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

Darong Dai† Guoqiang Tian‡

January 9, 2020

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 level and migration cost are the private information of each worker. It shows reasonable conditions under which the first-order approach applies, so the ironing surgery of Brett and Weymark (2017) is not always needed to guarantee the sufficient condition for incentive compatibility. Under quasilinear-in-consumption preferences, the tax schedule proposed by the median skill type is shown to be the Condorcet winner. While such schedule features negative marginal tax rates for low skills, it features positive ones for high skills with small migration elasticities; the marginal tax rates at the bottom and top skill levels cannot be unambiguously signed. Moreover, it identifies the conditions under which migration induces uniformly higher or lower equilibrium marginal tax rates for both low and high skills than the counterparts in autarky.

Keywords: Inter-jurisdiction migration; Redistributive taxation; Nonlinear income tax; Majority voting; Median voter.

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

*We benefit from the conversation and communication with John Weymark, who provided constructive suggestions that have substantially improved the quality of this study. We also thank the helpful feedback and comments from Yonghong An, Pedro Bento, Laurent Bouton, Klaus Desmet, Simona Fabrizi, Andrew T. Foerster, Andrew Glover, Dennis W. Jansen, Claus Thustrup Kreiner, Quan Li, Jason Lindo, Liqun Liu, Chen-Yu Pan, Tatevik Sekhposyan, Sang-Chul Suh, Kei-Mu Yi, Yuzhe Zhang, Sarah Zubairy, and the participants of Macro Research Group and Macro Student Conference (October 2017) at Texas A&M University, the Midwest Economic Theory Conference at SMU (November 2017), the 14th Meeting of the Society for Social Choice and Welfare (June 2018), the 2018 Asian Meeting of the Econometric Society, the 20th Annual Meeting of the Association for Public Economic Theory (July 2019), and the Joint Congress of the European Economic Association and the Econometric Society (August 2019). The usual disclaimer applies.

†Correspondence author. Institute for Advanced Research (IAR), Key Laboratory of Mathematical Economics (SUFE), Ministry of Education of China, Shanghai University of Finance and Economics, Shanghai 200433, People's Republic of China. E-mails: dai.darong@sufe.edu.cn; darongdai@hotmail.com

‡Department of Economics, Texas A&M University, College Station, TX 77843, USA. Institute for Advanced Studies in Finance and Economics, Hubei University of Economics, Wuhan 430205, People's Republic of China. E-mail: gtian@tamu.edu

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. Indeed, Kleven et al. (2013), Kleven et al. (2014) and Akcigit et al. (2016) estimate large migration elasticities, which could be larger than one, with respect to the income tax rate for these types of workers. As a kind of policy response in the context of more intense global competition, both the US and the UK plan to adopt skills-based immigration systems that shall end free movement but are supposed to attract the brightest and best from the rest of the world.

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; Blumkin et al., 2015) that builds on the seminal work of Mirrlees (1971) focuses on the normative perspective. The following issue is, however, not well addressed. How shall the schedule of redistributive taxation look like when workers can vote both in the ballot box and with their feet? The present article can be thought of as providing a political-economic counterpart to the normative literature.[1] Here the equilibrium tax schedule in each jurisdiction shall be selected by the pairwise majority rule. A real political economy that features labor mobility and majority voting is Switzerland that consists of 26 cantons, each of which determines its own income tax rate via direct democracy; meanwhile, there are lots of inter-canton migrations.

We consider a model of the economy consisting of two jurisdictions, between which workers can move by paying certain amount of migration costs. In each jurisdiction, workers differ in both skill levels and migration costs. While the ex ante distributions are common knowledge, the values of any worker's skill level and migration costs are only known to herself. As usual, taxation is based on the residence principle.[2] 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, the reservation utilities of workers, the ex post skill distribution as well as the tax base are endogenously determined. As such, there tends to be a complex interaction between migration and taxation in equilibrium.

As in Röell (2012), Bohn and Stuart (2013), Brett and Weymark (2016, 2017, 2019), and Dai (2019), we are interested in selfishly optimal income tax schedules. Each worker can be viewed as a citizen candidate who can propose an income tax schedule, or an allocation of incomes and consumptions for all types under the Taxation Principle, that

[1] In fact, we observe that in the Western democracies political factors are likely to considerably affect the structure of tax and transfer system in the coming years. The rise of extreme right parties in the EU in the backdrop of labor migrants and the concerns about the implications for the sustainability of the welfare state and the debate on the desirable marginal tax rates levied on top earners especially in the US in the backdrop of the looming national elections and the proposals of democratic candidates like Bernie Sanders and Sen. Elizabeth Warren are two manifestations. We also observe some trends that are in favor of the wealthy population, which may reflect political lobbying and the role of certain die-hard constituencies. As a caveat for a theoretical paper, it must adopt a political economy model that is tractable to certain degree and hence has some of these issues unexplored.

[2] The only exceptions are the US and Israel, where the citizens pay domestic income tax based on their global income.

maximizes the utility of her own type. Then, the pairwise majority rule is used to select the one that is going to be implemented in equilibrium.[3] Each worker proposes an income tax schedule as if she were representing the government. Following the mechanism design approach, each proposer must design incentive-compatible allocations satisfying the budget (or resource) constraint that has already accounted for individual participation constraints.

Under quasilinear-in-consumption preferences, the tax schedule proposed by the median skill type turns out to be the Condorcet winner in the majority rule equilibrium. 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, as established by Brett and Weymark (2017) in the case of a closed economy. Thus, governments subject to direct democracy and majority rule redistribute from the poor and rich toward the middle class.[4] This prediction extends the Director's Law (see Stigler, 1970) to the political-economic circumstance with inter-jurisdictional tax competition induced by cross-border labor mobility.

The current tax schedule exhibits the following characteristics. First, the sign of the marginal tax rates at the bottom and top skill levels is ambiguous. If their respective tax liability is sufficiently large, it is positive at the bottom while is negative at the top. If, however, the tax liability is sufficiently small, it is negative at the bottom while is positive at the top. Second, marginal tax rates are negative for low skills, and hence they receive positive transfers, but for high skills there is an endogenously determined and skill-dependent threshold of migration elasticity such that they face positive marginal tax rates if their migration elasticities are below this threshold; otherwise, they could face negative ones.[5] As such, the migration threat from high skills does matter in terms of income redistribution. Third, it creates two potential downward discontinuities, one at the skill level of the proposer and the other at the bottom skill level, in the resulting income schedule. In consequence, two bridges over two endogenous bunching regions that include, respectively, the skill level of the proposer and the bottom skill level must be built to iron the income schedule such that the second-order incentive compatibility condition is satisfied, namely that the resource allocated to a higher skilled worker shall be no smaller than that allocated to a lower skilled worker.

This paper is related to three strands of literature. First, it relates to the literature studying selfishly optimal nonlinear taxation determined by the pairwise majority rule,[6] such as Röell (2012), Bohn and Stuart (2013), and Brett and Weymark (2016, 2017, 2019). By adding the location choice for workers, both the reservation utility of the standard

[3]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.

[4]It 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.

[5]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).

[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.

participation constraint and the ex post skill distribution are endogenously determined. The current analysis is the most close to that of Brett and Weymark (2017), whose novel predictions have been generalized or overturned in the following five aspects.

Firstly, instead of showing that the first-order approach always fails to satisfy the sufficient condition for truth-telling under asymmetric information, it applies under certain conditions in the current model, and hence the ironing surgery is not always needed. Secondly, marginal tax rates for high skills are not always positive, and they are rather negative under large migration elasticities. Thirdly, the marginal tax rate at the endpoints of skill distribution is not always equal to zero, and it could be either negative or positive under certain conditions. Fourthly, in addition the potential downward discontinuity at the skill level of the proposer, there may be another one at the bottom skill level. Fifthly, the marginal tax rates of low and high skills could be uniformly higher or lower, and also the level of redistribution of the equilibrium taxation schedule could be either higher or lower, than the counterparts in the autarky equilibrium.

Second, it relates to the literature analyzing how the change of skill distribution affects equilibrium tax rates and 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 socially optimal income tax policies. The present paper adopts the political-economic 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 complementary to the literature.

Third, it relates to the literature studying taxation under the joint consideration of labor mobility and voting. The literature either assumes away asymmetric information (e.g., Cremer and Pestieau, 1998; Hindriks, 2001), restricts attention to a 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 the probabilistic voting in a representative democracy (e.g., Brett, 2016). Mobility and voting are allowed in the model of Morelli et al. (2012), but income taxation is determined by a benevolent social planner and only the constitutional choice is determined by majority voting. They, therefore, still follow the normative approach in terms of the design of taxation policies.

The rest of the paper is organized as follows. Section 2 sets up the model of the economy. Section 3 derives and characterizes selfishly optimal nonlinear income tax schedules. Section 4 establishes the voting equilibrium. Section 5 identifies qualitatively the effects of migration on the equilibrium marginal tax rates. Section 6 concludes with some remarks. Proofs are relegated to the Appendix.

2 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) and Dai et al. (2019) to consider such 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 the amount of labor supply. A worker has nonnegative consumption c that is also her after-tax income, namely

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

Following Diamond (1998), Blumkin et al. (2015) and Brett and Weymark (2017), preferences over consumption and labor supply are represented by the quasilinear-in-consumption utility function, $\tilde{u}(c, l; m) = c - h(l) - \mathbb{I} \cdot m$, which is common to all workers with $\mathbb{I} = 1$ if she decides to migrate and $\mathbb{I} = 0$ otherwise.[8] 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- 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: optimal labor supply on the intensive margin and optimal residence choice on the extensive margin.

[7]The simpler assumption is that migration costs and skill levels are independently distributed, as adopted by Morelli et al. (2012) and Bierbrauer et al. (2013). Assuming that migration costs distribute identically and independently across skill levels, Blumkin et al. (2015) also show that the migration elasticity is increasing with respect to the skill level, which seems to be consistent with the empirical finding of Doquier and Marfouk (2006) and Simula and Trannoy (2010) that individuals with a higher skill level are more likely to migrate.

[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). Under risk-neutral preferences, c actually could be interpreted as a nonnegative wealth-transfer from the government (or the mechanism designer).

2.1 Intensive Margin

If a worker decides to stay in jurisdiction A, then she maximizes (3) subject to $\mathbb{I} = 0$ and (2), yielding $T'(y(w)) = 1 - (1/w)h'(y(w)/w)$ 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 indirect (or gross) 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 $U(w) = \max_{w' \in [\underline{w}, \overline{w}]} c(w') - h(y(w')/w)$, for all $w \in [\underline{w}, \overline{w}]$. The first-order incentive compatibility (FOIC) condition is thus given by

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

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.

2.2 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 notation, we let $\tilde{f}(w) \equiv \phi(\Delta(w); w)$ and $\tilde{\theta}(w) \equiv \theta(\Delta(w); w)$. Also, the ex post skill distribution function is defined as

$$\Gamma(\underline{w}, w) \equiv \int_{\underline{w}}^{w} \tilde{f}(t)dt, \tag{10}$$

for all $w \in [\underline{w}, \overline{w}]$.

3 Selfishly Optimal Nonlinear Income Tax Schedules

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

$$\int_{\underline{w}}^{\overline{w}} [y(w) - c(w)] \tilde{f}(w) dw \geq 0, \tag{11}$$

where we have used (2). Under the quasilinear-in-consumption assumption, (11) 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 $\tilde{f}$ depends on the taxation policy chosen by the opponent jurisdiction, the strategic interaction between competing governments is fully captured by this budget constraint (or resource constraint).

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 (11), 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. We use the equilibrium concept of Bierbrauer et al. (2013) modified to account for the different way that the tax schedules are determined. That is, the tax schedule of each jurisdiction generates an allocation that satisfies the resource feasibility and incentive constraints and wins a pairwise majority contest among the tax schedules that are selfishly optimal given the other jurisdiction's schedule.

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

$$\max_{\{c(w), y(w)\}_{w \in [\underline{w}, \overline{w}]}} U(k) \quad \text{subject to } (5), (6), (7) \text{ and } (11), \tag{12}$$

taking as given the allocation schedule in the opponent jurisdiction. By (12), the resulting allocation schedule is indeed *selfishly optimal* for the proposer.

3.1 First-Order Approach

We consider first the first-order approach by which the SOIC condition (7) is ignored. Formally, problem (12) is relaxed as

$$\max_{\{c(w), y(w)\}_{w \in [\underline{w}, \overline{w}]}} U(k) \quad \text{subject to } (5), (6) \text{ and } (11). \tag{13}$$

In order to solve problem (13), we first give the following lemma.

[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.

Lemma 3.1 *The optimal schedule of before-tax incomes $y(\cdot)$ for type k's problem (13) is obtained by solving the following unconstrained maximization problem:*

$$\begin{aligned}\max_{y(\cdot)} \quad & \frac{1}{\Gamma(\underline{w},\overline{w})}\int_{\underline{w}}^{\overline{w}} \left\{ \left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \tilde{f}(w) - \frac{y(w)}{w^2}h'\left(\frac{y(w)}{w}\right)\Gamma(w,\overline{w})\right\} dw \\ & + \int_{\underline{w}}^{k} \frac{y(w)}{w^2}h'\left(\frac{y(w)}{w}\right) dw.\end{aligned} \tag{14}$$

Proof. See the Appendix. ■

By setting $k = \underline{w}$, the solution to (14) is the maxi-min income schedule, denoted by $y^R(\cdot)$; and by setting $k = \overline{w}$, the solution is the maxi-max income schedule, denoted by $y^M(\cdot)$. It is straightforward that the income schedule that solves program (14) coincides with the maxi-max solution for individuals with skill levels smaller than k and coincides with the maxi-min solution for individuals with skill levels larger than k. Moreover, noting that both $\Gamma(\underline{w},\overline{w})$ and $\Gamma(w,\overline{w})$ are functions of $y(\cdot)$, the first-order conditions are to be more involved than the counterparts in the case of autarky.

Using Lemma 3.1, we now establish the following two lemmas.

Lemma 3.2 *By setting $k = \underline{w}$ in problem (14), the maximin income schedule $\left\{y^R(w)\right\}_{w\in(\underline{w},\overline{w}]}$ solves the equation*

$$\begin{aligned} & \overbrace{\left[1 - \frac{1}{w}h'\left(\frac{y(w)}{w}\right)\right]\tilde{f}(w)}^{\textit{intensive margin}} \\ & - \overbrace{\left[\frac{1}{w^2}h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^3}h''\left(\frac{y(w)}{w}\right)\right]\left[\Gamma(\underline{w},\overline{w}) - \Gamma(\underline{w},w)\right]}^{\textit{intensive margin}} \\ & + \underbrace{\left[y(w) - h\left(\frac{y(w)}{w}\right) - \frac{y(w)}{w^2}h'\left(\frac{y(w)}{w}\right) - U(\underline{w})\right]\frac{\partial \tilde{f}(w)}{\partial y(w)}}_{\textit{extensive margin}} = 0,\end{aligned} \tag{15}$$

in which

$$\frac{\partial \tilde{f}(w)}{\partial y(w)} = \begin{cases} g_-(\Delta(w)|w)f_-(w)n_-\frac{\partial U(w)}{\partial y(w)} & \textit{for } \Delta(w) \geq 0, \\ g(-\Delta(w)|w)f(w)\frac{\partial U(w)}{\partial y(w)} & \textit{for } \Delta(w) \leq 0 \end{cases} \tag{16}$$

with

$$\frac{\partial U(w)}{\partial y(w)} = \frac{1}{w^2}h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^3}h''\left(\frac{y(w)}{w}\right), \tag{17}$$

for $\forall w \in (\underline{w},\overline{w}]$. Then, $y^R(\underline{w})$ is obtained by balancing the government budget constraint.

Proof. See the Appendix. ■

By (16) and (17), we have $\partial\tilde{f}(w)/\partial y(w) > 0$, and hence under the maxi-min income schedule an increase of $y(w)$ induces labor inflows of skill levels $w \in (\underline{w},\overline{w}]$.

Lemma 3.3 *By setting $k = \overline{w}$ in problem (14), the maximax income schedule $\left\{y^M(w)\right\}_{w\in(\underline{w},\overline{w})}$ solves the equation*

$$\underbrace{\left[1 - \frac{1}{w}h'\left(\frac{y(w)}{w}\right)\right]\tilde{f}(w) + \left[\frac{1}{w^2}h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^3}h''\left(\frac{y(w)}{w}\right)\right]\Gamma(\underline{w},w)}_{\textit{intensive margin}} = 0; \tag{18}$$

and $\underline{y}^M(\underline{w})$ solves the equation

$$\begin{aligned}&\overbrace{\left[1-\frac{1}{\underline{w}}h'\left(\frac{y(\underline{w})}{\underline{w}}\right)\right]\tilde{f}(\underline{w})}^{\text{intensive margin}}\\ &+\underbrace{\left[y(\underline{w})-h\left(\frac{y(\underline{w})}{\underline{w}}\right)-\frac{y(\underline{w})}{\underline{w}^2}h'\left(\frac{y(\underline{w})}{\underline{w}}\right)-U(\underline{w})\right]\frac{\partial\tilde{f}(\underline{w})}{\partial y(\underline{w})}}_{\text{extensive margin}}=0,\end{aligned}\tag{19}$$

in which

$$\frac{\partial\tilde{f}(\underline{w})}{\partial y(\underline{w})}=\begin{cases}g_-(\Delta(\underline{w})|\underline{w})f_-(\underline{w})n_-\frac{\partial U(\underline{w})}{\partial y(\underline{w})} & \text{for } \Delta(\underline{w})\geq 0,\\ g(-\Delta(\underline{w})|\underline{w})f(\underline{w})\frac{\partial U(\underline{w})}{\partial y(\underline{w})} & \text{for } \Delta(\underline{w})\leq 0\end{cases}\tag{20}$$

with

$$\frac{\partial U(\underline{w})}{\partial y(\underline{w})}=-\left[\frac{1}{\underline{w}^2}h'\left(\frac{y(\underline{w})}{\underline{w}}\right)+\frac{y(\underline{w})}{\underline{w}^3}h''\left(\frac{y(\underline{w})}{\underline{w}}\right)\right].\tag{21}$$

Then, $y^M(\overline{w})$ is obtained by balancing the government budget constraint.

Proof. See the Appendix. ■

The counterparts of (15) and (18)-(19) derived by Brett and Weymark (2017) in the case of autarky are given, respectively, as follows:

$$\begin{aligned}&\left[1-\frac{1}{w}h'\left(\frac{y(w)}{w}\right)\right]f(w)\ -\left[\frac{1}{w^2}h'\left(\frac{y(w)}{w}\right)+\frac{y(w)}{w^3}h''\left(\frac{y(w)}{w}\right)\right][1-F(w)]=0,\\ &\left[1-\frac{1}{w}h'\left(\frac{y(w)}{w}\right)\right]f(w)\ +\left[\frac{1}{w^2}h'\left(\frac{y(w)}{w}\right)+\frac{y(w)}{w^3}h''\left(\frac{y(w)}{w}\right)\right]F(w)=0.\end{aligned}$$

As such, their intuition for the first-order conditions can be slightly modified to account for the migration response of individuals considered in the present context.

Recall that the proposer of type k wishes to maximize the utility of her own type, so the left hand side of equation (18), denoted LHS^M, captures the additional utility that individuals of type k can gain by increasing $y(w)$ by one unit for individuals of type $w < k$. At the maxima, this value must be zero. In LHS^M, $\tilde{f}(w)$ gives the extra units of resources that can be diverted from individuals of type w to the type-k individuals. Meanwhile, incentive compatibility must be restored after this increase. Given that individuals of type $w < k$ are distorted upwards, such a redistribution of resources is constrained by upward incentive compatibility conditions that prevent individuals of lower types from mimicking types above them. To this end, each individual of type w can be given $(1/w)h'(y(w)/w)$ additional units of consumption such that she has no incentive to mimic any other type. The first term of LHS^M thus gives the net additional resources that type-k individuals can divert from type-w individuals while restoring incentive compatibility. As is obvious, such a change placed on type w does not affect the incentives of types above w. On the other hand, this change slackens the upward incentive constraints for types below w, so type-k individuals can reclaim from these individuals the amount of resources given by the second term of LHS^M.

Departing from the case considered by Brett and Weymark (2017), here the participation constraints on the extensive margin must be taken into account given that $\tilde{f}(w)$

is a function of $y(w)$. In fact, equation (18) is derived from equation (49) given in the Appendix:

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

Since we have $\partial U(w)/\partial y(w) = \partial U(\overline{w})/\partial y(w) = 0$ for any $w \in (\underline{w}, \overline{w})$ at the solution, we get from (16) that $\partial\tilde{f}(w)/\partial y(w) = 0$ at the maxima for any $w \in (\underline{w}, \overline{w})$, which explains why we have (18) in Lemma 3.3. That is, for individuals of type $w \in (\underline{w}, k)$, the first-order effect of the change of $y(w)$ on the endogenous skill density $\tilde{f}(w)$ degenerates at the maxima.

The interpretation of equation (19) is similar to that of equation (18) except that the first-order effect of the change of $y(\underline{w})$ on the endogenous skill density $\tilde{f}(\underline{w})$ no longer degenerates at the maxima. In fact, it follows from equations (20)-(21) that the first-order effect is negative, namely that the increase of $y(\underline{w})$ drives out some type-$\underline{w}$ workers at the maxima. Using (2), (5) and (6), equation (19) can be rewritten as:

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

in which, due to the reduction of type-$\underline{w}$ workers, $T^M\left(y(\underline{w})\right)[\partial\tilde{f}(\underline{w})/\partial y(\underline{w})] < 0$ measures the revenue loss of the proposer, whereas $-U'(\underline{w})[\partial\tilde{f}(\underline{w})/\partial y(\underline{w})] > 0$ measures the gain of the proposer because of paying less information rent.

The proposer of type k also wishes to extract resources from higher types, so such a redistribution must be constrained by downward incentive compatibility conditions. The left hand side of equation (15), denoted LHS^R, gives the additional resources that individuals of type k can secure for themselves by one unit increase of $y(w)$ for individuals of type $w > k$. Once again, such a value must be zero at the maxima. The first term of LHS^R can be interpreted in the same way as that of LHS^M. Since individuals of types higher than w face downward incentive constraints, the second term of LHS^R gives the amount of additional consumption diverted to individuals of types higher than w such that they have no incentive to mimic any other type. Using (2), (5) and (6) again, the third term of LHS^R that captures the change on the extensive margin can be rewritten as:

$$\left[T^R(y(w))+U(w)-U'(w)-U(\underline{w})\right]\frac{\partial\tilde{f}(w)}{\partial y(w)}.$$

As it follows from (16)-(17) that $\partial\tilde{f}(w)/\partial y(w) > 0$, $T^R(y(w))[\partial\tilde{f}(w)/\partial y(w)]$ measures the revenue gain due to the increase of tax base, $U'(w)[\partial\tilde{f}(w)/\partial y(w)]$ measures the information rent paid to these new immigrants, and $[U(w)-U(\underline{w})][\partial\tilde{f}(w)/\partial y(w)] > 0$ measures a sort of positive-externality gain from attracting immigrants of skill level w rather than those of the bottom skill level.

In consequence, solving problem (13) leads to the following result.

Proposition 3.1 *The selfishly optimal schedule of pre-tax incomes proposed by any worker of type $k \in (\underline{w}, \overline{w})$ is given by*

$$y(w) = \begin{cases} \underline{y}^M(w) & \text{for } w = \underline{w}, \\ y^M(w) & \text{for } w \in (\underline{w}, k), \\ y^R(w) & \text{for } w \in (k, \overline{w}]. \end{cases} \tag{22}$$

Proof. See the Appendix. ■

Since we focus on selfishly optimal income tax schedules, a proposer of type k wishes to redistribute incomes (or resources) from all other types toward her own type. To this end, for types above her own, she optimally proposes the maximin income schedule, whereas for types below her own, she optimally proposes the maximax income schedule.

By applying the formula of marginal tax rate given in (4) to Lemmas 3.2 and 3.3 and using Proposition 3.1, we summarize the prediction as the following theorem.

Theorem 3.1 *The selfishly optimal income tax schedule proposed by any worker of type $k \in (\underline{w}, \overline{w})$ is given by*

$$\tau(w) = \begin{cases} \underline{\tau}^M(w) & \text{for } w = \underline{w}, \\ \tau^M(w) & \text{for } w \in (\underline{w}, k), \\ \tau^R(w) & \text{for } w \in (k, \overline{w}] \end{cases} \tag{23}$$

in which these marginal tax rates are expressed as

$$\underline{\tau}^M(w) = -\frac{\partial \tilde{f}(w)}{\partial y(w)} \frac{1}{\tilde{f}(w)} \left[y(w) - h\left(\frac{y(w)}{w}\right) - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) - U(\underline{w}) \right], \tag{24}$$

$$\tau^M(w) = -\frac{\Gamma(\underline{w}, w)}{w\tilde{f}(w)} \left[\frac{1}{w} h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^2} h''\left(\frac{y(w)}{w}\right) \right], \tag{25}$$

and

$$\begin{aligned} \tau^R(w) = {} & \frac{\Gamma(\underline{w}, \overline{w}) - \Gamma(\underline{w}, w)}{w\tilde{f}(w)} \left[\frac{1}{w} h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^2} h''\left(\frac{y(w)}{w}\right) \right] \\ & - \frac{\partial \tilde{f}(w)}{\partial y(w)} \frac{1}{\tilde{f}(w)} \left[y(w) - h\left(\frac{y(w)}{w}\right) - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) - U(\underline{w}) \right]. \end{aligned} \tag{26}$$

In the case of autarky considered by Brett and Weymark (2017), the marginal tax rates are obtained as follows:

$$\begin{aligned} \hat{\tau}^M(w) &= -\frac{F(w)}{wf(w)} \left[\frac{1}{w} h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^2} h''\left(\frac{y(w)}{w}\right) \right], \\ \hat{\tau}^R(w) &= \frac{1 - F(w)}{wf(w)} \left[\frac{1}{w} h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^2} h''\left(\frac{y(w)}{w}\right) \right]. \end{aligned} \tag{27}$$

As is obvious, the marginal tax rates given by (24)-(26) differ from those given by (27) in two important aspects. First, as the ex post skill distribution is endogenously determined as a function of income and consumption, the migration decision on the extensive margin imposes non-trivial effects on these tax rates. Second, in addition to the discontinuity between the maximax tax schedule and the maximin tax schedule, we show that there may exist another discontinuity at the bottom skill level within the maximax tax schedule.

To further characterize the marginal tax rates given by Theorem 3.1, we obtain the following proposition.

Proposition 3.2 *Regarding the sign of these marginal tax rates given by (24)-(26), we have the following predictions.*

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

$$\underline{\tau}^M(\underline{w}) \begin{cases} < 0 & \text{for } T^M(y(\underline{w})) < U'(\underline{w}), \\ = 0 & \text{for } T^M(y(\underline{w})) = U'(\underline{w}), \\ > 0 & \text{for } T^M(y(\underline{w})) > U'(\underline{w}). \end{cases} \tag{28}$$

(ii) $\tau^M(w) < 0$ for all $w \in (\underline{w}, k)$.

(iii) For workers of type $w \in (k, \overline{w})$, if the tax liability satisfies $T^R(y(w)) \leq U'(w) + U(\underline{w}) - U(w)$, then $\tau^R(w) > 0$; otherwise, there is a threshold of the elasticity of migration, written as $\tilde{\theta}_\star(w) \equiv \frac{c(w)[\Gamma(\underline{w},\overline{w})-\Gamma(\underline{w},w)]}{[T^R(y(w))+U(w)-U'(w)-U(\underline{w})]\tilde{f}(w)}$, such that

$$\tau^R(w) \begin{cases} < 0 & \text{for } \tilde{\theta}(w) > \tilde{\theta}_\star(w), \\ = 0 & \text{for } \tilde{\theta}(w) = \tilde{\theta}_\star(w), \\ > 0 & \text{for } \tilde{\theta}(w) < \tilde{\theta}_\star(w). \end{cases} \tag{29}$$

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

$$\tau^R(\overline{w}) \begin{cases} < 0 & \text{for } T^R(y(\overline{w})) > U'(\overline{w}) + U(\underline{w}) - U(\overline{w}), \\ = 0 & \text{for } T^R(y(\overline{w})) = U'(\overline{w}) + U(\underline{w}) - U(\overline{w}), \\ > 0 & \text{for } T^R(y(\overline{w})) < U'(\overline{w}) + U(\underline{w}) - U(\overline{w}), \end{cases} \tag{30}$$

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

Proof. See the Appendix. ∎

Instead of showing that the marginal tax rate at the bottom skill level is always equal to zero, as suggested by Brett and Weymark (2017) in the case of autarky, it is so only when the tax liability of the lowest skilled workers is equal to a critical value; otherwise, it is positive (respectively, negative) if the tax liability is greater (respectively, smaller) than this critical value. We provide some intuition of (28). In the first-order condition (19), by one unit increase of $y(\underline{w})$, the proposer of type k diverted a net amount of resources from type-$\underline{w}$ workers on the intensive margin, and also on the extensive margin she realized a net amount of resources from driving out some type-$\underline{w}$ workers. At the maxima, the sum of these two amounts of resources must be zero, and the sign of the marginal tax rate, $\underline{\tau}^M(\underline{w})$, is actually the sign of the former net amount of resources. On the extensive margin, if the per unit loss of tax revenue, denoted $T^M(y(\underline{w}))$, is smaller than the per unit information rent saved, denoted $U'(\underline{w})$, then the sign of the latter net amount of resources is positive. As a result, the sign of the former net amount of resources must be negative, giving rise to $\underline{\tau}^M(\underline{w}) < 0$. The intuition of claims (ii)-(iv) follows from the same reasoning.

The sign of maximin marginal tax rates, given by (29)-(30), departs from that of Brett and Weymark (2017) in two ways. First, instead of showing that the marginal tax rate is always equal to zero for the highest skilled workers, we show that it is so only when their tax liability is equal to the critical value $U'(\overline{w}) + U(\underline{w}) - U(\overline{w})$; otherwise, it is

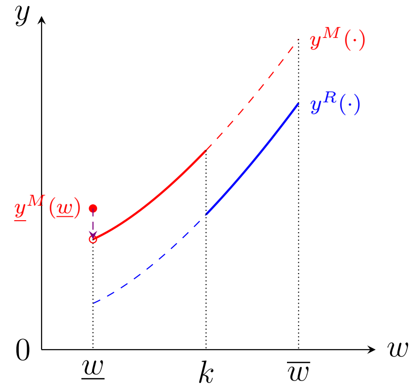

Figure 1: The case with two downward discontinuities in the income schedule.

negative above the threshold while is positive below the threshold. Second, instead of showing that the marginal tax rate is always positive for workers of types higher than the type of the proposer, it is positive when either their tax liability is below the threshold $U'(w)+U(\underline{w})-U(w)$, or their tax liability is above the threshold and also their migration elasticity is below an endogenously determined threshold.

The following proposition identifies when the first-order approach is appropriate for dealing with the incentive compatibility issue.

Proposition 3.3 *For any proposer of type $k \in (\underline{w}, \overline{w})$, the income schedule given by (22) may have the following discontinuities.*

(i) There is a downward discontinuity at $w = k$ for the following two cases: (i-a) $T^R(y(w)) \leq U'(w)+U(\underline{w})-U(w)$; (i-b) $T^R(y(w)) > U'(w)+U(\underline{w})-U(w)$ and $\tilde{\theta}(w) < \tilde{\theta}^*(w)$, in which $\tilde{\theta}^*(w) \equiv \frac{c(w)\Gamma(\underline{w},\overline{w})}{[T^R(y(w))+U(w)-U'(w)-U(\underline{w})]\tilde{f}(w)}$.

(ii) There is no such downward discontinuity at $w = k$ if $T^R(y(w)) > U'(w)+U(\underline{w})-U(w)$ and $\tilde{\theta}(w) \geq \tilde{\theta}^*(w)$.

(iii) There is a downward discontinuity at $w = \underline{w}$ if $T^M(y(\underline{w})) < U'(\underline{w})$.

(iv) There is either no discontinuity or an upward discontinuity at $w = \underline{w}$ if $T^M(y(\underline{w})) \geq U'(\underline{w})$.

Proof. See the Appendix. ■

In case (i), namely either the maximin tax liability is smaller than the threshold $U'(w)+U(\underline{w})-U(w)$, or the maximin tax liability is above the threshold $U'(w)+U(\underline{w})-U(w)$ and the migration elasticity is smaller than the threshold $\tilde{\theta}^*(w)$, then there is a downward discontinuity at the skill level of the proposer in the income schedule (see Figure 1), which obviously violates the SOIC condition (7). In case (iii), namely the maximax tax liability of the lowest skilled workers is smaller than the threshold $U'(\underline{w})$, then there is a downward discontinuity at the bottom skill level (see also Figure 1). In consequence, only under the joint conditions given in cases (ii) and (iv) will the first-order approach apply, and hence the SOIC condition could be safely ignored in solving problem (12). Indeed, the introduction of tax-driven migration may overturn the prediction of Brett and Weymark (2017) that there is always a downward discontinuity at the type of the proposer and that this discontinuity is unique.

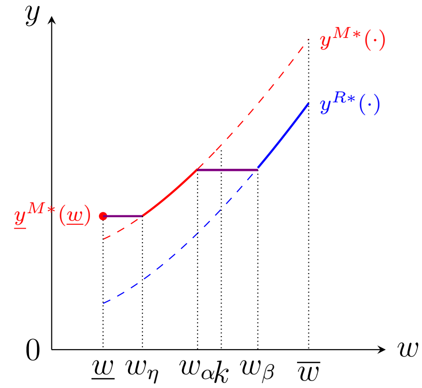

Figure 2: The income schedule with two bridges.

Using Proposition 3.2, for skills above the skill level of the proposer, condition (i-a) implies that all of them face positive tax rates, yielding downward distortions on their incomes; condition (i-b) implies that those with migration elasticities smaller than $\tilde{\theta}_\star(w)$ face positive tax rates, those with a migration elasticity of $\tilde{\theta}_\star(w)$ face a zero tax rate, and those with migration elasticities belonging to the interval $(\tilde{\theta}_\star(w), \tilde{\theta}^*(w))$ face negative tax rates; condition (ii) implies that all of them face negative tax rates, yielding upward distortions on their incomes. As claim (i) of Proposition 3.2 shows that workers of skills below the type of the proposer always face negative tax rates, only under condition (ii) of Proposition 3.3 can it be possible that the downward discontinuity at the skill level of the proposer disappears, and be even possible that there is an upward discontinuity at the skill level of the proposer. Claims (iii)-(iv) can be analogously analyzed.

3.2 The Complete Solution

Following Brett and Weymark (2016, 2017), if the 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 $y^{M*}(\cdot)$ and $y^{R*}(\cdot)$ 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 a bridge, which includes her own type between the maximax and maximin parts of this schedule, as well as a bridge that starts from the bottom skill type and ends within the maximax income schedule (see Figure 2).

Based on Proposition 3.3, the following result is established.

Theorem 3.2 *For any proposer of type $k \in (\underline{w}, \overline{w})$, let the following conditions be satisfied:*

- $T^R(y(w)) \leq U'(w) + U(\underline{w}) - U(w)$, *or* $T^R(y(w)) > U'(w) + U(\underline{w}) - U(w)$ *and* $\tilde{\theta}(w) < \tilde{\theta}^*(w)$, *for all* $w > k$;
- $T^M(y(\underline{w})) < U'(\underline{w})$.

Then, the selfishly optimal schedule of pre-tax incomes that is incentive compatible is given as follows:

$$y^*(w) = \begin{cases} \underline{y}^{M*}(\underline{w}) & \text{for } w \in [\underline{w}, w_\eta] \text{ if } w_\eta \le w_\alpha, \\ y^{M*}(w) & \text{for } w \in (w_\eta, w_\alpha), \\ y^{M*}(w_\alpha) & \text{for } w \in [w_\alpha, w_\beta] \text{ if } w_\alpha > \underline{w}, \\ y^{R*}(w_\beta) & \text{for } w \in [w_\alpha, w_\beta] \text{ if } w_\beta < \overline{w}, \\ y^{R*}(w) & \text{for } w \in (w_\beta, \overline{w}], \end{cases} \tag{31}$$

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. See the Appendix. ■

The proof of Theorem 3.2, though very lengthy, is just an application of the ironing surgery developed by Brett and Weymark (2017). Basically, we first fix the bridge endpoints w_η, w_α and w_β, and let $y^*(\underline{w}, w_\eta)$ and $y^*(w_\alpha, w_\beta)$ denote the optimal before-tax income levels of these two bridges over skill intervals $[\underline{w}, w_\eta]$ and $[w_\alpha, w_\beta]$, respectively. It is easy to show that the optimal value of the bridge endpoint w_η is the solution to equation $y^{M*}(w) = \underline{y}^{M*}(\underline{w})$, or $w_\eta = (y^{M*})^{-1}(\underline{y}^{M*}(\underline{w}))$. In the proof of the following Lemma 3.5, we shall establish the equations that implicitly solve for these two endpoints, w_α and w_β. Now, the selfishly optimal income schedule of proposer $k \in (w_\alpha, w_\beta)$ is obtained by solving the following problem:

$$\max_{y(\cdot)} \left(\int_{w_\eta}^{w_\alpha} \left\{ \left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \frac{\tilde{f}(w)}{\Gamma(\underline{w}, \overline{w})} + \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \left[1 - \frac{\Gamma(w, \overline{w})}{\Gamma(\underline{w}, \overline{w})} \right] \right\} \cdot \mathbb{I}_{\{w_\eta \le w_\alpha\}} dw \right.$$
$$\left. + \int_{w_\beta}^{\overline{w}} \left\{ \left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \frac{\tilde{f}(w)}{\Gamma(\underline{w}, \overline{w})} - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \frac{\Gamma(w, \overline{w})}{\Gamma(\underline{w}, \overline{w})} \right\} \cdot \mathbb{I}_{\{\overline{w} \ge w_\beta\}} dw \right),$$

in which $\mathbb{I}$ denotes the indicator function. To complete the proof, we just need to show that $y^*(w_\alpha, w_\beta) = y^{M*}(w_\alpha)$ for $w_\alpha > \underline{w}$ and $y^*(w_\alpha, w_\beta) = y^{R*}(w_\beta)$ for $w_\beta < \overline{w}$ by applying the same reasoning used by Brett and Weymark (2017) to prove that the resulting $y^*(\cdot)$ must be continuous over the interval $[\underline{w}, \overline{w}]$.

The following two lemmas characterize how the endogenously determined bridge endpoints w_η, w_α and w_β change with the type of the proposer.

Lemma 3.4 *The bridge endpoint $w_\eta(k)$ is decreasing in the type, k, of the proposer.*

Proof. See the Appendix. ■

Lemma 3.5 *If the following condition holds:*

$$\frac{\mathrm{d}\Gamma(\psi(w_\beta), w_\beta)}{\mathrm{d}w_\beta} \le 0 \quad \text{and} \quad \frac{\mathrm{d}\Gamma(w_\alpha, \varphi(w_\alpha))}{\mathrm{d}w_\alpha} \le 0, \tag{32}$$

in which

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

and

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

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

Proof. See the Appendix. ■

Though the basic idea of the proof is brought from Brett and Weymark (2017), the ex post skill distribution makes the details much more involved than theirs. Combining with Figure 2, Lemma 3.5 provides the condition under which the bridge moves upwards as the skill level of the proposer increases, which is useful for proving the following Theorem 4.1. The functions $\psi(\cdot)$ and $\varphi(\cdot)$ are well-defined as the income schedule $y^*(\cdot)$ is non-decreasing over the support, as shown by Theorem 3.2. Moreover, the technical restriction of condition (32) on the ex post skill distribution means that the measure (or mass) of skills that are bunched with the proposer in the selfishly optimal income schedule is non-increasing in the skill levels at the endpoints of the bridge.

4 The Voting Equilibrium

Following the common practice in the 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 tax schedule chosen as being determined, albeit indirectly, by a pairwise majority voting process.[11] As endogenous population raises conceptual difficulties when determining who will do the voting, we assume it is the ex ante residents who shall vote. The types of ex post residents cannot be explicitly identified, so having them do the voting seems to be problematic.[12] It is also practically reasonable to invoke the "citizen criterion", namely let the initial population residing in the country under a laissez-faire regime do the voting, as suggested by Simula and Trannoy (2012) and Blumkin et al. (2015) in the normative counterpart.[13] In each round, workers vote over two arbitrarily-selected alternatives. The one that survives all rounds becomes 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 written as

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

[11]It is a well-known strategyproof mechanism.

[12]For example, if a worker proposes a schedule that results in some types only locating in the other jurisdiction, then those types actually don't get to vote on this proposal. Moreover, if the schedules proposed by two different proposers result in different sets of types being residents, then it is more difficult to determine which types should get to vote.

[13]In some developed countries, immigrants, especially newcomers, cannot participate into the political process. As such, only full-fledged citizens can participate in collective decision-making. However, many countries adopt a delayed instead of an immediate assimilation policy, so there is in general a minimum number of years of residency as a prerequisite condition for being granted citizenship, which for instance goes from three (e.g., Netherlands, Australia and Canada) to ten (e.g., Switzerland).

Theorem 4.1 *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 the Appendix. ■

As in Brett and Weymark (2017), we establish the existence of a Condorcet winner in the current political economy. Even though the skill distribution is endogenous in this model, one would note that the ex post skill distribution per se is type independent. That is, it does not differ from proposer to proposer. 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.

Theorem 4.1 provides a theoretical support for the empirical finding of Corneo and Neher (2015) who 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.

5 The Effect of Migration on Equilibrium Tax Schedule

In the present economy, individual migration decision is made by taking into account potential tax policy differentials of two competing jurisdictions, and meanwhile tax policies are subject to mobile tax bases, leading to a complex interaction between migration and taxation in equilibrium. The assumption of quasilinear-in-consumption preferences, nevertheless, enables us to characterize the nontrivial effect of migration on the equilibrium tax schedules that enacted the selfishly optimal wishes of the endogenous median voters of these two jurisdictions.

To this end, we shall compare our marginal tax rates to those derived in autarky by Brett and Weymark (2017). As shown in Theorem 3.1, migration affects marginal tax rates through endogenizing the skill distribution that is a key part of the tax formula and also is the determinant of the equilibrium median skill level. So, migration affects both the distortion level and the redistribution scale. In what follows, we let w_m denote the median skill level of the ex ante skill distribution, $F(w)$, and let $\tilde{w}_m$ denote that of the ex post distribution, $\Gamma(\underline{w}, w)$. We have identified the conditions determining the relative magnitude of $\tilde{w}_m$ and w_m in Appendix B.[14]

Using the tax formulas established in Theorem 3.1 and the tax formulas given by (27), we give the following lemma.

[14] Essentially, these conditions relate to 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.

Lemma 5.1 *Let $\tilde{w}_m = w_m$, then we have the following predictions.*

(i) For all $w \in (\underline{w}, w_m)$, we have

$$\hat{\tau}^M(w) \begin{cases} < \tau^M(w) & \text{for } F(w)/f(w) > \Gamma(\underline{w}, w)/\tilde{f}(w), \\ = \tau^M(w) & \text{for } F(w)/f(w) = \Gamma(\underline{w}, w)/\tilde{f}(w), \\ > \tau^M(w) & \text{for } F(w)/f(w) < \Gamma(\underline{w}, w)/\tilde{f}(w). \end{cases} \tag{33}$$

(ii) If $[1 - F(w)]/f(w) \geq [\Gamma(\underline{w}, \overline{w}) - \Gamma(\underline{w}, w)]/\tilde{f}(w)$ and $T^R(y(w)) \geq U'(w) + U(\underline{w}) - U(w)$, then $\hat{\tau}^R(w) \geq \tau^R(w)$; if $[1 - F(w)]/f(w) \leq [\Gamma(\underline{w}, \overline{w}) - \Gamma(\underline{w}, w)]/\tilde{f}(w)$ and $T^R(y(w)) \leq U'(w) + U(\underline{w}) - U(w)$, then $\hat{\tau}^R(w) \leq \tau^R(w)$; if $[1 - F(w)]/f(w) < [\Gamma(\underline{w}, \overline{w}) - \Gamma(\underline{w}, w)]/\tilde{f}(w)$ and $T^R(y(w)) > U'(w) + U(\underline{w}) - U(w)$, then

$$\hat{\tau}^R(w) \begin{cases} < \tau^R(w) & \text{for } \tilde{\theta}(w) < \Theta^R(w), \\ = \tau^R(w) & \text{for } \tilde{\theta}(w) = \Theta^R(w), \\ > \tau^R(w) & \text{for } \tilde{\theta}(w) > \Theta^R(w); \end{cases} \tag{34}$$

if $[1 - F(w)]/f(w) > [\Gamma(\underline{w}, \overline{w}) - \Gamma(\underline{w}, w)]/\tilde{f}(w)$ and $T^R(y(w)) < U'(w) + U(\underline{w}) - U(w)$, then

$$\hat{\tau}^R(w) \begin{cases} < \tau^R(w) & \text{for } \tilde{\theta}(w) > \Theta^R(w), \\ = \tau^R(w) & \text{for } \tilde{\theta}(w) = \Theta^R(w), \\ > \tau^R(w) & \text{for } \tilde{\theta}(w) < \Theta^R(w); \end{cases} \tag{35}$$

in which

$$\Theta^R(w) \equiv \frac{c(w)}{T^R(y(w)) + U(w) - U'(w) - U(\underline{w})} \left[\frac{\Gamma(\underline{w}, \overline{w}) - \Gamma(\underline{w}, w)}{\tilde{f}(w)} - \frac{1 - F(w)}{f(w)} \right]. \tag{36}$$

(iii) If $T^R(y(w)) \leq U'(w) + U(\underline{w}) - U(w)$, then $\hat{\tau}^M(w) < \tau^R(w)$; if $T^R(y(w)) > U'(w) + U(\underline{w}) - U(w)$, then

$$\hat{\tau}^M(w) \begin{cases} < \tau^R(w) & \text{for } \tilde{\theta}(w) < \Theta^{MR}(w), \\ = \tau^R(w) & \text{for } \tilde{\theta}(w) = \Theta^{MR}(w), \\ > \tau^R(w) & \text{for } \tilde{\theta}(w) > \Theta^{MR}(w), \end{cases} \tag{37}$$

in which

$$\Theta^{MR}(w) \equiv \frac{c(w)}{T^R(y(w)) + U(w) - U'(w) - U(\underline{w})} \left[\frac{\Gamma(\underline{w}, \overline{w}) - \Gamma(\underline{w}, w)}{\tilde{f}(w)} + \frac{F(w)}{f(w)} \right]. \tag{38}$$

(iv) For all $w \in (\underline{w}, w_m)$, we have $\tau^M(w) < \hat{\tau}^R(w)$.

Proof. See the Appendix. ■

Assuming the same median skill level for the ex ante and ex post skill distributions, Lemma 5.1 identifies the conditions such that we can compare the marginal tax rates in the migration equilibrium to those in the autarky equilibrium for almost all skill levels. For skills below the median skill level, only the level of migration matters, namely, whether the jurisdiction under consideration faces a net labor inflow or outflow, as shown in part (i). While for skills above the median skill level, the level and the elasticity of migration, as well as the tax liability in the migration equilibrium are all relevant, as shown in part (ii). In fact, we give two thresholds of migration elasticity, $\Theta^R(w)$ and $\Theta^{MR}(w)$, and a threshold of the tax liability in the migration equilibrium, $U'(w) + U(\underline{w}) - U(w)$. Parts (iii) and (iv) are used to characterize the relative positions of maximin and maximax income schedules under migration and autarky, respectively, which are especially useful when the ex post median skill level differs from the ex ante one. Based on Proposition 3.3, we need to consider separately the following two cases.

5.1 The Case with a Downward Discontinuity at the Median Skill Level

Using Lemma 5.1, we give the following two propositions that characterize the qualitative effects of migration imposed on the equilibrium pre-tax income schedule derived by adopting the first-order approach and characterized by part (i) of Proposition 3.3.

Proposition 5.1 *Suppose $T^M(y(\underline{w})) < U'(\underline{w})$ and there is a downward discontinuity of the pre-tax income schedule at the median voter of the migration equilibrium. Let one of the following conditions hold:*

(a) $F(w)/f(w) < \Gamma(\underline{w}, w)/\tilde{f}(w)$ for all $w \in [\underline{w}, w_m]$, and $T^R(y(w)) < U'(w) + U(\underline{w}) - U(w)$, $[1 - F(w)]/f(w) > [\Gamma(\underline{w}, \overline{w}) - \Gamma(\underline{w}, w)]/\tilde{f}(w)$ and $\tilde{\theta}(w) < \Theta^R(w)$ for all $w \in (w_m, \overline{w}]$.

(b) $F(w)/f(w) < \Gamma(\underline{w}, w)/\tilde{f}(w)$ for all $w \in [\underline{w}, w_m]$, and $T^R(y(w)) > U'(w) + U(\underline{w}) - U(w)$, $[1 - F(w)]/f(w) \geq [\Gamma(\underline{w}, \overline{w}) - \Gamma(\underline{w}, w)]/\tilde{f}(w)$ and $\tilde{\theta}(w) < \min\{\Theta^{MR}(w), \tilde{\theta}^*(w)\}$ for all $w \in (w_m, \overline{w}]$.

(c) $F(w)/f(w) < \Gamma(\underline{w}, w)/\tilde{f}(w)$ for all $w \in [\underline{w}, w_m]$, and $T^R(y(w)) > U'(w) + U(\underline{w}) - U(w)$, $[1 - F(w)]/f(w) < [\Gamma(\underline{w}, \overline{w}) - \Gamma(\underline{w}, w)]/\tilde{f}(w)$ and $\Theta^R(w) < \tilde{\theta}(w) < \min\{\Theta^{MR}(w), \tilde{\theta}^*(w)\}$ for all $w \in (w_m, \overline{w}]$.

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 claim (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 claim (i) still holds, and workers of type $w \in (w_m, \tilde{w}_m]$ face even lower tax rates than when $\tilde{w}_m = w_m$.

Proof. First, it is easy to see from Propositions 3.2 and 3.3, (36) and (38) that $\Theta^R(w) < \tilde{\theta}_\star(w) < \min\{\Theta^{MR}(w), \tilde{\theta}^*(w)\}$. The relative magnitude of $\Theta^{MR}(w)$ and $\tilde{\theta}^*(w)$ is in general ambiguous. Using claim (i) of Lemma 5.1 and Proposition 3.3(iii), we get $\hat{\tau}^M(w) > \tau^M(w)$ whenever $T^M(y(\underline{w})) < U'(\underline{w})$ and $F(w)/f(w) < \Gamma(\underline{w}, w)/\tilde{f}(w)$ for all $w \in (\underline{w}, w_m]$, as desired in all of the three conditions. Furthermore, using (27), we have $0 > \hat{\tau}^M(w) >$

$\tau^M(w)$. Using (35), the remaining requirements of condition (a) lead to $\hat{\tau}^R(w) > \tau^R(w)$. Using claim (ii) of Proposition 3.2, the requirement of $T^R(y(w)) < U'(w) + U(\underline{w}) - U(w)$ in condition (a) leads to $\tau^R(w) > 0$, and hence condition (a) implies that $\hat{\tau}^R(w) > \tau^R(w) > 0 > \hat{\tau}^M(w) > \tau^M(w)$, as desired in claim (i).

Using claim (ii) of Lemma 5.1, the requirements of $T^R(y(w)) > U'(w) + U(\underline{w}) - U(w)$ and $[1 - F(w)]/f(w) \geq [\Gamma(\underline{w}, \overline{w}) - \Gamma(\underline{w}, w)]/\tilde{f}(w)$ in condition (b) imply that $\hat{\tau}^R(w) > \tau^R(w)$. If $\tilde{\theta}(w) \leq \tilde{\theta}_\star(w)$, then we get from claim (ii) of Proposition 3.2 that $\tau^R(w) > 0$, and hence condition (b) also implies that $\hat{\tau}^R(w) > \tau^R(w) > 0 > \hat{\tau}^M(w) > \tau^M(w)$, as desired in claim (i). Departing from condition (a), here the migration elasticity is allowed to be greater than the threshold $\tilde{\theta}_\star(w)$, which means that it is possible that $\tau^R(w) < 0$ based on claim (ii) of Proposition 3.2. However, using claim (iii) of Lemma 5.1, it is still guaranteed that $0 > \tau^R(w) > \hat{\tau}^M(w)$ for $\tilde{\theta}_\star(w) < \tilde{\theta}(w) < \min\{\Theta^{MR}(w), \tilde{\theta}^*(w)\}$. Using (27), we have in this case that $\hat{\tau}^R(w) > 0 > \tau^R(w) > \hat{\tau}^M(w) > \tau^M(w)$. Once again, claim (i) follows.

Using claim (ii) of Lemma 5.1, the requirements of $T^R(y(w)) > U'(w) + U(\underline{w}) - U(w)$, $[1 - F(w)]/f(w) < [\Gamma(\underline{w}, \overline{w}) - \Gamma(\underline{w}, w)]/\tilde{f}(w)$ and $\Theta^R(w) < \tilde{\theta}(w)$ for all $w \in (w_m, \overline{w}]$ in condition (c) imply that $\hat{\tau}^R(w) > \tau^R(w)$. If $\Theta^R(w) < \tilde{\theta}(w) \leq \tilde{\theta}_\star(w)$, then we get from claim (ii) of Proposition 3.2 that $\tau^R(w) > 0$, and hence condition (c) also implies that $\hat{\tau}^R(w) > \tau^R(w) > 0 > \hat{\tau}^M(w) > \tau^M(w)$, as desired in claim (i). Similar to condition (b), here the migration elasticity is allowed to be greater than the threshold $\tilde{\theta}_\star(w)$, which leads to the possible prediction of $\tau^R(w) < 0$ based on claim (ii) of Proposition 3.2. However, using claim (iii) of Lemma 5.1, it is still guaranteed that $0 > \tau^R(w) > \hat{\tau}^M(w)$ for $\tilde{\theta}_\star(w) < \tilde{\theta}(w) < \min\{\Theta^{MR}(w), \tilde{\theta}^*(w)\}$. Using (27), we have in this case that $\hat{\tau}^R(w) > 0 > \tau^R(w) > \hat{\tau}^M(w) > \tau^M(w)$. Once again, claim (i) follows. Finally, applying claim (i) and Figure 3, claims (ii)-(iii) of this proposition are immediate. ■

For the jurisdiction under consideration, Proposition 5.1 identifies the conditions under which workers tend to face lower marginal tax rates in the migration equilibrium than in the autarky equilibrium. The key is to guarantee that $\hat{\tau}^R(w) > \tau^R(w)$ for high skills while $\hat{\tau}^M(w) > \tau^M(w)$ for low skills. Here we further require that $\tau^R(w) > \hat{\tau}^M(w)$, namely that the maximin tax rate in the migration equilibrium is greater than the maximax tax rate in the autarky equilibrium, which makes a relevant difference when the ex ante and ex post median skill levels are not the same one, as illustrated by Figures 3 and 4. Elaborating further, this is a somewhat reasonable requirement because $\hat{\tau}^M(w) < 0$ always holds true while $\tau^R(w)$ is likely to be positive, as shown in Proposition 3.2.

All the three conditions require that this jurisdiction faces net labor outflow in low skills, namely that $F(w)/f(w) < \Gamma(\underline{w}, w)/\tilde{f}(w)$ for all $w \in (\underline{w}, w_m]$. Concerning high skills, they impose different requirements in terms of the direction of net labor flow, the elasticity of migration and the level of tax liability in the migration equilibrium. Specifically, condition (a) states that there is net labor inflow, the migration elasticity is bounded above by $\Theta^R(w)$, and the tax liability is bounded above. Condition (b) states that there is net labor inflow, the migration elasticity is bounded above by $\min\{\Theta^{MR}(w), \tilde{\theta}^*(w)\}$, and the tax liability is bounded below. Condition (c) states that there is net labor outflow, the migration elasticity is bounded below by $\Theta^R(w)$ and above by $\min\{\Theta^{MR}(w), \tilde{\theta}^*(w)\}$, and the tax liability is bounded below.

Regardless of whether inter-jurisdiction migration is allowed or not, recall that the equilibrium tax schedule features that the median voter extracts resources from all other types towards his own type, the balanced budget constraint (or resource constraint) must be satisfied, and that the low skills receive transfers instead of paying taxes. It is particularly worthwhile noting that the prediction of negative marginal tax rates for low skills is not due to, and also has not been overturned by, the introduction of migration.

Claim (i) under condition (a) could be interpreted as follows. First of all, even holding the same amount of resources the median voter extracts in the autarky equilibrium, the net labor outflow of low skills and the net labor inflow of high skills could support the claim of $\hat{\tau}^R(w) > \tau^R(w) > 0 > \hat{\tau}^M(w) > \tau^M(w)$, namely that each low-skill worker receives more transfers and each high-skill worker pays less taxes than in the autarky equilibrium. The ex post balanced budget constraint still holds by, for example, adjusting the amount of resources the median voter extracts. Such a claim could be due to that, in an asymmetric migration equilibrium, the median voter of the opponent jurisdiction chooses a more redistributive taxation scheme than what this median voter has chosen, namely that he imposes higher taxes in high skills which drives out some high skills while allocates more transfers to low skills which attracts some low skills. The requirement of a relatively low tax liability on each high-skill worker, namely $T^R(y(w)) < U'(w) + U(\underline{w}) - U(w)$, could be the joint effect of adverse selection under asymmetric information and the net labor inflow of high skills who face positive marginal tax rates. In terms of the former fact, the median voter must tolerate some information rent extracted by high skills, while in terms of the latter fact, ceteris paribus, a larger tax base justifies a lower level of tax liability facing each worker.

The requirement of $T^R(y(w)) < U'(w) + U(\underline{w}) - U(w)$ in condition (a) diverges from that of $T^R(y(w)) > U'(w)+U(\underline{w})-U(w)$ in condition (b) because we have $\tilde{\theta}(w) < \Theta^R(w)$ in condition (a) but $\tilde{\theta}(w) < \min\{\Theta^{MR}(w), \tilde{\theta}^*(w)\}$ in condition (b), for all $w \in (w_m, \overline{w}]$. For those with migration elasticities belonging to the set $(\tilde{\theta}_\star(w), \min\{\Theta^{MR}(w), \tilde{\theta}^*(w)\})$, they actually receive transfers under condition (b), so their incentive compatibility constraints are relaxed relative to the counterparts under condition (a). Moreover, although there is net labor inflow of high skills in condition (b), the ex post tax base is not necessarily larger than in the autarky equilibrium because only skills with migration elasticities smaller than the threshold $\tilde{\theta}_\star(w)$ pay positive taxes, thus rationalizing a lower bound of the tax liability.

Claim (i) under condition (c) could be attributed to that the median voter of the opponent jurisdiction implements even lower tax rates on high skills such that this jurisdiction faces net labor outflow of high skills, which is justified by relatively large migration elasticities, namely, $\Theta^R(w) < \tilde{\theta}(w)$ for all $w \in (w_m, \overline{w}]$. This median voter must impose lower tax rates on high skills than he would impose in the autarky equilibrium; otherwise, even more high skills shall choose to relocate at the opponent jurisdiction, cutting more the tax base facing him. By enlarging the tax base, ceteris paribus, the opponent jurisdiction is able to provide more generous transfers to low skills, resulting in a net labor inflow of low skills. As is obvious, the combination of the net labor outflow of high skills and lower tax rates implies that this jurisdiction collects less taxes in this migration equilibrium than in the autarky equilibrium. Given that the balanced budget constraint must be satisfied, this median voter is likely to be worse off in terms of resource extraction under migration, depending on the relative magnitude of the net labor outflow of low skills and the increase of transfers each worker receives.

In addition, for types belonging to $(\tilde{w}_m, w_m]$, as shown in case (ii) of Proposition 5.1, they face higher tax rates than in autarky because the median voter under migration is poorer than that in autarky. They belong to the low-income class in autarky while belong to the high-income class under migration, so their status changes from receiving transfers

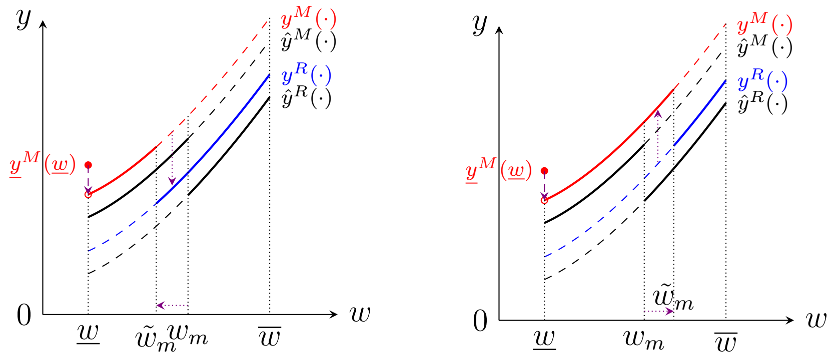

Figure 3: The case with migration inducing lower marginal tax rates than autarky.

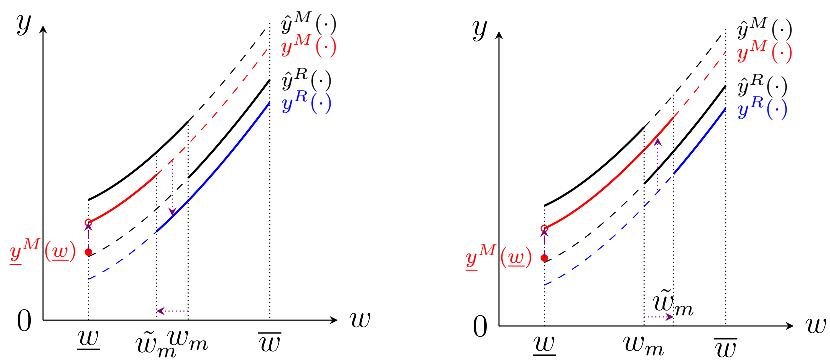

Figure 4: The case with migration inducing higher marginal tax rates than autarky.

to paying taxes.

Proposition 5.2 *Suppose $T^M(y(\underline{w})) \geq U'(\underline{w})$ and there is a downward discontinuity of the pre-tax income schedule at the median voter of the migration equilibrium. Let one of the following conditions hold:*

(a) $F(w)/f(w) > \Gamma(\underline{w}, w)/\tilde{f}(w)$ for all $w \in [\underline{w}, w_m]$, and $T^R(y(w)) \leq U'(w) + U(\underline{w}) - U(w)$ and $[1 - F(w)]/f(w) < [\Gamma(\underline{w}, \overline{w}) - \Gamma(\underline{w}, w)]/\tilde{f}(w)$ for all $w \in (w_m, \overline{w}]$.

(b) $F(w)/f(w) > \Gamma(\underline{w}, w)/\tilde{f}(w)$ for all $w \in [\underline{w}, w_m]$, and $T^R(y(w)) < U'(w) + U(\underline{w}) - U(w)$ and $[1 - F(w)]/f(w) \leq [\Gamma(\underline{w}, \overline{w}) - \Gamma(\underline{w}, w)]/\tilde{f}(w)$ for all $w \in (w_m, \overline{w}]$.

(c) $F(w)/f(w) > \Gamma(\underline{w}, w)/\tilde{f}(w)$ for all $w \in [\underline{w}, w_m]$, and $T^R(y(w)) > U'(w) + U(\underline{w}) - U(w)$, $[1 - F(w)]/f(w) < [\Gamma(\underline{w}, \overline{w}) - \Gamma(\underline{w}, w)]/\tilde{f}(w)$ and $\tilde{\theta}(w) < \Theta^R(w)$ for all $w \in (w_m, \overline{w}]$.

(d) $F(w)/f(w) > \Gamma(\underline{w}, w)/\tilde{f}(w)$ for all $w \in [\underline{w}, w_m]$, and $T^R(y(w)) < U'(w) + U(\underline{w}) - U(w)$, $[1 - F(w)]/f(w) > [\Gamma(\underline{w}, \overline{w}) - \Gamma(\underline{w}, w)]/\tilde{f}(w)$ and $\tilde{\theta}(w) > \Theta^R(w)$ for all $w \in (w_m, \overline{w}]$.

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 claim (i) still holds, and 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 claim (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. The proof is quite similar to that of the previous proposition and hence the detail is omitted to economize on the space. Basically, using Lemma 5.1, Proposition 3.2 and (27), we get $\tau^R(w) > \hat{\tau}^R(w) > 0 > \tau^M(w) > \hat{\tau}^M(w)$ under one of the four conditions, then claim (i) holds true. Using claim (i) and Figure 4, claims (ii)-(iii) are immediate. ■

All the four conditions in Proposition 5.2 require that this jurisdiction faces net labor inflow in low skills, namely that $F(w)/f(w) > \Gamma(\underline{w}, w)/\tilde{f}(w)$ for all $w \in (\underline{w}, w_m]$. Concerning high skills, these conditions impose different requirements in terms of the direction of net labor flow, the elasticity of migration and the level of tax liability in the migration equilibrium. Specifically, conditions (a) and (b) state that there is net labor outflow and the tax liability is bounded above, whereas no restrictions are imposed on the elasticity of migration. There is just a minor difference of these two conditions. Condition (c) states that there is net labor outflow, the migration elasticity is bounded above by $\Theta^R(w)$, and the tax liability is bounded below. Condition (d) states that there is net labor inflow, the migration elasticity is bounded below by $\Theta^R(w)$, and the tax liability is bounded above.

In an asymmetric migration equilibrium, claim (i) under conditions (a) and (b) must be attributable to a less redistributive taxation policy chosen by the opponent jurisdiction. Compared to its tax policy in the autarky equilibrium, a relatively more redistributive tax policy shall drive out some high skills and result in the net labor outflow of high skills. Meanwhile, even the transfers allocated to low skills are smaller than those in the autarky equilibrium, as long as they are higher than those of the opponent jurisdiction, then it must face a labor inflow of low skills. In addition, the combination of migration threat and the incentive to mimic low skills requires that the tax liability imposed on high skills must be bounded above. The interpretation of claim (i) under conditions (b) and (c) can be analogously obtained. In particular, for types belonging to $(w_m, \tilde{w}_m]$, as shown in case (iii) of Proposition 5.2, they, however, face lower tax rates than in autarky because the median voter under migration is richer than that in autarky. They belong to the high-income class in autarky while belong to the low-income class under migration, so their status changes from paying taxes to receiving transfers.

We now proceed to the qualitative characterization under the complete solution given by Theorem 3.2. In fact, we have established the following two corollaries.

Corollary 5.1 *Suppose $T^M(y(\underline{w})) < U'(\underline{w})$ holds. Let one of the conditions in Proposition 5.1 hold, then the effects of migration under the complete solution can be identified as follows.*

(i) If the left endpoint of the bridge is sufficiently close to the left endpoint of the bridge of Brett and Weymark (2017), then migration induces lower marginal tax rates for all skills.

(ii) If the left endpoint of the bridge is sufficiently smaller than that of Brett and Weymark (2017), then migration induces lower marginal tax rates for all but some skills in the left neighbourhood of the ex ante median skill level.

Proof. By the proof of Theorem 3.2 and Proposition 5.1, the relationship between the income schedule under migration and that in autarky can be illustrated by Figure 5, in which the black thick curve represents the income schedule derived by Brett and Weymark (2017) after ironing the downward discontinuity at the ex ante median skill level, with the bridge endpoints denoted by $\hat{w}_\alpha$ and $\hat{w}_\beta$. Also, w_α, w_β denote the bridge endpoints endogenously determined in the proof of Lemma 3.5. Immediately, result (i) follows from the left one of Figure 5, and result (ii) follows from the right one. ■

As shown in Figure 5, if the left endpoint, w_α, is sufficiently close to the left endpoint, $\hat{w}_\alpha$, of the bridge of Brett and Weymark (2017), then the current bridge is completely above theirs. If, however, w_α is sufficiently smaller than $\hat{w}_\alpha$, then there must be two intersections of the present income schedule and theirs. As such, the left component of their bridge is above the present one, implying lower marginal tax rates for the corresponding skill levels than ours. Since the equations determining w_α and $\hat{w}_\alpha$ are highly nonlinear equations, their relationship cannot be explicitly identified.

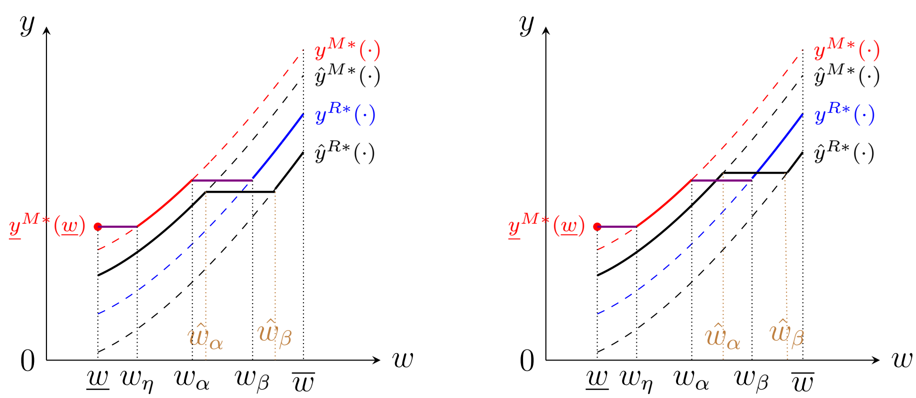

Figure 5: Graphic Illustration of Corollary 5.1.

Corollary 5.2 *Suppose $T^M(y(\underline{w})) \geq U'(\underline{w})$ holds. Let one of the conditions in Proposition 5.2 and $T^M(y(\underline{w})) \geq U'(\underline{w})$ hold, then the effects of migration under the complete solution can be identified as follows.*

(i) If the left endpoint of the bridge is sufficiently close to the left endpoint of the bridge of Brett and Weymark (2017), then migration induces higher marginal tax rates for all skills.

(ii) If the left endpoint of the bridge is sufficiently larger than that of Brett and Weymark (2017), then migration induces higher marginal tax rates for all but some skills in the left neighbourhood of the ex post median skill level.

Proof. Making use of Figure 6, the proof is similar to that of Corollary 5.1 and hence is omitted. ■

The main economic intuition of Propositions 5.1 and 5.2 still holds for these two corollaries. Taking into account the SOIC conditions makes it important to compare the set of skills bunched with the ex ante median skill level and the set of skills bunched with the ex post median skill level. Here the shift of middle class stems from the shift of the skill level of the median voter. The insight concerning the political economy approach to

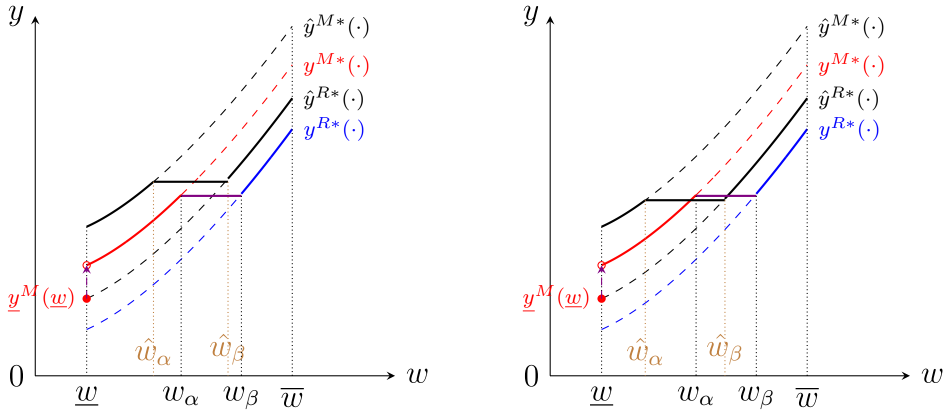

Figure 6: Graphic Illustration of Corollary 5.2.

income redistribution subject to migration is the following: the complex interaction between migration and majority voting may lead to a poorer or richer middle class than that under a benevolent social planner, yielding the possibility of departing from the conventional wisdom claiming that geographic mobility *always* limits the ability of government to redistribute incomes via a tax-transfer system (see Stigler, 1957).

5.2 The Case without a Downward Discontinuity at the Median Skill Level

We now proceed to identify the effect of migration in the case corresponding to part (ii) of Proposition 3.3. As the downward discontinuity at the median skill level only occurs in the income schedule of the autarky equilibrium, the analysis seems to be more straightforward than that of the previous subsection. We summarize the main results in the following proposition.

Proposition 5.3 *Suppose* $T^R(y(w)) > U'(w) + U(\underline{w}) - U(w)$ *and* $\tilde{\theta}(w) \geq \tilde{\theta}^*(w)$ *for all* $w \in (\tilde{w}_m, \overline{w}]$*, then the following statements are true.*

- *(i) If* $F(w)/f(w) < \Gamma(\underline{w}, w)/\tilde{f}(w)$ *for all* $w \in (\underline{w}, \tilde{w}_m)$ *and* $T^M(y(\underline{w})) < U'(\underline{w})$*, then workers of all skill types face lower tax rates than in autarky.*
- *(ii) It is impossible that all skill types face higher tax rates than in autarky.*
- *(iii) If* $F(w)/f(w) > \Gamma(\underline{w}, w)/\tilde{f}(w)$ *for all* $w \in (\underline{w}, \tilde{w}_m)$ *and* $T^M(y(\underline{w})) \geq U'(\underline{w})$*, then it is possible that all low skills receive less transfers while all high skills pay less taxes than in autarky.*

Proof. Noting the fact that the maximax income schedule is above the maximin income schedule in the autarky equilibrium while the maximax income schedule is not above the maximin income schedule in the migration equilibrium, claim (i) follows from applying part (i) of Lemma 5.1 and part (i) of Proposition 3.2. Under this fact, for all skill types to face higher tax rates under migration than in autarky, it must be that the maximin tax rates in the migration equilibrium are larger than those in the autarky equilibrium, which however cannot be satisfied by using part (ii) of Lemma 5.1 under the assumption of $T^R(y(w)) > U'(w) + U(\underline{w}) - U(w)$ and $\tilde{\theta}(w) \geq \tilde{\theta}^*(w)$ for all $w \in (\tilde{w}_m, \overline{w}]$. As such, we must have higher maximin tax rates in the autarky equilibrium than those in the migration equilibrium. Applying again the fact given at the beginning, claim (iii) is immediate by using part (i) of Lemma 5.1 and part (i) of Proposition 3.2. ■

If high skills face negative marginal tax rates in the migration equilibrium, then, ceteris paribus, they must be strictly better off under migration than in autarky, which gives the main insight of Proposition 5.3. While for low skills, although they face negative marginal tax rates regardless of whether migration is allowed or not, they could be either better off or worse off under migration than in autarky, depending on the direction of equilibrium labor flow. If the jurisdiction faces net labor outflow of low skills, then they are better off under migration, as shown in claim (i); otherwise, they are likely to be worse off under migration, as shown in claim (iii). The basic intuition is that each low skill worker becomes a negative externality to the other low skills as long as they are ready to receive transfers.

6 Conclusion

In this paper we have examined the feature of redistributive taxation when voters are geographically mobile at the expense of some unobserved migration costs. 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 majority voting and inter-jurisdiction migration, which hence makes the level of distortion and redistribution tend to deviate from that in autarky. The basic intuition is the following. For low skills who receive transfers under migration and autarky, each low skill worker becomes a negative externality to the other low skills, which hence makes migration or net labor flow per se relevant in terms of taxation. For high skills, if they pay positive taxes under migration and autarky, then ceteris paribus each one becomes a positive externality to the others, which again makes migration relevant. Moreover, if some high skills turn out to receive transfers under migration, then the effect of migration becomes more significant. We provide the sufficient and necessary conditions associated to the elasticity and level of migration to show the departure of the equilibrium level of redistribution under migration to that in autarky.

The selfish median voter faces the tradeoff between maximizing resources extracted from other types and maximizing resources available for extraction. For low skills, she transfers a positive amount of resources in the selfishly optimal tax schedule, which holds regardless of whether inter-jurisdiction migration is allowed or not; for high skills, especially those with high migration abilities, she 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 she benefits the most from such tax schedule, which is actually socially desirable when the middle class consists of the major part of a society, both equity and efficiency concerns are taken into account seriously under such type of institutional arrangement.

For future research, our model can be modified or extended in at least three directions. First, as quasilinear-in-labor preferences are often used in the income taxation literature, a parallel analysis can be conducted under such preferences, and novel implications for the design of redistributive taxation may emerge. Second, instead of proposing selfishly optimal tax schedules, we may expect voters of certain skills to exhibit other-regarding or pro-social preferences. And third, by imposing specific distribution functions of skills and migration costs as well as specific correlations between these two unobservable variables, one could investigate possible sorting types in the voting equilibrium.

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] 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.

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

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

[6] 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.

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

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

[9] 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.

[10] 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.

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

[12] Brett, C. and J. A. Weymark [2019]: "Majority Rule and Selfishly Optimal Nonlinear Income Tax Schedules with Discrete Skill Levels." *Social Choice and Welfare*, forthcoming.

[13] 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.

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

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

[16] Dai, D. [2019]: "Voting over Selfishly Optimal Tax Schedules: Can Pigouvian Tax Redistribute Income?" Working Paper, Institute for Advanced Research (IAR), Shanghai University of Finance and Economics.

[17] Dai, D., W. Gao and G. Tian [2019]: "Relativity, Mobility, and Optimal Nonlinear Income Taxation in an Open Economy." Working Paper, Institute for Advanced Research (IAR), Shanghai University of Finance and Economics.

[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] 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.

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

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

[22] 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.

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

[24] 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.

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

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

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

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

[29] 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.

[30] 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.

[31] 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.

[32] 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.

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

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

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

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

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

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

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

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

[41] 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.

[42] 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.

[43] 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.

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

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

Appendix: Proofs

Proof of Lemma 3.1. By using (6), we have

$$U(w) = U(\underline{w}) + \int_{\underline{w}}^{w} \frac{y(t)}{t^2} h'\left(\frac{y(t)}{t}\right) dt. \tag{39}$$

Integrating over the ex post support of the skill distribution yields

$$\int_{\underline{w}}^{\overline{w}} U(w)\tilde{f}(w)dw = U(\underline{w}) \int_{\underline{w}}^{\overline{w}} \tilde{f}(w)dw + \int_{\underline{w}}^{\overline{w}} \left[\int_{\underline{w}}^{w} \frac{y(t)}{t^2} h'\left(\frac{y(t)}{t}\right) dt\right] \tilde{f}(w)dw. \tag{40}$$

Reversing the order of integration in (40) gives rise to

$$\int_{\underline{w}}^{\overline{w}} U(w)\tilde{f}(w)dw = U(\underline{w}) \int_{\underline{w}}^{\overline{w}} \tilde{f}(w)dw + \int_{\underline{w}}^{\overline{w}} \frac{y(t)}{t^2} h'\left(\frac{y(t)}{t}\right) \left[\int_{t}^{\overline{w}} \tilde{f}(w)dw\right] dt. \tag{41}$$

Also, it follows from (5) that

$$\int_{\underline{w}}^{\overline{w}} U(w)\tilde{f}(w)dw = \int_{\underline{w}}^{\overline{w}} c(w)\tilde{f}(w)dw - \int_{\underline{w}}^{\overline{w}} h\left(\frac{y(w)}{w}\right) \tilde{f}(w)dw. \tag{42}$$

Applying the equality form of (11) to (42) shows that

$$\int_{\underline{w}}^{\overline{w}} U(w)\tilde{f}(w)dw = \int_{\underline{w}}^{\overline{w}} y(w)\tilde{f}(w)dw - \int_{\underline{w}}^{\overline{w}} h\left(\frac{y(w)}{w}\right) \tilde{f}(w)dw. \tag{43}$$

Combining (41) and (43) leads us to

$$\begin{aligned} U(\underline{w}) \int_{\underline{w}}^{\overline{w}} \tilde{f}(w)dw = & \int_{\underline{w}}^{\overline{w}} y(w)\tilde{f}(w)dw - \int_{\underline{w}}^{\overline{w}} h\left(\frac{y(w)}{w}\right) \tilde{f}(w)dw \\ & - \int_{\underline{w}}^{\overline{w}} \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \left[\int_{w}^{\overline{w}} \tilde{f}(t)dt\right] dw. \end{aligned} \tag{44}$$

Applying (10), we can rewrite (44) as

$$U(\underline{w}) = \frac{1}{\Gamma(\underline{w}, \overline{w})} \int_{\underline{w}}^{\overline{w}} \left\{ \left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \tilde{f}(w) - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \Gamma(w, \overline{w}) \right\} dw. \tag{45}$$

Substituting (45) into (39) and setting $w = k$, then the maximand in (14) is established. ■

Proof of Lemma 3.2. By setting $k = \underline{w}$, the maximand of problem (14) is hence given by (45). It is straightforward that the corresponding maximization problem can be solved point-wise. By letting $\partial U(\underline{w})/\partial y(w) = 0$ and rearranging the algebra, we obtain

$$\begin{aligned} & \left[1 - \frac{1}{w} h'\left(\frac{y(w)}{w}\right)\right] \tilde{f}(w) + \left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \frac{\partial \tilde{f}(w)}{\partial y(w)} - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \frac{\partial \Gamma(w, \overline{w})}{\partial y(w)} \\ = & \, U(\underline{w}) \frac{\partial \Gamma(\underline{w}, \overline{w})}{\partial y(w)} + \left[\frac{1}{w^2} h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^3} h''\left(\frac{y(w)}{w}\right)\right] \Gamma(w, \overline{w}). \end{aligned} \tag{46}$$

As is obvious by (10) that

$$\frac{\partial \Gamma(\underline{w}, \overline{w})}{\partial y(w)} = \frac{\partial \Gamma(w, \overline{w})}{\partial y(w)} = \frac{\partial \tilde{f}(w)}{\partial y(w)}, \tag{47}$$

applying which to (46) and rearranging the algebra, the desired (15) is hence established. By setting $k = w$ in the maximand of problem (14), then, for $\forall w \in (\underline{w}, \overline{w})$, (17) is immediate by evaluating

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

at the maximin income schedule. Also, by using (8), (16) is immediate. ■

Proof of Lemma 3.3. By setting $k = \overline{w}$, the maximand of problem (14) can be written as

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

The maximization problem can be solved point-wise. Applying $\partial U(\overline{w})/\partial y(w) = 0$ and (47) to (48) and rearranging the algebra, we obtain

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

Noting from (39) that $\partial U(w)/\partial y(w) = \partial U(\overline{w})/\partial y(w) = 0$ for all $w \in (\underline{w}, \overline{w})$ whenever evaluated at the maximax income schedule. Applying this to (16) reveals that $\partial \tilde{f}(w)/\partial y(w) = 0$ for all $w \in (\underline{w}, \overline{w})$, thus (18) follows immediately from (49). Moreover, it follows from (48) that the first order condition can be expressed as:

$$\frac{\partial U(\overline{w})}{\partial y(w)} = \frac{\partial U(\underline{w})}{\partial y(w)} + \frac{\partial}{\partial y(w)} \int_{\underline{w}}^{\overline{w}} \frac{y(t)}{t^2} h'\left(\frac{y(t)}{t}\right) dt = 0,$$

evaluating which at $y(w) = y(\underline{w})$ immediately gives (21). Evaluating (16) at $w = \underline{w}$ gives (20). Then, we obtain (19) by evaluating (49) at $w = \underline{w}$. ■

Proof of Proposition 3.1. By using the maximization problem (14) stated in Lemma 3.1, 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 (22) follows from a direct application of Lemmas 3.2 and 3.3. ■

Proof of Proposition 3.2. First, (28) in part (i) is immediate by applying (20)-(21) to the tax formula of (24). Part (ii) is also immediate by (25). By using the chain rule of calculus, (8) and (9), we have

$$\frac{\partial \tilde{f}(w)}{\partial y(w)} \frac{1}{\tilde{f}(w)} = \frac{\partial \tilde{f}(w)}{\partial \Delta(w)} \frac{c(w)}{\tilde{f}(w)} \frac{\partial U(w)}{\partial y(w)} \frac{1}{c(w)} = \tilde{\theta}(w) \frac{\partial U(w)}{\partial y(w)} \frac{1}{c(w)}. \tag{50}$$

Then we get from (50), (17), (26), (2), (5)-(6) and the condition

$$y(w) - h\left(\frac{y(w)}{w}\right) - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) = T^R(y(w)) + U(w) - U'(w) > U(\underline{w}) \tag{51}$$

that

$$\tau^R(w) = \underbrace{\frac{\partial U(w)}{\partial y(w)}}_{>0} \left\{ \underbrace{\frac{\Gamma(w, \overline{w})}{\tilde{f}(w)}}_{>0} - \underbrace{\frac{\tilde{\theta}(w)}{c(w)} \left[y(w) - h\left(\frac{y(w)}{w}\right) - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) - U(\underline{w}) \right]}_{>0} \right\},$$

by which assertion (29) is immediate. It follows from (6) that $U(\overline{w}) > U(\underline{w})$. Making use of (51) again shows the desired assertion (30). ■

Proof of Proposition 3.3. We shall complete the proof in three steps.

Step 1. It follows from (25) and (26) that

$$\begin{aligned} \tau^M(w) - \tau^R(w) = {} & \frac{\partial \tilde{f}(w)}{\partial y(w)} \frac{1}{\tilde{f}(w)} \left[y(w) - h\left(\frac{y(w)}{w}\right) - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) - U(\underline{w}) \right] \\ & - \frac{\Gamma(\underline{w}, \overline{w})}{w \tilde{f}(w)} \left[\frac{1}{w} h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^2} h''\left(\frac{y(w)}{w}\right) \right]. \end{aligned} \tag{52}$$

Applying (51) to (52), we immediately get $\tau^M(w) < \tau^R(w)$ for $T^R(y(w)) \leq U'(w) + U(\underline{w}) - U(w)$. Then we have either $\tau^M(w) < 0 < \tau^R(w)$ or $\tau^M(w) < \tau^R(w) \leq 0$. If $\tau^M(w) < 0 < \tau^R(w)$, then under tax schedule $\tau^M(\cdot)$ each type-w worker has her income distorted upward compared to the full-information solution, whereas her income is distorted downward compared to the full-information solution under tax schedule $\tau^R(\cdot)$. If $\tau^M(w) < \tau^R(w) \leq 0$, then each type-$w$ worker has her income distorted upward compared to the full-information solution under both tax schedules, but the magnitude of distortion is bigger under tax schedule $\tau^M(\cdot)$. Thus, no matter which case we consider, we see an upward discontinuity of the income schedule, as desired in part (i-a).

Step 2. If, however, $T^R(y(w)) > U'(w) + U(\underline{w}) - U(w)$, then we get from applying (50) and (51) to (52) that

$$\begin{aligned} & \tau^M(w) - \tau^R(w) \\ & = \underbrace{\frac{\partial U(w)}{\partial y(w)}}_{>0} \left\{ \underbrace{\frac{\tilde{\theta}(w)}{c(w)} \left[y(w) - h\left(\frac{y(w)}{w}\right) - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) - U(\underline{w}) \right]}_{>0} - \underbrace{\frac{\Gamma(\underline{w}, \overline{w})}{\tilde{f}(w)}}_{>0} \right\}, \end{aligned}$$

by which we arrive at the following result:

$$\tau^M(w) \begin{cases} < \tau^R(w) & \text{for } \tilde{\theta}(w) < \tilde{\theta}^*(w), \\ = \tau^R(w) & \text{for } \tilde{\theta}(w) = \tilde{\theta}^*(w), \\ > \tau^R(w) & \text{for } \tilde{\theta}(w) > \tilde{\theta}^*(w), \end{cases} \tag{53}$$

in which the critical value of migration elasticity is given by

$$\tilde{\theta}^*(w) = \frac{c(w)\Gamma(\underline{w}, \overline{w})}{[y(w) - h(y(w)/w) - (y(w)/w^2)h'(y(w)/w) - U(\underline{w})]\, \tilde{f}(w)}. \tag{54}$$

Using (29), (53)-(54) and $\tilde{\theta}^*(w) > \tilde{\theta}_\star(w)$, we have the following results: $0 > \tau^M(w) > \tau^R(w)$ if $\tilde{\theta}(w) > \tilde{\theta}^*(w)$, $0 > \tau^M(w) = \tau^R(w)$ if $\tilde{\theta}(w) = \tilde{\theta}^*(w)$, $\tau^M(w) < 0 \le \tau^R(w)$ if $\tilde{\theta}(w) \le \tilde{\theta}_\star(w) < \tilde{\theta}^*(w)$, and $\tau^M(w) < \tau^R(w) < 0$ if $\tilde{\theta}_\star(w) < \tilde{\theta}(w) < \tilde{\theta}^*(w)$. By applying the same reasoning used to prove part (i-a), the desired assertions in parts (i-b) and (ii) follows.

Step 3. As w approaches $\underline{w}$ from the above, we get from equation (25) and the continuity of $\tau^M(w)$ over the interval $(\underline{w}, k)$ that $\tau^M(\underline{w}) = 0$. As a result, under the marginal tax rate $\tau^M(\underline{w})$ each type-$\underline{w}$ worker has her income non-distorted as in the full-information solution. We thus have by using equations (24), (20), (21), (2), (5) and (6) that

$$\underline{\tau}^M(\underline{w}) - \tau^M(\underline{w}) = -\underbrace{\frac{\partial \tilde{f}(\underline{w})}{\partial y(\underline{w})} \frac{1}{\tilde{f}(\underline{w})}}_{>0} \cdot \left[T^M(y(\underline{w})) - U'(\underline{w})\right].$$

If $T^M(y(\underline{w})) > U'(\underline{w})$, then we immediately have $\underline{\tau}^M(\underline{w}) > \tau^M(\underline{w}) = 0$, yielding that under the marginal tax rate $\underline{\tau}^M(\underline{w})$ each type-$\underline{w}$ worker has her income distorted downward compared to the full-information solution. We thus see an upward discontinuity of the income schedule at the bottom skill level, as desired in claim (iv). If, in the contrast, $T^M(y(\underline{w})) < U'(\underline{w})$, then we have $\underline{\tau}^M(\underline{w}) < \tau^M(\underline{w}) = 0$, yielding a downward discontinuity of the income schedule at the bottom skill level, as desired in claim (iii). ■

Proof of Theorem 3.2. We shall complete the proof in four steps.

Step 1. Under the conditions given in Theorem 3.2, we get from Proposition 3.3 that there will be two downward discontinuities in the income schedule, and hence we need to build two bridges such that the resulting income schedule satisfies the SOIC condition. Obviously, the left endpoint of the first bridge is just $\underline{w}$. Let's fix first the other bridge endpoints w_η, w_α and w_β that shall be endogenously determined, and let $y^*(\underline{w}, w_\eta)$ and $y^*(w_\alpha, w_\beta)$ denote the optimal before-tax income levels on the bridges over skill intervals $[\underline{w}, w_\eta]$ and $[w_\alpha, w_\beta]$, respectively. Without loss of generality, we assume that the bridge in the middle 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 $y^{R*}(\cdot)$. 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.

Step 2. We now equivalently rewrite the maximand of problem (14) as follows:

$$
\begin{aligned}
U(k) = & \int_{\underline{w}}^{k} \left\{ \left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \frac{\tilde{f}(w)}{\Gamma(\underline{w}, \overline{w})} + \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \left[1 - \frac{\Gamma(w, \overline{w})}{\Gamma(\underline{w}, \overline{w})} \right] \right\} dw \\
& + \int_{k}^{\overline{w}} \left\{ \left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \frac{\tilde{f}(w)}{\Gamma(\underline{w}, \overline{w})} - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \frac{\Gamma(w, \overline{w})}{\Gamma(\underline{w}, \overline{w})} \right\} dw.
\end{aligned}
\tag{55}
$$

Taking into account the bunching possibility, (55) should be modified as follows:

$$
\begin{aligned}
U^*(k) = & \int_{\underline{w}}^{k} \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w | w \notin \mathcal{B}^M\}} dw \\
& + \int_{\underline{w}}^{k} \tilde{\Phi}^{M*}(w, y(w), \Gamma(w_-, w_+), \Gamma(w_-, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w | w \in \mathcal{B}^M\}} dw \\
& + \int_{k}^{\overline{w}} \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w | w \notin \mathcal{B}^R\}} dw \\
& + \int_{k}^{\overline{w}} \tilde{\Phi}^{R*}(w, y(w), \Gamma(w_-, w_+), \Gamma(w_+, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w | w \in \mathcal{B}^R\}} dw,
\end{aligned}
\tag{56}
$$

in which

$$
\begin{aligned}
& \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \\
\equiv & \left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \frac{\tilde{f}(w)}{\Gamma(\underline{w}, \overline{w})} + \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \left[1 - \frac{\Gamma(w, \overline{w})}{\Gamma(\underline{w}, \overline{w})} \right]; \\
& \tilde{\Phi}^{M*}(w, y(w), \Gamma(w_-, w_+), \Gamma(w_-, \overline{w}), \Gamma(\underline{w}, \overline{w})) \\
\equiv & \left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \frac{\Gamma(w_-, w_+)}{\Gamma(\underline{w}, \overline{w})} + \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \left[1 - \frac{\Gamma(w_-, \overline{w})}{\Gamma(\underline{w}, \overline{w})} \right]
\end{aligned}
\tag{57}
$$

and

$$
\begin{aligned}
& \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \\
\equiv & \left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \frac{\tilde{f}(w)}{\Gamma(\underline{w}, \overline{w})} - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \frac{\Gamma(w, \overline{w})}{\Gamma(\underline{w}, \overline{w})}; \\
& \tilde{\Phi}^{R*}(w, y(w), \Gamma(w_-, w_+), \Gamma(w_+, \overline{w}), \Gamma(\underline{w}, \overline{w})) \\
\equiv & \left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \frac{\Gamma(w_-, w_+)}{\Gamma(\underline{w}, \overline{w})} - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \frac{\Gamma(w_+, \overline{w})}{\Gamma(\underline{w}, \overline{w})}
\end{aligned}
\tag{58}
$$

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

$$\begin{aligned}
\max_{y(\cdot)} \Bigg\{ & \int_{\underline{w}}^{w_\eta} \tilde{\Phi}^{M*}(w, y(w), \Gamma(\underline{w}, w_\eta), \Gamma(\underline{w}, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta > \underline{w}\}} dw \\
& + \int_{w_\eta}^{w_\alpha} \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta \leq w_\alpha\}} dw \\
& + \int_{w_\alpha}^{k} \tilde{\Phi}^{M*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\alpha, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}} dw \\
& + \int_{k}^{w_\beta} \tilde{\Phi}^{R*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\}} dw \\
& + \int_{w_\beta}^{\overline{w}} \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{\overline{w} \geq w_\beta\right\}} dw \Bigg\},
\end{aligned} \tag{59}$$

subject to

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

In consequence, problem (59) can be simplified as the following unconstrained maximization problem:

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

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

$$\begin{aligned}
& \frac{\partial \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial y(w)} \\
& + \frac{\partial \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial \tilde{f}(w)} \frac{\partial \tilde{f}(w)}{\partial y(w)} \\
& + \frac{\partial \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial \Gamma(w, \overline{w})} \frac{\partial \Gamma(w, \overline{w})}{\partial y(w)} \\
& + \frac{\partial \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial \Gamma(\underline{w}, \overline{w})} \frac{\partial \Gamma(\underline{w}, \overline{w})}{\partial y(w)} = 0 \quad \text{for } \forall w \in (w_\eta, w_\alpha),
\end{aligned} \tag{60}$$

and

$$
\begin{aligned}
&\frac{\partial\tilde{\Phi}^{R*}(w,y(w),\tilde{f}(w),\Gamma(w,\overline{w}),\Gamma(\underline{w},\overline{w}))}{\partial y(w)}\\
&+\frac{\partial\tilde{\Phi}^{R*}(w,y(w),\tilde{f}(w),\Gamma(w,\overline{w}),\Gamma(\underline{w},\overline{w}))}{\partial\tilde{f}(w)}\frac{\partial\tilde{f}(w)}{\partial y(w)}\\
&+\frac{\partial\tilde{\Phi}^{R*}(w,y(w),\tilde{f}(w),\Gamma(w,\overline{w}),\Gamma(\underline{w},\overline{w}))}{\partial\Gamma(w,\overline{w})}\frac{\partial\Gamma(w,\overline{w})}{\partial y(w)}\\
&+\frac{\partial\tilde{\Phi}^{R*}(w,y(w),\tilde{f}(w),\Gamma(w,\overline{w}),\Gamma(\underline{w},\overline{w}))}{\partial\Gamma(\underline{w},\overline{w})}\frac{\partial\Gamma(\underline{w},\overline{w})}{\partial y(w)}=0\quad\text{for }\forall w\in(w_\beta,\overline{w}].
\end{aligned}\tag{61}
$$

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

Step 4. 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 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 $[w_\beta,\overline{w}]$. 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 $y^*(\underline{w},w_\eta)\equiv\underline{y}^{M*}(\underline{w})$, the desired income schedule given by (31) is therefore established. ■

Proof of Lemma 3.4. We get from (56), (59), (20) and (21) that

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

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

Proof of Lemma 3.5. We shall complete the proof in five steps.

Step 1. 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

3.2 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

$$\psi(w_\beta) \equiv \begin{cases} (y^{M*})^{-1}(y^{R*}(w_\beta)) & \text{if } w_\alpha > \underline{w}, \\ w_\alpha & \text{if } w_\alpha = \underline{w}. \end{cases} \tag{62}$$

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

$$\begin{aligned} &\Xi(w_\beta; k) \\ \equiv &\int_{w_\eta}^{\psi(w_\beta)} \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta \le w_\alpha\} \cap \{y(w) = y^{M*}(w)\}} dw \\ &+ \int_{\psi(w_\beta)}^{k} \tilde{\Phi}^{M*}(w, y(w), \Gamma(\psi(w_\beta), w_\beta), \Gamma(\psi(w_\beta), \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\} \cap \left\{y(w) = y^{R*}(w_\beta)\right\}} dw \\ &+ \int_{k}^{w_\beta} \tilde{\Phi}^{R*}(w, y(w), \Gamma(\psi(w_\beta), w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\} \cap \left\{y(w) = y^{R*}(w_\beta)\right\}} dw \\ &+ \int_{w_\beta}^{\overline{w}} \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{\overline{w} \ge w_\beta\right\} \cap \{y(w) = y^{R*}(w)\}} dw. \end{aligned} \tag{63}$$

Using (63), the first-order condition with respect to w_β can be derived as

$$\Psi_1 + \Psi_2(k) + \Psi_3(k) + \Psi_4 = 0, \tag{64}$$

in which

$$\begin{aligned} &\Psi_1 = \\ &\frac{\mathrm{d}\psi(w_\beta)}{\mathrm{d}w_\beta} \Big\{ \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta \le w_\alpha\} \cap \{y(w) = y^{M*}(w)\} \cap \left\{w = \psi(w_\beta)\right\}} \\ &- \tilde{\Phi}^{M*}(w, y(w), \Gamma(\psi(w_\beta), w_\beta), \Gamma(\psi(w_\beta), \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\} \cap \left\{y(w) = y^{R*}(w_\beta)\right\} \cap \left\{w = \psi(w_\beta)\right\}} \Big\}, \end{aligned} \tag{65}$$

$$\begin{aligned} &\Psi_2(k) = \\ &\int_{\psi(w_\beta)}^{k} \frac{\partial \tilde{\Phi}^{M*}(w, y^{R*}(w_\beta), \Gamma(\psi(w_\beta), w_\beta), \Gamma(\psi(w_\beta), \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial y^{R*}(w_\beta)} \frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta} \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}} dw \\ &+ \int_{\psi(w_\beta)}^{k} \frac{\partial \tilde{\Phi}^{M*}(w, y^{R*}(w_\beta), \Gamma(\psi(w_\beta), w_\beta), \Gamma(\psi(w_\beta), \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial \Gamma(\psi(w_\beta), w_\beta)} \frac{\partial \Gamma(\psi(w_\beta), w_\beta)}{\partial w_\beta} \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}} dw \\ &+ \int_{\psi(w_\beta)}^{k} \frac{\partial \tilde{\Phi}^{M*}(w, y^{R*}(w_\beta), \Gamma(\psi(w_\beta), w_\beta), \Gamma(\psi(w_\beta), \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial \Gamma(\psi(w_\beta), \overline{w})} \frac{\partial \Gamma(\psi(w_\beta), \overline{w})}{\partial w_\beta} \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}} dw \\ &+ \int_{\psi(w_\beta)}^{k} \frac{\partial \tilde{\Phi}^{M*}(w, y^{R*}(w_\beta), \Gamma(\psi(w_\beta), w_\beta), \Gamma(\psi(w_\beta), \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial \Gamma(\underline{w}, \overline{w})} \frac{\partial \Gamma(\underline{w}, \overline{w})}{\partial w_\beta} \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}} dw \\ &\equiv \int_{\psi(w_\beta)}^{k} [\Psi_{21}(w) + \Psi_{22}(w) + \Psi_{23}(w) + \Psi_{24}(w)] \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}} dw, \end{aligned} \tag{66}$$

$$
\begin{aligned}
&\Psi_3(k) = \\
&\int_k^{w_\beta} \frac{\partial \tilde{\Phi}^{R*}(w, y^{R*}(w_\beta), \Gamma(\psi(w_\beta), w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial y^{R*}(w_\beta)} \frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta} \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\}} dw \\
&+ \int_k^{w_\beta} \frac{\partial \tilde{\Phi}^{R*}(w, y^{R*}(w_\beta), \Gamma(\psi(w_\beta), w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial \Gamma(\psi(w_\beta), w_\beta)} \frac{\partial \Gamma(\psi(w_\beta), w_\beta)}{\partial w_\beta} \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\}} dw \\
&+ \int_k^{w_\beta} \frac{\partial \tilde{\Phi}^{R*}(w, y^{R*}(w_\beta), \Gamma(\psi(w_\beta), w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial \Gamma(w_\beta, \overline{w})} \frac{\partial \Gamma(w_\beta, \overline{w})}{\partial w_\beta} \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\}} dw \\
&+ \int_k^{w_\beta} \frac{\partial \tilde{\Phi}^{R*}(w, y^{R*}(w_\beta), \Gamma(\psi(w_\beta), w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial \Gamma(\underline{w}, \overline{w})} \frac{\partial \Gamma(\underline{w}, \overline{w})}{\partial w_\beta} \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\}} dw \\
&\equiv \int_k^{w_\beta} \left[\Psi_{31}(w) + \Psi_{32}(w) + \Psi_{33}(w) + \Psi_{34}(w)\right] \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\}} dw,
\end{aligned} \tag{67}
$$

and

$$
\begin{aligned}
\Psi_4 = &\ \tilde{\Phi}^{R*}(w, y(w), \Gamma(\psi(w_\beta), w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\} \cap \left\{y(w) = y^{R*}(w_\beta)\right\} \cap \left\{w = w_\beta\right\}} \\
&- \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{\overline{w} \geq w_\beta\right\} \cap \{y(w) = y^{R*}(w)\} \cap \left\{w = w_\beta\right\}}.
\end{aligned} \tag{68}
$$

Step 2. By using (57) and (58), we can have

$$
\begin{aligned}
&\tilde{\Phi}^{M*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\alpha, \overline{w}), \Gamma(\underline{w}, \overline{w})) \\
&\quad \equiv \left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \frac{\Gamma(w_\alpha, w_\beta)}{\Gamma(\underline{w}, \overline{w})} + \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right)\left[1 - \frac{\Gamma(w_\alpha, \overline{w})}{\Gamma(\underline{w}, \overline{w})}\right], \\
&\tilde{\Phi}^{R*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w})) \\
&\quad \equiv \left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \frac{\Gamma(w_\alpha, w_\beta)}{\Gamma(\underline{w}, \overline{w})} - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \frac{\Gamma(w_\beta, \overline{w})}{\Gamma(\underline{w}, \overline{w})},
\end{aligned} \tag{69}
$$

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

By using (62), (69), (57) and (58), we can rewrite (65) as

$$
\begin{aligned}
\Psi_1 =& \frac{\mathrm{d}\psi(w_\beta)}{\mathrm{d}w_\beta} \Big\{\tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta \leq w_\alpha\} \cap \left\{y(w) = y^{R*}(w_\beta)\right\} \cap \left\{w = \psi(w_\beta)\right\}} \\
&- \tilde{\Phi}^{M*}(w, y(w), \Gamma(w, w_\beta), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\} \cap \left\{y(w) = y^{R*}(w_\beta)\right\} \cap \left\{w = \psi(w_\beta)\right\}}\Big\} \\
=& \frac{\mathrm{d}\psi(w_\beta)}{\mathrm{d}w_\beta} \Big\{\tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \\
&- \tilde{\Phi}^{M*}(w, y(w), \Gamma(w, w_\beta), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w}))\Big\} \cdot \mathbb{I}_{\{w_\eta \leq w_\alpha\} \cap \{w_\alpha > \underline{w}\} \cap \left\{y(w) = y^{R*}(w_\beta)\right\} \cap \left\{w = \psi(w_\beta)\right\}} \\
=& \frac{\mathrm{d}\psi(w_\beta)}{\mathrm{d}w_\beta} \left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \frac{\tilde{f}(w) - \Gamma(w, w_\beta)}{\Gamma(\underline{w}, \overline{w})} \cdot \mathbb{I}_{\{w_\alpha \geq w_\eta\} \cap \left\{y(w) = y^{R*}(w_\beta)\right\} \cap \left\{w = \psi(w_\beta)\right\}} \\
=& \frac{\mathrm{d}\psi(w_\beta)}{\mathrm{d}w_\beta} \left[y^{R*}(w_\beta) - h\left(\frac{y^{R*}(w_\beta)}{\psi(w_\beta)}\right)\right] \frac{\tilde{f}(\psi(w_\beta)) - \Gamma(\psi(w_\beta), w_\beta)}{\Gamma(\underline{w}, \overline{w})} \cdot \mathbb{I}_{\{w_\alpha \geq w_\eta\}}.
\end{aligned} \tag{70}
$$

By using (66) and (69), we have

$$\begin{aligned}\Psi_{21}(w) = &\left\{ \left[1 - \frac{1}{w}h'\left(\frac{y^{R*}(w_\beta)}{w}\right)\right] \frac{\Gamma(\psi\left(w_\beta\right), w_\beta)}{\Gamma(\underline{w}, \overline{w})} \right. \\ &\left. + \left[\frac{1}{w^2}h'\left(\frac{y^{R*}(w_\beta)}{w}\right) + \frac{y^{R*}(w_\beta)}{w^3}h''\left(\frac{y^{R*}(w_\beta)}{w}\right)\right]\left[1 - \frac{\Gamma(\psi\left(w_\beta\right), \overline{w})}{\Gamma(\underline{w}, \overline{w})}\right]\right\} \frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta},\end{aligned} \tag{71}$$

$$\begin{aligned}\Psi_{22}(w) = &\left[y^{R*}(w_\beta) - h\left(\frac{y^{R*}(w_\beta)}{w}\right)\right] \frac{1}{\Gamma(\underline{w}, \overline{w})} \\ &\times \left\{\tilde{f}(w_\beta) - \tilde{f}\left(\psi(w_\beta)\right)\frac{\mathrm{d}\psi(w_\beta)}{\mathrm{d}w_\beta} + \left[\int_{\psi(w_\beta)}^{w_\beta} \frac{\partial \tilde{f}(w)}{\partial y^{R*}(w_\beta)} \frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta} dw\right]\right\},\end{aligned} \tag{72}$$

$$\begin{aligned}\Psi_{23}(w) = &\frac{y^{R*}(w_\beta)}{w^2}h'\left(\frac{y^{R*}(w_\beta)}{w}\right)\frac{1}{\Gamma(\underline{w}, \overline{w})} \\ &\times \left[\tilde{f}\left(\psi(w_\beta)\right)\frac{\mathrm{d}\psi(w_\beta)}{\mathrm{d}w_\beta} - \int_{\psi(w_\beta)}^{w_\beta} \frac{\partial \tilde{f}(w)}{\partial y^{R*}(w_\beta)} \frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta} dw\right],\end{aligned} \tag{73}$$

and

$$\begin{aligned}\Psi_{24}(w) = &\left\{\frac{y^{R*}(w_\beta)}{w^2}h'\left(\frac{y^{R*}(w_\beta)}{w}\right)\Gamma(\psi(w_\beta), \overline{w}) - \left[y^{R*}(w_\beta) - h\left(\frac{y^{R*}(w_\beta)}{w}\right)\right]\Gamma(\psi(w_\beta), w_\beta)\right\} \\ &\times \frac{1}{\Gamma(\underline{w}, \overline{w})^2}\int_{\psi(w_\beta)}^{w_\beta} \frac{\partial \tilde{f}(w)}{\partial y^{R*}(w_\beta)} \frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta} dw.\end{aligned} \tag{74}$$

Similarly, By using (67) and (69), we have

$$\begin{aligned}\Psi_{31}(w) = &\left\{ \left[1 - \frac{1}{w}h'\left(\frac{y^{R*}(w_\beta)}{w}\right)\right] \frac{\Gamma(\psi\left(w_\beta\right), w_\beta)}{\Gamma(\underline{w}, \overline{w})} \right. \\ &\left. - \left[\frac{1}{w^2}h'\left(\frac{y^{R*}(w_\beta)}{w}\right) + \frac{y^{R*}(w_\beta)}{w^3}h''\left(\frac{y^{R*}(w_\beta)}{w}\right)\right]\frac{\Gamma(w_\beta, \overline{w})}{\Gamma(\underline{w}, \overline{w})}\right\} \frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta},\end{aligned} \tag{75}$$

$$\begin{aligned}\Psi_{32}(w) = &\left[y^{R*}(w_\beta) - h\left(\frac{y^{R*}(w_\beta)}{w}\right)\right] \frac{1}{\Gamma(\underline{w}, \overline{w})} \\ &\times \left\{\tilde{f}(w_\beta) - \tilde{f}\left(\psi(w_\beta)\right)\frac{\mathrm{d}\psi(w_\beta)}{\mathrm{d}w_\beta} + \left[\int_{\psi(w_\beta)}^{w_\beta} \frac{\partial \tilde{f}(w)}{\partial y^{R*}(w_\beta)} \frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta} dw\right]\right\},\end{aligned} \tag{76}$$

$$\Psi_{33}(w) = \frac{y^{R*}(w_\beta)}{w^2}h'\left(\frac{y^{R*}(w_\beta)}{w}\right)\frac{\tilde{f}\left(w_\beta\right)}{\Gamma(\underline{w}, \overline{w})}, \tag{77}$$

and

$$\begin{aligned}\Psi_{34}(w) = &\left\{\frac{y^{R*}(w_\beta)}{w^2}h'\left(\frac{y^{R*}(w_\beta)}{w}\right)\Gamma(w_\beta, \overline{w}) - \left[y^{R*}(w_\beta) - h\left(\frac{y^{R*}(w_\beta)}{w}\right)\right]\Gamma(\psi(w_\beta), w_\beta)\right\} \\ &\times \frac{1}{\Gamma(\underline{w}, \overline{w})^2}\int_{\psi(w_\beta)}^{w_\beta} \frac{\partial \tilde{f}(w)}{\partial y^{R*}(w_\beta)} \frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta} dw.\end{aligned} \tag{78}$$

Finally, by using (69), we can rewrite (68) as

$$\begin{aligned}
\Psi_4 &= \tilde{\Phi}^{R*}(w, y(w), \Gamma(\psi(w), w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\} \cap \left\{y(w) = y^{R*}(w_\beta)\right\} \cap \left\{w = w_\beta\right\}} \\
&\quad - \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{\overline{w} \geq w_\beta\right\} \cap \{y(w) = y^{R*}(w)\} \cap \left\{w = w_\beta\right\}} \\
&= \left[\tilde{\Phi}^{R*}(w, y(w), \Gamma(\psi(w), w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) - \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w}))\right] \\
&\quad \times \mathbb{I}_{\left\{w_\beta < \overline{w}\right\} \cap \{y(w) = y^{R*}(w)\} \cap \left\{w = w_\beta\right\}} \\
&= \left\{\left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \frac{\Gamma\left(\psi(w), w\right) - \tilde{f}(w)}{\Gamma\left(\underline{w}, \overline{w}\right)}\right\} \mathbb{I}_{\left\{w_\beta < \overline{w}\right\} \cap \{y(w) = y^{R*}(w)\} \cap \left\{w = w_\beta\right\}}.
\end{aligned} \tag{79}$$

Step 3. Suppose $w_\alpha > \underline{w}$ holds. By continuity of income schedule $y^*(\cdot)$, we get from Theorem 3.2 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

$$\varphi(w_\alpha) \equiv \begin{cases} (y^{R*})^{-1}(y^{M*}(w_\alpha)) & \text{if } w_\beta < \overline{w}, \\ w_\beta & \text{if } w_\beta = \overline{w}. \end{cases} \tag{80}$$

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

$$\begin{aligned}
&\Xi(w_\alpha; k) \\
&\equiv \int_{w_\eta}^{w_\alpha} \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta \leq w_\alpha\} \cap \{y(w) = y^{M*}(w)\}} dw \\
&\quad + \int_{w_\alpha}^{k} \tilde{\Phi}^{M*}(w, y(w), \Gamma(w_\alpha, \varphi(w_\alpha)), \Gamma(w_\alpha, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\} \cap \{y(w) = y^{M*}(w_\alpha)\}} dw \\
&\quad + \int_{k}^{\varphi(w_\alpha)} \tilde{\Phi}^{R*}(w, y(w), \Gamma(w_\alpha, \varphi(w_\alpha)), \Gamma(\varphi(w_\alpha), \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\} \cap \{y(w) = y^{M*}(w_\alpha)\}} dw \\
&\quad + \int_{\varphi(w_\alpha)}^{\overline{w}} \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{\overline{w} \geq w_\beta\right\} \cap \{y(w) = y^{R*}(w)\}} dw.
\end{aligned} \tag{81}$$

Using (81), the first-order condition with respect to w_α can be derived as

$$\Lambda_1 + \Lambda_2(k) + \Lambda_3(k) + \Lambda_4 = 0, \tag{82}$$

in which

$$\begin{aligned}
\Lambda_1 &= \left[\tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) - \tilde{\Phi}^{M*}(w, y(w), \Gamma(w, \varphi(w)), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w}))\right] \\
&\quad \times \mathbb{I}_{\{w_\eta \leq w_\alpha\} \cap \{w_\alpha > \underline{w}\} \cap \{y(w) = y^{M*}(w)\} \cap \{w = w_\alpha\}} \\
&= \left\{\left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \frac{\tilde{f}(w) - \Gamma(w, \varphi(w))}{\Gamma\left(\underline{w}, \overline{w}\right)}\right\} \mathbb{I}_{\{w_\alpha \geq w_\eta\} \cap \{y(w) = y^{M*}(w)\} \cap \{w = w_\alpha\}},
\end{aligned} \tag{83}$$

$$\Lambda_2(k) = \int_{w_\alpha}^{k} \left[\Lambda_{21}(w) + \Lambda_{22}(w) + \Lambda_{23}(w) + \Lambda_{24}(w)\right] \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}} dw, \tag{84}$$

with

$$\Lambda_{21}(w) = \left\{ \left[1 - \frac{1}{w} h' \left(\frac{y^{M*}(w_\alpha)}{w} \right) \right] \frac{\Gamma(w_\alpha, \varphi(w_\alpha))}{\Gamma(\underline{w}, \overline{w})} + \left[\frac{1}{w^2} h' \left(\frac{y^{M*}(w_\alpha)}{w} \right) + \frac{y^{M*}(w_\alpha)}{w^3} h'' \left(\frac{y^{M*}(w_\alpha)}{w} \right) \right] \left[1 - \frac{\Gamma(w_\alpha, \overline{w})}{\Gamma(\underline{w}, \overline{w})} \right] \right\} \frac{\mathrm{d}y^{M*}(w_\alpha)}{\mathrm{d}w_\alpha}, \tag{85}$$

$$\Lambda_{22}(w) = \left[y^{M*}(w_\alpha) - h \left(\frac{y^{M*}(w_\alpha)}{w} \right) \right] \frac{1}{\Gamma(\underline{w}, \overline{w})} \times \left\{ \tilde{f}\left(\varphi(w_\alpha)\right) \frac{\mathrm{d}\varphi(w_\alpha)}{\mathrm{d}w_\alpha} - \tilde{f}(w_\alpha) + \left[\int_{w_\alpha}^{\varphi(w_\alpha)} \frac{\partial \tilde{f}(w)}{\partial y^{M*}(w_\alpha)} \frac{\mathrm{d}y^{M*}(w_\alpha)}{\mathrm{d}w_\alpha} dw \right] \right\}, \tag{86}$$

$$\Lambda_{23}(w) = \frac{y^{M*}(w_\alpha)}{w^2} h' \left(\frac{y^{M*}(w_\alpha)}{w} \right) \frac{1}{\Gamma(\underline{w}, \overline{w})} \times \left\{ \tilde{f}(w_\alpha) - \left[\int_{w_\alpha}^{\varphi(w_\alpha)} \frac{\partial \tilde{f}(w)}{\partial y^{M*}(w_\alpha)} \frac{\mathrm{d}y^{M*}(w_\alpha)}{\mathrm{d}w_\alpha} dw \right] \right\}, \tag{87}$$

and

$$\Lambda_{24}(w) = \left\{ \frac{y^{M*}(w_\alpha)}{w^2} h' \left(\frac{y^{M*}(w_\alpha)}{w} \right) \Gamma(w_\alpha, \overline{w}) - \left[y^{M*}(w_\alpha) - h \left(\frac{y^{M*}(w_\alpha)}{w} \right) \right] \Gamma(w_\alpha, \varphi(w_\alpha)) \right\} \times \frac{1}{\Gamma(\underline{w}, \overline{w})^2} \int_{w_\alpha}^{\varphi(w_\alpha)} \frac{\partial \tilde{f}(w)}{\partial y^{M*}(w_\alpha)} \frac{\mathrm{d}y^{M*}(w_\alpha)}{\mathrm{d}w_\alpha} dw; \tag{88}$$

$$\Lambda_3(k) = \int_k^{\varphi(w_\alpha)} \left[\Lambda_{31}(w) + \Lambda_{32}(w) + \Lambda_{33}(w) + \Lambda_{34}(w) \right] \cdot \mathbb{I}_{\left\{ w_\beta < \overline{w} \right\}} dw, \tag{89}$$

with

$$\Lambda_{31}(w) = \left\{ \left[1 - \frac{1}{w} h' \left(\frac{y^{M*}(w_\alpha)}{w} \right) \right] \frac{\Gamma(w_\alpha, \varphi(w_\alpha))}{\Gamma(\underline{w}, \overline{w})} - \left[\frac{1}{w^2} h' \left(\frac{y^{M*}(w_\alpha)}{w} \right) + \frac{y^{M*}(w_\alpha)}{w^3} h'' \left(\frac{y^{M*}(w_\alpha)}{w} \right) \right] \frac{\Gamma(\varphi(w_\alpha), \overline{w})}{\Gamma(\underline{w}, \overline{w})} \right\} \frac{\mathrm{d}y^{M*}(w_\alpha)}{\mathrm{d}w_\alpha}, \tag{90}$$

$$\Lambda_{32}(w) = \left[y^{M*}(w_\alpha) - h \left(\frac{y^{M*}(w_\alpha)}{w} \right) \right] \frac{1}{\Gamma(\underline{w}, \overline{w})} \times \left\{ \tilde{f}\left(\varphi(w_\alpha)\right) \frac{\mathrm{d}\varphi(w_\alpha)}{\mathrm{d}w_\alpha} - \tilde{f}(w_\alpha) + \left[\int_{w_\alpha}^{\varphi(w_\alpha)} \frac{\partial \tilde{f}(w)}{\partial y^{M*}(w_\alpha)} \frac{\mathrm{d}y^{M*}(w_\alpha)}{\mathrm{d}w_\alpha} dw \right] \right\}, \tag{91}$$

$$\Lambda_{33}(w) = \frac{y^{M*}(w_\alpha)}{w^2} h' \left(\frac{y^{M*}(w_\alpha)}{w} \right) \frac{\tilde{f}\left(\varphi(w_\alpha)\right)}{\Gamma(\underline{w}, \overline{w})} \frac{\mathrm{d}\varphi(w_\alpha)}{\mathrm{d}w_\alpha}, \tag{92}$$

and

$$
\begin{aligned}
&\Lambda_{34}(w) \\
= &\left\{ \frac{y^{M*}(w_\alpha)}{w^2} h'\left(\frac{y^{M*}(w_\alpha)}{w}\right) \Gamma(\varphi(w_\alpha), \overline{w}) - \left[y^{M*}(w_\alpha) - h\left(\frac{y^{M*}(w_\alpha)}{w}\right)\right] \Gamma(w_\alpha, \varphi(w_\alpha)) \right\} \\
&\times \frac{1}{\Gamma(\underline{w}, \overline{w})^2} \int_{w_\alpha}^{\varphi(w_\alpha)} \frac{\partial \tilde{f}(w)}{\partial y^{M*}(w_\alpha)} \frac{\mathrm{d}y^{M*}(w_\alpha)}{\mathrm{d}w_\alpha} dw;
\end{aligned}
\tag{93}
$$

and finally

$$
\begin{aligned}
\Lambda_4 =& \frac{\mathrm{d}\varphi(w_\alpha)}{\mathrm{d}w_\alpha} \times \\
&[\tilde{\Phi}^{R*}(w, y(w), \Gamma(w_\alpha, w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\} \cap \{y(w) = y^{M*}(w_\alpha)\} \cap \{w = \varphi(w_\alpha)\}} \\
&- \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{\overline{w} \geq w_\beta\right\} \cap \{y(w) = y^{M*}(w_\alpha)\} \cap \{w = \varphi(w_\alpha)\}}] \\
=& \left[\tilde{\Phi}^{R*}(w, y(w), \Gamma(w_\alpha, w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) - \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w}))\right] \\
&\times \mathbb{I}_{\left\{w_\beta < \overline{w}\right\} \cap \{y(w) = y^{M*}(w_\alpha)\} \cap \{w = \varphi(w_\alpha)\}} \cdot \frac{\mathrm{d}\varphi(w_\alpha)}{\mathrm{d}w_\alpha} \\
=& \left\{ \left[y^{M*}(w_\alpha) - h\left(\frac{y^{M*}(w_\alpha)}{\varphi(w_\alpha)}\right)\right] \frac{\Gamma\left(w_\alpha, \varphi(w_\alpha)\right) - \tilde{f}(\varphi(w_\alpha))}{\Gamma\left(\underline{w}, \overline{w}\right)} \right\} \frac{\mathrm{d}\varphi(w_\alpha)}{\mathrm{d}w_\alpha} \mathbb{I}_{\left\{w_\beta < \overline{w}\right\}}.
\end{aligned}
\tag{94}
$$

Step 4. We first consider the case with $w_\beta < \overline{w}$. It follows from (63) and (64) that

$$
\frac{\partial^2 \Xi(w_\beta; k)}{\partial w_\beta \partial k} = \frac{\mathrm{d}\Psi_2(k)}{\mathrm{d}k} + \frac{\mathrm{d}\Psi_3(k)}{\mathrm{d}k}. \tag{95}
$$

By using equations (71)-(77), (95) can be explicitly expressed as

$$
\begin{aligned}
&\frac{\mathrm{d}\Psi_2(k)}{\mathrm{d}k} + \frac{\mathrm{d}\Psi_3(k)}{\mathrm{d}k} \\
= &\left[\Psi_{21}(k) + \Psi_{22}(k) + \Psi_{23}(k) + \Psi_{24}(k)\right] - \left[\Psi_{31}(k) + \Psi_{32}(k) + \Psi_{33}(k) + \Psi_{34}(k)\right] \\
= &\left[\Psi_{21}(k) - \Psi_{31}(k)\right] + \underbrace{\left[\Psi_{22}(k) - \Psi_{32}(k)\right]}_{=0} + \left[\Psi_{23}(k) - \Psi_{33}(k)\right] + \left[\Psi_{24}(k) - \Psi_{34}(k)\right] \\
= &\left[\frac{1}{k^2} h'\left(\frac{y^{R*}(w_\beta)}{k}\right) + \frac{y^{R*}(w_\beta)}{k^3} h''\left(\frac{y^{R*}(w_\beta)}{k}\right)\right] \left[1 - \frac{\Gamma(\psi(w_\beta), \overline{w})}{\Gamma(\underline{w}, \overline{w})} + \frac{\Gamma(w_\beta, \overline{w})}{\Gamma(\underline{w}, \overline{w})}\right] \frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta} \\
&+ \frac{y^{R*}(w_\beta)}{k^2} h'\left(\frac{y^{R*}(w_\beta)}{k}\right) \frac{1}{\Gamma(\underline{w}, \overline{w})} \left[-\frac{\mathrm{d}\Gamma(\psi(w_\beta), w_\beta)}{\mathrm{d}w_\beta}\right] \\
&+ \frac{y^{R*}(w_\beta)}{k^2} h'\left(\frac{y^{R*}(w_\beta)}{k}\right) \frac{\Gamma(\psi(w_\beta), \overline{w}) - \Gamma(w_\beta, \overline{w})}{\Gamma(\underline{w}, \overline{w})^2} \int_{\psi(w_\beta)}^{w_\beta} \frac{\partial \tilde{f}(w)}{\partial y^{R*}(w_\beta)} \frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta} dw.
\end{aligned}
\tag{96}
$$

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

$$
\frac{\partial^2 \Xi(w_\beta; k)}{\partial w_\beta \partial k} = \frac{\mathrm{d}\Psi_2(k)}{\mathrm{d}k} + \frac{\mathrm{d}\Psi_3(k)}{\mathrm{d}k} > 0 \tag{97}
$$

whenever

$$\frac{\mathrm{d}\Gamma(\psi(w_\beta), w_\beta)}{\mathrm{d}w_\beta} \leq 0,$$

as desired. In particular, $\mathrm{d}\psi(w_\beta)/\mathrm{d}w_\beta > 0$ based on the construction of $\psi(\cdot)$ as well as the monotonicity of the income schedule. In the case of (97), $\Xi(w_\beta; k)$ is a supermodular function, and an application of Topkis Theorem (see Theorem 6.1 of Topkis (1978)) implies that $w_\beta(k)$ is nondecreasing in k.

Step 5. We now consider the case with $w_\alpha > \underline{w}$. It follows from (81) and (82) that

$$\frac{\partial^2 \Xi(w_\alpha; k)}{\partial w_\alpha \partial k} = \frac{\mathrm{d}\Lambda_2(k)}{\mathrm{d}k} + \frac{\mathrm{d}\Lambda_3(k)}{\mathrm{d}k}. \tag{98}$$

By using equations (84)-(92), (98) can be explicitly expressed as

$$\begin{aligned}
&\frac{\mathrm{d}\Lambda_2(k)}{\mathrm{d}k} + \frac{\mathrm{d}\Lambda_3(k)}{\mathrm{d}k} \\
=\ & [\Lambda_{21}(k) + \Lambda_{22}(k) + \Lambda_{23}(k) + \Lambda_{24}(k)] - [\Lambda_{31}(k) + \Lambda_{32}(k) + \Lambda_{33}(k) + \Lambda_{34}(k)] \\
=\ & [\Lambda_{21}(k) - \Lambda_{31}(k)] + \underbrace{[\Lambda_{22}(k) - \Lambda_{32}(k)]}_{=0} + [\Lambda_{23}(k) - \Lambda_{33}(k)] + [\Lambda_{24}(k) - \Lambda_{34}(k)] \\
=\ & \left[\frac{1}{k^2} h'\left(\frac{y^{M*}(w_\alpha)}{k}\right) + \frac{y^{M*}(w_\alpha)}{k^3} h''\left(\frac{y^{M*}(w_\alpha)}{k}\right)\right]\left[1 - \frac{\Gamma(w_\alpha, \overline{w})}{\Gamma(\underline{w}, \overline{w})} + \frac{\Gamma(\varphi(w_\alpha), \overline{w})}{\Gamma(\underline{w}, \overline{w})}\right] \frac{\mathrm{d}y^{M*}(w_\alpha)}{\mathrm{d}w_\alpha} \\
& + \frac{y^{M*}(w_\alpha)}{k^2} h'\left(\frac{y^{M*}(w_\alpha)}{k}\right) \frac{1}{\Gamma(\underline{w}, \overline{w})}\left[-\frac{\mathrm{d}\Gamma(w_\alpha, \varphi(w_\alpha))}{\mathrm{d}w_\alpha}\right] \\
& + \frac{y^{M*}(w_\alpha)}{k^2} h'\left(\frac{y^{M*}(w_\alpha)}{k}\right) \frac{\Gamma(w_\alpha, \overline{w}) - \Gamma(\varphi(w_\alpha), \overline{w})}{\Gamma(\underline{w}, \overline{w})^2} \int_{w_\alpha}^{\varphi(w_\alpha)} \frac{\partial \tilde{f}(w)}{\partial y^{M*}(w_\alpha)} \frac{\mathrm{d}y^{M*}(w_\alpha)}{\mathrm{d}w_\alpha} dw.
\end{aligned} \tag{99}$$

Since $\mathrm{d}y^{M*}(w_\alpha)/\mathrm{d}w_\alpha > 0$ by assumption, we have by (99) that

$$\frac{\partial^2 \Xi(w_\alpha; k)}{\partial w_\alpha \partial k} = \frac{\mathrm{d}\Lambda_2(k)}{\mathrm{d}k} + \frac{\mathrm{d}\Lambda_3(k)}{\mathrm{d}k} > 0 \tag{100}$$

whenever

$$\frac{\mathrm{d}\Gamma(w_\alpha, \varphi(w_\alpha))}{\mathrm{d}w_\alpha} \leq 0,$$

as desired. In the case of (100), $\Xi(w_\alpha; k)$ is a supermodular function, and an application of Topkis Theorem (see Theorem 6.1 of Topkis (1978)) again implies that $w_\alpha(k)$ is nondecreasing in k. ■

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

Step 1. Based on Proposition 3.3, we consider first the case with a downward discontinuity at the type of the proposer, namely the income schedules with the SOIC condition being violated. Let's consider two alternative proposers of types k_1 and k_2, for $k_1 < k_2$. Since the income schedules they proposed 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,

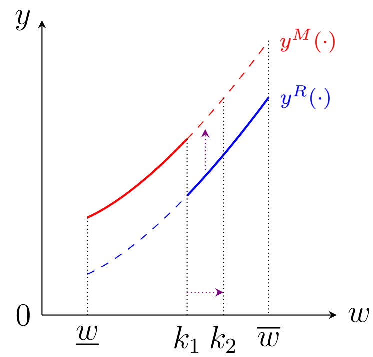

Figure 7: The type of proposer changes from k_1 to k_2 when SOIC is violated.

if the proposer changes from type k_1 to type k_2, all workers of types belonging 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 7). 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 proposed for her. Formally,

$$U(w, w) \ \geq \ U(w, k) \quad \text{for } \forall w, k \in [\underline{w}, \overline{w}]. \tag{101}$$

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 (101), we have

$$\begin{aligned} U(w, k_1) \ &= \ U(k_1, k_1) - \int_w^{k_1} h'\left(\frac{y(t, k_1)}{t}\right) \frac{y(t, k_1)}{t^2} dt \\ &\geq \ U(k_1, k_2) - \int_w^{k_1} h'\left(\frac{y(t, k_1)}{t}\right) \frac{y(t, k_1)}{t^2} dt. \end{aligned} \tag{102}$$

Similarly, we can get by (6) that

$$U(w, k_2) \ = \ U(k_1, k_2) - \int_w^{k_1} h'\left(\frac{y(t, k_2)}{t}\right) \frac{y(t, k_2)}{t^2} dt. \tag{103}$$

Solving for $U(k_1, k_2)$ from (103) and inserting it into (102) produces

$$U(w, k_1) - U(w, k_2) \ \geq \ \int_w^{k_1} \left[h'\left(\frac{y(t, k_2)}{t}\right) \frac{y(t, k_2)}{t^2} - h'\left(\frac{y(t, k_1)}{t}\right) \frac{y(t, k_1)}{t^2}\right] dt. \tag{104}$$

Since h is strictly increasing and convex, we hence have by using (104) that $U(w, k_1) \geq U(w, k_2)$, which combined with (101) reveals that

$$U(w, w) \ \geq \ U(w, k_1) \ \geq \ U(w, k_2), \quad \forall w < k_1 < k_2. \tag{105}$$

Step 3. Second, for the case with $w > k_1 > k_2$, we also get by using (6) and (101) that

$$\begin{aligned} U(w, k_1) &= U(k_1, k_1) + \int_{k_1}^{w} h'\left(\frac{y(t, k_1)}{t}\right) \frac{y(t, k_1)}{t^2} dt \\ &\geq U(k_1, k_2) + \int_{k_1}^{w} h'\left(\frac{y(t, k_1)}{t}\right) \frac{y(t, k_1)}{t^2} dt. \end{aligned} \tag{106}$$

Similarly,

$$U(w, k_2) = U(k_1, k_2) + \int_{k_1}^{w} h'\left(\frac{y(t, k_2)}{t}\right) \frac{y(t, k_2)}{t^2} dt. \tag{107}$$

Making use of (106) and (107) gives rise to

$$U(w, k_1) - U(w, k_2) \geq \int_{k_1}^{w} \left[h'\left(\frac{y(t, k_1)}{t}\right) \frac{y(t, k_1)}{t^2} - h'\left(\frac{y(t, k_2)}{t}\right) \frac{y(t, k_2)}{t^2} \right] dt. \tag{108}$$

Applying the same reasoning used in step 2 to (108), we arrive at

$$U(w, w) \geq U(w, k_1) \geq U(w, k_2), \quad \forall w > k_1 > k_2. \tag{109}$$

Accordingly, (105) combined with (109) reveals that the preference of any given type of worker is indeed (weakly) single-peaked on the set of types. Applying the Black's Median Voter Theorem (see Black, 1948), the desired assertion hence follows. After ironing the downward discontinuity by building a bridge, we establish the income schedule that meets the SOIC condition in Theorem 3.2. Using Lemma 3.5, it is immediate that the reasoning used above applies as well, and hence the desired assertion holds true for the complete solution of the tax design problem.

Step 4. We now move to the case with the SOIC condition being met under the first-order approach. As shown by Figure 8, if the proposer changes from type k_1 to type k_2, all workers of types belonging to set $[k_1, k_2]$ are strictly worse off in terms of pre-tax income while all other workers with the remaining types are neutral to this change. We hence have that $y(w, k_1) \geq y(w, k_2)$ for $\forall w, k_1, k_2 \in [\underline{w}, \overline{w}]$, and $y(w, k_1) > y(w, k_2)$ for $\forall w \in [k_1, k_2]$. In addition, condition (101) still applies here. It is easy to show that we still have (104) for $w < k_1 < k_2$ and (108) for $w > k_1 > k_2$. Since the right hand side of inequalities (104) and (108) is equal to zero, the assertion established above holds true in the current case. ■

Proof of Lemma 5.1. Using (25) and (27), we obtain

$$\hat{\tau}^M(w) - \tau^M(w) = -\left[\frac{1}{w^2} h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^3} h''\left(\frac{y(w)}{w}\right)\right] \left[\frac{F(w)}{f(w)} - \frac{\Gamma(\underline{w}, w)}{\tilde{f}(w)}\right],$$

from which assertion (i) immediately follows.

Using (26), (27) and (17), we obtain

$$\begin{aligned} \hat{\tau}^R(w) - \tau^R(w) = &\frac{\partial U(w)}{\partial y(w)} \left[\frac{1 - F(w)}{f(w)} - \frac{\Gamma(\underline{w}, \overline{w}) - \Gamma(\underline{w}, w)}{\tilde{f}(w)}\right] \\ &+ \frac{\partial U(w)}{\partial y(w)} \frac{\tilde{\theta}(w)}{c(w)} \left[T^R(y(w)) + U(w) - U'(w) - U(\underline{w})\right], \end{aligned}$$

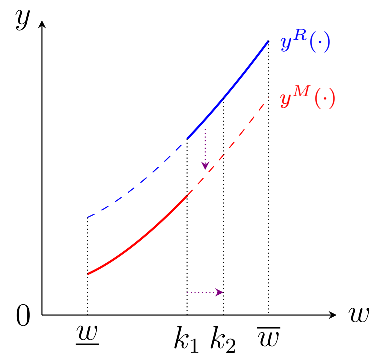

Figure 8: The type of proposer changes from k_1 to k_2 when SOIC is met.

from which assertion (ii) follows.

Applying (27) and (26) reveals that

$$\begin{aligned}&\hat{\tau}^M(w) - \tau^R(w)\\ &= -\frac{\partial U(w)}{\partial y(w)}\left\{\frac{\Gamma(\underline{w},\overline{w}) - \Gamma(\underline{w},w)}{\tilde{f}(w)} + \frac{F(w)}{f(w)} - \frac{\tilde{\theta}(w)}{c(w)}\left[T^R(y(w)) + U(w) - U'(w) - U(\underline{w})\right]\right\},\end{aligned}$$

by which we can establish assertion (iii).

Applying (25) and (27) reveals that

$$\hat{\tau}^R(w) - \tau^M(w) = \left[\frac{1}{w^2}h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^3}h''\left(\frac{y(w)}{w}\right)\right]\cdot\left[\frac{\Gamma(\underline{w},w)}{\tilde{f}(w)} + \frac{1-F(w)}{f(w)}\right],$$

by which assertion (iv) is immediate. ■

Appendix B: The Relative Magnitude of Ex Ante and Ex Post Median Skill Levels

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

$$\Gamma(\underline{w},\overline{w}) = \int_{\underline{w}}^{\overline{w}} \tilde{f}(w)dw \;=\; \int_{\underline{w}}^{w_m} \tilde{f}(w)dw + \int_{w_m}^{\overline{w}} \tilde{f}(w)dw. \tag{110}$$

By using (8), we get the right-hand terms of (110) as

$$\int_{\underline{w}}^{w_m} \tilde{f}(w)dw = \frac{1}{2} \;+\; \underbrace{L^I([\underline{w},w_m]) - L^O([\underline{w},w_m])}_{L^{NI}([\underline{w},w_m]) \;=\; \text{net labor inflow}} \tag{111}$$

and

$$\int_{w_m}^{\overline{w}} \tilde{f}(w)dw = \frac{1}{2} \;+\; \underbrace{L^I([w_m,\overline{w}]) - L^O([w_m,\overline{w}])}_{L^{NI}([w_m,\overline{w}]) \;=\; \text{net labor inflow}}, \tag{112}$$

in which the measures of labor inflows are defined as

$$L^I([\underline{w}, w_m]) \equiv \int_{\{w\in[\underline{w},w_m]|\Delta(w)\geq 0\}} G_-(\Delta(w)|w)f_-(w)n_-dw,$$
$$L^I([w_m, \overline{w}]) \equiv \int_{\{w\in[w_m,\overline{w}]|\Delta(w)\geq 0\}} G_-(\Delta(w)|w)f_-(w)n_-dw, \tag{113}$$

and the measures of labor outflows are defined as

$$L^O([\underline{w}, w_m]) \equiv \int_{\{w\in[\underline{w},w_m]|\Delta(w)\leq 0\}} G(-\Delta(w)|w)f(w)dw,$$
$$L^O([w_m, \overline{w}]) \equiv \int_{\{w\in[w_m,\overline{w}]|\Delta(w)\leq 0\}} G(-\Delta(w)|w)f(w)dw. \tag{114}$$

By using (110)-(114), 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 6.1 *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 6.2 *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 6.3 *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 6.1-6.3 consider three possible cases corresponding to three possible ex post measures of workers of all skill levels.

Proposition 6.1 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 6.2-6.3 in the similar way.