underline{w},w_m]) > 0$ *and* $L^{NI}([w_m,\overline{w}]) < 0$*, then* $\tilde{w}_m < w_m$*; (c) If* $L^{NI}([\underline{w},w_m]) < 0$ *and* $L^{NI}([w_m,\overline{w}]) > 0$*, then* $\tilde{w}_m > w_m$.

Proposition 7.4 *Suppose* $\Gamma(\underline{w},\overline{w}) > 1$. *We have: (a) If* $L^{NI}([\underline{w},w_m]) = 0$ *and* $L^{NI}([w_m,\overline{w}]) > 0$*, then* $\tilde{w}_m > w_m$*; (b) If* $L^{NI}([\underline{w},w_m]) > 0$ *and* $L^{NI}([w_m,\overline{w}]) = 0$*, then* $\tilde{w}_m < w_m$*; (c) If* $L^{NI}([\underline{w},w_m]) > 0$ *and* $L^{NI}([w_m,\overline{w}]) > 0$*, then* $\tilde{w}_m < w_m$ *for* $L^{NI}([\underline{w},w_m]) > L^{NI}([w_m,\overline{w}])$*,* $\tilde{w}_m = w_m$ *for* $L^{NI}([\underline{w},w_m]) = L^{NI}([w_m,\overline{w}])$*, and* $\tilde{w}_m > w_m$ *for* $L^{NI}([\underline{w},w_m]) < L^{NI}([w_m,\overline{w}])$*; (d) If* $L^{NI}([\underline{w},w_m]) > 0$ *and* $L^{NI}([w_m,\overline{w}]) < 0$*, then* $\tilde{w}_m < w_m$*; (e) If* $L^{NI}([\underline{w},w_m]) < 0$ *and* $L^{NI}([w_m,\overline{w}]) > 0$*, then* $\tilde{w}_m > w_m$.

Proposition 7.5 *Suppose* $\Gamma(\underline{w},\overline{w}) < 1$. *We have: (a) If* $L^{NI}([\underline{w},w_m]) = 0$ *and* $L^{NI}([w_m,\overline{w}]) < 0$*, then* $\tilde{w}_m < w_m$*; (b) If* $L^{NI}([\underline{w},w_m]) < 0$ *and* $L^{NI}([w_m,\overline{w}]) = 0$*, then* $\tilde{w}_m > w_m$*; (c) If* $L^{NI}([\underline{w},w_m]) < 0$ *and* $L^{NI}([w_m,\overline{w}]) < 0$*, then* $\tilde{w}_m < w_m$ *for* $L^{NI}([\underline{w},w_m]) > L^{NI}([w_m,\overline{w}])$*,* $\tilde{w}_m = w_m$ *for* $L^{NI}([\underline{w},w_m]) = L^{NI}([w_m,\overline{w}])$*, and* $\tilde{w}_m > w_m$ *for* $L^{NI}([\underline{w},w_m]) < L^{NI}([w_m,\overline{w}])$*; (d) If* $L^{NI}([\underline{w},w_m]) > 0$ *and* $L^{NI}([w_m,\overline{w}]) < 0$*, then* $\tilde{w}_m < w_m$*; (e) If* $L^{NI}([\underline{w},w_m]) < 0$ *and* $L^{NI}([w_m,\overline{w}]) > 0$*, then* $\tilde{w}_m > w_m$.

To identity the relation between ex ante and ex post median skill levels, we divide the ex post population of workers into two groups: the first group of workers with skill levels lower than the ex ante median skill level and the second group of workers with skill levels higher than the ex ante median skill level. Propositions 7.3-7.5 consider three possible cases corresponding to three possible ex post measures of workers of all skill levels.

Proposition 7.3 considers the case that migrations do not change the total measure of workers. Then we have three possible subcases. Subcase (a) shows that labor inflow and labor outflow cancel each other for both groups, and hence the median skill level should be the same under the same total measure. Subcase (b) shows that the first group faces positive net labor inflow while the second group faces positive net labor outflow, hence the position of ex post median skill level should move towards the left direction under the same total measure, leading to a smaller median skill level than the ex ante one. Subcase (c) shows that the first group faces positive net labor outflow while the second group faces positive net labor inflow, hence the position of ex post median skill level should move towards the right direction under the same total measure, leading to a larger median skill level than the ex ante one. We can analyze Propositions 7.4-7.5 in the similar way.