ESSAYS IN NONLINEAR LABOR INCOME TAXATION WITH TAX-DRIVEN MIGRATIONS

A Dissertation

by

DARONG DAI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Guoqiang Tian
Committee Members,	Dennis W. Jansen
	Yuzhe Zhang
	Quan Li
Head of Department,	Timothy Gronberg

May 2018

Major Subject: Economics

ABSTRACT

This dissertation includes two essays in nonlinear labor income taxation with tax-driven migrations. In the first essay, we study optimal nonlinear income taxation in an open economy with migration possibilities and social comparisons. Recent evidence suggests that globalization has not just reduced the barriers to international labor mobility but also induced more cross-country comparisons. In an open economy with tax-driven migrations and consumption externalities motivated by altruism or jealousy, we derive an optimal tax formula that subsumes existing ones obtained under maximin social objective and additively separable utility, and identify the sign of second-best marginal tax rates for all skill levels. We establish thresholds of the elasticity and level of migration to determine when relativity and inequality are complementary (or substitutive) in shaping the optimal top tax rates. These thresholds are in general different between altruism-type and jealousy-type relativity. Surprisingly, there exist reasonable combinations of relativity, mobility and inequality such that tax competition results in higher equilibrium top tax rates than proposed in autarky. Also, under both Nash and Stackelberg tax competition, we have the following numerical finding by plugging realistic parameter values in our tax formula. If the migration probability of top-income workers is around 50%, then the country facing labor inflow (respectively, outflow) of these types of workers implements around 10% lower (respectively, higher) top tax rates than suggested by the autarky equilibrium which does not allow for migration possibilities.

In the second essay, we study majority voting over selfishly optimal nonlinear income tax schedules proposed by a continuum of workers who can migrate between two competing jurisdictions at the expense of some migration cost. Both skill and migration cost are the private information of each worker. Assuming quasilinear-in-consumption preferences, the tax schedule proposed by the median skill type is the Condorcet winner that redistributes incomes from the rich and poor toward the middle. While it features negative marginal tax rates for low skills, it features positive marginal tax rates for high skills who have elasticities of migration smaller than a threshold. In a comparison with the autarky economy, we establish the skill-dependent threshold of migration elasticity for all types of workers. If their migration elasticities are higher than their respective threshold, then migrations induce lower marginal tax rates than does autarky; otherwise migrations induce higher marginal tax rates for the jurisdiction facing net labor inflow in low skills while net labor outflow in high skills. Counterfactual simulations using empirical parameter estimates show that eliminating migrations in the U.S. would generate top tax rates over 20% higher than the 42.5% that was actually implemented.

DEDICATION

To my mother, my father, and my wife.

ACKNOWLEDGMENTS

I would like to express my sincere appreciation and gratitude to my advisor, Dr. Guoqiang Tian, for his academic guidance and enthusiastic encouragement throughout this research and during these past five years. I would like to thank the members of my dissertation committee, Dr. Dennis Jansen, Dr. Yuzhe Zhang and Dr. Quan Li, not only for their time and extreme patience, but also for their intellectual contributions to my development as an economist. Finally, helpful comments and suggestions from Yonghong An, Pedro Bento, Laurent Bouton, Klaus Desmet, Ian Dew-Becker, Andrew T. Foerster, Li Gan, Amy Glass, Andrew Glover, Silvana Krasteva, Jason Lindo, Liqun Liu, Fernando Luco, Jonathan Meer, Tatevik Sekhposyan, Sang-Chul Suh, Manuelita Ureta, Kei-Mu Yi, Sarah Zubairy and participants of Macro Research Group and Macro Student Conference (October 2017) at Texas A&M, and of the Midwest Economic Theory Conference at SMU (November 2017) are gratefully acknowledged.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Guoqiang Tian (chair), Professor Dennis W. Jansen, Professor Yuzhe Zhang of the Department of Economics, and Professor Quan Li of the Department of Political Science. Portions of this dissertation were conducted with Wenzheng Gao and Guoqiang Tian.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by Department of Economics at Texas A&M University.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

1. INTRODUCTION

This dissertation studies redistributive labor income taxation with tax-driven migrations. We consider a continuum of workers who differ in both migration costs and labor productivities that are assumed to be their private information. We hence follow the mechanism design approach and aim to design redistributive-taxation policies satisfying the participation constraint, incentive-compatibility constraints and the balanced government budget constraint.

In the first essay[1], we study redistributive taxation policies in an open economy with taking into account both migrations and cross-country social comparisons. This study is motivated by recent evidence that globalization has not just reduced the barriers to international labor mobility but also induced more cross-country social comparisons. For example, [3], [4] and [5] estimate large migration elasticities with respect to income tax rate for highly skilled workers. The mobility of taxpayers thus induces tax competition between countries. Using survey-data for countries in Western Europe, Becchetti et al. [6] find that the contribution of cross-country comparisons to well-being increased from the early 1970s to 2002. In fact, since Veblen [7], economists (see, [8], [9], [10], [11], [12] and [13]) recognize that the well-being of economic agents depends on relative consumption in addition to absolute consumption, no matter they are motivated by altruism or jealousy. So, taxing consumption externalities seems to be welfare enhancing as any other Pigouvian tax. We focus on the tool of labor income tax and design optimal nonlinear tax schedules to deal with income inequality and consumption externality, including positive externality driven by altruism and negative externality driven by jealousy. By setting up a framework taking into account both consumption relativity and labor mobility, we are allowed to design more realistic income tax schedules. In addition, we can analyze how relativity and mobility affect optimal income tax rates under any given degree of inequality. In particular, for top-income workers, we ask how income inequality and consumption relativity *together* determine their tax rates, and how the resulting joint effect changes across different types and magnitudes of cross-country labor

[1]This chapter is based on my unpublished working paper [2].

flows. To our knowledge, this study represents the first attempt to address these interesting issues in a unified framework.

In the second essay, we follow the political-economic approach and study majority voting over selfishly optimal nonlinear income tax schedules proposed by a continuum of workers (see, [14]) who can migrate between two competing jurisdictions at the expense of some migration cost. To analyze how the possibility of geographic mobility affects the design of redistributive taxation, the literature (see, e.g., [15], [16], [17], [18], [1]) that builds on the seminal work of Mirrlees [19] focuses on the normative perspective. Little attention has been paid to the positive perspective. We hence address this question: how would the schedule of redistributive taxation look like when workers can vote both in the ballot box and with their feet? In particular, answering this question allows us to reexamine the conventional wisdom (see [20]) claiming that geographic mobility limits the ability of government to redistribute incomes via a tax-transfer system. To our knowledge, the answer is not yet well established. Indeed, the literature either assumes away asymmetric information (e.g., [21], [22]), restricts attention to flat tax (e.g., [22]) and special connections between skills and migration costs in a two-type setting that rules out countervailing incentives (e.g., [23]), or focuses on probabilistic voting in a representative democracy (e.g., [24]). These simplifications make it possible to obtain sharp predictions, whereas reasonable doubts about the generality and robustness of their predictions may arise.

To achieve our goal, we consider an economy consisting of two jurisdictions. Following [25], [26], [27] and [14], 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 that maximizes the utility of her own type. Then, the pairwise majority rule is used to select the tax schedule that is going to be implemented in equilibrium. By assuming quasilinear-in-consumption preferences, we show that the tax schedule proposed by the median skill type is a Condorcet winner in the majority-rule equilibrium, which provides support for the empirical finding of [28]. In addition, the equilibrium tax schedule exhibits the following characteristics. First, it coincides with the maximax tax schedule for types below the median skill level and coincides with the maximin tax schedule for types above the median skill level. Second, marginal tax rates are negative for low incomes, whereas for higher incomes there is an endogenously determined and skill-dependent threshold of the elasticity of migration such that they are negative only when migration elasticity is higher than this threshold, otherwise they are nonnegative. Third, it creates three potential discontinuities, one at the skill level of the proposer and the other two at the endpoints of skill distribution, in the resulting income schedule. And fourth, by allowing for inter-jurisdictional migrations that endogenize the ex post skill distribution and median skill level, the resulting level of distortion and redistribution deviates from that in an autarky economy without allowing for migration possibilities. Counterfactual simulations using empirical parameter estimates show that eliminating migrations in the U.S. would generate top tax rates over 20% higher than the 42.5% (e.g., [29]) that was actually implemented.

2. RELATIVITY, MOBILITY, AND OPTIMAL NONLINEAR INCOME TAXATION IN AN OPEN ECONOMY

2.1 Introduction

Recent evidence suggests that globalization has not just reduced the barriers to international labor mobility but also induced more cross-country social comparisons.[1] For example, [3], [4] and [5] estimate large migration elasticities with respect to tax rate for highly skilled workers. The mobility of taxpayers thus induces tax competition between countries.[2] Using survey-data for countries in Western Europe, Becchetti et al. [6] find that the contribution of cross-country comparisons to well-being increased from the early 1970s to 2002. In fact, since Veblen [7], economists[3] recognize that the well-being of economic agents depends on relative consumption in addition to absolute consumption, no matter they are motivated by altruism or jealousy. So, taxing consumption externalities seems to be welfare enhancing as any other Pigouvian tax.

In the optimal income taxation literature, economists either focus on how consumption relativity and income inequality together shape the optimal tax schedules under a single government (e.g., [31], [8], [32]) or focus on how labor mobility and income inequality together shape the optimal tax schedules with two competing governments (e.g., [17], [1], [33]). Even though these studies have provided some insightful predictions, the first strand of literature neglects cross-border effects and assumes away the possibility that people may have endogenous outside options and hence their reservation utilities may be endogenously determined in the optimal tax design problem, while the second strand uses the optimality of income tax schedules that is biased and hence misleading for policy suggestions as people do exhibit altruism- and jealousy-type preferences in reality.[4]

In this paper we focus on the tool of labor income tax and design optimal nonlinear tax schedules to deal with income inequality and consumption externality, including positive externality

[1] Piketty [30] even argues that cross-country social comparisons seem to constitute an important part of the motivation behind Thatcher's and Reagan's drastic income tax reductions in the early 1980s.

[2] A recent example is the "Tax Cuts and Jobs Act" signed into law by President Trump.

[3] See, [9], [10], [11], [12] and [13].

[4] See, e.g., [34], [35], and [36] for evidences.

driven by altruism and negative externality driven by jealousy, in an open economy with international labor mobility. By setting up a framework taking into account both consumption relativity and labor mobility, we are allowed to design more realistic income tax schedules. In addition, we can analyze how relativity and mobility affect optimal income tax rates under any given degree of inequality. In particular, for top-income workers, we ask how income inequality and consumption relativity *together* determine their tax rates, and how the resulting joint effect changes across different types and magnitudes of cross-country labor flows. To our knowledge, our paper represents the first attempt to address these interesting issues in a unified framework.

For this purpose, we focus on income tax schedules that competing governments find optimal to implement in two types of non-cooperative equilibrium: Nash and Stackelberg. We start with the Nash solution in which each country takes the strategy of the opponent country as given. Each government fully internalizes consumption externalities affecting workers within its own country, but ignores the externalities affecting the opponent country.

As argued by Aronsson and Johansson-Stenman [37], Nash competition is not necessarily the most realistic one since the ability to commit to pubic policy may differ among countries. We thus analyze a Stackelberg equilibrium where one country acts as the leader with the opponent country acting as the follower. As is canonical, the leader shall recognize the behavioral responses of the follower and take into account the externalities it causes to the follower. This, accordingly, implies that optimal tax schedules in these two types of equilibrium are in general different for the leader country.

In each country, workers differ in both skills and migration costs, the continuous distributions of which are common knowledge while the values of which are assumed to be their private information. We thus follow the mechanism design approach. Taking as given income taxes implemented in both countries, workers make individual decisions along two margins: the allocation of one-unit time between work and leisure on the intensive margin, and the location choice on the extensive margin. To make the analysis more transparent, we restrict attention to the most redistributive social objective, maximin[5], in the spirit of Rawls [39]. As a result, after taking into account individual responses, each government designs incentive-compatible allocations such that the utility of the worst-off is maximized and the public-sector budget constraint is satisfied. In particular, provided the endogenous location choice, workers actually have an endogenous reservation utility that depends on the tax policy of the opponent country. Throughout, taxes can only be conditioned on income and are levied according to the residence principle.

We characterize the best response of each government and obtain a formula determining optimal marginal tax rates. The optimal tax formula obtained by Oswald [40] and Kanbur and Tuomala [8] (K&T hereafter) for a closed economy is augmented by a migration effect that changes the Pigouvian-tax term and Mirrleesian-tax term, leading to a much more comprehensive formula. In addition, as in [1], we derive an optimal tax formula under the useful benchmark called the Tiebout-best, in which workers' skills are assumed to be common knowledge while migration costs remain private information. By eliminating incentive-compatibility constraints, the maximization problem of tax design becomes much simpler. In fact, we explicitly solve for Tiebout-best tax liabilities and Tiebout-best marginal tax rates.

Under jealousy-type relativity, the tax schedule has these characteristics. First, if second-best tax liabilities are no higher than Tiebout-best tax liabilities, then second-best marginal tax rates in both Nash and Stackelberg equilibrium are positive over the entire income range but the endpoints. Second, we identify a group of (sufficient) conditions such that second-best tax liabilities are lower than Tiebout-best tax liabilities for almost all skills. Third, for workers of the lowest skill, second-best marginal tax rate is positive and also is higher than that under the Tiebout-best, implying a downward distortion relative to the Tiebout best. Fourth, for workers of the highest skill, marginal tax rate is the same under the second-best and the Tiebout-best, which may be interpreted as a version of "no distortion at the top". And fifth, for the leader country under Stackelberg competition, marginal tax rates are always higher than those under Nash competition, which however cannot be directly carried over to the case with altruism-type relativity.

[5] As is demonstrated by [38], focusing on the maximin objective significantly simplifies the analytical analysis of the optimal income tax structure.

Under altruism-type relativity, the tax schedule has these characteristics. First, for workers of the lowest skill, second-best marginal tax rate is higher than Tiebout-best tax rate, implying a downward distortion relative to the Tiebout best. Second, for workers of the highest skill, marginal tax rate is the same under the second-best and the Tiebout-best, and is strictly negative if the ex ante skill distribution is bounded above. And third, we identify a group of conditions such that: (a) the Mirrleesian-tax term in the tax formula is positive over the entire income range but the endpoints; (b) for workers of the highest skill, second-best tax liability is lower than Tiebout-best tax liability, with the difference representing their information rent; and (c) for workers of the lowest skill, second-best tax liability is positive with a lower bound.

To analyze how relativity and inequality together shape the marginal tax rate imposed on top-income workers, we obtain a closed-form formula of the optimal asymptotic marginal tax rate. We show that the *elasticity* and *level* of migration are two key variables in determining whether or not relativity and inequality play a complementary role in shaping the optimal top tax rate. More importantly, we have identified relevant thresholds of these two variables under both types of consumption relativity, enabling us to make sharp predictions regarding the composite effect of relativity and inequality placed on equilibrium top tax rates.

By using realistic parameter values from empirical studies, we simulate these tax rates in both types of equilibrium and compare them to those under the K&T-formula. In both Nash and Stackelberg equilibrium, the country with large labor inflow imposes much lower while the country with large labor outflow imposes much higher tax rates than suggested by K&T. Also, there are reasonable combinations of parameters measuring relativity, inequality and mobility such that tax competition induces higher tax rates than in autarky. This finding yields an important departure from the common prediction regarding the competition effect on equilibrium tax rate. As an implication for open economies, these results reveal that normative public policy recommendations on redistributive income taxes must take between-country tax competition, tax-driven migrations and relative consumption concerns seriously, otherwise workers are likely to face welfare losses or the economy is likely to face efficiency losses.

Our study is related to the literature studying optimal nonlinear income taxation in an open economy, such as [15], [16], [41], [1], [37] and [42]. The major difference between these studies and our paper is that we focus on examining how the interplay of relativity and inequality determines the optimal nonlinear income tax schedule and meanwhile how the joint effect of relativity and inequality is modified by tax-driven migrations, which are ignored by these studies. They, except [37], completely ignore the effect of relative consumption concern placed on the design of Mirrlees income taxes. As numerically illustrated in Section 2.5, relative consumption concern does result in quantitatively significant effects on the optimal marginal tax rates and hence should not be ignored. Though [37] consider both tax competition and relative consumption concerns, cross-border labor mobility is not allowed there, whereas we show that migrations can shape the tax-competition effect and hence equilibrium tax rates in an important way. As such, our study shows the importance of simultaneously taking into account tax-driven migrations and relative consumption concerns in designing optimal nonlinear income taxes and hence extends the literature towards a more realistic tax design.

The remainder of the paper is organized as follows. Section 2.2 sets up the model. Section 2.3 derives the optimal tax formula in Nash equilibrium and establishes some qualitative properties. Section 2.4 derives the optimal tax formula in Stackelberg equilibrium and establishes some qualitative properties. Section 2.5 provides some numerical examples regarding the optimal asymptotic marginal tax rates and compares our results with those calculated using K&T-formula. Section 2.6 concludes. All formal proofs are relegated to Appendix A.

2.2 The Model

We consider an economy consisting of two countries, indexed by $i \in \{A, B\}$. The measure of workers in country i is normalized to 1, while that of the opponent country $-i$ is denoted by n_{-i}, for $0 < n_{-i} \leq 1$. Each worker is characterized by three characteristics: her native country $i \in \{A, B\}$, her skill $w \in [\underline{w}, \overline{w}]$ with $0 < \underline{w} < \overline{w} \leq \infty$, and the migration cost $m \in \mathbb{R}^+$ she supports if she decides to live abroad. If a worker faces an infinitely large migration cost, then she is immobile. Following [1], we do not make any restriction on the correlation between skills and migration costs.

The skill density function in country i, $f_i(w) = F_i'(w) > 0$, is assumed to be differentiable for all $w \in [\underline{w}, \overline{w}]$ and is single-peaked, with a mode at w_m. For each skill w, $g_i(m|w)$ denotes the conditional density of the migration cost and $G_i(m|w) = \int_0^m g_i(x|w)dx$ the conditional cumulative distribution function. The initial joint density of (m, w) is thus $g_i(m|w)f_i(w)$ while $G_i(m|w)f_i(w)$ is the mass of workers of skill w with migration costs lower than m.

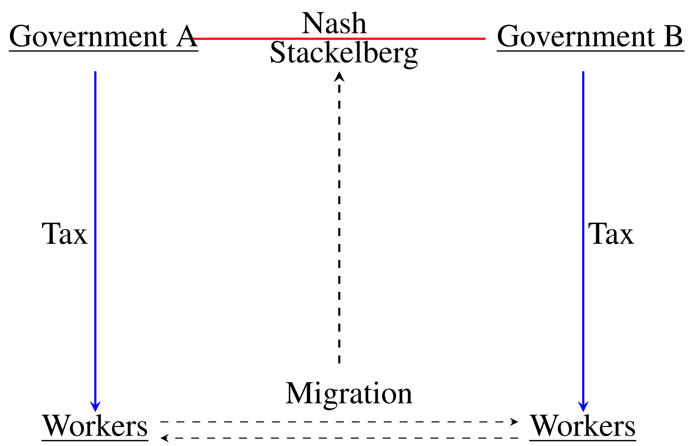

Figure 2.1: Agents and Relationships

Following [19], governments do not observe workers' types (w, m) and can only condition transfers on earnings y via an income tax function, $T_i(\cdot)$, for $i = A, B$. By assumption, taxes are levied according to the residence principle. In an open economy with international labor mobility, migration threat actually induces tax competition between these two governments, and we consider both Nash and Stackelberg competition (see Figure 2.1).

2.2.1 Individual Choices

Assume that all workers have the same additively separable utility function. So, for a worker of type (w, m) in country i:

$$u(c_i(w), l_i(w); \mu_i, \mu_{-i}, m) = v(c_i(w)) - h(l_i(w)) + \psi(\mu_i, \mu_{-i}) - \mathbb{I} \cdot m, \tag{2.1}$$

where c_i is consumption, l_i is labor (and $1 - l_i$ is leisure), $\mathbb{I}$ is equal to 1 if she decides to migrate and to 0 otherwise, μ_i is a domestic comparison consumption level, and μ_{-i} is a cross-country comparison consumption level, with $v' > 0 \geq v''$, $h' > 0$ and $h'' > 0$. Following common practice[6], comparison consumption levels are constructed as follows:

$$\mu_i = \int_{\underline{w}}^{\overline{w}} c_i(w) f_i(w) dw, \tag{2.2}$$

for $i \in \{A, B\}$. For later use, we give the following two assumptions.

Assumption 2.2.1 (Bounded Jealousy)**.** *For* $\psi_i(\mu_i, \mu_{-i}) \equiv \partial\psi/\partial\mu_i < 0$*,* $\psi_{-i}(\mu_i, \mu_{-i}) \equiv \partial\psi/\partial\mu_{-i} < 0$*,* $i \in \{A, B\}$ *and* $w \in [\underline{w}, \overline{w}]$*, we have* $\max\{|\psi_i(\mu_i, \mu_{-i})|, |\psi_{-i}(\mu_i, \mu_{-i})|\} < v'(c_i(w))$.

Assumption 2.2.2 (Bounded Altruism)**.** *For* $\psi_i(\mu_i, \mu_{-i}) > 0$*,* $\psi_{-i}(\mu_i, \mu_{-i}) > 0$*,* $i \in \{A, B\}$ *and* $w \in [\underline{w}, \overline{w}]$*, we have* $\max\{\psi_i(\mu_i, \mu_{-i}), \psi_{-i}(\mu_i, \mu_{-i})\} < v'(c_i(w))$.

Assumptions 2.2.1-2.2.2 state that the utility contribution of relative consumption is strictly smaller than that of absolute consumption, no matter the worker exhibits jealousy- or altruism-type relative consumption. These assumptions are consistent with general intuition as well as real data (see [13]).

The worker obtains her income from wages, with income denoted by $y_i \equiv wl_i(w) \geq 0$. Her budget constraint is thus:

$$c_i(w) = y_i(w) - T_i(y_i(w)). \tag{2.3}$$

Each worker is assumed to be small relative to the whole economy, and hence she takes μ_i and μ_{-i} as exogenously given. If she stays in country i, she maximizes (2.1) subject to $\mathbb{I} = 0$ and (2.3), yielding the first-order condition:

$$\frac{h'(l_i(w))}{wv'(c_i(w))} = 1 - T_i'(y_i(w)). \tag{2.4}$$

We denote by $U_i(w)$ her indirect utility.

[6]See, e.g., [40], [8] and [37].

We now proceed to her migration decision. As is obvious, migration occurs if and only if $m < U_{-i}(w) - U_i(w)$. As in [1], after combining the migration decisions made by workers born in both countries, the mass of residents of skill w in country i can be written as:

$$\phi_i(\Delta_i(w); w) \equiv \begin{cases} f_i(w) + G_{-i}(\Delta_i(w)|w) f_{-i}(w) n_{-i} & \text{for } \Delta_i(w) \geq 0, \\ (1 - G_i(-\Delta_i(w)|w)) f_i(w) & \text{for } \Delta_i(w) \leq 0. \end{cases} \tag{2.5}$$

with $\Delta_i(w) \equiv U_i(w) - U_{-i}(w)$. To ensure that $\phi_i(\cdot; w)$ is differentiable, we impose the technical restriction that $g_i(0|w) f_i(w) = g_{-i}(0|w) f_{-i}(w) n_{-i}$, which is verified when the two countries are symmetric or when there is a fixed cost of migration, namely $g_i(0|w) = g_{-i}(0|w) = 0$. We can then define the semi-elasticity of migration and the elasticity of migration, respectively, as:

$$\eta_i(\Delta_i(w); w) \equiv \frac{\partial \phi_i(\Delta_i(w); w)}{\partial \Delta_i} \frac{1}{\phi_i(\Delta_i(w); w)} \tag{2.6}$$

and

$$\theta_i(\Delta_i(w); w) \equiv c_i(w) \eta_i(\Delta_i(w); w). \tag{2.7}$$

For later use, and also to save on notations, we let $\tilde{f}_i(w) \equiv \phi_i(\Delta_i(w); w)$, $\tilde{\eta}_i(w) \equiv \eta_i(\Delta_i(w); w)$ and $\tilde{\theta}_i(w) \equiv \theta_i(\Delta_i(w); w)$.

2.2.2 Governments

In country $i \in \{A, B\}$, a benevolent government designs the tax system to maximize the welfare of the worst-off workers. By using (2.1) and (2.4), it is easy to show that $U_i(\underline{w}) = \min\{U_i(w) : w \in [\underline{w}, \overline{w}]\}$. That is, the worst-off are exactly those workers with wage rate $\underline{w}$ at the bottom of the skill distribution.

We choose maximin as the social objective due to the following considerations. First, many jobs of the workers of the lowest skills are at the bottom of global value chain and characterized as low-paid, insecure and dangerous (see [43]). Second, they have the lowest migration (or foot-voting) ability, as migration rates increase in skill (see [44]). Third, especially for those in developed countries, the worst-off may be even worse in an open economy because they may lose jobs in the global competition with those workers of the lowest skills in developing countries. And fourth, as a normative criterion, maximin is a crucial principle in achieving the social justice suggested by Rawls [39].

As is canonical, each government faces two sorts of constraints. The first is the fiscal budget constraint:

$$\int_{\underline{w}}^{\overline{w}} T_i(y_i(w))\phi_i(U_i(w) - U_{-i}(w); w)dw \geq R, \tag{2.8}$$

where $R \geq 0$ is an exogenous revenue requirement. As $v_c(\cdot) > 0$, (2.8) must be binding. In particular, here the participation constraint has been incorporated into the fiscal budget constraint (2.8) through the ex post skill density ϕ_i. The second is the set of incentive-compatibility constraints:

$$v(c_i(w)) - h(y_i(w)/w) \geq v(c_i(w')) - h(y_i(w')/w) \;\; \forall w, w' \in [\underline{w}, \overline{w}]. \tag{2.9}$$

The necessary conditions for (2.9) to be satisfied are:

$$\dot{U}_i(w) = h'(l_i(w))\frac{l_i(w)}{w} \;\; \forall w \in [\underline{w}, \overline{w}], \tag{2.10}$$

which gives the first-order incentive compatibility (FOIC) conditions. Sufficiency is guaranteed by the second-order incentive compatibility (SOIC) conditions, $\dot{y}_i(w) \geq 0$. If $\dot{y}_i(w) > 0$, then the first-order approach is appropriate.

As a result, optimal tax design is equivalent to solve the following maximization problem:

$$\max_{\{U_i(w), l_i(w), \mu_i\}} U_i(\underline{w})$$

subject to (2.2), (2.8), (2.10) and $\dot{y}_i(w) \geq 0$.

2.3 Nash Equilibrium

2.3.1 Optimal Tax Formula

We state the first major result in the following theorem.

Theorem 2.3.1. *In a Nash equilibrium with $\dot{y}_i(w) > 0$, the second-best marginal tax rates verify:*

$$\frac{T_i'(y_i(w))}{1 - T_i'(y_i(w))} = \overbrace{\frac{\gamma_i}{\lambda_i}\frac{f_i(w)}{\tilde{f}_i(w)}}^{\textit{Pigouvian-type tax}} + \overbrace{\mathcal{A}_i(w)\mathcal{B}_i(w)\mathcal{C}_i(w)}^{\textit{Mirrleesian-type tax}} \tag{2.11}$$

where: $\mathcal{A}_i(w) \equiv 1 + [l_i(w)h''(l_i)/h'(l_i)]$, $\mathcal{B}_i(w) \equiv \left[\tilde{F}_i(\overline{w}) - \tilde{F}_i(w)\right]/w\tilde{f}_i(w)$,

$$\mathcal{C}_i(w) \equiv \frac{v'(c_i(w))\int_w^{\overline{w}} \left\{\frac{1}{v'(c_i(t))}\left[1 + \frac{\gamma_i}{\lambda_i}\frac{f_i(t)}{\tilde{f}_i(t)}\right] - T_i(y_i(t))\tilde{\eta}_i(t)\right\}\tilde{f}_i(t)dt}{\tilde{F}_i(\overline{w}) - \tilde{F}_i(w)} \tag{2.12}$$

and

$$\frac{\gamma_i}{\lambda_i} = \frac{-\int_{\underline{w}}^{\overline{w}} \frac{\psi_i(\mu_i,\mu_{-i})}{v'(c_i(w))}\tilde{f}_i(w)dw}{1 + \int_{\underline{w}}^{\overline{w}} \frac{\psi_i(\mu_i,\mu_{-i})}{v'(c_i(w))}f_i(w)dw} \tag{2.13}$$

with $\tilde{F}_i(w) \equiv \int_{\underline{w}}^{w} \tilde{f}_i(t)dt$ denoting the ex post skill distribution in country $i \in \{A, B\}$. Moreover, if $T_i'(y_i(w))$ is non-increasing in w, then the SOIC conditions are not binding, namely $\dot{y}_i(w) > 0$ holds.

Proof. See Appendix A. □

Our optimal tax formula (2.11) differs from the classic one derived by Diamond [45] and Saez [46] in three ways: (i) the ex post density $\tilde{f}_i(\cdot)$ of taxpayers replaces the ex ante density $f_i(\cdot)$, (ii) tax liability $T_i(y_i(\cdot))$ enters term $\mathcal{C}_i(w)$ as a tax level effect, and (iii) there is a Pigouvian tax used to correct consumption externalities. Also, (i) and (ii) constitute new features relative to K&T. As is clear soon, these differences lead to qualitative and quantitative analyses much more challenging than what we have seen in the literature.

To intuitively interpret the optimal tax formula (2.11), we investigate the effects of a small tax reform, as shown in Figure 2.2, in a unilaterally deviating country i: the second-best marginal

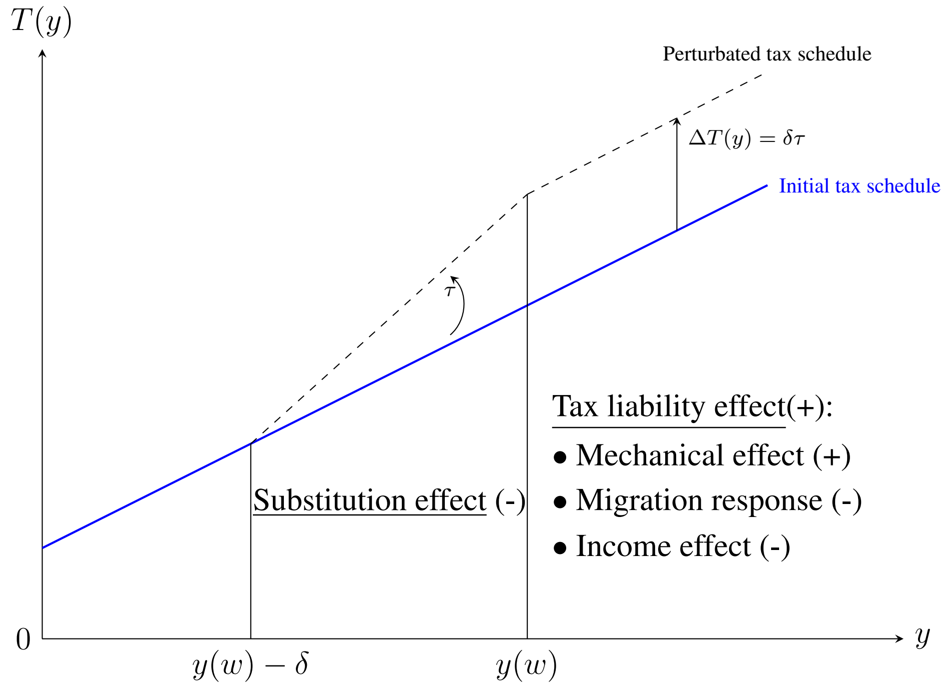

Figure 2.2: A Small Tax Reform Perturbation (e.g., [1])

tax rates $T_i'(y_i(w))$ are uniformly increased by a small amount $\tau > 0$ on the income interval $[y_i(w) - \delta, y_i(w)]$ for some small constant $\delta > 0$. As a consequence, tax liabilities above $y_i(w)$ are uniformly increased by $\delta\tau$. This gives rise to the following effects.

First, a worker with income in $[y_i(w) - \delta, y_i(w)]$ responds to the rise in the marginal tax rate by a substitution effect between leisure and labor, which hence reduces the taxes she pay. Second, each worker with skills above w faces a lump-sum increase $\delta\tau$ in her tax liability, which is called the mechanical effect in the literature (e.g., [46]). Since the unilateral rise in tax liability reduces her indirect utility in the deviating country, compared to its competitor, the number of labor outflow increases and hence the number of taxpayers with skills above w decreases. Following [1], we define the tax liability effect as the sum of the mechanical and migration effects for all skill levels above w. And third, the increase in tax liability tightens the consumption budget, and hence it follows from (2.12) that income effect will in turn reduce the positive mechanical effect. Since the optimal tax formula (2.11) is derived based on the Nash equilibrium, any unilateral deviation we consider cannot induce any first-order effect on the tax revenue of the deviating country. This implies that the tax liability effect must be positive so that the substitution effect is offset by the tax liability effect.

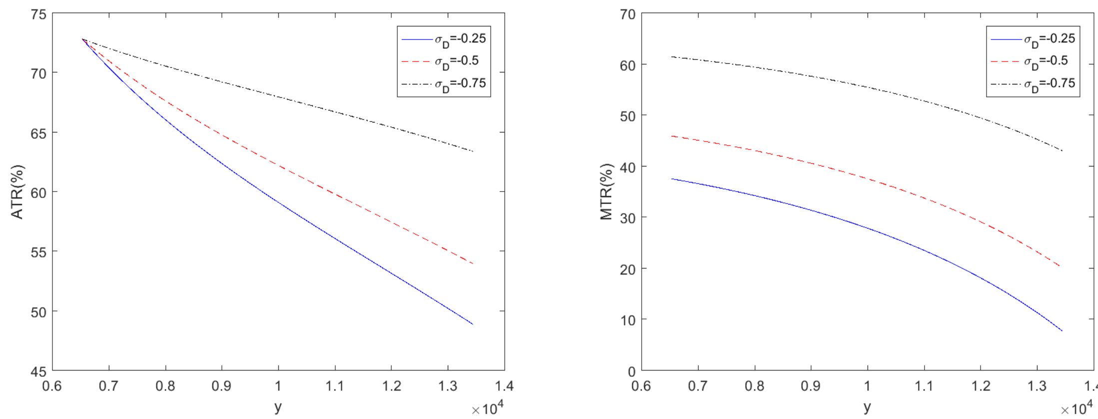

Figure 2.3: ATR and MTR in a Symmetric Nash equilibrium

To see how relativity changes the average tax rate (ATR)[7] and marginal tax rate (MTR), we also numerically solve the optimal tax formula (2.11) under the following assumptions (see Figure 2.3). First, these two countries are assumed to symmetric. Second, following [47], we put the mode $w_m = \$19,800$ and the highest skill level $\overline{w} = \$40,748$ with workers within this income interval having a Pareto income distribution, with density function $f(w) = aw_m^a/w^{a+1}$ for $w_m \leq w \leq \overline{w}$. Third, we use the quasilinear-in-consumption preference with a constant elasticity of labor supply, formally $u_i = c_i - l_i^{1+\frac{1}{\varepsilon}} / \left(1 + \frac{1}{\varepsilon}\right) + \sigma_D \mu_i + \sigma_F \mu_{-i}$ with $\sigma_D, \sigma_F \in (-1, 0)$. And fourth, also similar to the distribution assumption used by [47], we let the conditional distribution of migration costs be logistic:

$$G(0|w) = \frac{\text{exp}(-\chi w)}{1 + \text{exp}(-\chi w)} \;\; \text{for} \;\; \chi \in (0, 1).$$

Parameter values for simulation are given by $\varepsilon = 0.25$, $a = 2$,[8] $\chi = 0.5$, $l = 0.33$ and $\sigma_D \in$

[7]Since it is impossible to solve for a formula of ATR in the current context, we rely on numerical simulation to see the shape of ATR and how it changes with respect to the change of the degree of consumption relativity.

[8]See, e.g., [29], [8] and [47].

$(-1, 0)$. It follows from Figure 2.3 that both ATR and MTR increase as the degree of relative consumption concern $|\sigma_D|$ increases, for any $w \in [w_m, \overline{w}]$.

2.3.2 Qualitative Properties

To derive the qualitative properties of the optimal tax formula established in Theorem 2.3.1, we follow the approach developed by [47] and start by considering the same problem as in the second best, except that skills w are common knowledge, so migration costs m remain private information. Using the same terminology as used by [1], we call this benchmark *the Tiebout best*.

Lemma 2.3.1. *In a Nash equilibrium, we have the following predictions:*

(i) The Tiebout-best tax liabilities are given by

$$T_i^*(y_i(w)) = \frac{1}{v'(c_i(w))\tilde{\eta}_i(w)} \left[1 + \frac{\gamma_i}{\lambda_i} \frac{f_i(w)}{\tilde{f}_i(w)}\right]$$

for $\forall i \in \{A, B\}$, $\forall w \in (\underline{w}, \overline{w}]$, *with an upward jump discontinuity at* $\underline{w}$.

(ii) The Tiebout-best marginal tax rates verify:

$$\frac{T_i^{*'}(y_i(w))}{1 - T_i^{*'}(y_i(w))} = \frac{\gamma_i}{\lambda_i} \frac{f_i(w)}{\tilde{f}_i(w)} \quad \forall w \in [\underline{w}, \overline{w}]$$

with γ_i/λ_i *given in Theorem 2.3.1.*

Proof. See Appendix A. □

Under jealousy-type consumption comparison, it follows from (2.13) that $\gamma_i/\lambda_i > 0$. So for all skills but the bottom skill, the Tiebout-best tax liabilities under jealousy are strictly decreasing in the elasticity of migration, as shown in part (i). In addition, if the revenue requirement R is sufficiently small, then it follows from the fiscal-budget constraint (2.8) that the worst-off workers receive net transfers in the Tiebout-best economy. However, we have $\gamma_i/\lambda_i < 0$ under altruism-type consumption comparison, so the Tiebout-best tax liabilities are strictly decreasing in the elasticity of migration and the worst-off workers receive net transfers only when $\tilde{f}_i(w)/f_i(w) > -\gamma_i/\lambda_i$, namely either the amount of labor flow is bounded below or the degree of consumption comparison is bounded above. As shown in part (ii), the Tiebout-best marginal tax rates are used for correcting consumption externality as well as attracting labor inflow. In particular, tax rates are strictly positive under jealousy while strictly negative under altruism. Also, the ex ante to ex post density ratio $f_i(w)/\tilde{f}_i(w)$ and jealousy comparison impose a complementary effect while this ratio and altruism comparison impose a substitutive effect on the Tiebout-best tax liabilities and tax rates.

The following proposition gives a complete characterization of the second-best tax schedules under jealousy-type consumption comparison.

Proposition 2.3.1. *Suppose Assumption 2.2.1 holds, then the optimal tax structure in the Nash equilibrium has the following characteristics:*

(i) If $T_i(y_i(w)) \leq T_i^*(y_i(w))$ for $\forall w \in (\underline{w}, \overline{w})$, then $T_i'(y_i(w)) > 0$ for $\forall w \in (\underline{w}, \overline{w})$.

(ii) $T_i'(y_i(\underline{w})) > T_i^{*'}(y_i(\underline{w})) > 0$ and $T_i'(y_i(\overline{w})) = T_i^{*'}(y_i(\overline{w})) > 0$ for $\overline{w} < \infty$.

(iii) If $h(\cdot)$ is isoelastic, $f_i(w)/\tilde{f}_i(w)$ is decreasing in w, and $T_i(y_i(w)) \leq T_i^*(y_i(w))$ for $\forall w \in [\underline{w}, \overline{w}]$, then we have:

- **(a)** *$T_i'(y_i(w))$ is decreasing for $w \leq w_m$; and*
- **(b)** *$T_i'(y_i(w))$ is decreasing for $w > w_m$ when $wf_i(w)$ is non-decreasing in w.*

(iv) If $f_i(w)/\tilde{f}_i(w)$ is non-increasing in w, $-\frac{v''(c_i(w))\dot{c}_i(w)}{v'(c_i(w))} \leq \frac{\dot{\tilde{\eta}}_i(w)}{\tilde{\eta}_i(w)}$, and there exists a $\tilde{w} \in (\underline{w}, \overline{w})$ such that $T_i'(y_i(\tilde{w})) \geq 0$, then

$$T_i(y_i(w)) < \begin{cases} T_i^*(y_i(w)) & \text{for } \underline{w} < w \leq \tilde{w}, \\ \frac{1}{v'(c_i(w))\tilde{\eta}_i(w)}\left[1 + \frac{\gamma_i}{\lambda_i}\frac{f_i(w)}{\tilde{f}_i(w)}\right] & \text{for } w = \underline{w}. \end{cases}$$

Proof. See Appendix A. □

Parts (i)-(ii) identify (sufficient) conditions such that the second-best marginal tax rates are strictly positive over the entire income distribution. In particular, if tax liabilities are bounded above by the Tiebout-best tax liabilities, then the second-best marginal tax rates are strictly positive for almost all skills. Part (iii) identifies conditions such that the SOIC conditions are not binding, namely the first-order approach is reliable. Part (iv) identifies a sufficient condition such that the second-best tax liabilities are bounded above by the Tiebout-best tax liabilities.

The following proposition gives a complete characterization of the second-best tax schedules under altruism-type consumption comparison.

Proposition 2.3.2. *Suppose Assumption 2.2.2 holds, then the optimal tax structure in the Nash equilibrium has the following characteristics:*

(i) $T_i'(y_i(\underline{w})) > T_i^{*'}(y_i(\underline{w}))$ *and* $T_i'(y_i(\overline{w})) = T_i^{*'}(y_i(\overline{w})) < 0$ *for* $\overline{w} < \infty$.

(ii) Let $\Gamma_i(w) \equiv 1 + \frac{\gamma_i}{\lambda_i}\frac{f_i(w)}{\tilde{f}_i(w)}$. *If* $\Gamma_i(w) \geq 0$ *and* $\dot{\Gamma}_i(w) \geq \Gamma_i(w)\frac{\dot{\tilde{\eta}}_i(w)}{\tilde{\eta}_i(w)}$, *then we have*

(a) $\mathcal{A}_i(w)\mathcal{B}_i(w)\mathcal{C}_i(w) > 0$ *for* $\forall w \in (\underline{w}, \overline{w})$;

(b) $T_i(y_i(\overline{w})) < T_i^*(y_i(\overline{w}))$;

(c) $T_i(y_i(\underline{w})) > \Gamma_i(\underline{w})/[v'(c_i(\underline{w}))\tilde{\eta}_i(\underline{w})]$.

Proof. See Appendix A. □

For the worst-off workers, the second-best marginal tax rates under altruism comparison are strictly higher than the corresponding Tiebout-best marginal tax rates. For the top-income workers with bounded skill distribution, the second-best marginal tax rates under altruism comparison are equal to the corresponding Tiebout-best marginal tax rates and are strictly negative. Part (ii) identifies mild conditions such that the following conclusions hold: (a) the Mirrleesian-type tax in the optimal tax formula (2.11) is strictly positive for all skills but the endpoints; (b) for top-income workers, the second-best tax liabilities are strictly smaller than the Tiebout-best tax liabilities; and (c) there is a lower bound of the second-best tax liabilities for the worst-off workers.

The following proposition establishes a closed-form formula of the optimal asymptotic tax rates (or the tax rates placed on top-income workers) under certain conditions.

Proposition 2.3.3. *Suppose economic environments satisfy the following conditions:*

(a) $v'(\cdot) = 1$, *namely quasilinear-in-consumption preferences;*

(b) $h(\cdot)$ *is isoelastic with elastic coefficient* $\varepsilon > 0$*;*

(c) $F_i(w)$ *is a Pareto distribution with* $\overline{w} = \infty$ *and Pareto index* $a_i > 1$.

Then, the optimal asymptotic marginal tax rate (AMTR) in a Nash equilibrium is:

$$T_i'(y_i(\infty)) = \frac{\frac{\gamma_i}{\lambda_i}\alpha_i(\infty) + \left[1 + \frac{\gamma_i}{\lambda_i}\alpha_i(\infty)\right](1+\varepsilon)(1/a_i)}{1 + \frac{\gamma_i}{\lambda_i}\alpha_i(\infty) + \left[1 + \frac{\gamma_i}{\lambda_i}\alpha_i(\infty) + \tilde{\theta}_i(\infty)\right](1+\varepsilon)(1/a_i)},$$

with $\tilde{\theta}_i(\infty) \equiv \lim_{w\uparrow\infty} \tilde{\theta}_i(w) \geq 0$ *and* $\alpha_i(\infty) \equiv \lim_{w\uparrow\infty} \frac{f_i(w)}{\tilde{f}_i(w)} \geq 0$.

Proof. See Appendix A. □

Given that the optimal tax formula (2.11) is quite complicated, restrictions (a)-(c) must be tolerated for explicitly solving for the optimal AMTR. In fact, conditions (a)-(b) are widely used in the literature of optimal taxation, and Pareto distribution is an empirically supported assumption for high-income workers. In the current context, the optimal AMTR is a continuously differentiable function of five important variables: the degree of consumption comparison γ_i/λ_i, the measure of labor flow $\alpha_i(\infty)$, the elasticity of labor supply ε, the degree of income inequality $1/a_i$, and the elasticity of migration $\tilde{\theta}_i(\infty)$. In particular, AMTR is strictly decreasing in the elasticity of migration.

The following two propositions characterize the composition effect of consumption relativity and income inequality on the optimal AMTR. The composition effect can be completely different under alternative circumstances.

Proposition 2.3.4. *Suppose* $\psi(\mu_i, \mu_{-i}) = \sigma_D \mu_i + \tilde{\psi}(\mu_{-i})$ *for a constant* $\sigma_D \in (-1, 0)$, $F_i(w) = F_{-i}(w)$ *and* $\partial \tilde{F}_i(\infty)/\partial a_i = 0$, *then we have the following predictions.*

(i) If $\tilde{\theta}_i(\infty) \leq 1$, *then* $\frac{\partial^2 T_i'(y_i(\infty))}{\partial(-\sigma_D)\partial(1/a_i)} < 0$.

(ii) If $\tilde{\theta}_i(\infty) > 1$, then

$$\frac{\partial^2 T_i'(y_i(\infty))}{\partial\left(-\sigma_D\right)\partial(1/a_i)}\begin{cases} < 0 & \textit{for } \alpha_i(\infty) < \left(\frac{\lambda_i}{\gamma_i}\right)\frac{1+\tilde{\theta}_i(\infty)+\frac{1+\varepsilon}{a_i}\left[1+\tilde{\theta}_i(\infty)^2\right]}{\left(1+\frac{1+\varepsilon}{a_i}\right)\left[\tilde{\theta}_i(\infty)-1\right]}; \\ > 0 & \textit{for } \alpha_i(\infty) > \left(\frac{\lambda_i}{\gamma_i}\right)\frac{1+\tilde{\theta}_i(\infty)+\frac{1+\varepsilon}{a_i}\left[1+\tilde{\theta}_i(\infty)^2\right]}{\left(1+\frac{1+\varepsilon}{a_i}\right)\left[\tilde{\theta}_i(\infty)-1\right]}. \end{cases}$$

Proof. See Appendix A. □

Proposition 2.3.4 analyzes the case with jealousy-type consumption comparison. If the elasticity of migration is not greater than one, then relativity and inequality play a substitutive role in shaping AMTR. Precisely, the higher is inequality, the lower is the effect of relativity in raising AMTR; similarly, the higher is relativity, the lower is the effect of inequality in raising AMTR. However, if the elasticity of migration is greater than one, then relativity and inequality play a substitutive role only when the ex post mass of top-income workers is greater than a threshold, otherwise relativity and inequality play a complementary role in shaping the optimal AMTR.

Under similar assumptions, K&T show that relativity and inequality always play a substitutive role in a closed economy. We show that such a conclusion depends on the elasticity of migration and the level of migration in an open economy. So, Proposition 2.3.4 generalizes the prediction of K&T as a special case with $\tilde{\theta}_i(\infty) = 0$ and $\alpha_i(\infty) = 1$.

Proposition 2.3.5. *Suppose $\psi(\mu_i, \mu_{-i}) = \sigma_D \mu_i + \tilde{\psi}(\mu_{-i})$ for a constant $\sigma_D \in (0, 1)$, $F_i(w) = F_{-i}(w)$ and $\partial \tilde{F}_i(\infty)/\partial a_i = 0$, then we have the following predictions.*

(i) If $\tilde{\theta}_i(\infty) < 1$, then

$$\frac{\partial^2 T_i'(y_i(\infty))}{\partial\sigma_D\partial(1/a_i)}\begin{cases} < 0 & \textit{for } \left(-\frac{\lambda_i}{\gamma_i}\right)\frac{1+\frac{1+\varepsilon}{a_i}\left[1+\tilde{\theta}_i(\infty)\right]}{1+\frac{1+\varepsilon}{a_i}} < \alpha_i(\infty) < \left(-\frac{\lambda_i}{\gamma_i}\right)\frac{1+\tilde{\theta}_i(\infty)+\frac{1+\varepsilon}{a_i}\left[1+\tilde{\theta}_i(\infty)^2\right]}{\left(1+\frac{1+\varepsilon}{a_i}\right)\left[1-\tilde{\theta}_i(\infty)\right]}; \\ > 0 & \textit{otherwise.} \end{cases}$$

(ii) If $\tilde{\theta}_i(\infty) = 1$, then

$$\frac{\partial^2 T_i'(y_i(\infty))}{\partial\sigma_D \partial(1/a_i)} \begin{cases} < 0 & \textit{for } \alpha_i(\infty) > \left(-\frac{\lambda_i}{\gamma_i}\right) \frac{1+2\left(\frac{1+\varepsilon}{a_i}\right)}{1+\frac{1+\varepsilon}{a_i}}; \\ > 0 & \textit{for } \alpha_i(\infty) < \left(-\frac{\lambda_i}{\gamma_i}\right) \frac{1+2\left(\frac{1+\varepsilon}{a_i}\right)}{1+\frac{1+\varepsilon}{a_i}}. \end{cases}$$

(iii) If $\tilde{\theta}_i(\infty) > 1$, then

$$\frac{\partial^2 T_i'(y_i(\infty))}{\partial\sigma_D \partial(1/a_i)} \begin{cases} < 0 & \textit{for } \alpha_i(\infty) > \left(-\frac{\lambda_i}{\gamma_i}\right) \frac{1+\frac{1+\varepsilon}{a_i}\left[1+\tilde{\theta}_i(\infty)\right]}{1+\frac{1+\varepsilon}{a_i}}; \\ > 0 & \textit{for } \alpha_i(\infty) < \left(-\frac{\lambda_i}{\gamma_i}\right) \frac{1+\frac{1+\varepsilon}{a_i}\left[1+\tilde{\theta}_i(\infty)\right]}{1+\frac{1+\varepsilon}{a_i}}. \end{cases}$$

Proof. See Appendix A. □

Proposition 2.3.5 analyzes the composition effect of relativity and inequality on the optimal AMTR under altruism-type consumption comparison. Compared to the case with jealousy-type consumption comparison, the current predictions seem to be more subtle.

If the elasticity of migration is smaller than one, then relativity and inequality impose a substitutive effect on AMTR only when the ex post mass of top-income workers is bounded both below and above. Otherwise, they impose a complementary effect, namely the higher is inequality, the higher is the effect of relativity in raising AMTR; or the higher is relativity, the higher is the effect of inequality in raising AMTR. If the elasticity of migration is greater than or equal to one, then relativity and inequality impose a substitutive effect on AMTR only when the ex post mass of top-income workers is smaller than some threshold, and this threshold is heterogenous for different values of migration elasticity.

2.4 Stackelberg Equilibrium

2.4.1 Optimal Tax Formula

Without any loss of generality, we denote by i the leader country and $-i$ the follower country in the current Stackelberg game. We thus state the second major result in the following theorem.

Theorem 2.4.1. *In a Stackelberg equilibrium, the optimal tax formula is the same as that in the Nash equilibrium, except that:*

$$\frac{\gamma_i}{\lambda_i} = \frac{\int_{\underline{w}}^{\overline{w}} \left(\frac{\partial c_i(w)}{\partial \mu_i} + \frac{\partial c_i(w)}{\partial \mu_{-i}} \frac{\partial \mu_{-i}}{\partial \mu_i} \right) \tilde{f}_i(w) dw}{1 - \int_{\underline{w}}^{\overline{w}} \left(\frac{\partial c_i(w)}{\partial \mu_i} + \frac{\partial c_i(w)}{\partial \mu_{-i}} \frac{\partial \mu_{-i}}{\partial \mu_i} \right) f_i(w) dw}$$

with

$$\frac{\partial \mu_{-i}}{\partial \mu_i} = \frac{\int_{\underline{w}}^{\overline{w}} \frac{\partial c_{-i}(w)}{\partial \mu_i} f_{-i}(w) dw}{1 - \int_{\underline{w}}^{\overline{w}} \frac{\partial c_{-i}(w)}{\partial \mu_{-i}} f_{-i}(w) dw}$$

for the leader country i.

Proof. See Appendix A. □

Theorems 2.4.1 and 2.3.1 together demonstrate how the form of tax competition might affect optimal tax rates. Intuitively, since the leader country takes into account the behavioral response of the follower country in the dynamic Stackelberg game, it partially internalizes cross-country consumption externalities, namely the additional term $\partial \mu_{-i} / \partial \mu_i$ is in general different from zero.

2.4.2 Qualitative Properties

Using Theorem 2.4.1, the following corollary is immediate.

Corollary 2.4.1. *If $|\psi_i(\mu_i, \mu_{-i})| > |\psi_{-i}(\mu_i, \mu_{-i})|$ for country i, then the qualitative properties (of Nash equilibrium) established in Propositions 2.3.1-2.3.3 carry over to the current Stackelberg equilibrium.*

For additively separable functional forms of $\psi(\mu_i, \mu_{-i})$, condition $|\psi_i(\mu_i, \mu_{-i})| > |\psi_{-i}(\mu_i, \mu_{-i})|$ means that the degree of domestic consumption comparison is greater than that of cross-country consumption comparison for workers in country i. Given the real-life observation that people are more often to make status comparison with people who live in their social networks, this restriction can be regarded as reasonable.

Proposition 2.4.1. *If economic environments satisfy the following conditions:*

(a) The utility function of relative consumptions has the form:

$$\psi(\mu_i, \mu_{-i}) = \begin{cases} \sigma_D \mu_i + \sigma_F \mu_{-i} & \text{for country } i, \\ \sigma_D \mu_{-i} + \sigma_F \mu_i & \text{for country } -i \end{cases}$$

with coefficients $\sigma_D, \sigma_F \in (-1, 0) \cup (0, 1)$ *and* $|\sigma_F| + |\sigma_D| < 1$*;*

(b) $F_i(w) = F_{-i}(w)$*;*

(c) $\partial \tilde{F}_i(\infty)/\partial a_i = 0$*.*

Then, for the optimal AMTR of leader country i *in a Stackelberg equilibrium, the predictions established in Propositions 2.3.4-2.3.5 carry over to the current equilibrium.*

Proof. See Appendix A. □

Provided that we have assumed quasilinear-in-consumption preferences in solving for the optimal AMTR, condition (a) is thus a natural restriction. Condition (b) simplifies our analysis by eliminating asymmetry between these two countries, which however is not an essential requirement for establishing the current prediction. Condition (c) is a technical assumption mainly for the purpose of simplicity. The main message implied by Proposition 2.4.1 is that the composition effect of relativity and inequality imposed on the optimal AMTR is in general the same under both forms of tax competition, even though the corresponding AMTRs are in general different.

Proposition 2.4.2. *If Assumption 2.2.1 holds, then the government of the leader country imposes a higher marginal tax rate in the Stackelberg equilibrium than that in the Nash equilibrium.*

Proof. See Appendix A. □

Intuitively, since jealousy implies negative consumption externality and marginal tax rates strictly increase as externality increases, the leader country who (partially) internalizes cross-country consumption externalities imposes a higher tax rate than that it may impose in a simultaneous-move static game where no one internalizes cross-country consumption externalities.

2.5 Numerical Illustration

In this section we provide some numerical examples on the optimal AMTR established in Proposition 2.3.3. Although these exercises are very coarse, they indeed enable us to see quantitatively how large the difference on the optimal AMTR can be made by the effects of strategic tax competition and cross-country consumption comparisons, when compared to what K&T have obtained in a closed economy through ignoring these effects.

For simplicity, we use the linear utility function of relative consumptions shown in condition (a) of Proposition 2.4.1. The following tables present AMTRs for different parameter values, when the Pareto index $a_i = 2$ and 3, the coefficient of domestic relative consumption $\sigma_D = -0.5$, 0 and 0.5, and the elasticity of labor supply $\varepsilon = 0.25$, 0.33, and 0.5. We consider three elasticity scenarios. The first two with $\varepsilon = 0.25$ and 0.33 are realistic midrange estimates (see [48]), while $\varepsilon = 0.5$ is a little bit larger than the current average empirical estimates. We consider two inequality scenarios. The first one with $a_i = 2$ is based on the 2005 U.S. empirical income distribution (see [29]), while $a_i = 3$ is chosen to be larger than this realistic number to represent an experimental scenario with more equal income distribution.

Moreover, we consider three relativity scenarios. $\sigma_D = 0$ denotes the benchmark case without relative consumption, whereas $\sigma_D = -0.5$ measures the degree of jealousy and $\sigma_D = 0.5$ measures the degree of altruism. One of the key findings of the empirical research on relativity is that the estimated coefficient on income (consumption) and income comparison is statistically almost equal and opposite (see, e.g., [11]). Given the assumption of quasi-linear preferences, $\sigma_D = -0.5$ seems to be reasonable. In fact, it is consistent with the finding of [10] who use survey-experimental methods to see how much we care about absolute versus relative income and consumption. We choose $\sigma_D = 0.5$ for the sake of comparing with the case with $\sigma_D = -0.5$. By assumption, the degree of cross-country social comparisons is smaller than that of domestic social comparisons, so we let $\sigma_F^2 = 0.04$[9] in what follows. Following [18], we let the value of the elasticity of migration be 0.25, i.e., $\tilde{\theta}_i(\infty) = 0.25$. We summarize all realistic parameter values in Table 2.1.

[9] In fact, we have not found any realistic estimates of this parameter in empirical literature.

Table 2.1: Parameter Values

	Value	Description	Source/Target
$\tilde{\theta}_i(\infty)$	0.25	Migration elasticity	Piketty & Saez [18]
ε	(0.1,0.4)	Labor-supply elasticity	Saez et al. [48]
a_i	2	Pareto index	Diamond & Saez [29]
σ_D	-0.5	Domestic relativity	Clark et al. [13]
σ_F	-0.2	Cross-country relativity	$\lvert\sigma_F\rvert < \lvert\sigma_D\rvert$

If $\Delta_i(\infty) \geq 0$, namely top-income workers get an indirect utility in country i which is not less than that they can get in the other country $-i$, then the density ratio $\alpha_i(\infty)$ must not be greater than 1. That is, country i has the potential to attract more high-skill workers from the opponent country $-i$. Similarly, if $\Delta_i(\infty) \leq 0$, then the density ratio $\alpha_i(\infty)$ must not be smaller than 1. The following tables consider both cases.

2.5.1 A Comparison with K&T

In Tables 2.3-2.10, we use red numbers to denote the optimal AMTRs calculated using the formula of K&T. In both types of equilibrium, we obtain the following findings under different values of $\alpha_i(\infty)$. In particular, negative numbers imply that workers receive transfers, which occurs only when workers exhibit altruism-type relative consumption preferences.

First, for each given labor-supply elasticity and given degree of relative consumption, AMTR increases as inequality increases. Second, for each given degree of inequality and given degree of relative consumption, AMTR increases as elasticity increases. And third, for each given elasticity and given degree of inequality, AMTR significantly increases under jealousy and significantly declines under altruism, compared to the benchmark case without relative consumption concerns.

We summarize the economic mechanism in Table 2.2, in which the superscripts of MTR^O and MTR^C denote open-economy and closed-economy, respectively. In particular, we just consider the MTR of the leader country under Stackelberg tax competition. Essentially, as already shown in Table 2.2, relativity and migration are determinant factors in such a comparison.

Since no one internalizes the cross-country consumption externality under Nash competition,

Table 2.2: Economic Mechanism Governing Quantitative Findings

		Nash	Nash	Stack	Stack	
	Relativity effect	$=$	$=$	$>$	$>$	
	Migration effect	Small	Large	Small	Large	
MTR^O	Labor inflow	$<$	$<<$	$>$	$<<$	MTR^C
MTR^O	Labor outflow	$<$	$>>$	$>$	$>>$	MTR^C

the relativity effect on MTR is the same between an open economy and a closed economy. In contrast, as the leader country internalizes cross-country consumption externality under Stackelberg competition, the relativity effect on MTR implemented by the leader country in an open economy should be greater than that in a closed economy. Therefore, if there is no migration between countries, only the MTR implemented under Stackelberg competition should be higher than that implemented in a closed economy.

For comparing MTR^O and MTR^C under Nash, migration effect dominates relativity effect. If labor flow is small, no matter it is inflow or outflow, Nash competition implies a smaller MTR than that in a closed economy without any migration threat imposed on the government. Nevertheless, if labor flow is large, then migration effect is heterogenous between the case with labor inflow and the case with labor outflow. Precisely, large labor inflow must be induced by a much lower MTR compared to MTR^C, while large labor outflow must be induced by a much higher MTR compared to MTR^C.

For comparing MTR^O and MTR^C under Stackelberg, both relativity effect and migration effect matter. If labor flow is small, then relativity effect dominates migration effect for both the case with labor inflow and the case with labor outflow, implying that MTR^O under Stackelberg competition should be higher than MTR^C. However, if labor flow is large, then migration effect dominates relativity effect and it is heterogenous between the case with labor inflow and the case with labor outflow. Precisely, large labor inflow must be induced by a much lower MTR compared to MTR^C, while large labor outflow must be induced by a much higher MTR compared to MTR^C. As a result, under a large labor flow, the prediction is the same between Nash and Stackelberg tax competition.

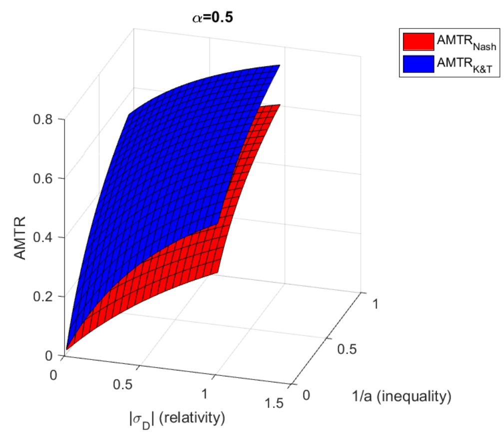

Figure 2.4: For $\alpha_i(\infty) = 0.5$, $\varepsilon = \tilde{\theta}_i(\infty) = 0.25$, $a_i > 1$, $\sigma_D \in (-1, 0)$

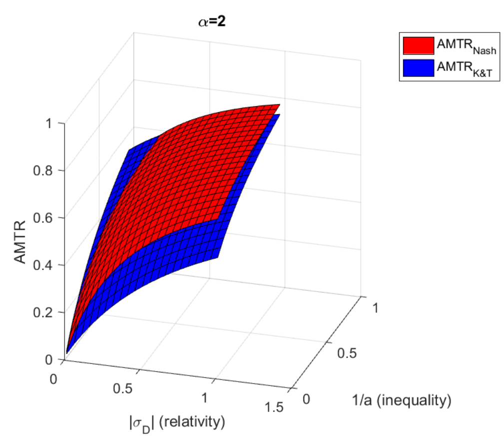

Figure 2.5: For $\alpha_i(\infty) = 2$, $\varepsilon = \tilde{\theta}_i(\infty) = 0.25$, $a_i > 1$, $\sigma_D \in (-1, 0)$

2.5.1.1 Nash vs. K&T

Tables 2.3-2.6 compare optimal AMTRs in Nash equilibrium with those in K&T. They show that the difference on AMTRs increases as the net level of migration increases, precisely as $\alpha_i(\infty)$ declines under $\Delta_i(\infty) \geq 0$ and as $\alpha_i(\infty)$ increases under $\Delta_i(\infty) \leq 0$. In particular, we can obtain under symmetry between these two countries that $\alpha_i(\infty) = 0.67 \iff Pr(m < U_i(\infty) - U_{-i}(\infty)) = 49\%$ and $\alpha_i(\infty) = 2.00 \iff Pr(m < U_{-i}(\infty) - U_i(\infty)) = 50\%$, namely the migration probability is around 50% at these values of $\alpha_i(\infty)$. Under jealousy-type relativity with $\Delta_i(\infty) \geq 0$, Nash AMTRs are always smaller than those in K&T (see Figure 2.4). However, if

$\Delta_i(\infty) \leq 0$, they are greater than those in K&T when $\alpha_i(\infty)$ is larger than some critical value (see Figure 2.5). Under altruism-type relativity with $\Delta_i(\infty) \geq 0$, they are always greater than those in K&T. However, if $\Delta_i(\infty) \leq 0$, they are always smaller than those in K&T. Also, if there is no relative consumption concern, then they are always smaller than those in K&T. Moreover, the migration effect is magnified by consumption comparison.

Table 2.3: AMTR (%) under $\Delta_i(\infty) \geq 0$ with $\alpha_i(\infty) = 0.95$

	$\varepsilon = 0.25$	$\varepsilon = 0.25$	$\varepsilon = 0.33$	$\varepsilon = 0.33$	$\varepsilon = 0.5$	$\varepsilon = 0.5$
	$a_i = 2$	$a_i = 3$	$a_i = 2$	$a_i = 3$	$a_i = 2$	$a_i = 3$
$\sigma_D = -0.5$	65.2,69.2	61.5,64.7	65.8,70.0	62.0,65.4	67.0,71.4	63.1,66.7
$\sigma_D = 0$	35.1,38.5	27.4,29.4	36.3,39.9	28.5,30.7	38.7,42.9	30.8,33.3
$\sigma_D = 0.5$	8.7,7.7	-3.0,-5.9	10.6,9.9	-1.3,-3.9	14.2,14.3	2.2,0.0

Table 2.4: AMTR (%) under $\Delta_i(\infty) \geq 0$ with $\alpha_i(\infty) = 0.67$

	$\varepsilon = 0.25$	$\varepsilon = 0.25$	$\varepsilon = 0.33$	$\varepsilon = 0.33$	$\varepsilon = 0.5$	$\varepsilon = 0.5$
	$a_i = 2$	$a_i = 3$	$a_i = 2$	$a_i = 3$	$a_i = 2$	$a_i = 3$
$\sigma_D = -0.5$	59.7,69.2	55.3,64.7	60.4,70.0	55.9,65.4	61.8,71.4	57.2,66.7
$\sigma_D = 0$	35.1,38.5	27.4,29.4	36.3,39.9	28.5,30.7	38.7,42.9	30.8,33.3
$\sigma_D = 0.5$	18.5,7.7	8.3,-5.9	20.1,9.9	9.8,-3.9	23.2,14.3	12.8,0.0

Table 2.5: AMTR (%) under $\Delta_i(\infty) \leq 0$ with $\alpha_i(\infty) = 1.05$

	$\varepsilon = 0.25$	$\varepsilon = 0.25$	$\varepsilon = 0.33$	$\varepsilon = 0.33$	$\varepsilon = 0.5$	$\varepsilon = 0.5$
	$a_i = 2$	$a_i = 3$	$a_i = 2$	$a_i = 3$	$a_i = 2$	$a_i = 3$
$\sigma_D = -0.5$	66.8,69.2	63.3,64.7	67.4,70.0	63.8,65.4	68.5,71.4	64.8,66.7
$\sigma_D = 0$	35.1,38.5	27.4,29.4	36.3,39.9	28.5,30.7	38.7,42.9	30.8,33.3
$\sigma_D = 0.5$	4.6,7.7	-7.7,-5.9	6.6,9.9	-5.9,-3.9	10.4,14.3	-2.2,0.0

Table 2.6: AMTR (%) under $\Delta_i(\infty) \leq 0$ with $\alpha_i(\infty) = 1.55$

	$\varepsilon = 0.25$	$\varepsilon = 0.25$	$\varepsilon = 0.33$	$\varepsilon = 0.33$	$\varepsilon = 0.5$	$\varepsilon = 0.5$
	$a_i = 2$	$a_i = 3$	$a_i = 2$	$a_i = 3$	$a_i = 2$	$a_i = 3$
$\sigma_D = -0.5$	73.1,69.2	70.3,64.7	73.6,70.0	70.7,65.4	74.5,71.4	71.5,66.7
$\sigma_D = 0$	35.1,38.5	27.4,29.4	36.3,39.9	28.5,30.7	38.7,42.9	30.8,33.3
$\sigma_D = 0.5$	-22.8,7.7	-40.0,-5.9	-20.1,9.9	-37.4,-3.9	-14.9,14.3	-32.4,0.0

2.5.1.2 *Stackelberg vs. K&T*

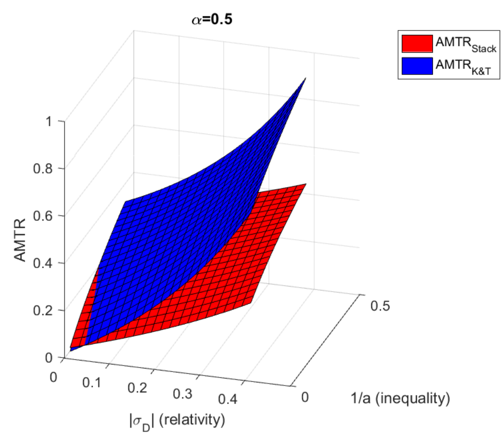

Figure 2.6: For $\alpha_i(\infty) = 0.5$, $\varepsilon = \tilde{\theta}_i(\infty) = 0.25$, $a_i > 1$, $\sigma_D < 0$, $\sigma_F^2 = 0.04$

Tables 2.7-2.10 compare optimal AMTRs in Stackelberg equilibrium, denoted by blue numbers, with those in K&T. The difference on AMTRs increases as the net level of migration increases, precisely as $\alpha_i(\infty)$ declines under $\Delta_i(\infty) \geq 0$ and as $\alpha_i(\infty)$ increases under $\Delta_i(\infty) \leq 0$. Under jealousy-type relativity with $\Delta_i(\infty) \leq 0$, Stackelberg AMTRs are in general larger than those in K&T (see Figure 2.7). However, if $\Delta_i(\infty) \geq 0$, they are smaller than those in K&T for $\alpha_i(\infty)$ smaller than some critical value (see Figure 2.6). Under altruism-type relativity with $\Delta_i(\infty) \geq 0$, they are always greater than those in K&T. However, if $\Delta_i(\infty) \leq 0$, they are always smaller than those in K&T.

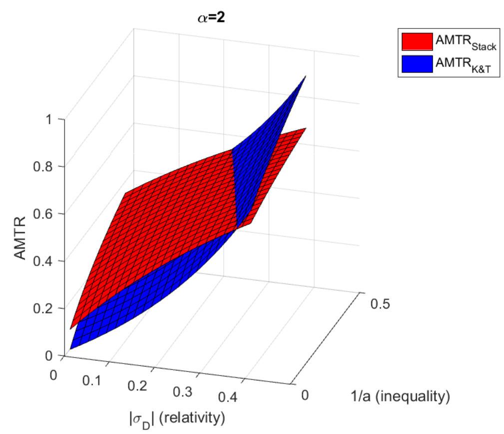

Figure 2.7: For $\alpha_i(\infty) = 2$, $\varepsilon = \tilde{\theta}_i(\infty) = 0.25$, $a_i > 1$, $\sigma_D < 0$, $\sigma_F^2 = 0.04$

Table 2.7: AMTR (%) under $\Delta_i(\infty) \geq 0$ with $\alpha_i(\infty) = 0.95$

	$\varepsilon = 0.25$	$\varepsilon = 0.25$	$\varepsilon = 0.33$	$\varepsilon = 0.33$	$\varepsilon = 0.5$	$\varepsilon = 0.5$
	$a_i = 2$	$a_i = 3$	$a_i = 2$	$a_i = 3$	$a_i = 2$	$a_i = 3$
$\sigma_D = -0.5$	70.5,69.2	67.3,64.7	71.0,70.0	67.8,65.4	71.9,71.4	68.7,66.7
$\sigma_D = 0$	37.4,38.5	30.0,29.4	38.5,39.9	31.1,30.7	40.8,42.9	33.2,33.3
$\sigma_D = 0.5$	10.0,7.7	-1.4,-5.9	11.9,9.9	0.3,-3.9	15.4,14.3	3.6,0.0

Table 2.8: AMTR (%) under $\Delta_i(\infty) \geq 0$ with $\alpha_i(\infty) = 0.55$

	$\varepsilon = 0.25$	$\varepsilon = 0.25$	$\varepsilon = 0.33$	$\varepsilon = 0.33$	$\varepsilon = 0.5$	$\varepsilon = 0.5$
	$a_i = 2$	$a_i = 3$	$a_i = 2$	$a_i = 3$	$a_i = 2$	$a_i = 3$
$\sigma_D = -0.5$	61.7,69.2	57.5,64.7	62.3,70.0	58.1,65.4	63.6,71.4	59.3,66.7
$\sigma_D = 0$	36.4,38.5	28.9,29.4	38.5,39.9	31.1,30.7	40.8,42.9	33.2,33.3
$\sigma_D = 0.5$	22.6,7.7	13.1,-5.9	24.1,9.9	14.5,-3.9	27.1,14.3	17.3,0.0

2.5.2 Nash vs. Stackelberg: the Leader Country

Tables 2.11-2.12 illustrate Proposition 2.4.2 by comparing AMTRs under these two types of equilibrium. As is obvious, no matter $\Delta_i(\infty) \geq 0$ or $\Delta_i(\infty) \leq 0$, these AMTRs in Nash equilibrium are in general smaller than those in Stackelberg equilibrium (see also Figures 2.8-2.9).

Table 2.9: AMTR (%) under $\Delta_i(\infty) \leq 0$ with $\alpha_i(\infty) = 1.05$

	$\varepsilon = 0.25$	$\varepsilon = 0.25$	$\varepsilon = 0.33$	$\varepsilon = 0.33$	$\varepsilon = 0.5$	$\varepsilon = 0.5$
	$a_i = 2$	$a_i = 3$	$a_i = 2$	$a_i = 3$	$a_i = 2$	$a_i = 3$
$\sigma_D = -0.5$	72.1,69.2	69.1,64.7	72.5,70.0	69.5,65.4	73.5,71.4	70.4,66.7
$\sigma_D = 0$	37.6,38.5	30.2,29.4	38.8,39.9	31.3,30.7	41.0,42.9	33.5,33.3
$\sigma_D = 0.5$	6.2,7.7	-5.9,-5.9	8.1,9.9	-4.1,-3.9	11.9,14.3	-0.5,0.0

Table 2.10: AMTR (%) under $\Delta_i(\infty) \leq 0$ with $\alpha_i(\infty) = 1.55$

	$\varepsilon = 0.25$	$\varepsilon = 0.25$	$\varepsilon = 0.33$	$\varepsilon = 0.33$	$\varepsilon = 0.5$	$\varepsilon = 0.5$
	$a_i = 2$	$a_i = 3$	$a_i = 2$	$a_i = 3$	$a_i = 2$	$a_i = 3$
$\sigma_D = -0.5$	78.0,69.2	75.7,64.7	78.4,70.0	76.1,65.4	79.1,71.4	76.7,66.7
$\sigma_D = 0$	38.7,38.5	31.5,29.4	38.8,39.9	31.3,30.7	41.0,42.9	33.5,33.3
$\sigma_D = 0.5$	-18.9,7.7	-35.4,-5.9	-16.4,9.9	-33.0,-3.9	-11.4,14.3	-28.1,0.0

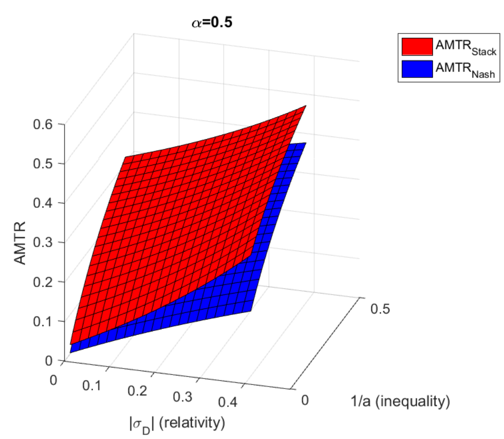

Figure 2.8: For $\alpha_i(\infty) = 0.5$, $\varepsilon = \tilde{\theta}_i(\infty) = 0.25$, $a_i > 1$, $\sigma_D < 0$, $\sigma_F^2 = 0.04$

2.6 Conclusion

In this essay, we develop a theoretical framework to analyze how the interplay of relative consumption concern and income inequality determines optimal income taxes in an international setting with two competing countries. We establish and qualitatively characterize nonlinear labor income tax schedules that competing Rawlsian governments should implement when workers hav-

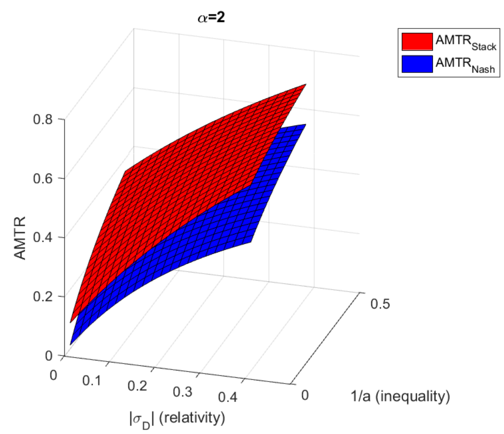

Figure 2.9: For $\alpha_i(\infty) = 2$, $\varepsilon = \tilde{\theta}_i(\infty) = 0.25$, $a_i > 1$, $\sigma_D < 0$, $\sigma_F^2 = 0.04$

Table 2.11: AMTR (%) under $\Delta_i(\infty) \geq 0$ with $\alpha_i(\infty) = 0.95$

	$\varepsilon = 0.25$	$\varepsilon = 0.25$	$\varepsilon = 0.33$	$\varepsilon = 0.33$	$\varepsilon = 0.5$	$\varepsilon = 0.5$
	$a_i = 2$	$a_i = 3$	$a_i = 2$	$a_i = 3$	$a_i = 2$	$a_i = 3$
$\sigma_D = -0.5$	65.2,70.5	61.5,67.3	65.8,71.0	62.0,67.8	67.0,71.9	63.1,68.7
$\sigma_D = 0$	35.1,37.4	27.4,30.0	36.3,38.5	28.5,31.1	38.7,40.8	30.8,33.2
$\sigma_D = 0.5$	8.7,10.0	-3.0,-1.4	10.6,11.9	-1.3,0.3	14.2,15.4	2.2,3.6

Table 2.12: AMTR (%) under $\Delta_i(\infty) \leq 0$ with $\alpha_i(\infty) = 1.05$

	$\varepsilon = 0.25$	$\varepsilon = 0.25$	$\varepsilon = 0.33$	$\varepsilon = 0.33$	$\varepsilon = 0.5$	$\varepsilon = 0.5$
	$a_i = 2$	$a_i = 3$	$a_i = 2$	$a_i = 3$	$a_i = 2$	$a_i = 3$
$\sigma_D = -0.5$	66.8,72.1	63.3,69.1	67.4,72.5	63.8,69.5	68.5,73.5	64.8,70.4
$\sigma_D = 0$	35.1,37.6	27.4,30.2	36.3,38.8	28.5,31.3	38.7,41.0	30.8,33.5
$\sigma_D = 0.5$	4.6,6.2	-7.7,-5.9	6.6,8.1	-5.9,-4.1	10.4,11.9	-2.2,-0.5

ing private information on skills and migration costs decide where to live and how much to work. In addition to the case where governments play Nash, we also examine the scenario where they play Stackelberg.

Firstly, we obtain an optimal tax formula that can be interpreted as a nontrivial generalization of those obtained by [45], [46], [8], [1] and [37]. Secondly, we numerically calculate optimal AMTRs under both types of equilibrium and compare them to those obtained using the formula of K&T, finding that the country with large labor inflow imposes a much smaller marginal tax rate and the country with large labor outflow imposes a much higher marginal tax rate than suggested by K&T. This finding holds for sufficiently various combinations of parameters measuring relative consumption, labor mobility and income inequality. Thirdly, for the case with quasilinear-in-consumption preferences and jealousy-type consumption comparison, we show that the leader country imposes a higher marginal tax rate in Stackelberg equilibrium than that it may impose in Nash equilibrium. And fourthly, we provide a complete characterization on how relativity and inequality together determine the optimal AMTR under both Nash and Stackelberg tax competition, finding that both the elasticity and level of migration are determinant for predicting when relativity and inequality are complementary or substitutive in shaping the optimal tax rates placed on top-income workers.

We, therefore, show that the optimal redistributive taxation policy for countries involved in globalization should not ignore these important effects resulted from tax-driven migrations as well as the interplay of relativity and inequality. Since alternative forms of tax competition generate heterogenous effects on optimal tax rates, the identification of the form of tax competition should be of practical relevance, which however awaits future research.

3. VOTING OVER SELFISHLY OPTIMAL INCOME TAX SCHEDULES WITH TAX-DRIVEN MIGRATIONS

3.1 Introduction

As barriers to labor mobility have been lowered and education and language skills have improved, governments are facing the challenge that the base of labor income tax is becoming more mobile. This is especially true for highly skilled workers. For example, [3], [4] and [5] estimate large migration elasticities with respect to income tax rate for these types of workers.

To analyze how the possibility of geographic mobility affects the design of redistributive taxation, the literature (see, e.g., [15], [16], [17], [18], [1]) that builds on the seminal work of Mirrlees [19] focuses on the normative perspective. Little attention has been paid to the positive perspective. We hence address this question: how would the schedule of redistributive taxation look like when workers can vote both in the ballot box and with their feet? In particular, answering this question allows us to reexamine the conventional wisdom (see [20]) claiming that geographic mobility limits the ability of government to redistribute incomes via a tax-transfer system.

To our knowledge, the answer is not yet well established. Indeed, the literature either assumes away asymmetric information (e.g., [21], [22]), restricts attention to flat tax (e.g., [22]) and special connections between skills and migration costs in a two-type setting that rules out countervailing incentives (e.g., [23]), or focuses on probabilistic voting in a representative democracy (e.g., [24]). These simplifications make it possible to obtain sharp predictions, whereas reasonable doubts about the generality and robustness of their predictions may arise.

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

To achieve our goal, we consider an economy consisting of two jurisdictions, between which workers born in each one can move by paying certain amount of migration cost. In each jurisdiction, workers differ in both skill level that measures their labor productivity and migration cost that measures their foot-voting capability. While the ex ante distribution of skill levels and the conditional distribution of migration costs are common knowledge, the values of any worker's labor productivity and migration cost are only known to herself. As usual, taxation is based on the residence principle[1], and we focus on the Nash competition induced by the mobile tax base between these two jurisdictions. Taking as given the income taxes implemented in both jurisdictions, workers make individual decisions along two margins: optimal labor supply on the intensive margin and optimal residence choice on the extensive margin. In particular, by allowing for location choice, both the reservation utilities of workers and the ex post skill distribution are endogenously determined. This feature, on one hand, enables us to design more realistic income tax schedules. On the other hand, it makes the qualitative analysis to be much less transparent.

Following [25], [26], [27] and [14], 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 that maximizes the utility of her own type. Then, the pairwise majority rule is used to select the tax schedule that is going to be implemented in equilibrium.[2] Importantly, each worker proposes an income tax schedule as if she were representing the government. Following the mechanism design approach, each proposer must take into account individual responses along both intensive and extensive margins and design incentive-compatible allocations satisfying the government budget constraint.

We rely on the first-order approach in the text and leave the complete solution to the tax design problem to Appendix B due to its complexity. By assuming quasilinear-in-consumption preferences, we show that the tax schedule proposed by the median skill type is a Condorcet winner in

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

[2] In a population with the majority consisting of "poor" individuals, Höchtl et al. [49] 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.

the majority rule equilibrium, which provides support for the empirical finding of [28]. They show by using survey data that most democracies implement the preferred redistribution of the median voter and also the probability to serve the median voter increases with the quality of democracy. Moreover, the tax rates implemented in the experiment of [50] closely track the preferences of the median skill worker, and the cross-national empirical evidence of [51] emphasizes the political channel as well as the middle class in determining the extent of redistribution.

The current tax schedule exhibits the following characteristics.

First, it coincides with the maximax tax schedule for types below the median skill level and coincides with the maximin tax schedule for types above the median skill level. We thus claim that governments under direct democracy and majority rule tend to redistribute from both the poor and the rich toward the middle class. This prediction not only extends the Director's Law (see [52]) to the circumstance with both labor mobility and inter-jurisdictional tax competition, but also provides theoretical support for the empirical finding of [53] that all Dutch political parties give a higher political weight to middle incomes than to the poor and the rich.

Second, marginal tax rates are negative for low incomes, whereas for higher incomes there is an endogenously determined and skill-dependent threshold of the elasticity of migration such that they are negative only when migration elasticity is higher than this threshold, otherwise they are nonnegative. To some extent, this result is consistent with our intuition about the feature of redistributive taxation. In addition, provided that income taxation in the United States is based on citizenship principle other than residence principle, the elasticities of migration for 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 higher incomes (see Congressional Budget Office [54]).

Third, it creates three potential discontinuities, one at the skill level of the proposer and the other two at the endpoints of skill distribution, in the resulting income schedule. In the voting equilibrium, there always exists a downward discontinuity at the median skill level. Importantly, we identify the indirect utility level of the least skilled as the determinant factor of the type of the other two discontinuities. If it lies between a negative threshold and zero, then both of them are downward. If it lies above zero, then one is upward at the lowest skill level and the other is downward at the highest skill level. If, however, it lies below the negative threshold, then one is downward at the lowest skill level and the other is upward at the highest skill level. We hence need to build more than one bridges when we iron the tax schedule to satisfy the monotonicity constraint placed on incomes, namely the second-order sufficient condition for incentive compatibility. As such, truth-telling allocation calls for bunching not only in the middle-income class, but also either in the low-income class or in the high-income class.

Fourth, by allowing for inter-jurisdictional migrations that endogenize the ex post skill distribution and median skill level, the resulting level of distortion and redistribution deviates from that under a single government. To be precise, we directly compare our tax schedule to that of [14] derived in autarky and establish the following predictions under the same median skill level. If migration elasticities are large for both low and high skills, then migrations induce lower marginal tax rates and hence smaller redistributions than does autarky. If, however, their migration elasticities are small, then the jurisdiction facing net labor inflow in low skills while net labor outflow in high skills imposes higher tax rates on both low and high skills. So, high skills pay more taxes while low skills receive less transfers than in autarky. If the median skill level is different between migration equilibrium and autarky equilibrium, then these predictions hold for all skills but those between these two median skill levels. The direction of change depends on which median voter is richer (or poorer).

To complement the theoretical analysis, the model is simulated with empirical parameter estimates based on the U.S. data. Since the literature suggests that tax-driven migrations be more likely to occur in the population of high skills (e.g., [44]) and the upper part of empirical income distribution be well fitted by Pareto distribution (e.g., [46]), our simulations focus on top-income workers. Following [55] and [29], we set the ex post Pareto index to be 1.5. For the ex ante Pareto index in autarky (without migrations), our numerical experiments allow it to have varying values around 1.5. The ex post and ex ante degrees of income inequality are not necessarily the same. To do counterfactual simulations, we manipulate the value of another parameter, which is an approximation of the elasticity of utility with respect to pre-tax income, in our tax formula such that it generates the current 42.5% top U.S. marginal tax rate (see [29]). Our calculation reveals that there can be large differences of marginal tax rate between our prediction and that in autarky. For example, for a labor supply elasticity of 0.25 (e.g., [48]), a migration elasticity of 0.25 (e.g., [1]) and an ex ante Pareto index of 1.5, the autarky equilibrium generates a top tax rate 34.4% higher than the implemented 42.5%. Everything else equal, if the labor supply elasticity increases to the highest possible estimate 0.4 (e.g., [48]), the autarky equilibrium still generates a top tax rate 27.5% higher than the implemented 42.5%. Naturally, these numbers of differences should only be interpreted within the current economic context and be taken as providing a rough assessment of the quantitative significance of migrations in terms of reducing equilibrium distortions imposed on top-income workers.

The rest of the paper is organized as follows. Section 3.2 discusses the relation of our paper to the literature. Section 3.3 sets up the model of the economy. Section 3.4 derives and characterizes selfishly optimal nonlinear income tax schedules. Section 3.5 establishes the voting equilibrium. Section 3.6 qualitatively and quantitatively identifies the effect of migrations on the equilibrium schedule of marginal tax rates. Section 3.7 concludes. Proofs are relegated to Appendix B.

3.2 Literature Review

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

First, it relates to the literature studying selfishly optimal nonlinear taxation determined by the majority rule,[3] such as [25], [26], [27] and [14]. While Röell [25] considers the case with a discrete skill distribution, the others use a continuum version of the Röell model. They all but [14] impose a minimum-utility constraint, but they all demonstrate that the schedule proposed by the median skill type is a Condorcet winner. By sacrificing the minimum-utility constraint and assuming quasilinear-in-consumption preferences, [14] realizes a complete characterization of

[3]Voting over selfishly optimal tax schedules has also been studied by [56] and [57], but the former study focuses on linear taxes and the latter study focuses on quadratic tax schedules.

selfishly optimal nonlinear tax schedules without calling for quite technical analysis.

By adding the location choice for workers, both the reservation utility of the standard participation constraint and the ex post skill distribution are endogenously determined. So, our extension makes the underlying tax design issue more realistic and indeed modifies the predictions obtained by [14] in three main ways.

Firstly, there tend to be more than one discontinuities in the tax schedule, and hence the ironing procedure used in deriving the complete solution seems to be much more involved. Secondly, marginal tax rates for higher incomes are not definitely positive, and they are rather negative for workers with sufficiently large elasticities of migration. And thirdly, migrations reduce distortions caused by marginal tax rates under a reasonable range of migration elasticities, and also the level of redistribution is likely to be either smaller or larger than suggested by [14], depending on whether the ex post median skill level lies at the right or at the left of the ex ante median skill level.

Second, it relates to the literature analyzing how the change of skill distribution affects equilibrium tax rates as well as the level of redistribution, such as [58], [23], [59], and [1]. No matter they use discrete or continuous skill distributions, they all follow the normative approach and focus on optimal income taxation. This paper follows the positive approach and stresses the fact that tax schedules are, directly or indirectly, chosen by self-interested voters in democracies. In this sense, it would be regarded as a complementary work to the literature.

3.3 The Model

We consider an economy consisting of two jurisdictions, called A and B. 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 she decides to relocate. In particular, if she faces an infinitely large migration cost, then she is immobile. Following [1], we do not make any restriction on the correlation between skills and migration costs. The simpler assumption is that migration costs and skill levels are independently distributed, as adopted by [41] and [42]. It, nevertheless, seems to be inconsistent with the empirical finding of [44] that migration rates increase in skill.

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 [19] and [1], 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 restrict our attention to Nash competition in which 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

$$y = wl, \tag{3.1}$$

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

$$c = y - T(y). \tag{3.2}$$

Preferences over consumption and labor supply are represented by the quasilinear-in-consumption utility function[4]

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

which is common to all workers with $\mathbb{I}$ being equal to 1 if she decides to migrate and to 0 otherwise.

[4]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., [60]).

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 (3.1), the utility function in terms of observable variables is written as

$$u(c, y; w, m) = c - h\left(\frac{y}{w}\right) - \mathbb{I} \cdot m. \tag{3.3}$$

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

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

3.3.1 Intensive Margin

If a worker decides to stay in jurisdiction A, then she maximizes (3.3) subject to $\mathbb{I} = 0$ and (3.2), yielding the first-order condition:

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

whenever T is differentiable. If it is not differentiable at some incomes, then the marginal tax rate is not well-defined. To avoid unnecessary technical issues, we follow [14] and directly define the function of marginal tax rate as

$$\tau(w) \equiv 1 - \frac{1}{w}h'\left(\frac{y(w)}{w}\right), \quad \forall w \in [\underline{w}, \overline{w}]. \tag{3.4}$$

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

We then define the resulting indirect utility as

$$U(w) \equiv c(w) - h\left(\frac{y(w)}{w}\right), \quad \forall w \in [\underline{w}, \overline{w}]. \tag{3.5}$$

Incentive compatibility requires that

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

The necessary condition is thus

$$U'(w) = h'\left(\frac{y(w)}{w}\right)\frac{y(w)}{w^2}, \ \ \forall w \in [\underline{w}, \overline{w}], \tag{3.6}$$

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

$$y'(w) \geq 0, \ \ \forall w \in [\underline{w}, \overline{w}]. \tag{3.7}$$

If constraint (3.7) does not bind, then the first-order approach is appropriate.

3.3.2 Migration Decision

For a worker of type (w, m) born in jurisdiction A, she will migrate to jurisdiction B if and only if $m < U_{-}(w) - U(w)$. As in [1], after combining the migration decisions made by workers born in both jurisdictions, the mass of residents of skill w in jurisdiction A can be written as:

$$\phi(\Delta(w); w) \equiv \begin{cases} f(w) + G_{-}(\Delta(w)|w) f_{-}(w) n_{-} & \text{for } \Delta(w) \geq 0, \\ (1 - G(-\Delta(w)|w)) f(w) & \text{for } \Delta(w) \leq 0 \end{cases} \tag{3.8}$$

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 [1], define the elasticity of migration as

$$\theta(\Delta(w); w) \equiv \frac{\partial \phi(\Delta(w); w)}{\partial \Delta(w)} \frac{c(w)}{\phi(\Delta(w); w)} \tag{3.9}$$

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

3.4 Selfishly Optimal Nonlinear Income Tax Schedules

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

$$\int_{\underline{w}}^{\overline{w}} [y(w) - c(w)]\tilde{f}(w)dw \geq 0, \tag{3.10}$$

where we have used (3.2). Provided the quasi-linearity in consumption, (3.10) must be binding. In particular, the participation constraint has been incorporated into this fiscal budget constraint through the term of ex post skill density $\tilde{f}$.

As in [27] and [14], each worker can propose an income tax schedule satisfying incentive compatibility constraints (3.6)-(3.7) and the government budget constraint (3.10), and then pairwise majority rule is used to determine which of these schedules shall be implemented. That is, each worker can be seen as a citizen candidate who may be elected as the representative agent of the government.

By applying the Taxation Principle[5] (see [61] and [62]) that allows us to restrict attention to simple direct mechanisms[6], for a worker of type $k \in [\underline{w}, \overline{w}]$, proposing an income tax schedule is equivalent to proposing an allocation schedule $\{c(w), y(w)\}_{w \in [\underline{w}, \overline{w}]}$ which solves the maximization problem

$$\max_{\{c(w),y(w)\}_{w \in [\underline{w}, \overline{w}]}} U(k) \quad \text{subject to } (3.5), (3.6), (3.7) \text{ and } (3.10). \tag{3.11}$$

By (3.11), the resulting allocation schedule is indeed *selfishly optimal* for the proposer.

We leave the complete solution to Appendix B, and here we rely on the first-order approach that considers a simpler while still useful case in which the SOIC condition (3.7) is ignored. Formally,

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

[6] If individual skills are drawn independently, [63] proves that the optimal sophisticated mechanism with strategic interdependence is a simple mechanism as long as individuals exhibit decreasing risk aversion.

problem (3.11) is relaxed as

$$\max_{\{c(w),y(w)\}_{w\in[\underline{w},\overline{w}]}} U(k) \quad \text{subject to } (3.5), (3.6) \text{ and } (3.10). \tag{3.12}$$

Solving problem (3.12) leads to the following theorem.

Theorem 3.4.1. *The selfishly optimal schedule of pre-tax incomes proposed by any worker of type* $k \in (\underline{w}, \overline{w})$ *is given by*

$$y(w) = \begin{cases} \underline{y}^M(w) & \textit{for } w = \underline{w}, \\ y^M(w) & \textit{for } w \in (\underline{w}, k), \\ y^R(w) & \textit{for } w \in (k, \overline{w}), \\ \overline{y}^R(w) & \textit{for } w = \overline{w}. \end{cases} \tag{3.13}$$

Proof. See Appendix B. □

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

By applying the formula (3.4) to Lemmas B.1.2 and B.1.3 stated in Appendix B and using Theorem 3.4.1, we summarize the resulting prediction as the second theorem.

Theorem 3.4.2. *The selfishly optimal income tax schedule proposed by any worker of type* $k \in (\underline{w}, \overline{w})$ *is given by*

$$\tau(w) = \begin{cases} \underline{\tau}^M(w) & \textit{for } w = \underline{w}, \\ \tau^M(w) & \textit{for } w \in (\underline{w}, k), \\ \tau^R(w) & \textit{for } w \in (k, \overline{w}), \\ \overline{\tau}^R(w) & \textit{for } w = \overline{w} \end{cases} \tag{3.14}$$

in which these marginal tax rates are explicitly given as

$$\underline{\tau}^M(\underline{w}) = -\frac{\partial \tilde{f}(\underline{w})}{\partial y(\underline{w})} \frac{1}{\tilde{f}(\underline{w})} \left[y(\underline{w}) - h\left(\frac{y(\underline{w})}{\underline{w}}\right) + \frac{y(\underline{w})}{\underline{w}} h'\left(\frac{y(\underline{w})}{\underline{w}}\right) - U(\underline{w}) \right], \tag{3.15}$$

$$\begin{aligned} \tau^M(w) = {} & \frac{\Gamma(w,\overline{w}) - \Gamma(\underline{w},\overline{w})}{w\tilde{f}(w)} \left[\frac{1}{w} h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^2} h''\left(\frac{y(w)}{w}\right) \right] \\ & - \frac{\partial \tilde{f}(w)}{\partial y(w)} \frac{1}{\tilde{f}(w)} \left[y(w) - h\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w} h'\left(\frac{y(w)}{w}\right) \right], \end{aligned} \tag{3.16}$$

$$\begin{aligned} \tau^R(w) = {} & \frac{\Gamma(w,\overline{w})}{w\tilde{f}(w)} \left[\frac{1}{w} h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^2} h''\left(\frac{y(w)}{w}\right) \right] \\ & - \frac{\partial \tilde{f}(w)}{\partial y(w)} \frac{1}{\tilde{f}(w)} \left[y(w) - h\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w} h'\left(\frac{y(w)}{w}\right) \right] \end{aligned} \tag{3.17}$$

and

$$\overline{\tau}^R(\overline{w}) = -\frac{\partial \tilde{f}(\overline{w})}{\partial y(\overline{w})} \frac{1}{\tilde{f}(\overline{w})} \left[y(\overline{w}) - h\left(\frac{y(\overline{w})}{\overline{w}}\right) - U(\underline{w}) \right], \tag{3.18}$$

with $\partial \tilde{f}(w)/\partial y(w)$ determined by equations (B.12) and (B.13) stated in Appendix B for $\forall w \in [\underline{w}, \overline{w}]$.

These marginal tax rates depart from those obtained by [14] in two important aspects. First, in addition to the discontinuity between the maximax tax schedule and the maximin tax schedule, we show that there may exist another two discontinuities: one at the lowest type within the maximax schedule and the other at the highest type within the maximin schedule. Second, as the ex post skill distribution is endogenously determined as a function of income and consumption, the migration decision along the extensive margin imposes non-trivial effects on these tax rates.

By using Theorem 3.4.2, we obtain the following two propositions.

Proposition 3.4.1. *Regarding the sign of these marginal tax rates given by (3.15)-(3.16), we have the following predictions.*

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

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

0*, of tax liability such that*

$$\underline{\tau}^M(\underline{w}) \begin{cases} < 0 & \textit{for } T^M(y(\underline{w})) > \widehat{T}^M(y(\underline{w})), \\ = 0 & \textit{for } T^M(y(\underline{w})) = \widehat{T}^M(y(\underline{w})), \\ > 0 & \textit{for } T^M(y(\underline{w})) < \widehat{T}^M(y(\underline{w})); \end{cases} \tag{3.19}$$

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

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

$$\underline{\tau}^M(\underline{w}) \begin{cases} < 0 & \textit{for } ATR^M(\underline{w}) > -1, \\ = 0 & \textit{for } ATR^M(\underline{w}) = -1, \\ > 0 & \textit{for } ATR^M(\underline{w}) < -1; \end{cases} \tag{3.20}$$

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

$$\underline{\tau}^M(\underline{w}) \begin{cases} < 0 & \textit{for } ATR^M(\underline{w}) < -1, \\ = 0 & \textit{for } ATR^M(\underline{w}) = -1, \\ > 0 & \textit{for } ATR^M(\underline{w}) > -1; \end{cases} \tag{3.21}$$

in which ATR denotes average tax rate.

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

Proof. See Appendix B. □

The sign of these maximax marginal tax rates departs from that of [14] as follows. Instead of showing that the marginal tax rate is always zero for the lowest skilled, we show that it is zero only when their tax liability is equal to a negative critical value or when their average tax rate is equal to -1. That is, it is more likely to be either strictly positive or strictly negative. In particular, if the elasticity of migration for the lowest skilled is small, it is negative when their average tax rate is relatively large and is positive when their average tax rate is relatively small, as shown in (3.20).

Proposition 3.4.2. *Regarding the sign of these marginal tax rates given by (3.17)-(3.18), we have the following predictions.*

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

$$\tau^R(w) \begin{cases} < 0 & \textit{for } \tilde{\theta}(w) > \tilde{\theta}_\star(w), \\ = 0 & \textit{for } \tilde{\theta}(w) = \tilde{\theta}_\star(w), \\ > 0 & \textit{for } \tilde{\theta}(w) < \tilde{\theta}_\star(w). \end{cases} \tag{3.22}$$

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

$$\tau^R(\overline{w}) \begin{cases} < 0 & \textit{for } T^R(y(\overline{w})) > U(\underline{w}) - U(\overline{w}), \\ = 0 & \textit{for } T^R(y(\overline{w})) = U(\underline{w}) - U(\overline{w}), \\ > 0 & \textit{for } T^R(y(\overline{w})) < U(\underline{w}) - U(\overline{w}), \end{cases} \tag{3.23}$$

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

Proof. See Appendix B. □

The sign of these maximin marginal tax rates departs from that of [14] in two ways. First, instead of showing that the marginal tax rate is always zero for the highest skilled, we show that it is zero only when their tax liability is equal to a negative critical value defined as the utility difference between the lowest skilled and the highest skilled. Second, instead of showing that the marginal tax rate is always positive for workers of types higher than the type k of the proposer, it is positive only when their elasticity of migration is smaller than an endogenously determined critical value.

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

By using Theorems 3.4.1 and 3.4.2, the following proposition is obtained.

Proposition 3.4.3. *For any proposer of type $k \in (\underline{w}, \overline{w})$, the proposed income schedule given by (3.13) may have the following discontinuities.*

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

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

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

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

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

Proof. See Appendix B. □

These discontinuities are graphically shown in Figure 3.1.

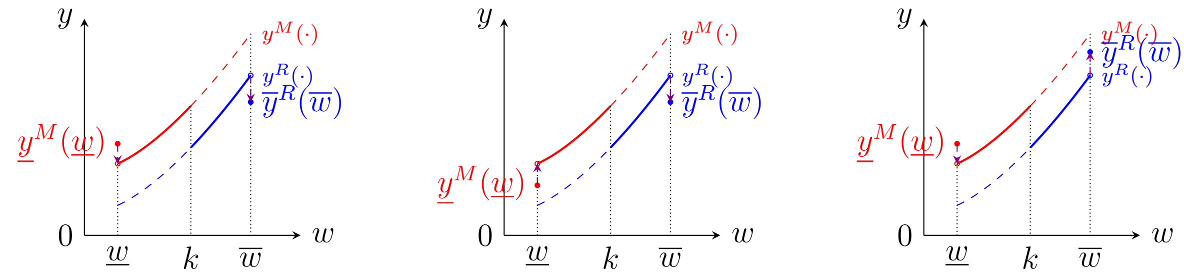

Figure 3.1: Graphic Illustration of Proposition 3.4.3

In an economy with an exogenous skill distribution, the tax schedule of [27] and [14] only exhibits the downward discontinuity at the type of the proposer, as shown in part (ii). In the current economy with an endogenous skill distribution, we show that there may be another two downward discontinuities at the lowest type and the highest type. Importantly, we identify the utility of the lowest skilled as the key variable determining whether or not these two discontinuities are indeed downward. In order to iron this tax schedule, as shown in Appendix B, it is useful to know when the SOIC condition is violated at the endpoints of the ex post skill distribution.

3.5 The Voting Equilibrium

Following one common practice in literature, majority rule is used to select the income tax schedule that shall be implemented. Each worker is assumed to have one vote. As argued by Roberts [64], if political parties in a democratic system choose the income tax schedule to maximize the likelihood of being elected then it is somewhat reasonable to view the chosen tax schedule as being determined, albeit indirectly, by a *pairwise majority voting* process. In each round, workers vote over two arbitrarily-selected alternatives. The one that survives all rounds is hence the winner.

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

$$U(w,k) = c(w,k) - h\left(\frac{y(w,k)}{w}\right). \tag{3.24}$$

We then have the following theorem.

Theorem 3.5.1. *The selfishly optimal income tax schedule for the median skill type is a Condorcet winner when pairwise majority voting is restricted to the income tax schedules that are selfishly optimal for some skill type.*

Proof. See Appendix B. □

We hence establish the existence of a Condorcet winner in the current political economy. In particular, as there is a continuum of tax schedules in our problem, the single-crossing condition used by [65] 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 [66] Median Voter Theorem.

3.6 Identifying the Effect of Migrations on Equilibrium Tax Schedule

3.6.1 Qualitative Characterization

In what follows, we let w_m denote the median skill level of the ex ante distribution $F(w)$, and let $\tilde{w}_m$ denote that of the ex post distribution $\Gamma(\underline{w}, w) = \int_{\underline{w}}^{w} \tilde{f}(t)dt$. To identify the effects of migrations placed on marginal tax rates, we shall compare our marginal tax rates to those of [14] derived in autarky. As is clear soon, migrations affect marginal tax rates through endogenizing the skill distribution that is a key part of the tax formula and is the determinant of the median skill level. So, migrations affect both the distortion level and the redistribution scale.

We summarize our main findings as two propositions.

Proposition 3.6.1. *Suppose*

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

and

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

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

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

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

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

Proof. See Appendix B and Figure 3.2.[7] □

Intuition: If migration elasticities of high skills are reasonably large, then their migration threat is strong enough to motivate the government to impose lower tax rates on them relative to the autarky in which migrations and tax competition are forbidden. Since high types face lower tax rates and also low types have reasonably large migration elasticities, the incentives for low types to mimic high types become stronger, which implies that the government needs to transfer more to them to prevent them from mimicking. This analysis explains why migrations under these conditions generally induce lower tax rates than suggested by the autarky equilibrium for both low and high skills. In addition, for types belonging to $(\tilde{w}_m, w_m]$ in case (ii) of Proposition 3.6.1,

[7] In these figures, $\hat{y}^R(\cdot)$ and $\hat{y}^M(\cdot)$ denote the maximin and maximax income schedules derived by [14].

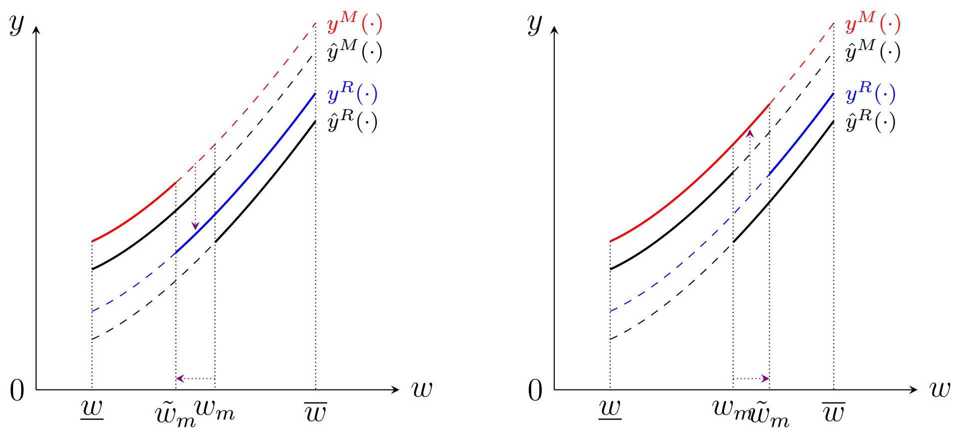

Figure 3.2: Graphic Illustration of Proposition 3.6.1

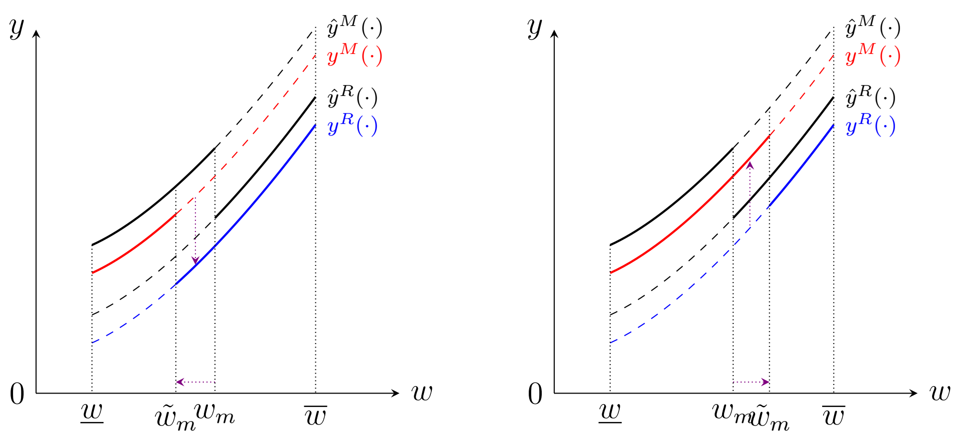

Figure 3.3: Graphic Illustration of Proposition 3.6.2

they face higher tax rates than in autarky because the median voter under migrations is poorer. They belong to the low-income class in autarky while belonging to the high-income class under migrations, so their status changes from receiving transfers to paying taxes.

Proposition 3.6.2. *Suppose*

$$\begin{cases} \tilde{\theta}(w) < \Theta^M(w) \text{ and } f(w)\Gamma(\underline{w}, w) < \tilde{f}(w)F(w) & \text{for } \forall w \in (\underline{w}, w_m], \\ \tilde{\theta}(w) < \Theta^R(w) \text{ and } f(w)[\Gamma(\underline{w}, \overline{w}) - \Gamma(\underline{w}, w)] > \tilde{f}(w)[1 - F(w)] & \text{for } \forall w \in (w_m, \overline{w}), \end{cases} \tag{3.25}$$

then we have the following predictions.

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

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

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

Proof. See Appendix B and Figure 3.3. □

Intuition: Assumption (3.25) provides three conditions: (a) migration elasticity is small for both low and high skills; (b) the jurisdiction under consideration faces net labor inflow in low skills but (c) net labor outflow in high skills. Condition (c) means that tax base shrinks relative to the autarky, so marginal tax rates imposed on remaining high skills must be higher than those in autarky, which is somehow guaranteed by condition (a). As now high skills face higher tax rates, low skills have weaker incentives to mimic them, which implies that, everything else equal, the information rents paid to low skills can be smaller. Also, for the same amount of transfers, condition (b) implies that each low-skill worker receives less than they may receive in autarky. As such, both high and low skills face higher marginal tax rates than in autarky. For workers between ex ante and ex post median skill levels, the reasoning used in analyzing Proposition 3.6.1 still applies.

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

3.6.2 Quantitative Simulation

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

It has been argued by [46] and [55] that Pareto distribution fits the empirical income distribution at high income levels reasonably well. Assuming un-truncated distributions, namely $\overline{w} = \infty$, with ex ante Pareto index $a > 1$ and ex post Pareto index $\tilde{a} > 1$, we can thus obtain

$$\frac{1 - F(w)}{wf(w)} = \frac{1}{a} \text{ and } \frac{\Gamma(w, \overline{w})}{w\tilde{f}(w)} = \frac{1}{\tilde{a}}, \tag{3.26}$$

which measure, respectively, the ex ante and the ex post degrees of income inequality. By using U.S. tax return micro data for 2005, Diamond and Saez [29] find that the empirical Pareto coefficient is approximately a constant around 1.5 for adjusted gross incomes higher than \$200,000, we hence set $\tilde{a} = 1.5$. Following [45], we assume that $h(l)$ takes the isoelastic form

$$h(l) = l^{1+\frac{1}{\varepsilon}} \Big/ \left(1 + \frac{1}{\varepsilon}\right) \tag{3.27}$$

for constant elasticity $\varepsilon > 0$. Saez et al. [48] survey the recent literature and conclude that the best available estimates range from 0.12 to 0.4 in the U.S. As usual, we take a central value of $\varepsilon = 0.25$ for the benchmark exercise. In addition, we, without loss of generality, normalize the upper bound of l to be 1 and focus on interior values, namely $l \in (0, 1)$. Following [1], we set the elasticity of migration for top-income workers to be $\tilde{\theta}(w) = 0.25$.

By applying (3.26) and (3.27), we have the formula of the difference between marginal tax rates as

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

in which

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

can be interpreted as an approximation of the elasticity of utility with respect to pre-tax income under quasilinear-in-consumption preferences. To make our tax formula generate the current 42.5% top U.S. marginal tax rate (see [29]), we consider two cases:

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

We give our findings under varying values of a in Table 3.1. As is obvious, there can be large differences of marginal tax rates between the current prediction and that in autarky.

Table 3.1: The δ_{MTR} (%) under Cases (a) and (b)

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

3.7 Conclusion

In this paper we have examined the feature of redistributive taxation when voters are geographically mobile at the expense of some unobservable 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. For simplicity, we restrict attention to Nash tax competition between these two jurisdictions. We have established the voting equilibrium under the majority rule and have fully characterized the equilibrium income tax schedule that would enact the wishes of median voters. The resulting redistributive policy highlights a dynamic interaction between voting-with-hand and voting-with-feet, which hence makes the level of distortion and redistribution tend to deviate from that occurs in autarky.

Given what we have established, the goal of this paper is achieved. For future research, our model can be modified or extended in at least two directions. First, as both quasilinear-in-consumption and quasilinear-in-labor preferences are used in taxation literature, a parallel analysis would be conducted under the second type of preferences, and potentially different implications for redistributive taxation would be worthwhile investigating. Second, instead of being completely motivated by self-interest, it is natural to expect voters of certain skills to exhibit other-regarding or pro-social preferences. Incorporating this concern into our model might make the resulting tax schedule a better approximation to reality.

4. SUMMARY AND CONCLUSIONS

We have conducted a theoretical analysis with some numerical illustrations to examine redistributive labor-income taxation policies. We follow the mechanism design approach, and emphasize the features of endogenous outside options facing heterogeneous workers as well as the endogenous interaction between tax rate and tax base.

The first essay develops a framework to analyze how the interplay of relative consumption concern and income inequality determines optimal redistributive tax policies in an international setting with two competing countries. We establish and characterize nonlinear labor income tax schedules that competing Rawlsian governments should implement when workers having private information on skills and migration costs decide where to live and how much to work. We consider both Nash and Stackelberg tax competition, and obtain four main results. First, we establish an optimal tax formula that generalizes existing ones obtained under maximin social objective and additively separable preferences. Second, we numerically calculate optimal asymptotic marginal tax rates under both types of equilibrium and compare them to those obtained in an autarky equilibrium, finding that the country with large labor inflow implements a much smaller marginal tax rate and the country with large labor outflow implements a much higher marginal tax rate than suggested by autarky-equilibrium predictions. Third, for the case with quasilinear-in-consumption preferences and jealousy-type consumption comparison, we show that the leader country implements a higher marginal tax rate in Stackelberg equilibrium than that it may implement in Nash equilibrium. And fourth, we provide a complete characterization on how relativity and inequality together determine the optimal top-tax rate under both types of tax competition, finding that both the elasticity and level of migration are determinant for predicting when relativity and inequality are complementary (or substitutive) in shaping the optimal tax rates placed on top-income workers.

The second essay investigates the feature of redistributive taxation when voters are heterogeneous in innate abilities and are geographically mobile at the expense of some unobservable migration costs. We consider two jurisdictions and restrict attention to Nash tax competition. We have established the voting equilibrium under the majority rule and have fully characterized the equilibrium income tax schedule that would enact the wishes of median voters. The resulting redistributive policy highlights a dynamic interaction between voting-with-hand and voting-with-feet, which hence makes the level of distortion and redistribution tend to deviate from that occurs in autarky. We extend the literature by identifying endogenous and skill-dependent bounds of migration elasticity based on which we predict that migrations induce lower marginal tax rates than suggested by an autarky equilibrium if workers' migration elasticities are bigger than their respective bound. Also, we depart from the conventional wisdom by predicting for the jurisdiction with net labor inflow in low skills while with net labor outflow in high skills that migrations induce higher marginal tax rates than suggested by an autarky equilibrium if workers' migration elasticities are smaller than their respective bound.

As a final remark of the second essay, we, without resorting to a benevolent social planner and just assuming that each voter is motivated by self-interest, still obtain an equilibrium tax schedule that achieves a desirable balance between equity and efficiency. We argue that private information and foot-voting possibility are key mechanisms leading to this prediction. For low skills, the median voter (or government) transfers a positive amount of resources to them to induce them to tell truth in equilibrium and hence minimize the informational rents; for high skills, especially those with high foot-voting capabilities, the median voter will not tax them that much as compared to the case in which they cannot migrate in order to restore a desirable tax base for reallocation. As such, even the median voter benefits the most from this tax schedule, both equity and efficiency concerns are taken into account seriously under this kind of institutional arrangement. In sum, it is the "majority-rule democracy" combined with "relevant private information" and "migration freedom" of heterogeneous voters that lead to a desirable balance between equity and efficiency among a population of completely self-interested voters.

REFERENCES

[1] E. Lehmann, L. Simula, and A. Trannoy, "Tax me if you can! optimal nonlinear income tax between competing governments," *The Quarterly Journal of Economics*, vol. 129, no. 4, pp. 1995–2030, 2014.

[2] D. Dai, W. Gao, and G. Tian, "Relativity, mobility and optimal nonlinear income taxation in an open economy," *Working Paper*, 2017.

[3] H. J. Kleven, C. Landais, and E. Saez, "Taxation and international migration of superstars: Evidence from the european football market," *American Economic Review*, vol. 103, no. 5, pp. 1892–1924, 2013.

[4] H. J. Kleven, C. Landais, E. Saez, and E. Schultz, "Migration and wage effects of taxing top earners: Evidence from the foreigners' tax scheme in denmark," *The Quarterly Journal of Economics*, vol. 129, no. 1, pp. 333–378, 2014.

[5] U. Akcigit, S. Baslandze, and S. Stantcheva, "Taxation and the international mobility of inventors," *American Economic Review*, vol. 106, no. 10, pp. 2930–2981, 2016.

[6] L. Becchetti, S. Castriota, L. Corrado, and E. G. Ricca, "Beyond the joneses: Inter-country income comparisons and happiness," *The Journal of Socio-Economics*, vol. 45, no. C, pp. 187–195, 2013.

[7] T. Veblen, *The theory of the leisure class: An economic study in the evolution of institutions*. Macmillan, 1899.

[8] R. Kanbur and M. Tuomala, "Relativity, inequality and optimal nonlinear income taxation," *International Economic Review*, vol. 54, no. 4, pp. 1199–1217, 2013.

[9] J. S. Duesenberry, *Income, saving and the theory of consumer behaviour*. Harvard University Press, 1949.

[10] F. Alpizar, F. Carlsson, and O. Johansson-Stenman, "How much do we care about absolute versus relative income and consumption?," *Journal of Economic Behavior and Organization*, vol. 56, no. 3, pp. 405–421, 2005.

[11] E. F. P. Luttmer, "Neighbors as negatives: Relative earnings and well-being," *The Quarterly Journal of Economics*, vol. 120, no. 3, pp. 963–1002, 2005.

[12] F. W. Andersson, "Is concern for relative consumption a function of relative consumption?," *Journal of Socio-Economics*, vol. 37, no. 1, pp. 353–364, 2008.

[13] A. Clark, P. Frijters, and M. A. Shields, "Relative income, happiness, and utility: An explanation for the easterlin paradox and other puzzles," *Journal of Economic Literature*, vol. 46, no. 1, pp. 95–144, 2008.

[14] C. Brett and J. A. Weymark, "Voting over selfishly optimal nonlinear income tax schedules," *Games and Economic Behavior*, vol. 101, no. C, pp. 172–188, 2017.

[15] J. A. Mirrlees, "Migration and optimal income taxes," *Journal of Public Economics*, vol. 18, no. 3, pp. 319–341, 1982.

[16] L. Simula and A. Trannoy, "Optimal income tax under the threat of migration by top-income earners," *Journal of Public Economics*, vol. 94, no. 1-2, pp. 163–173, 2010.

[17] L. Simula and A. Trannoy, "Shall we keep the highly skilled at home? the optimal income tax perspective," *Social Choice and Welfare*, vol. 39, no. 4, pp. 751–782, 2012.

[18] T. Piketty and E. Saez, "Optimal labor income taxation," *Handbook of Public Economics*, vol. 5, no. 7, pp. 391–474, 2013.

[19] J. A. Mirrlees, "An exploration in the theory of optimum income taxation," *The Review of Economic Studies*, vol. 38, no. 2, pp. 175–208, 1971.

[20] G. J. Stigler, "The tenable range of functions of local government," *Federal Expenditure Policy for Economic Growth and Stability: Papers Submitted by Panelists Appearing Before the Subcommittee on Fiscal Policy*, 1957.

[21] H. Cremer and P. Pestieau, "Social insurance, majority voting and labor mobility," *Journal of Public Economics*, vol. 68, no. 3, pp. 397–420, 1998.

[22] J. Hindriks, "Mobility and redistributive politics," *Journal of Public Economic Theory*, vol. 3, no. 1, pp. 95–120, 2001.

[23] J. Hamilton and P. Pestieau, "Optimal income taxation and the ability distribution: Implications for migration equilibria," *International Tax and Public Finance*, vol. 12, no. 1, pp. 29–45, 2005.

[24] C. Bretta, *Probabilistic voting over nonlinear income taxes with international migration*. Department of Economics, Mount Allison University, 2016.

[25] A. A. Röell, "Voting over nonlinear income tax schedules," *Working Paper*, 2012.

[26] H. Bohn and C. Stuart, "Revenue extraction by median voters," *Working Paper*, 2013.

[27] C. Brett and J. A. Weymark, "Voting over selfishly optimal nonlinear income tax schedules with a minimum-utility constraint," *Journal of Mathematical Economics*, vol. 67, no. C, pp. 18–31, 2016.

[28] G. Corneo and F. Neher, "Democratic redistribution and rule of the majority," *European Journal of Political Economy*, vol. 40, no. PA, pp. 96–109, 2015.

[29] P. Diamond and E. Saez, "The case for a progressive tax: From basic research to policy recommendations," *Journal of Economic Perspectives*, vol. 25, no. 4, pp. 165–190, 2011.

[30] T. Piketty, *Capital in the twenty-first century*. Belknap Press, Cambridge, MA, 2014.

[31] F. Alvarez-Cuadrado and N. V. Long, "Envy and inequality," *The Scandinavian Journal of Economics*, vol. 114, no. 3, pp. 949–973, 2012.

[32] T. Aronsson and O. Johansson-Stenman, "Positional preferences in time and space: Optimal income taxation with dynamic social comparisons," *Journal of Economic Behavior and Organization*, vol. 101, no. C, pp. 1–23, 2014.

[33] V. Lipatov and A. Weichenrieder, “Welfare and labor supply implications of tax competition for mobile labor,” *Social Choice and Welfare*, vol. 45, no. 2, pp. 457–477, 2015.

[34] C. Fong, “Social preferences, self-interest, and the demand for redistribution,” *Journal of Public Economics*, vol. 82, no. 2, pp. 225–246, 2001.

[35] E. Fehr and K. M. Schmidt, “The economics of fairness, reciprocity and altruism-experimental evidence and new theories,” *Handbook of the Economics of Giving, Altruism and Reciprocity*, vol. 1, no. 8, pp. 615–691, 2006.

[36] O. Heffetz and R. H. Frank, “Preferences for status: Evidence and economic implications,” *Handbook of Social Economics*, vol. 1, no. 3, pp. 69–91, 2011.

[37] T. Aronsson and O. Johansson-Stenman, “Keeping up with the joneses, the smiths and the tanakas: On international tax coordination and social comparisons,” *Journal of Public Economics*, vol. 131, no. C, pp. 71–86, 2015.

[38] R. Boadway and L. Jacquet, “Optimal marginal and average income taxation under maximin,” *Journal of Economic Theory*, vol. 143, no. 1, pp. 425–441, 2008.

[39] J. Rawls, *A theory of justice*. Harvard University Press, 1971.

[40] A. J. Oswald, “Altruism, jealousy and the theory of optimal non-linear taxation,” *Journal of Public Economics*, vol. 20, no. 1, pp. 77–87, 1983.

[41] F. Bierbrauer, C. Brett, and J. A. Weymark, “Strategic nonlinear income tax competition with perfect labor mobility,” *Games and Economic Behavior*, vol. 82, no. C, pp. 292–311, 2013.

[42] T. Blumkin, E. Sadka, and Y. Shem-Tov, “International tax competition: Zero tax rate at the top re-established,” *International Tax and Public Finance*, vol. 22, no. 5, pp. 760–776, 2015.

[43] G. Gereffi and X. Luo, “Risks and opportunities of participation in global value chains (english),” *Policy Research Working Paper*, vol. 1, no. WPS6847, 2014.

[44] F. Docquier and A. Marfouk, “International migration by educational attainment (1990-2000),” *International Migration, Remittances and Development*, 2006.

[45] P. A. Diamond, “Optimal income taxation: An example with a u-shaped pattern of optimal marginal tax rates,” *The American Economic Review*, vol. 88, no. 1, pp. 83–95, 1998.

[46] E. Saez, “Using elasticities to derive optimal income tax rates,” *Review of Economic Studies*, vol. 68, no. 1, pp. 205–229, 2001.

[47] L. Jacquet, E. Lehmann, and B. Van der Linden, “Optimal redistributive taxation with both extensive and intensive responses,” *Journal of Economic Theory*, vol. 148, no. 5, pp. 1770–1805, 2013.

[48] E. Saez, J. Slemrod, and S. H. Geertz, “The elasticity of taxable income with respect to marginal tax rates: A critical review,” *Journal of Economic Literature*, vol. 50, no. 1, pp. 3–50, 2012.

[49] W. Höchtl, R. Sausgruber, and J.-R. Tyran, “Inequality aversion and voting on redistribution,” *European Economic Review*, vol. 56, no. 7, pp. 1406–1421, 2012.

[50] M. Agranov and T. R. Palfrey, “Equilibrium tax rates and income redistribution: A laboratory study,” *Journal of Public Economics*, vol. 130, no. C, pp. 45–58, 2015.

[51] K. Gründler and S. Köllner, “Determinants of governmental redistribution: Income distribution, development levels, and the role of perceptions,” *Journal of Comparative Economics*, vol. 45, no. 4, pp. 930–962, 2017.

[52] G. J. Stigler, “Director’s law of public income redistribution,” *The Journal of Law & Economics*, vol. 13, no. 1, pp. 1–10, 1970.

[53] B. Jacobs, E. L. W. Jongen, and F. T. Zoutman, “Revealed social preferences of dutch political parties,” *Journal of Public Economics*, vol. 156, no. C, pp. 81–100, 2017.

[54] C. B. Office, “Effective marginal tax rates for low- and moderate-income workers,” no. 4149, 2012.

[55] A. B. Atkinson, T. Piketty, and E. Saez, “Top incomes in the long run of history,” *Journal of Economic Literature*, vol. 49, no. 1, pp. 3–71, 2011.

[56] A. H. Meltzer and S. F. Richard, “A rational theory of the size of government,” *Journal of Political Economy*, vol. 89, no. 5, pp. 914–927, 1981.

[57] J. Hindriks and P. De Donder, “The politics of redistributive social insurance,” *Journal of Public Economics*, vol. 87, no. 12, pp. 2639–2660, 2003.

[58] M. Leite-Monteiro, “Redistributive policy with labour mobility across countries,” *Journal of Public Economics*, vol. 65, no. 2, pp. 229–244, 1997.

[59] C. Brett and J. A. Weymark, “How optimal nonlinear income taxes change when the distribution of the population changes,” *Journal of Public Economics*, vol. 95, no. 11, pp. 1239–1247, 2011.

[60] J. Gruber and E. Saez, “The elasticity of taxable income: Evidence and implications,” *Journal of Public Economics*, vol. 84, no. 1, pp. 1–32, 2002.

[61] P. J. Hammond, “Straightforward individual incentive compatibility in large economies,” *The Review of Economic Studies*, vol. 46, no. 2, pp. 263–282, 1979.

[62] R. Guesnerie, *A contribution to the pure theory of taxation*. Cambridge University Press, Cambridge, 1995.

[63] F. J. Bierbrauer, “On the optimality of optimal income taxation,” *Journal of Economic Theory*, vol. 146, no. 5, pp. 2105–2116, 2011.

[64] K. W. S. Roberts, “Voting over income tax schedules,” *Journal of Public Economics*, vol. 8, no. 3, pp. 329–340, 1977.

[65] J. S. Gans and M. Smart, “Majority voting with single-crossing preferences,” *Journal of Public Economics*, vol. 59, no. 2, pp. 219–237, 1996.

[66] D. Black, “On the rationale of group decision-making,” *Journal of Political Economy*, vol. 56, no. 1, pp. 23–34, 1948.

[67] D. M. Topkis, “Minimizing a submodular function on a lattice,” *Operations Research*, vol. 26, no. 2, pp. 305–321, 1978.

APPENDIX A

RELATIVITY, MOBILITY, AND OPTIMAL NONLINEAR INCOME TAXATION IN AN OPEN ECONOMY

A.1 Proofs

Proof of Theorem 2.3.1: We shall complete the proof in 3 steps.

Step 1. Given the FOC (2.4) of individual choice, the indirect utility of a type-w worker in country $i \in \{A, B\}$ can be written as

$$U_i(w) = v(\varphi_i(U_i(w), l_i(w), \mu_i, \mu_{-i})) - h(l_i(w)) + \psi(\mu_i, \mu_{-i}), \tag{A.1}$$

where we treat individual consumption $c_i(w)$ as an implicit function of $U_i(w)$, $l_i(w)$, μ_i, μ_{-i}, and equivalently rewrite it as $\varphi_i(\cdot)$. By applying the Implicit Function Theorem, we get from (A.1) that

$$\frac{\partial \varphi_i}{\partial l_i} = \frac{h'(l_i(w))}{v'(c_i(w))}, \; \frac{\partial \varphi_i}{\partial U_i} = \frac{1}{v'(c_i(w))}, \; \frac{\partial \varphi_i}{\partial \mu_i} = -\frac{\psi_i(\mu_i, \mu_{-i})}{v'(c_i(w))} \text{ and } \frac{\partial \varphi_i}{\partial \mu_{-i}} = -\frac{\psi_{-i}(\mu_i, \mu_{-i})}{v'(c_i(w))}. \tag{A.2}$$

Step 2. For expositional purposes, we follow the first-order approach and ignore the SOIC conditions. After deriving the solutions, then we can verify whether the SOIC conditions are binding or not. The corresponding Lagrangian is written as follows:

$$\begin{aligned}
&\mathcal{L}_i(\{U_i(w), l_i(w)\}_{w\in[\underline{w},\overline{w}]}, \mu_i; \lambda_i, \gamma_i, \{\varsigma_i(w)\}_{w\in[\underline{w},\overline{w}]}) \\
= \;& U_i(\underline{w}) + \lambda_i \int_{\underline{w}}^{\overline{w}} \left\{ [wl_i(w) - \varphi_i(U_i(w), l_i(w), \mu_i, \mu_{-i})]\phi_i(U_i(w) - U_{-i}(w); w) - \frac{R}{\overline{w} - \underline{w}} \right\} dw \\
&+ \int_{\underline{w}}^{\overline{w}} \varsigma_i(w) \left[h'(l_i(w)) \frac{l_i(w)}{w} - \dot{U}_i(w) \right] dw \\
&+ \gamma_i \left[\mu_i - \int_{\underline{w}}^{\overline{w}} \varphi_i(U_i(w), l_i(w), \mu_i, \mu_{-i}) f_i(w) dw \right]
\end{aligned} \tag{A.3}$$

where $\lambda_i > 0$ is the multiplier associated with the binding budget constraint (2.8), $\varsigma_i(w)$ is the multiplier associated with the FOIC conditions (2.10), and γ_i is the multiplier associated with the comparison consumption constraint (2.2). Integrating by parts, we obtain

$$\int_{\underline{w}}^{\overline{w}} \varsigma_i(w)\dot{U}_i(w)dw = \varsigma_i(\overline{w})U_i(\overline{w}) - \varsigma_i(\underline{w})U_i(\underline{w}) - \int_{\underline{w}}^{\overline{w}} \dot{\varsigma}_i(w)U_i(w)dw. \tag{A.4}$$

Plugging (A.4) in (A.3) gives rise to

$$\begin{aligned}
&\mathcal{L}_i(\{U_i(w), l_i(w)\}_{w\in[\underline{w},\overline{w}]}, \mu_i; \lambda_i, \gamma_i, \{\varsigma_i(w)\}_{w\in[\underline{w},\overline{w}]}) \\
=\ & U_i(\underline{w}) + \lambda_i \int_{\underline{w}}^{\overline{w}} \left\{ [wl_i(w) - \varphi_i(U_i(w), l_i(w), \mu_i, \mu_{-i})]\phi_i(U_i(w) - U_{-i}(w); w) - \frac{R}{\overline{w} - \underline{w}} \right\} dw \\
& + \varsigma_i(\underline{w})U_i(\underline{w}) - \varsigma_i(\overline{w})U_i(\overline{w}) + \int_{\underline{w}}^{\overline{w}} \left[\varsigma_i(w)h'(l_i(w))\frac{l_i(w)}{w} + \dot{\varsigma}_i(w)U_i(w) \right] dw \\
& + \gamma_i \left[\mu_i - \int_{\underline{w}}^{\overline{w}} \varphi_i(U_i(w), l_i(w), \mu_i, \mu_{-i}) f_i(w)dw \right].
\end{aligned}$$

Assuming that there is no bunching of workers of different skills and the existence of an interior solution, applying (A.2) shows that the necessary conditions can be written as follows:

$$\begin{aligned}
\frac{\partial \mathcal{L}_i}{\partial l_i(w)} = \lambda_i \left[w - \frac{h'(l_i(w))}{v'(c_i(w))} \right] \tilde{f}_i(w) - \gamma_i \frac{h'(l_i(w))}{v'(c_i(w))} f_i(w) \\
+ \frac{\varsigma_i(w)h'(l_i(w))}{w} \left[1 + \frac{l_i(w)h''(l_i(w))}{h'(l_i(w))} \right] = 0 \ \ \forall w \in [\underline{w}, \overline{w}],
\end{aligned} \tag{A.5}$$

$$\begin{aligned}
\frac{\partial \mathcal{L}_i}{\partial U_i(w)} = -\frac{\lambda_i \tilde{f}_i(w)}{v'(c_i(w))} + \lambda_i T_i(y_i(w))\tilde{\eta}_i(w)\tilde{f}_i(w) \\
- \frac{\gamma_i f_i(w)}{v'(c_i(w))} + \dot{\varsigma}_i(w) = 0 \ \ \forall w \in (\underline{w}, \overline{w}),
\end{aligned} \tag{A.6}$$

$$\frac{\partial \mathcal{L}_i}{\partial U_i(\underline{w})} = 1 + \varsigma_i(\underline{w}) = 0, \tag{A.7}$$

$$\frac{\partial \mathcal{L}_i}{\partial U_i(\overline{w})} = -\varsigma_i(\overline{w}) = 0, \tag{A.8}$$

$$\frac{\partial \mathcal{L}_i}{\partial \mu_i} = \lambda_i \int_{\underline{w}}^{\overline{w}} \frac{\psi_i(\mu_i, \mu_{-i})}{v'(c_i(w))} \tilde{f}_i(w)dw + \gamma_i \left[1 + \int_{\underline{w}}^{\overline{w}} \frac{\psi_i(\mu_i, \mu_{-i})}{v'(c_i(w))} f_i(w)dw \right] = 0. \tag{A.9}$$

Using (A.6), we get

$$\frac{\dot{\varsigma}_i(w)}{\lambda_i} = \frac{\gamma_i}{\lambda_i}\frac{f_i(w)}{v'(c_i(w))} + \frac{\tilde{f}_i(w)}{v'(c_i(w))} - T_i(y_i(w))\tilde{\eta}_i(w)\tilde{f}_i(w).$$

Integrating on both sides of this equation and using the transversality condition (A.8), we obtain

$$-\frac{\varsigma_i(w)}{\lambda_i} = \frac{\gamma_i}{\lambda_i}\int_w^{\overline{w}} \frac{f_i(t)}{v'(c_i(t))}dt + \int_w^{\overline{w}} \left[\frac{1}{v'(c_i(t))} - T_i(y_i(t))\tilde{\eta}_i(t)\right]\tilde{f}_i(t)dt. \tag{A.10}$$

Rearranging (A.5) via using FOC (2.4), we have

$$\frac{T_i'(y_i(w))}{1-T_i'(y_i(w))} = \frac{\gamma_i}{\lambda_i}\frac{f_i(w)}{\tilde{f}_i(w)} - \frac{\varsigma_i(w)}{\lambda_i}\frac{v'(c_i(w))}{w\tilde{f}_i(w)}\left[1+\frac{l_i(w)h''(l_i(w))}{h'(l_i(w))}\right] \ \ \forall w \in [\underline{w}, \overline{w}]. \tag{A.11}$$

Substituting (A.10) into (A.11) gives the first-order conditions characterizing the optimum marginal tax rates, with γ_i/λ_i determined by solving (A.9).

Step 3. To derive a sufficient condition for the optimal marginal tax profile to satisfy the SOIC conditions, we rewrite the FOC (2.4) as

$$\underbrace{\frac{v'(c_i(w))}{h'(y_i(w)/w)}}_{\text{LHS}} = \underbrace{\frac{1}{w[1-T_i'(y_i(w))]}}_{\text{RHS}}. \tag{A.12}$$

Noting that

$$\frac{\text{dLHS}}{\text{d}w} = \frac{v''(c_i(w))\dot{c}_i(w)}{h'(y_i(w)/w)} - \frac{v'(c_i(w))h''(y_i(w)/w)[w\dot{y}_i(w)-y_i(w)]}{[wh'(y_i(w)/w)]^2}$$

and $c_i(w) = y_i(w) - T_i(y_i(w)) \Rightarrow \dot{c}_i(w) = \dot{y}_i(w)[1-T_i'(y_i(w))]$, thus

$$\frac{\text{dLHS}}{\text{d}w} < 0 \implies \dot{y}_i(w) > 0. \tag{A.13}$$

Also, noting that

$$\frac{\mathrm{dRHS}}{\mathrm{d}w} = -\left\{w\left[1 - T_i'(y_i(w))\right]\right\}^2 \left[1 - T_i'(y_i(w)) - w\frac{\mathrm{d}T_i'(y_i(w))}{\mathrm{d}w}\right],$$

we thus arrive at

$$\frac{\mathrm{d}T_i'(y_i(w))}{\mathrm{d}w} \leq 0 \implies \frac{\mathrm{dRHS}}{\mathrm{d}w} < 0. \tag{A.14}$$

Therefore, (A.12) combined with (A.13) and (A.14) implies that

$$\frac{\mathrm{d}T_i'(y_i(w))}{\mathrm{d}w} \leq 0 \implies \dot{y}_i(w) > 0,$$

as desired. **QED**

Proof of Lemma 2.3.1: The Lagrangian of government i's problem can be expressed as

$$\begin{aligned}
&\mathcal{L}_i(\{U_i(w), l_i(w)\}_{w\in[\underline{w},\overline{w}]}, \mu_i; \lambda_i, \gamma_i) \\
= \ & U_i(\underline{w}) + \lambda_i \int_{\underline{w}}^{\overline{w}} \left\{[wl_i(w) - \varphi_i(U_i(w), l_i(w), \mu_i, \mu_{-i})]\phi_i(U_i(w) - U_{-i}(w); w) - \frac{R}{\overline{w} - \underline{w}}\right\} dw \\
&+ \gamma_i \left[\mu_i - \int_{\underline{w}}^{\overline{w}} \varphi_i(U_i(w), l_i(w), \mu_i, \mu_{-i}) f_i(w) dw\right].
\end{aligned}$$

where $\lambda_i > 0$ is the multiplier associated with the binding budget constraint (2.8) and γ_i is the multiplier associated with the comparison consumption constraint (2.2). Assuming that there is no bunching of workers of different skills and the existence of an interior solution, applying (A.2) gives these necessary conditions:

$$\frac{\partial \mathcal{L}_i}{\partial U_i(w)} = \lambda_i \left[T_i(y_i(w))\tilde{\eta}_i(w) - \frac{1}{v'(c_i(w))}\right] \tilde{f}_i(w) - \gamma_i \frac{f_i(w)}{v'(c_i(w))} = 0 \ \ \forall w \in (\underline{w}, \overline{w}], \tag{A.15}$$

$$\frac{\partial \mathcal{L}_i}{\partial l_i(w)} = \lambda_i \left(w - \frac{h'(l_i(w))}{v'(c_i(w))}\right) \tilde{f}_i(w) - \gamma_i \frac{h'(l_i(w)) f_i(w)}{v'(c_i(w))} = 0 \ \ \forall w \in [\underline{w}, \overline{w}], \tag{A.16}$$

and

$$\frac{\partial \mathcal{L}_i}{\partial \mu_i} = \lambda_i \int_{\underline{w}}^{\overline{w}} \frac{\psi_i(\mu_i, \mu_{-i})}{v'(c_i(w))} \tilde{f}_i(w) dw + \gamma_i \left[1 + \int_{\underline{w}}^{\overline{w}} \frac{\psi_i(\mu_i, \mu_{-i})}{v'(c_i(w))} f_i(w) dw \right] = 0. \tag{A.17}$$

By using (A.15), we obtain the Tiebout-best tax liabilities. By using (A.16) and the FOC (2.4), we obtain the Tiebout-best marginal tax rates. The ratio γ_i/λ_i is determined by (A.17). As is obvious, the least productive workers receive a transfer determined by the government's budget constraint. Therefore, the optimal tax function is discontinuous at $w = \underline{w}$. **QED**

Proof of Proposition 2.3.1: We shall complete the proof in 6 steps.

Step 1. By Assumption 2.2.1, $\gamma_i/\lambda_i > 0$ is guaranteed. Given that $v(\cdot)$ is strictly increasing and $h(\cdot)$ is strictly increasing and convex, for (i) to hold it suffices to show that $\mathcal{C}_i(w) \geq 0$ for $\forall w \in (\underline{w}, \overline{w})$. Therefore, by directly comparing the formulas of marginal tax rates established in Theorem 2.3.1 and Lemma 2.3.1, claim (i) is immediate.

Step 2. By applying the transversality condition (A.7) to equation (A.11), it is easy to see that

$$\frac{T_i'(y_i(\underline{w}))}{1 - T_i'(y_i(\underline{w}))} > \frac{\gamma_i}{\lambda_i} \frac{f_i(\underline{w})}{\tilde{f}_i(\underline{w})} > 0$$

under Assumption 2.2.1. Similarly, for a bounded skill distribution with $\overline{w} < \infty$, applying the transversality condition (A.8) to equation (A.11) gives

$$\frac{T_i'(y_i(\overline{w}))}{1 - T_i'(y_i(\overline{w}))} = \frac{\gamma_i}{\lambda_i} \frac{f_i(\overline{w})}{\tilde{f}_i(\overline{w})} > 0$$

under Assumption 2.2.1. By using Lemma 2.3.1 again, the required assertion (ii) follows.

Step 3. Suppose $h(\cdot)$ takes the isoelastic form, then $\mathcal{A}_i(w)$ is a positive constant. Suppose the first-order approach is valid, namely the SOIC conditions are not binding in the Nash equilibrium, then we have that $v'(\cdot)$ is strictly deceasing in w as $v(\cdot)$ is assumed to be strictly concave. With single-peaked skill distributions, $1/wf_i(w)$ always decreases before the mode w_m. Beyond the mode, it either increases or decreases, depending on how rapidly $f_i(w)$ falls with w. A sufficient, though not necessary, condition for decreasing $T_i'(\cdot)$ over the entire skill distribution is that aggregate skills $wf_i(w)$ are non-decreasing beyond the mode w_m. Also, noting from term $\mathcal{C}_i(w)$ that

$$\frac{\mathrm{d}}{\mathrm{d}w}\int_w^{\overline{w}}\left\{\frac{1}{v'(c_i(t))}\left[1+\frac{\gamma_i}{\lambda_i}\frac{f_i(t)}{\tilde{f}_i(t)}\right]-T_i(y_i(t))\tilde{\eta}_i(t)\right\}\tilde{f}_i(t)dt$$
$$=\left\{T_i(y_i(w))-\frac{1}{v'(c_i(w))\tilde{\eta}_i(w)}\left[1+\frac{\gamma_i}{\lambda_i}\frac{f_i(w)}{\tilde{f}_i(w)}\right]\right\}\tilde{\eta}_i(w)\tilde{f}_i(w),$$

hence an application of Lemma 2.3.1 completes the proof of claim (iii).

Step 4. Define

$$\Sigma_i(w)\equiv-\frac{\varsigma_i(w)}{\lambda_i}=\frac{\gamma_i}{\lambda_i}\int_w^{\overline{w}}\frac{f_i(t)}{v'(c_i(t))}dt+\int_w^{\overline{w}}\left[\frac{1}{v'(c_i(t))}-T_i(y_i(t))\tilde{\eta}_i(t)\right]\tilde{f}_i(t)dt,$$

then the signs of $\Sigma_i(w)$ and $\varsigma_i(w)$ are opposite. As $\varsigma_i(w)$ is differentiable, $\Sigma_i(w)$ is differentiable as well and we have

$$\Sigma_i'(w)=\left\{T_i(y_i(w))-\frac{1}{v'(c_i(w))\tilde{\eta}_i(w)}\left[1+\frac{\gamma_i}{\lambda_i}\frac{f_i(w)}{\tilde{f}_i(w)}\right]\right\}\tilde{\eta}_i(w)\tilde{f}_i(w)\equiv\xi_i(w)\tilde{\eta}_i(w)\tilde{f}_i(w), \tag{A.18}$$

which implies that $\Sigma_i'(w)$ and $\xi_i(w)$ have the same sign. Note that

$$\frac{\mathrm{d}}{\mathrm{d}w}\left\{\frac{1}{v'(c_i(w)\tilde{\eta}_i(w))}\left[1+\frac{\gamma_i}{\lambda_i}\frac{f_i(w)}{\tilde{f}_i(w)}\right]\right\}$$
$$=\underbrace{-\frac{v''(c_i(w))\dot{c}_i(w)}{[v'(c_i(w))]^2\tilde{\eta}_i(w)}\left[1+\frac{\gamma_i}{\lambda_i}\frac{f_i(w)}{\tilde{f}_i(w)}\right]}_{>0}$$
$$-\frac{\dot{\tilde{\eta}}_i(w)}{v'(c_i(w))[\tilde{\eta}_i(w)]^2}\left[1+\frac{\gamma_i}{\lambda_i}\frac{f_i(w)}{\tilde{f}_i(w)}\right]+\underbrace{\frac{1}{v'(c_i(w))\tilde{\eta}_i(w)}\frac{\gamma_i}{\lambda_i}\frac{\mathrm{d}[f_i(w)/\tilde{f}_i(w)]}{\mathrm{d}w}}_{\leq 0}$$

under Assumption 2.2.1 and the assumptions that $\dot{y}_i(w)>0$ and $f_i(w)/\tilde{f}_i(w)$ is non-increasing in

w, thus the condition $-\frac{v''(c_i(w))\dot{c}_i(w)}{v'(c_i(w))} \leq \frac{\dot{\tilde{\eta}}_i(w)}{\tilde{\eta}_i(w)}$ is sufficient for

$$\frac{\mathrm{d}}{\mathrm{d}w}\left\{\frac{1}{v'(c_i(w)\tilde{\eta}_i(w))}\left[1+\frac{\gamma_i}{\lambda_i}\frac{f_i(w)}{\tilde{f}_i(w)}\right]\right\} \leq 0. \tag{A.19}$$

If we assume that $\Sigma_i(w) \geq 0$, then we get from the optimal tax formula in Theorem 2.3.1 that $T_i'(y_i(w)) > 0$. Then applying (A.18) and (A.19) shows that $\xi_i'(w) > 0$ given $\Sigma_i(w) \geq 0$.

Step 5. Assume that there exists a $\tilde{w} \in (\underline{w}, \overline{w})$ such that $\Sigma_i(\tilde{w}) \geq 0$. Then we have two cases to consider in what follows, namely either $\Sigma_i'(\tilde{w}) \geq 0$ or $\Sigma_i'(\tilde{w}) < 0$. If $\Sigma_i'(\tilde{w}) \geq 0$, then we have both $\xi_i(\tilde{w}) \geq 0$ and $\xi_i'(\tilde{w}) > 0$. So the continuity of $\xi_i(w)$ with respect to w implies that there is an open interval with lower bound $\tilde{w}$ such that $\xi_i(\cdot) > 0$, and hence $\Sigma_i'(\cdot) > 0$, on this interval. $\Sigma_i(\cdot)$ is thus positive and strictly increasing on this interval. Without loss of generality, let $(\tilde{w}, \hat{w})$ be a maximal interval on which $\Sigma_i'(w) > 0$ with $\tilde{w} < \hat{w} \leq \overline{w}$. As a consequence, $0 \leq \Sigma_i(\tilde{w}) < \Sigma(\hat{w})$, which implies that $\xi_i'(w) > 0$ for $\forall w \in [\tilde{w}, \hat{w}]$. As a result, $0 \leq \xi_i(\tilde{w}) < \xi_i(\hat{w})$, which leads us to $\Sigma_i'(\hat{w}) > 0$ by using (A.18). Therefore, $\Sigma_i(\cdot)$ is increasing on $[\tilde{w}, \overline{w}]$ given that $\Sigma_i'(\tilde{w}) \geq 0$. We know from the transversality condition (A.8) that $\Sigma_i(\overline{w}) = -\varsigma_i(\overline{w})/\lambda_i = 0$. As we have already shown that $0 \leq \Sigma_i(\tilde{w}) < \Sigma_i(\overline{w})$, an immediate contradiction occurs. We, accordingly, claim that $\Sigma_i'(\tilde{w}) \geq 0$ does not hold true.

Step 6. Given that we have shown that $\Sigma_i'(\tilde{w}) < 0$ for the chosen $\tilde{w}$, we thus have $\xi_i(\tilde{w}) < 0$ by (A.18) and $\xi_i'(\tilde{w}) > 0$. Similarly, the continuity of $\xi_i(w)$ with respect to w implies that there is an open interval with upper bound $\tilde{w}$ such that $\xi_i(\cdot) < 0$, and hence $\Sigma_i'(\cdot) < 0$, on this interval. $\Sigma_i(\cdot)$ is thus positive and strictly decreasing on this interval. Without loss of generality, let $(w^*, \tilde{w})$ be a maximal interval on which $\Sigma_i'(w) < 0$ with $\underline{w} \leq w^* < \tilde{w}$. In consequence, $0 \leq \Sigma_i(\tilde{w}) < \Sigma(w^*)$, which implies that $\xi_i'(w) > 0$ for $\forall w \in [w^*, \tilde{w}]$. As a result, $0 > \xi_i(\tilde{w}) > \xi_i(w^*)$, which leads us to $\Sigma_i'(w^*) < 0$ by using (A.18). Therefore, $\Sigma_i(\cdot)$ will not stop decreasing until reaching the lower bound $\underline{w}$, namely $\Sigma_i(\cdot)$ will be decreasing on $[\underline{w}, \tilde{w}]$. We know from the transversality condition (A.7) that $\Sigma_i(\underline{w}) = -\varsigma_i(\underline{w})/\lambda_i > 0$. Since we have already shown that $0 \leq \Sigma_i(\tilde{w}) < \Sigma_i(\underline{w})$, thus the transversality condition is fulfilled in this case. By using (A.18) and Lemma 2.3.1 again, the required assertion (iv) follows. **QED**

Proof of Proposition 2.3.2: We shall complete the proof in 5 steps.

Step 1. It is easy to verify that $\gamma_i/\lambda_i < 0$ under Assumption 2.2.2. Thus, comparing the optimal tax formulas in Theorem 2.3.1 and Lemma 2.3.1, it is immediate that $T_i'(y_i(\overline{w})) = T_i^{*'}(y_i(\overline{w})) < 0$ for $\overline{w} < \infty$. Noting from (A.7) that $\varsigma_i(\underline{w}) = -1$, thus $\mathcal{A}_i(\underline{w})\mathcal{B}_i(\underline{w})\mathcal{C}_i(\underline{w}) > 0$, which implies that $T_i'(y_i(\underline{w})) > T_i^{*'}(y_i(\underline{w}))$ by using Lemma 2.3.1 again. The proof of claim (i) is hence complete.

Step 2. Note that

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}w}\left\{\frac{1}{v'(c_i(w)\tilde{\eta}_i(w))}\left[1+\frac{\gamma_i}{\lambda_i}\frac{f_i(w)}{\tilde{f}_i(w)}\right]\right\} \\
&= \underbrace{-\frac{v''(c_i(w))\dot{c}_i(w)}{[v'(c_i(w))]^2\tilde{\eta}_i(w)}\left[1+\frac{\gamma_i}{\lambda_i}\frac{f_i(w)}{\tilde{f}_i(w)}\right]}_{\geq 0 \text{ given } \Gamma_i(w)\geq 0} \\
&\quad \underbrace{-\frac{\dot{\tilde{\eta}}_i(w)}{v'(c_i(w))[\tilde{\eta}_i(w)]^2}\left[1+\frac{\gamma_i}{\lambda_i}\frac{f_i(w)}{\tilde{f}_i(w)}\right]+\frac{1}{v'(c_i(w))\tilde{\eta}_i(w)}\frac{\gamma_i}{\lambda_i}\frac{\mathrm{d}[f_i(w)/\tilde{f}_i(w)]}{\mathrm{d}w}}_{\equiv\, \Xi_i(w)},
\end{aligned} \tag{A.20}$$

we get by rearranging the terms that

$$\Xi_i(w) \geq 0 \iff \dot{\Gamma}_i(w) \geq \Gamma_i(w)\frac{\dot{\tilde{\eta}}_i(w)}{\tilde{\eta}_i(w)}. \tag{A.21}$$

It follows from (A.18) that

$$\xi_i'(w) = T_i'(y_i(w))\dot{y}_i(w) - \frac{\mathrm{d}}{\mathrm{d}w}\left\{\frac{1}{v'(c_i(w)\tilde{\eta}_i(w))}\left[1+\frac{\gamma_i}{\lambda_i}\frac{f_i(w)}{\tilde{f}_i(w)}\right]\right\}. \tag{A.22}$$

We now show that, under Assumption 2.2.2, $\Sigma_i(w) \leq 0$ implies $\xi_i'(w) < 0$, with $\Sigma_i(w)$ defined in the proof of Proposition 2.3.1. Assume $\Sigma_i(w) \leq 0$, then we get from the optimal tax formula established in Theorem 2.3.1 that $T_i'(y_i(w))\dot{y}_i(w) < 0$ under Assumption 2.2.2. Therefore, it follows from (A.20)-(A.22) and the assumption $\dot{y}_i(w) > 0$ under the first-order approach that $\xi_i'(w) < 0$, as desired.

Step 3. Here we prove this claim: if there exists a $\tilde{w} \in [\underline{w}, \overline{w})$ such that $\Sigma_i(\tilde{w}) \leq 0$ and $\Sigma_i'(\tilde{w}) \leq 0$, then $\Sigma_i(\cdot)$ is decreasing on the closed interval $[\tilde{w}, \overline{w}]$. Given these assumptions, then we have both $\xi_i(\tilde{w}) \leq 0$ and $\xi_i'(\tilde{w}) < 0$. So the continuity of $\xi_i(w)$ with respect to w implies that there is an open interval with lower bound $\tilde{w}$ such that $\xi_i(\cdot) < 0$, and hence $\Sigma_i'(\cdot) < 0$, on this interval. $\Sigma_i(\cdot)$ is thus negative and strictly decreasing on this interval. Without loss of generality, let $(\tilde{w}, \hat{w})$ be a maximal interval on which $\Sigma_i'(w) < 0$ with $\tilde{w} < \hat{w} \leq \overline{w}$. As a consequence, $\Sigma(\hat{w}) < \Sigma_i(\tilde{w}) \leq 0$, which implies that $\xi_i'(w) < 0$ for $\forall w \in [\tilde{w}, \hat{w}]$. Consequently, $\xi_i(\hat{w}) < \xi_i(\tilde{w}) \leq 0$, which leads us to $\Sigma_i'(\hat{w}) < 0$ by using (A.18). Therefore, $\Sigma_i(\cdot)$ is decreasing on $[\tilde{w}, \overline{w}]$, as desired.

Step 4. Here we prove another claim: if there exists a $\tilde{w} \in (\underline{w}, \overline{w}]$ such that $\Sigma_i(\tilde{w}) \leq 0$ and $\Sigma_i'(\tilde{w}) \geq 0$, then $\Sigma_i(\cdot)$ is increasing on the closed interval $[\underline{w}, \tilde{w}]$. Given these assumptions, we thus have $\xi_i(\tilde{w}) \geq 0$ by (A.18) and $\xi_i'(\tilde{w}) < 0$. Similarly, the continuity of $\xi_i(w)$ with respect to w implies that there is an open interval with upper bound $\tilde{w}$ such that $\xi_i(\cdot) > 0$, and hence $\Sigma_i'(\cdot) > 0$, on this interval. $\Sigma_i(\cdot)$ is thus negative and strictly increasing on this interval. Without loss of generality, let $(w^*, \tilde{w})$ be a maximal interval on which $\Sigma_i'(w) > 0$ with $\underline{w} \leq w^* < \tilde{w}$. In consequence, $\Sigma(w^*) < \Sigma_i(\tilde{w}) \leq 0$, which implies that $\xi_i'(w) < 0$ for $\forall w \in [w^*, \tilde{w}]$. Consequently, $0 \leq \xi_i(\tilde{w}) < \xi_i(w^*)$, which leads us to $\Sigma_i'(w^*) > 0$ by using (A.18). Therefore, $\Sigma_i(\cdot)$ will not stop increasing until reaching the lower bound $\underline{w}$, namely $\Sigma_i(\cdot)$ will be increasing on $[\underline{w}, \tilde{w}]$, as desired.

Step 5. We now show that $\Sigma_i(w) > 0$, and hence $\mathcal{A}_i(w)\mathcal{B}_i(w)\mathcal{C}_i(w) > 0$, on $(\underline{w}, \overline{w})$. In fact, we prove this result by means of contradiction. It follows from the transversality condition (A.8) that $\Sigma_i(\overline{w}) = -\varsigma_i(\overline{w})/\lambda_i = 0$. If there exists a $\tilde{w} \in [\underline{w}, \overline{w})$ such that $\Sigma_i(\tilde{w}) \leq 0$ and $\Sigma_i'(\tilde{w}) \leq 0$, then we have proven in Step 3 that $\Sigma_i(\cdot)$ is decreasing on the closed interval $[\tilde{w}, \overline{w}]$. This means that $\Sigma_i(\overline{w}) < \Sigma_i(\tilde{w}) \leq 0$, yielding an immediate contradiction. Also, it follows from the transversality condition (A.7) that $\Sigma_i(\underline{w}) = -\varsigma_i(\underline{w})/\lambda_i > 0$. If there exists a $\tilde{w} \in (\underline{w}, \overline{w}]$ such that $\Sigma_i(\tilde{w}) \leq 0$ and $\Sigma_i'(\tilde{w}) \geq 0$, then we have proven in Step 4 that $\Sigma_i(\cdot)$ is increasing on the closed interval $[\underline{w}, \tilde{w}]$. This means that $\Sigma_i(\underline{w}) < \Sigma_i(\tilde{w}) \leq 0$, yielding an immediate contradiction. We can thus conclude that $\Sigma_i(w) > 0$ for $\forall w \in (\underline{w}, \overline{w})$ and also $\Sigma_i'(\underline{w}) > 0 > \Sigma_i'(\overline{w})$. By using Lemma 2.3.1 and (A.18) again, then the proof of claim (ii) is complete. **QED**

Proof of Proposition 2.3.3: It follows from condition (b) that $\mathcal{A}_i(w) = 1 + \varepsilon$, a fixed positive constant. The ex post skill distribution term $\mathcal{B}_i(w)$ can be decomposed through

$$\mathcal{B}_i(w) = \frac{1 - F_i(w)}{w f_i(w)} \cdot \frac{\frac{\tilde{F}_i(\infty) - \tilde{F}_i(w)}{1 - F_i(w)}}{\tilde{f}_i(w)/f_i(w)}.$$

By condition (c), we have $\frac{1-F_i(w)}{w f_i(w)} = 1/a_i$. By L'Hôpital's rule, we obtain

$$\lim_{w\uparrow\infty} \frac{\tilde{F}_i(\infty) - \tilde{F}_i(w)}{1 - F_i(w)} = \lim_{w\uparrow\infty} \frac{\tilde{f}_i(w)}{f_i(w)}.$$

As a result, $\lim_{w\uparrow\infty} \mathcal{B}_i(w) = 1/a_i$. By using the definition of the elasticity of migration and conditions (a) and (c), term $\mathcal{C}_i(w)$ can be rewritten as

$$\mathcal{C}_i(w) = \frac{\int_w^\infty \left[1 + \frac{\gamma_i}{\lambda_i}\frac{f_i(t)}{\tilde{f}_i(t)} - \frac{T_i(y_i(t))}{y_i(t) - T_i(y_i(t))}\tilde{\theta}_i(t)\right] \tilde{f}_i(t)dt}{\tilde{F}_i(\infty) - \tilde{F}_i(w)}.$$

Thus, making use of the L'Hôpital's rule again shows that

$$\lim_{w\uparrow\infty} \mathcal{C}_i(w) = 1 + \frac{\gamma_i}{\lambda_i}\alpha_i(\infty) - \frac{T_i'(y_i(\infty))}{1 - T_i'(y_i(\infty))}\tilde{\theta}_i(\infty).$$

So, we get from the optimal tax formula derived in Theorem 2.3.1 that

$$\frac{T_i'(y_i(\infty))}{1 - T_i'(y_i(\infty))} = \frac{\gamma_i}{\lambda_i}\alpha_i(\infty) + (1+\varepsilon)\frac{1}{a_i}\left[1 + \frac{\gamma_i}{\lambda_i}\alpha_i(\infty) - \frac{T_i'(y_i(\infty))}{1 - T_i'(y_i(\infty))}\tilde{\theta}_i(\infty)\right],$$

rearranging the algebra of which gives the desired optimal asymptotic tax rate. **QED**

Proof of Proposition 2.3.4: We shall complete the proof in 3 steps.

Step 1. By applying condition (a) assumed in Proposition 2.3.3 and the assumption $\psi_i(\mu_i, \mu_{-i}) =$

$\sigma_D \in (-1, 0)$ to equation (2.13) produces

$$\frac{\gamma_i}{\lambda_i} = \frac{-\sigma_D}{1+\sigma_D}\tilde{F}_i(\infty) > 0,$$

in which it is unnecessary that $\tilde{F}_i(\infty) = 1$. Also, if $F_i(w) = F_{-i}(w)$, then by using the definition of $\tilde{f}_i(w)$ we obtain $\partial\alpha_i(\infty)/\partial(1/a_i) = 0$. Therefore, as long as $\partial\tilde{F}_i(\infty)/\partial(1/a_i) = 0$, we must have $\partial\left(\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)\right)/\partial(1/a_i) = 0$. In addition, it follows from the definition of $\tilde{\theta}_i(w)$ that $\partial\tilde{\theta}_i(\infty)/\partial(1/a_i) = 0$. Finally, it is straightforward that $\partial(\gamma_i/\lambda_i)/\partial(-\sigma_D) > 0$.

Step 2. Using the established formula of $T_i'(y_i(\infty))$, we have

$$\frac{\partial T_i'(y_i(\infty))}{\partial\left(\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)\right)} = \frac{[1+(1+\varepsilon)(1/a_i)]\left[1+(1+\varepsilon)(1/a_i)\tilde{\theta}_i(\infty)\right]}{\left\{1+\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)+(1+\varepsilon)(1/a_i)\left[1+\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)+\tilde{\theta}_i(\infty)\right]\right\}^2},$$

by which we hence obtain

$$\begin{aligned}&\frac{\partial^2 T_i'(y_i(\infty))}{\partial\left(\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)\right)\partial(1/a_i)}\\ =&\frac{1+\varepsilon}{\left\{1+\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)+(1+\varepsilon)(1/a_i)\left[1+\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)+\tilde{\theta}_i(\infty)\right]\right\}^3}\\ &\times\left(\left[1+\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)+\tilde{\theta}_i(\infty)\right]\left(1+\frac{1+\varepsilon}{a_i}\right)\left[\tilde{\theta}_i(\infty)-1\right]-\tilde{\theta}_i(\infty)\left\{1+\tilde{\theta}_i(\infty)\left[1+\frac{2(1+\varepsilon)}{a_i}\right]\right\}\right).\end{aligned}$$

Thus if the following condition holds true:

$$\left[1+\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)+\tilde{\theta}_i(\infty)\right]\left[\tilde{\theta}_i(\infty)-1\right] \leq 0,$$

then the cross-partial derivative is negative for any $\tilde{\theta}_i(\infty) \in (0, 1]$. It is easy to verify that this condition holds for $\tilde{\theta}_i(\infty) \leq 1$ with any $\sigma_D \in (-1, 0)$, as desired in part (i).

Step 3. If, however, $\tilde{\theta}_i(\infty) > 1$, then we see that

$$\left[1+\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)+\tilde{\theta}_i(\infty)\right]\left(1+\frac{1+\varepsilon}{a_i}\right)\left[\tilde{\theta}_i(\infty)-1\right] < \tilde{\theta}_i(\infty)\left\{1+\tilde{\theta}_i(\infty)\left[1+\frac{2(1+\varepsilon)}{a_i}\right]\right\}$$

is equivalent to

$$1+\frac{\gamma_i}{\lambda_i}\alpha_i(\infty) \;<\; \underbrace{\frac{\tilde{\theta}_i(\infty)\left\{2+\frac{1+\varepsilon}{a_i}\left[1+\tilde{\theta}_i(\infty)\right]\right\}}{\left(1+\frac{1+\varepsilon}{a_i}\right)\left[\tilde{\theta}_i(\infty)-1\right]}}_{>0}.$$

Also, noting that

$$\frac{\tilde{\theta}_i(\infty)\left\{2+\frac{1+\varepsilon}{a_i}\left[1+\tilde{\theta}_i(\infty)\right]\right\}}{\left(1+\frac{1+\varepsilon}{a_i}\right)\left[\tilde{\theta}_i(\infty)-1\right]}-1=\frac{1+\tilde{\theta}_i(\infty)+\frac{1+\varepsilon}{a_i}\left[1+\tilde{\theta}_i(\infty)^2\right]}{\left(1+\frac{1+\varepsilon}{a_i}\right)\left[\tilde{\theta}_i(\infty)-1\right]} > 0,$$

the desired assertion in part (ii) follows. **QED**

Proof of Proposition 2.3.5: We shall complete the proof in 3 steps.

Step 1. If $\tilde{\theta}_i(\infty) < 1$, then for the two determinant terms of

$$\operatorname{sgn}\left\{\frac{\partial^2 T_i'(y_i(\infty))}{\partial\left(\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)\right)\partial(1/a_i)}\right\}$$

established in Step 2 of Proposition 2.3.4 we see the following facts: for the first term we have

$$1+\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)+(1+\varepsilon)(1/a_i)\left[1+\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)+\tilde{\theta}_i(\infty)\right] > 0$$

is equivalent to

$$\alpha_i(\infty) \;<\; \underbrace{\left(-\frac{\lambda_i}{\gamma_i}\right)\frac{1+\frac{1+\varepsilon}{a_i}\left[1+\tilde{\theta}_i(\infty)\right]}{1+\frac{1+\varepsilon}{a_i}}}_{>0}$$

and also for the second term we have

$$\left[1+\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)+\tilde{\theta}_i(\infty)\right]\left(1+\frac{1+\varepsilon}{a_i}\right)\left[\tilde{\theta}_i(\infty)-1\right] > \tilde{\theta}_i(\infty)\left\{1+\tilde{\theta}_i(\infty)\left[1+\frac{2(1+\varepsilon)}{a_i}\right]\right\}$$

is equivalent to

$$\alpha_i(\infty) \;\; > \;\; \underbrace{\left(-\frac{\lambda_i}{\gamma_i}\right)\frac{1+\tilde{\theta}_i(\infty)+\frac{1+\varepsilon}{a_i}\left[1+\tilde{\theta}_i(\infty)^2\right]}{\left(1+\frac{1+\varepsilon}{a_i}\right)\left[1-\tilde{\theta}_i(\infty)\right]}}_{>0}.$$

Noting that

$$\frac{1+\frac{1+\varepsilon}{a_i}\left[1+\tilde{\theta}_i(\infty)\right]}{1+\frac{1+\varepsilon}{a_i}} > \frac{1+\tilde{\theta}_i(\infty)+\frac{1+\varepsilon}{a_i}\left[1+\tilde{\theta}_i(\infty)^2\right]}{\left(1+\frac{1+\varepsilon}{a_i}\right)\left[1-\tilde{\theta}_i(\infty)\right]}$$

is equivalent to

$$0 > 2\tilde{\theta}_i(\infty)\left(1+\frac{1+\varepsilon}{a_i}\right)\left[1+\frac{1+\varepsilon}{a_i}\tilde{\theta}_i(\infty)\right],$$

which however is an immediate contradiction, so we must have

$$\frac{1+\frac{1+\varepsilon}{a_i}\left[1+\tilde{\theta}_i(\infty)\right]}{1+\frac{1+\varepsilon}{a_i}} < \frac{1+\tilde{\theta}_i(\infty)+\frac{1+\varepsilon}{a_i}\left[1+\tilde{\theta}_i(\infty)^2\right]}{\left(1+\frac{1+\varepsilon}{a_i}\right)\left[1-\tilde{\theta}_i(\infty)\right]},$$

by which and the facts that $\gamma_i/\lambda_i < 0$ for $\sigma_D \in (0,1)$ and $\partial(\gamma_i/\lambda_i)/\partial\sigma_D < 0$ the desired assertion in part (i) follows.

Step 2. If $\tilde{\theta}_i(\infty) = 1$, then we get from Step 2 of Proposition 2.3.4 that

$$\frac{\partial^2 T_i'(y_i(\infty))}{\partial\left(\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)\right)\partial(1/a_i)} = \frac{-2(1+\varepsilon)\left(1+\frac{1+\varepsilon}{a_i}\right)}{\left\{1+\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)+(1+\varepsilon)(1/a_i)\left[2+\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)\right]\right\}^3}.$$

Also, since we have that

$$1+\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)+(1+\varepsilon)(1/a_i)\left[2+\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)\right] > 0$$

is equivalent to

$$\alpha_i(\infty) \;<\; \underbrace{\left(-\frac{\lambda_i}{\gamma_i}\right)\frac{1+2\frac{1+\varepsilon}{a_i}}{1+\frac{1+\varepsilon}{a_i}}}_{>0},$$

the required assertion in part (ii) immediately follows.

Step 3. If $\tilde{\theta}_i(\infty) > 1$, then for the second term determining

$$\operatorname{sgn}\left\{\frac{\partial^2 T_i'(y_i(\infty))}{\partial\left(\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)\right)\partial(1/a_i)}\right\}$$

we have that

$$\left[1+\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)+\tilde{\theta}_i(\infty)\right]\left(1+\frac{1+\varepsilon}{a_i}\right)\left[\tilde{\theta}_i(\infty)-1\right] > \tilde{\theta}_i(\infty)\left\{1+\tilde{\theta}_i(\infty)\left[1+\frac{2(1+\varepsilon)}{a_i}\right]\right\}$$

is equivalent to

$$\alpha_i(\infty) \;<\; \underbrace{\left(-\frac{\lambda_i}{\gamma_i}\right)\frac{1+\tilde{\theta}_i(\infty)+\frac{1+\varepsilon}{a_i}\left[1+\tilde{\theta}_i(\infty)^2\right]}{\left(1+\frac{1+\varepsilon}{a_i}\right)\left[1-\tilde{\theta}_i(\infty)\right]}}_{<0},$$

which however is an immediate contradiction. As a result, we must have

$$\left[1+\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)+\tilde{\theta}_i(\infty)\right]\left(1+\frac{1+\varepsilon}{a_i}\right)\left[\tilde{\theta}_i(\infty)-1\right] < \tilde{\theta}_i(\infty)\left\{1+\tilde{\theta}_i(\infty)\left[1+\frac{2(1+\varepsilon)}{a_i}\right]\right\}.$$

Thus, using the first determinant term $1+\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)+(1+\varepsilon)(1/a_i)\left[1+\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)+\tilde{\theta}_i(\infty)\right]$, the required assertion in part (iii) follows. **QED**

Proof of Theorem 2.4.1: As usual, we derive the Stackelberg equilibrium by using backward induction. Thus, the Lagrangian of the follower country $-i$ is the same as in the case when these two countries play Nash, while the Lagrangian for the leader country i is different and reads as follows:

$$
\begin{aligned}
&\mathcal{L}_i(\{U_i(w), l_i(w)\}_{w\in[\underline{w},\overline{w}]}, \mu_i; \lambda_i, \gamma_i, \{\varsigma_i(w)\}_{w\in[\underline{w},\overline{w}]}) \\
=\; & U_i(\underline{w}) + \lambda_i \int_{\underline{w}}^{\overline{w}} \left\{ [wl_i(w) - \varphi_i(U_i(w), l_i(w), \mu_i, \mu_{-i}(\mu_i))]\phi_i(U_i(w) - U_{-i}(w); w) - \frac{R}{\overline{w} - \underline{w}} \right\} dw \\
& + \varsigma_i(\underline{w})U_i(\underline{w}) - \varsigma_i(\overline{w})U_i(\overline{w}) + \int_{\underline{w}}^{\overline{w}} \left[\varsigma_i(w)h'(l_i(w))\frac{l_i(w)}{w} + \dot{\varsigma}_i(w)U_i(w) \right] dw \\
& + \gamma_i \left[\mu_i - \int_{\underline{w}}^{\overline{w}} \varphi_i(U_i(w), l_i(w), \mu_i, \mu_{-i}(\mu_i)) f_i(w) dw \right].
\end{aligned}
$$

Note that

$$
\mu_{-i} = \int_{\underline{w}}^{\overline{w}} \varphi_{-i}(U_{-i}(w), l_{-i}(w), \mu_{-i}, \mu_i) f_{-i}(w) dw,
$$

making use of the Implicit Function Theorem produces

$$
\frac{\partial \mu_{-i}}{\partial \mu_i} = \frac{\int_{\underline{w}}^{\overline{w}} \frac{\partial \varphi_{-i}}{\partial \mu_i} f_{-i}(w) dw}{1 - \int_{\underline{w}}^{\overline{w}} \frac{\partial \varphi_{-i}}{\partial \mu_{-i}} f_{-i}(w) dw}. \tag{A.23}
$$

Assuming that there is no bunching of workers of different skills and the existence of an interior solution, then all of these first-order necessary conditions of Lagrangian $\mathcal{L}_i$ are the same as those in the proof of Theorem 2.3.1 but

$$
\begin{aligned}
\frac{\partial \mathcal{L}_i}{\partial \mu_i} =\; & -\lambda_i \int_{\underline{w}}^{\overline{w}} \left(\frac{\partial \varphi_i}{\partial \mu_i} + \frac{\partial \varphi_i}{\partial \mu_{-i}} \frac{\partial \mu_{-i}}{\partial \mu_i} \right) \tilde{f}_i(w) dw \\
& + \gamma_i \left[1 - \int_{\underline{w}}^{\overline{w}} \left(\frac{\partial \varphi_i}{\partial \mu_i} + \frac{\partial \varphi_i}{\partial \mu_{-i}} \frac{\partial \mu_{-i}}{\partial \mu_i} \right) f_i(w) dw \right] = 0,
\end{aligned}
$$

where $\partial \mu_{-i}/\partial \mu_i$ is given by equation (A.23). **QED**

Proof of Proposition 2.4.1: Applying condition (a) and (A.2) to equation (A.23) shows that

$$
\frac{\partial \mu_{-i}}{\partial \mu_i} = \frac{-\sigma_F}{1 + \sigma_D},
$$

substituting which into the formula of γ_i/λ_i shown in Theorem 2.4.1 reveals that

$$\frac{\gamma_i}{\lambda_i} = \frac{\sigma_F^2 - (1+\sigma_D)\sigma_D}{(1+\sigma_D)^2 - \sigma_F^2}\tilde{F}_i(\overline{w}).$$

We thus obtain

$$\frac{\partial(\gamma_i/\lambda_i)}{\partial\sigma_D} = -\frac{\sigma_F^2 + (1+\sigma_D)^2}{\left[(1+\sigma_D)^2 - \sigma_F^2\right]^2}\tilde{F}_i(\overline{w}) < 0$$

and

$$\frac{\partial(\gamma_i/\lambda_i)}{\partial\sigma_F^2} = \frac{1+\sigma_D}{\left[(1+\sigma_D)^2 - \sigma_F^2\right]^2}\tilde{F}_i(\overline{w}) > 0.$$

As a result, using chain rule, Corollary 2.4.1 and Proposition 2.3.3 gives rise to

$$\frac{\partial^2 T_i'(y_i(\infty))}{\partial\sigma_D \partial(1/a_i)} = \underbrace{\frac{\partial^2 T_i'(y_i(\infty))}{\partial\left(\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)\right)\partial(1/a_i)}}_{<0} \cdot \underbrace{\alpha_i(\infty)\frac{\partial(\gamma_i/\lambda_i)}{\partial\sigma_D}}_{<0} \; > \; 0$$

and

$$\frac{\partial^2 T_i'(y_i(\infty))}{\partial\sigma_F^2 \partial(1/a_i)} = \underbrace{\frac{\partial^2 T_i'(y_i(\infty))}{\partial\left(\frac{\gamma_i}{\lambda_i}\alpha_i(\infty)\right)\partial(1/a_i)}}_{<0} \cdot \underbrace{\alpha_i(\infty)\frac{\partial(\gamma_i/\lambda_i)}{\partial\sigma_F^2}}_{>0} \; < \; 0$$

for $\forall\sigma_D \in (-1, 0)$ and $\tilde{\theta}_i(\infty) < 1$, as desired. For the other cases, we can similarly show that the predictions of Nash equilibrium carry over to the current Stackelberg equilibrium. **QED**

Proof of Proposition 2.4.2: It follows from Theorem 2.4.1 that

$$\frac{\partial(\gamma_i/\lambda_i)}{\partial(\partial\mu_{-i}/\partial\mu_i)} = \frac{\Lambda}{\left[1 - \int_{\underline{w}}^{\overline{w}}\left(\frac{\partial c_i(w)}{\partial\mu_i} + \frac{\partial c_i(w)}{\partial\mu_{-i}}\frac{\partial\mu_{-i}}{\partial\mu_i}\right) f_i(w)dw\right]^2}$$

where

$$\Lambda \equiv \left[1 - \int_{\underline{w}}^{\overline{w}} \frac{\partial c_i(w)}{\partial\mu_i} f_i(w)dw\right] \int_{\underline{w}}^{\overline{w}} \frac{\partial c_i(w)}{\partial\mu_{-i}} \tilde{f}_i(w)dw + \int_{\underline{w}}^{\overline{w}} \frac{\partial c_i(w)}{\partial\mu_{-i}} f_i(w)dw \int_{\underline{w}}^{\overline{w}} \frac{\partial c_i(w)}{\partial\mu_i} \tilde{f}_i(w)dw.$$

If Assumption 2.2.1 holds, then $\Lambda > 0$, and hence

$$\frac{\partial(\gamma_i/\lambda_i)}{\partial(\partial\mu_{-i}/\partial\mu_i)} > 0.$$

Also, using Theorem 2.4.1 and Assumption 2.2.1 again gives rise to $\partial\mu_{-i}/\partial\mu_i > 0$. Since we get from the optimal tax formula in Theorem 2.3.1 that optimal marginal tax rates strictly increase in γ_i/λ_i and $\partial\mu_{-i}/\partial\mu_i = 0$ in the Nash equilibrium, the required assertion accordingly follows. **QED**

APPENDIX B

VOTING OVER SELFISHLY OPTIMAL INCOME TAX SCHEDULES WITH TAX-DRIVEN MIGRATIONS

B.1 Proofs

Proof of Theorem 3.4.1: We shall complete the proof in 4 steps.

Step 1. In order to solve problem (3.12), we first give the following lemma.

Lemma B.1.1. *The optimal schedule of before-tax incomes $y(\cdot)$ for type k's problem (3.12) is obtained by solving the following unconstrained maximization problem*

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

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

Proof. By using (3.6), we have

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

Integrating over the ex post support of the skill distribution yields

$$\int_{\underline{w}}^{\overline{w}}U(w)\tilde{f}(w)dw = U(\underline{w})\int_{\underline{w}}^{\overline{w}}\tilde{f}(w)dw + \int_{\underline{w}}^{\overline{w}}\left[\int_{\underline{w}}^{w}\frac{y(t)}{t^2}h'\left(\frac{y(t)}{t}\right)dt\right]\tilde{f}(w)dw. \tag{B.3}$$

Reversing the order of integration in (B.3) gives rise to

$$\int_{\underline{w}}^{\overline{w}}U(w)\tilde{f}(w)dw = U(\underline{w})\int_{\underline{w}}^{\overline{w}}\tilde{f}(w)dw + \int_{\underline{w}}^{\overline{w}}\frac{y(t)}{t^2}h'\left(\frac{y(t)}{t}\right)\left[\int_{t}^{\overline{w}}\tilde{f}(w)dw\right]dt. \tag{B.4}$$

Also, it follows from (3.5) that

$$\int_{\underline{w}}^{\overline{w}} U(w)\tilde{f}(w)dw = \int_{\underline{w}}^{\overline{w}} c(w)\tilde{f}(w)dw - \int_{\underline{w}}^{\overline{w}} h\left(\frac{y(w)}{w}\right)\tilde{f}(w)dw. \tag{B.5}$$

Applying the equality form of (3.10) to (B.5) shows that

$$\int_{\underline{w}}^{\overline{w}} U(w)\tilde{f}(w)dw = \int_{\underline{w}}^{\overline{w}} y(w)\tilde{f}(w)dw - \int_{\underline{w}}^{\overline{w}} h\left(\frac{y(w)}{w}\right)\tilde{f}(w)dw. \tag{B.6}$$

Combining (B.4) and (B.6) leads us to

$$\begin{aligned} U(\underline{w})\int_{\underline{w}}^{\overline{w}} \tilde{f}(w)dw = & \int_{\underline{w}}^{\overline{w}} y(w)\tilde{f}(w)dw \\ & - \int_{\underline{w}}^{\overline{w}} h\left(\frac{y(w)}{w}\right)\tilde{f}(w)dw - \int_{\underline{w}}^{\overline{w}} \frac{y(w)}{w^2}h'\left(\frac{y(w)}{w}\right)\left[\int_{w}^{\overline{w}} \tilde{f}(t)dt\right]dw. \end{aligned} \tag{B.7}$$

Define

$$\Gamma(w,\overline{w}) \equiv \int_{w}^{\overline{w}} \tilde{f}(t)dt, \tag{B.8}$$

then we can rewrite (B.7) as

$$U(\underline{w}) = \frac{1}{\Gamma(\underline{w},\overline{w})}\int_{\underline{w}}^{\overline{w}} \left\{\left[y(w) - h\left(\frac{y(w)}{w}\right)\right]\tilde{f}(w) - \frac{y(w)}{w^2}h'\left(\frac{y(w)}{w}\right)\Gamma(w,\overline{w})\right\}dw. \tag{B.9}$$

Substituting (B.9) into (B.2) and setting $w = k$, then the maximand in (B.1) is established.

Step 2. We prove the following lemma.

Lemma B.1.2. *By setting $k = \underline{w}$ in problem (B.1), the maximin income schedule, denoted by*

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

is obtained as follows.

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

$$\left[1-\frac{1}{w}h'\left(\frac{y(w)}{w}\right)\right]\tilde{f}(w)+\left[y(w)-h\left(\frac{y(w)}{w}\right)+\frac{y(w)}{w}h'\left(\frac{y(w)}{w}\right)\right]\frac{\partial\tilde{f}(w)}{\partial y(w)}$$
$$=\left[\frac{1}{w^2}h'\left(\frac{y(w)}{w}\right)+\frac{y(w)}{w^3}h''\left(\frac{y(w)}{w}\right)\right]\Gamma(w,\overline{w}) \tag{B.10}$$

and

$$\left[1-\frac{1}{\overline{w}}h'\left(\frac{y(\overline{w})}{\overline{w}}\right)\right]\tilde{f}(\overline{w})+\left[y(\overline{w})-h\left(\frac{y(\overline{w})}{\overline{w}}\right)-U(\underline{w})\right]\frac{\partial\tilde{f}(\overline{w})}{\partial y(\overline{w})}=0, \tag{B.11}$$

respectively, in which

$$\frac{\partial\tilde{f}(w)}{\partial y(w)}=\begin{cases}g_-(\Delta(w)|w)f_-(w)n_-\frac{\partial U(w)}{\partial y(w)} & \textit{for } \Delta(w)\geq 0,\\ g(-\Delta(w)|w)f(w)\frac{\partial U(w)}{\partial y(w)} & \textit{for } \Delta(w)\leq 0\end{cases} \tag{B.12}$$

with

$$\frac{\partial U(w)}{\partial y(w)}=\frac{1}{w}h'\left(\frac{y(w)}{w}\right)+\frac{y(w)}{w^2}h''\left(\frac{y(w)}{w}\right) \tag{B.13}$$

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

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

Proof. By setting $k=\underline{w}$, the maximand of problem (B.1) is hence given by (B.9). It is straightforward that the corresponding maximization problem can be solved point-wise, and we just need to consider two cases.

Case I: $w\in(\underline{w},\overline{w})$. To solve the first-order condition $\partial U(\underline{w})/\partial y(w)=0$ using (B.9), it is sufficient to solve the equation

$$\left[1-\frac{1}{w}h'\left(\frac{y(w)}{w}\right)\right]\tilde{f}(w)+\left[y(w)-h\left(\frac{y(w)}{w}\right)\right]\frac{\partial\tilde{f}(w)}{\partial y(w)}-\frac{y(w)}{w^2}h'\left(\frac{y(w)}{w}\right)\frac{\partial\Gamma(w,\overline{w})}{\partial y(w)}$$
$$=\left[\frac{1}{w^2}h'\left(\frac{y(w)}{w}\right)+\frac{y(w)}{w^3}h''\left(\frac{y(w)}{w}\right)\right]\Gamma(w,\overline{w}). \tag{B.14}$$

By applying the formula of integration by parts to (B.8), we have

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

by which we obtain

$$\frac{\partial\Gamma(w,\overline{w})}{\partial y(w)}=-w\frac{\partial\tilde{f}(w)}{\partial y(w)}\ \text{ and }\ \frac{\partial\Gamma(w,\overline{w})}{\partial y(\overline{w})}=\overline{w}\frac{\partial\tilde{f}(\overline{w})}{\partial y(\overline{w})}. \tag{B.15}$$

Applying (B.15) to (B.14) and rearranging the algebra, the desired (B.10) is hence established. By setting $k = w$ in the maximand of problem (B.1), then for $\forall w \in (\underline{w}, \overline{w})$ (B.13) is immediate by applying the formula of integration by parts. Also, by using (3.8), (B.12) is immediate.

Case II: $w = \overline{w}$. By using (B.9) and the formula of integration by parts again, we have the following first-order condition:

$$\frac{\partial U(\underline{w})}{\partial y(\overline{w})}=-\frac{U(\underline{w})}{\Gamma(\underline{w},\overline{w})}\frac{\partial\Gamma(\underline{w},\overline{w})}{\partial y(\overline{w})}$$
$$+\frac{\overline{w}}{\Gamma(\underline{w},\overline{w})}\left\{\left[1-\frac{1}{\overline{w}}h'\left(\frac{y(\overline{w})}{\overline{w}}\right)\right]\tilde{f}(\overline{w})+\left[y(\overline{w})-h\left(\frac{y(\overline{w})}{\overline{w}}\right)\right]\frac{\partial\tilde{f}(\overline{w})}{\partial y(\overline{w})}\right\}=0. \tag{B.16}$$

By applying (B.15) to (B.16) and rearranging the algebra, equation (B.11) is thus obtained. In addition, by setting $k = \overline{w}$ in the maximand of problem (B.1) and making use of the formula of integration by parts, we obtain

$$
\begin{aligned}
\frac{\partial U(\overline{w})}{\partial y(\overline{w})} = & -\frac{U(\underline{w})}{\Gamma(\underline{w},\overline{w})}\frac{\partial \Gamma(\underline{w},\overline{w})}{\partial y(\overline{w})} \\
& + \frac{\overline{w}}{\Gamma(\underline{w},\overline{w})}\left\{\left[1-\frac{1}{\overline{w}}h'\left(\frac{y(\overline{w})}{\overline{w}}\right)\right]\tilde{f}(\overline{w}) + \left[y(\overline{w}) - h\left(\frac{y(\overline{w})}{\overline{w}}\right)\right]\frac{\partial \tilde{f}(\overline{w})}{\partial y(\overline{w})}\right\} \\
& + \overline{w}\left[\frac{1}{\overline{w}^2}h'\left(\frac{y(\overline{w})}{\overline{w}}\right) + \frac{y(\overline{w})}{\overline{w}^3}h''\left(\frac{y(\overline{w})}{\overline{w}}\right)\right].
\end{aligned}
\tag{B.17}
$$

By applying (B.16) to (B.17), then we arrive at that (B.13) also holds for $w = \overline{w}$.

Step 3. We also need the following lemma.

Lemma B.1.3. *By setting $k = \overline{w}$ in problem (B.1), the maximax income schedule, denoted by*

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

is obtained as follows.

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

$$
\begin{aligned}
&\left[1-\frac{1}{w}h'\left(\frac{y(w)}{w}\right)\right]\tilde{f}(w) + \left[y(w) - h\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w}h'\left(\frac{y(w)}{w}\right)\right]\frac{\partial \tilde{f}(w)}{\partial y(w)} \\
&\qquad = \left[\frac{1}{w^2}h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^3}h''\left(\frac{y(w)}{w}\right)\right]\left[\Gamma(w,\overline{w}) - \Gamma(\underline{w},\overline{w})\right]
\end{aligned}
\tag{B.18}
$$

and

$$
\left[1-\frac{1}{\underline{w}}h'\left(\frac{y(\underline{w})}{\underline{w}}\right)\right]\tilde{f}(\underline{w}) + \left[y(\underline{w}) - h\left(\frac{y(\underline{w})}{\underline{w}}\right) + \frac{y(\underline{w})}{\underline{w}}h'\left(\frac{y(\underline{w})}{\underline{w}}\right) - U(\underline{w})\right]\frac{\partial \tilde{f}(\underline{w})}{\partial y(\underline{w})} = 0,
\tag{B.19}
$$

respectively, in which $\partial \tilde{f}(w)/\partial y(w)$ is determined by equations (B.12) and (B.13) for $\forall w \in [\underline{w},\overline{w})$.

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

Proof. By setting $k = \overline{w}$, the maximand of problem (B.1) can be written as

$$U(\overline{w}) = \frac{1}{\Gamma(\underline{w}, \overline{w})} \int_{\underline{w}}^{\overline{w}} \left\{ \left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \tilde{f}(w) - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \Gamma(w, \overline{w}) \right\} dw \\ + \int_{\underline{w}}^{\overline{w}} \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) dw. \tag{B.20}$$

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

Case I: $w \in (\underline{w}, \overline{w})$. To solve the first-order condition $\partial U(\overline{w})/\partial y(w) = 0$ using (B.20), it is sufficient to solve the equation

$$\left[1 - \frac{1}{w} h'\left(\frac{y(w)}{w}\right)\right] \tilde{f}(w) + \left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \frac{\partial \tilde{f}(w)}{\partial y(w)} - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \frac{\partial \Gamma(w, \overline{w})}{\partial y(w)} \\ = \left[\frac{1}{w^2} h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^3} h''\left(\frac{y(w)}{w}\right)\right] [\Gamma(w, \overline{w}) - \Gamma(\underline{w}, \overline{w})]. \tag{B.21}$$

By using (B.15) and rearranging the algebra, (B.18) follows from (B.21). Similar to the proof of Lemma B.1.2, it is easy to verify that $\left\{\partial \tilde{f}(w)/\partial y(w)\right\}_{w \in (\underline{w}, \overline{w})}$ is determined by equations (B.12) and (B.13).

Case II: $w = \underline{w}$. By using (B.20) and the formula of integration by parts, we have the first-order condition:

$$\frac{\partial U(\overline{w})}{\partial y(\underline{w})} = - \frac{U(\underline{w})}{\Gamma(\underline{w}, \overline{w})} \frac{\partial \Gamma(\underline{w}, \overline{w})}{\partial y(\underline{w})} \\ - \frac{\underline{w}}{\Gamma(\underline{w}, \overline{w})} \left\{ \left[1 - \frac{1}{\underline{w}} h'\left(\frac{y(\underline{w})}{\underline{w}}\right)\right] \tilde{f}(\underline{w}) + \left[y(\underline{w}) - h\left(\frac{y(\underline{w})}{\underline{w}}\right)\right] \frac{\partial \tilde{f}(\underline{w})}{\partial y(\underline{w})} \right\} \\ + \frac{\underline{w}}{\Gamma(\underline{w}, \overline{w})} \left\{ \left[\frac{1}{\underline{w}^2} h'\left(\frac{y(\underline{w})}{\underline{w}}\right) + \frac{y(\underline{w})}{\underline{w}^3} h''\left(\frac{y(\underline{w})}{\underline{w}}\right)\right] \Gamma(\underline{w}, \overline{w}) + \frac{y(\underline{w})}{\underline{w}^2} h'\left(\frac{y(\underline{w})}{\underline{w}}\right) \frac{\partial \Gamma(\underline{w}, \overline{w})}{\partial y(\underline{w})} \right\} \\ - \underline{w} \left[\frac{1}{\underline{w}^2} h'\left(\frac{y(\underline{w})}{\underline{w}}\right) + \frac{y(\underline{w})}{\underline{w}^3} h''\left(\frac{y(\underline{w})}{\underline{w}}\right)\right] = 0. \tag{B.22}$$

By applying (B.15) and rearranging the algebra, (B.22) can be simplified as being the desired

(B.19). In order to solve for the term $\partial \tilde{f}(\underline{w})/\partial y(\underline{w})$ appearing in (B.22), we get from (B.9) that

$$\begin{aligned}\frac{\partial U(\underline{w})}{\partial y(\underline{w})} =& -\frac{U(\underline{w})}{\Gamma(\underline{w},\overline{w})}\frac{\partial \Gamma(\underline{w},\overline{w})}{\partial y(\underline{w})} + \underline{w}\left[\frac{1}{\underline{w}^2}h'\left(\frac{y(\underline{w})}{\underline{w}}\right) + \frac{y(\underline{w})}{\underline{w}^3}h''\left(\frac{y(\underline{w})}{\underline{w}}\right)\right] \\ & - \frac{\underline{w}}{\Gamma(\underline{w},\overline{w})}\left\{\left[1-\frac{1}{\underline{w}}h'\left(\frac{y(\underline{w})}{\underline{w}}\right)\right]\tilde{f}(\underline{w}) + \left[y(\underline{w}) - h\left(\frac{y(\underline{w})}{\underline{w}}\right)\right]\frac{\partial \tilde{f}(\underline{w})}{\partial y(\underline{w})}\right\} \\ & + \frac{\underline{w}}{\Gamma(\underline{w},\overline{w})}\frac{y(\underline{w})}{\underline{w}^2}h'\left(\frac{y(\underline{w})}{\underline{w}}\right)\frac{\partial \Gamma(\underline{w},\overline{w})}{\partial y(\underline{w})}.\end{aligned} \tag{B.23}$$

Applying (B.22) to (B.23) results in the desired equation (B.13) for $w = \underline{w}$.

Step 4. By using the maximization problem (B.1) stated in Lemma B.1.1, it is easy to show that

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

and

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

for $\forall k \in (\underline{w}, \overline{w})$. Therefore, the desired income schedule (3.13) follows from a direct application of Lemmas B.1.2 and B.1.3. **QED**

Proof of Proposition 3.4.1: We shall complete the proof in 3 steps.

Step 1. Making use of (3.2) and (3.5) reveals that

$$y(\underline{w}) - h\left(\frac{y(\underline{w})}{\underline{w}}\right) - U(\underline{w}) = T^M(y(\underline{w})). \tag{B.24}$$

Plugging (B.24) in (3.15) produces

$$\underline{\tau}^M(\underline{w}) = \underbrace{-\frac{\partial \tilde{f}(\underline{w})}{\partial y(\underline{w})}\frac{1}{\tilde{f}(\underline{w})}}_{<0}\left[T^M(y(\underline{w})) + \frac{y(\underline{w})}{\underline{w}}h'\left(\frac{y(\underline{w})}{\underline{w}}\right)\right], \tag{B.25}$$

in which we have used (B.12), (B.13) and the strict convexity of h. Also, by applying (3.6), (B.25) can be equivalently written as

$$\underline{\tau}^M(\underline{w}) = -\frac{\partial \tilde{f}(\underline{w})}{\partial y(\underline{w})} \frac{1}{\tilde{f}(\underline{w})} \left[T^M(y(\underline{w})) + \underline{w} U'(\underline{w}) \right]. \tag{B.26}$$

So, the required assertion (3.19) follows from (B.25) and (B.26).

Step 2. In fact, we can further rewrite (B.25) as

$$\begin{aligned} \underline{\tau}^M(\underline{w}) &= -\frac{\partial \tilde{f}(\underline{w})}{\partial y(\underline{w})} \frac{y(\underline{w})}{\tilde{f}(\underline{w})} \left[\frac{T^M(y(\underline{w}))}{y(\underline{w})} + \frac{1}{\underline{w}} h' \left(\frac{y(\underline{w})}{\underline{w}} \right) \right] \\ &= -\frac{\partial \tilde{f}(\underline{w})}{\partial U(\underline{w})} \frac{\partial U(\underline{w})}{\partial y(\underline{w})} \frac{y(\underline{w})}{\tilde{f}(\underline{w})} \left[\frac{T^M(y(\underline{w}))}{y(\underline{w})} + \frac{1}{\underline{w}} h' \left(\frac{y(\underline{w})}{\underline{w}} \right) \right] \\ &= -\frac{\partial \tilde{f}(\underline{w})}{\partial \Delta(\underline{w})} \frac{c(\underline{w})}{\tilde{f}(\underline{w})} \cdot \frac{\partial U(\underline{w})}{\partial y(\underline{w})} \cdot \frac{y(\underline{w})}{c(\underline{w})} \left[\frac{T^M(y(\underline{w}))}{y(\underline{w})} + \frac{1}{\underline{w}} h' \left(\frac{y(\underline{w})}{\underline{w}} \right) \right] \\ &= -\tilde{\theta}(\underline{w}) \cdot \mathrm{MU}_y(\underline{w}) \cdot \mathrm{Q}_{y,c}(\underline{w}) \left[\mathrm{ATR}^M(\underline{w}) + \frac{1}{\underline{w}} h' \left(\frac{y(\underline{w})}{\underline{w}} \right) \right], \end{aligned} \tag{B.27}$$

in which we have used the chain rule of calculus, (3.8) and (3.9). Applying (3.4) to (B.27) reveals that

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

rearranging the algebra of which results in

$$\underline{\tau}^M(\underline{w}) = -\frac{\tilde{\theta}(\underline{w}) \cdot \mathrm{MU}_y(\underline{w}) \cdot \mathrm{Q}_{y,c}(\underline{w})}{1 - \tilde{\theta}(\underline{w}) \cdot \mathrm{MU}_y(\underline{w}) \cdot \mathrm{Q}_{y,c}(\underline{w})} \left[\mathrm{ATR}^M(\underline{w}) + 1 \right] \tag{B.28}$$

whenever $\tilde{\theta}(\underline{w}) \cdot \mathrm{MU}_y(\underline{w}) \cdot \mathrm{Q}_{y,c}(\underline{w}) \neq 1$. As a result, assertions (3.20) and (3.21) are obtained by using (B.28).

Step 3. We get from Lemma B.1.1 that

$$\Gamma(w, \overline{w}) - \Gamma(\underline{w}, \overline{w}) = -\int_{\underline{w}}^{w} \tilde{f}(t) dt = -\Gamma(\underline{w}, w) < 0, \;\; \forall w > \underline{w}. \tag{B.29}$$

Also, we can show that

$$
\begin{aligned}
& y(w) - h\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w} h'\left(\frac{y(w)}{w}\right) \geq 0 \\
\Leftrightarrow \; & w \geq \frac{h(l(w))}{l(w)} \left[1 - \frac{l(w)h'(l(w))}{h(l(w))}\right].
\end{aligned} \tag{B.30}
$$

Let's define $H(l) \equiv lh'(l) - h(l)$. We have $H'(l) = lh''(l) > 0$ for $\forall l > 0$, which implies that $H(l)$ is strictly increasing in l and hence $H(l) > H(0) = 0$. In other words, $lh'(l)/h(l) > 1$, and hence the right hand side of the second inequality of (B.30) is negative. Given that $\underline{w} > 0$ by assumption, we conclude by using (B.30) again that

$$
y(w) - h\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w} h'\left(\frac{y(w)}{w}\right) > 0. \tag{B.31}
$$

So combining (3.16), (B.29), (B.31), (B.12), (B.13) and the strict convexity of h all together gives rise to the desired assertion (ii). **QED**

Proof of Proposition 3.4.2: We shall complete the proof in 2 steps.

Step 1. By using the chain rule of calculus, (3.8) and (3.9), we have

$$
\begin{aligned}
& \frac{\partial \tilde{f}(w)}{\partial y(w)} \frac{1}{\tilde{f}(w)} \left[h\left(\frac{y(w)}{w}\right) - y(w) - \frac{y(w)}{w} h'\left(\frac{y(w)}{w}\right)\right] \\
= & \frac{\partial \tilde{f}(w)}{\partial \Delta(w)} \frac{c(w)}{\tilde{f}(w)} \frac{\partial U(w)}{\partial y(w)} \frac{1}{c(w)} \left[h\left(\frac{y(w)}{w}\right) - y(w) - \frac{y(w)}{w} h'\left(\frac{y(w)}{w}\right)\right] \\
= & \tilde{\theta}(w) \frac{\partial U(w)}{\partial y(w)} \frac{1}{c(w)} \left[h\left(\frac{y(w)}{w}\right) - y(w) - \frac{y(w)}{w} h'\left(\frac{y(w)}{w}\right)\right].
\end{aligned} \tag{B.32}
$$

Then we get from (B.31), (B.32) and (3.17) that

$$
\tau^R(w) = \underbrace{\frac{\partial U(w)}{\partial y(w)}}_{>0} \left\{ \underbrace{\frac{\Gamma(w, \overline{w})}{w \tilde{f}(w)}}_{>0} - \underbrace{\frac{\tilde{\theta}(w)}{c(w)} \left[y(w) - h\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w} h'\left(\frac{y(w)}{w}\right)\right]}_{>0} \right\}, \tag{B.33}
$$

by which assertion (3.22) is immediate.

Step 2. It follows from (3.6) that $U(\overline{w}) > U(\underline{w})$. Making use (3.2) and (3.5) shows that

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

by which the desired assertion (3.23) follows. **QED**

Proof of Proposition 3.4.3: We shall complete the proof in 3 steps.

Step 1. As w approaches $\underline{w}$ from the above, we get from (3.16) that

$$\tau^M(\underline{w}) = -\frac{\partial \tilde{f}(\underline{w})}{\partial y(\underline{w})}\frac{1}{\tilde{f}(\underline{w})}\left[y(\underline{w}) - h\left(\frac{y(\underline{w})}{\underline{w}}\right) + \frac{y(\underline{w})}{\underline{w}}h'\left(\frac{y(\underline{w})}{\underline{w}}\right)\right], \tag{B.34}$$

in which we have used the continuity of $\tau^M(w)$ over the interval $(\underline{w}, k)$. Comparing this formula (B.34) with (3.15) reveals that

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

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

Step 2. It follows from (3.16) and (3.17) that

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

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

Step 3. As w approaches $\overline{w}$ from the below, we get from (3.17) that

$$\tau^R(\overline{w}) = -\frac{\partial \tilde{f}(\overline{w})}{\partial y(\overline{w})} \frac{1}{\tilde{f}(\overline{w})} \left[y(\overline{w}) - h\left(\frac{y(\overline{w})}{\overline{w}} \right) + \frac{y(\overline{w})}{\overline{w}} h' \left(\frac{y(\overline{w})}{\overline{w}} \right) \right], \tag{B.35}$$

in which we have used the continuity of $\tau^R(w)$ over the interval $(k, \overline{w})$. Comparing this formula (B.35) with (3.18) reveals that

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

which combined with (3.6) and the same reasoning used in step 1 leads us to the desired assertion shown in part (iii). **QED**

Proof of Theorem 3.5.1: We shall complete the proof in 3 steps.

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

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

$$U(w, w) \;\geq\; U(w, k) \quad \text{for } \forall w, k \in [\underline{w}, \overline{w}]. \tag{B.36}$$

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 [14].

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 (3.6) and (B.36), we have

$$\begin{aligned} U(w, k_1) \;&=\; U(k_1, k_1) - \int_w^{k_1} h'\left(\frac{y(t, k_1)}{t}\right) \frac{y(t, k_1)}{t^2} dt \\ &\geq\; U(k_1, k_2) - \int_w^{k_1} h'\left(\frac{y(t, k_1)}{t}\right) \frac{y(t, k_1)}{t^2} dt. \end{aligned} \tag{B.37}$$

Similarly, we can get by (3.6) that

$$U(w, k_2) \;=\; U(k_1, k_2) - \int_w^{k_1} h'\left(\frac{y(t, k_2)}{t}\right) \frac{y(t, k_2)}{t^2} dt. \tag{B.38}$$

Solving for $U(k_1, k_2)$ from (B.38) and inserting it into (B.37) produces

$$U(w, k_1) - U(w, k_2) \;\geq\; \int_w^{k_1} \left[h'\left(\frac{y(t, k_2)}{t}\right) \frac{y(t, k_2)}{t^2} - h'\left(\frac{y(t, k_1)}{t}\right) \frac{y(t, k_1)}{t^2}\right] dt. \tag{B.39}$$

Since h is strictly increasing and convex, we hence have by using (B.39) that $U(w, k_1) \geq U(w, k_2)$, which combined with (B.36) reveals that

$$U(w, w) \;\geq\; U(w, k_1) \;\geq\; U(w, k_2), \quad \forall w < k_1 < k_2. \tag{B.40}$$

<u>*Step 3.*</u> Second, for the case with $w > k_1 > k_2$, we also get by using (3.6) and (B.36) that

$$\begin{aligned} U(w, k_1) \;&=\; U(k_1, k_1) + \int_{k_1}^{w} h'\left(\frac{y(t, k_1)}{t}\right)\frac{y(t, k_1)}{t^2}dt \\ &\geq\; U(k_1, k_2) + \int_{k_1}^{w} h'\left(\frac{y(t, k_1)}{t}\right)\frac{y(t, k_1)}{t^2}dt. \end{aligned} \tag{B.41}$$

Similarly,

$$U(w, k_2) \;=\; U(k_1, k_2) + \int_{k_1}^{w} h'\left(\frac{y(t, k_2)}{t}\right)\frac{y(t, k_2)}{t^2}dt. \tag{B.42}$$

Making use of (B.41) and (B.42) gives rise to

$$U(w, k_1) - U(w, k_2) \;\geq\; \int_{k_1}^{w}\left[h'\left(\frac{y(t, k_1)}{t}\right)\frac{y(t, k_1)}{t^2} - h'\left(\frac{y(t, k_2)}{t}\right)\frac{y(t, k_2)}{t^2}\right]dt. \tag{B.43}$$

By applying the same reasoning used in step 2 to (B.43), we arrive at

$$U(w, w) \;\geq\; U(w, k_1) \;\geq\; U(w, k_2), \quad \forall w > k_1 > k_2. \tag{B.44}$$

Accordingly, (B.40) combined with (B.44) reveals that the preference of any given type of worker is indeed (weakly) single-peaked on the set of types. By applying the Black's Median Voter Theorem (see [66]), the desired assertion hence follows. **QED**

Proof of Propositions 3.6.1 and 3.6.2: To prove these two propositions, we just need to prove the following lemma.

Lemma B.1.4. *Suppose $\tilde{w}_m = w_m$, then we have the following predictions.*

(i) We have $\tau^R(w) > \hat{\tau}^M(w) > \tau^M(w)$ *if* $\Theta^M(w) < \tilde{\theta}(w) < \Theta^{MR}(w)$ *and* $\hat{\tau}^R(w) > \tau^R(w) > \hat{\tau}^M(w)$ *if* $\Theta^R(w) < \tilde{\theta}(w) < \Theta^{MR}(w)$*, in which* $\hat{\tau}^M(w)$ *and* $\hat{\tau}^R(w)$ *denote the maximax and maximin marginal tax rates derived by [14] and* $\Theta^M(w)$*,* $\Theta^R(w)$ *and* $\Theta^{MR}(w)$ *are respectively given by (B.53), (B.56) and (B.59).*

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

$$\begin{cases} \frac{f(w)}{F(w)} < \frac{\tilde{f}(w)}{\Gamma(\underline{w},w)} & \text{for } \forall w \in (\underline{w}, w_m], \\ \frac{f(w)}{1-F(w)} > \frac{\tilde{f}(w)}{\Gamma(\underline{w},\overline{w})-\Gamma(\underline{w},w)} & \text{for } \forall w \in (w_m, \overline{w}), \end{cases} \tag{B.45}$$

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

We shall complete the proof this lemma in 5 steps.

Step 1. The tax formulas of [14] are given as follows:

$$\hat{\tau}^R(w) = \frac{1-F(w)}{wf(w)}\left[\frac{1}{w}h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^2}h''\left(\frac{y(w)}{w}\right)\right], \tag{B.46}$$

$$\hat{\tau}^M(w) = -\frac{F(w)}{wf(w)}\left[\frac{1}{w}h'\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w^2}h''\left(\frac{y(w)}{w}\right)\right]. \tag{B.47}$$

By using our notation given by (B.13), (B.46) and (B.47) can be rewritten as

$$\hat{\tau}^R(w) = \frac{1-F(w)}{wf(w)}\frac{\partial U(w)}{\partial y(w)}, \tag{B.48}$$

$$\hat{\tau}^M(w) = -\frac{F(w)}{wf(w)}\frac{\partial U(w)}{\partial y(w)}. \tag{B.49}$$

Step 2. We can use (B.29) to rewrite (3.16) as

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

Substituting (B.32) into (B.50) and using (B.13) and (B.31), we can rewrite $\tau^M(w)$ as

$$\tau^M(w) = \underbrace{\frac{\partial U(w)}{\partial y(w)}}_{>0} \left\{ \underbrace{\frac{\tilde{\theta}(w)}{c(w)} \left[h\left(\frac{y(w)}{w}\right) - y(w) - \frac{y(w)}{w} h'\left(\frac{y(w)}{w}\right) \right]}_{<0} - \underbrace{\frac{\Gamma(\underline{w}, w)}{w\tilde{f}(w)}}_{>0} \right\}. \tag{B.51}$$

By using (B.49) and (B.51), we obtain

$$\begin{aligned} &\hat{\tau}^M(w) - \tau^M(w) \\ &\quad = \underbrace{\frac{\partial U(w)}{\partial y(w)}}_{>0} \left\{ \frac{\tilde{\theta}(w)}{c(w)} \left[y(w) - h\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w} h'\left(\frac{y(w)}{w}\right) \right] - \left[\frac{F(w)}{wf(w)} - \frac{\Gamma(\underline{w}, w)}{w\tilde{f}(w)} \right] \right\} \end{aligned} \tag{B.52}$$

for $\forall w \in (\underline{w}, w_m]$. We first define by using (B.52) that

$$\Theta^M(w) \equiv \frac{c(w)}{wl(w) - h(l(w)) + l(w)h'(l(w))} \left[\frac{F(w)}{wf(w)} - \frac{\Gamma(\underline{w}, w)}{w\tilde{f}(w)} \right]. \tag{B.53}$$

So using (B.52) and (B.53) gives rise to

$$\hat{\tau}^M(w) \begin{cases} < \tau^M(w) & \text{if } \tilde{\theta}(w) < \Theta^M(w), \\ = \tau^M(w) & \text{if } \tilde{\theta}(w) = \Theta^M(w), \\ > \tau^M(w) & \text{if } \tilde{\theta}(w) > \Theta^M(w). \end{cases} \tag{B.54}$$

Since by definition we have $\tilde{\theta}(w) > 0$, hence $\hat{\tau}^M(w) \leq \tau^M(w)$ predicted by (B.54) additionally requires that

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

as desired by (B.45).

Step 3. By using (B.48) and (B.33), we obtain

$$\hat{\tau}^R(w) - \tau^R(w)$$
$$= \underbrace{\frac{\partial U(w)}{\partial y(w)}}_{>0} \left\{ \frac{\tilde{\theta}(w)}{c(w)} \left[y(w) - h\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w} h'\left(\frac{y(w)}{w}\right) \right] - \left[\frac{\Gamma(w, \overline{w})}{w\tilde{f}(w)} - \frac{1 - F(w)}{wf(w)} \right] \right\} \tag{B.55}$$

for $\forall w \in (w_m, \overline{w})$. We first define by using (B.55) that

$$\Theta^R(w) \equiv \frac{c(w)}{wl(w) - h(l(w)) + l(w)h'(l(w))} \left[\frac{\Gamma(w, \overline{w})}{w\tilde{f}(w)} - \frac{1 - F(w)}{wf(w)} \right]. \tag{B.56}$$

Using (B.55) and (B.56) gives rise to

$$\hat{\tau}^R(w) \begin{cases} < \tau^R(w) & \text{if } \tilde{\theta}(w) < \Theta^R(w), \\ = \tau^R(w) & \text{if } \tilde{\theta}(w) = \Theta^R(w), \\ > \tau^R(w) & \text{if } \tilde{\theta}(w) > \Theta^R(w). \end{cases} \tag{B.57}$$

Since by definition we have $\tilde{\theta}(w) > 0$, hence $\hat{\tau}^R(w) \leq \tau^R(w)$ predicted by (B.57) additionally requires that

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

as desired by (B.45).

Step 4. Applying (B.49) and (B.33) reveals that

$$\hat{\tau}^M(w) - \tau^R(w)$$
$$= \underbrace{-\frac{\partial U(w)}{\partial y(w)}}_{<0} \left\{ \underbrace{\frac{\Gamma(w, \overline{w})}{w\tilde{f}(w)} + \frac{F(w)}{wf(w)}}_{>0} - \frac{\tilde{\theta}(w)}{c(w)} \left[y(w) - h\left(\frac{y(w)}{w}\right) + \frac{y(w)}{w} h'\left(\frac{y(w)}{w}\right) \right] \right\} \tag{B.58}$$

for $\forall w \in (w_m, \overline{w})$. We first define by using (B.58) that

$$\Theta^{MR}(w) \equiv \frac{c(w)}{wl(w) - h(l(w)) + l(w)h'(l(w))} \left[\frac{\Gamma(w, \overline{w})}{w\tilde{f}(w)} + \frac{F(w)}{wf(w)} \right]. \tag{B.59}$$

Thus, using (B.58) and (B.59) gives rise to

$$\hat{\tau}^M(w) \begin{cases} < \tau^R(w) & \text{if } \tilde{\theta}(w) < \Theta^{MR}(w), \\ = \tau^R(w) & \text{if } \tilde{\theta}(w) = \Theta^{MR}(w), \\ > \tau^R(w) & \text{if } \tilde{\theta}(w) > \Theta^{MR}(w). \end{cases} \tag{B.60}$$

<u>*Step 5.*</u> Applying (B.48) and (B.51) reveals that

$$\begin{aligned} &\hat{\tau}^R(w) - \tau^M(w) \\ &= \underbrace{\frac{\partial U(w)}{\partial y(w)}}_{>0} \left\{ \underbrace{\frac{\Gamma(\underline{w}, w)}{w\tilde{f}(w)} + \frac{1 - F(w)}{wf(w)}}_{>0} - \underbrace{\frac{\tilde{\theta}(w)}{c(w)} \left[h\left(\frac{y(w)}{w}\right) - y(w) - \frac{y(w)}{w} h'\left(\frac{y(w)}{w}\right) \right]}_{<0} \right\} \end{aligned} \tag{B.61}$$

for $\forall w \in (\underline{w}, w_m]$. Then, it is straightforward by (B.61) that $\hat{\tau}^R(w) > \tau^M(w)$. Finally, it is easy to verify by using equations (B.53), (B.56) and (B.59) that $\Theta^M(w) < \Theta^{MR}(w)$ and $\Theta^R(w) < \Theta^{MR}(w)$ for $\forall w \in (\underline{w}, \overline{w})$, as desired in part (i) of Lemma B.1.4. **QED**

B.2 The Complete Solution of Tax Design

Following [27] and [14], if either the maximin or the maximax income schedule obtained using the first-order approach fails to satisfy the SOIC condition (3.7), then it is necessary to bunch all types in a decreasing part of the schedule with some types who are in an increasing part. This kind of surgery is known as ironing, and any bunching region must be a closed interval.

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

Theorem B.2.1. *For any proposer of type $k \in (\underline{w}, \overline{w})$, the selfishly optimal schedule of pre-tax incomes, denoted by $y^*(\cdot)$, is given as follows.*

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

$$y^*(w) = \begin{cases} \underline{y}^{M*}(\underline{w}) & \textit{for } w \in [\underline{w}, w_\eta] \textit{ if } w_\eta \leq w_\alpha, \\ y^{M*}(w) & \textit{for } w \in (w_\eta, w_\alpha), \\ y^{M*}(w_\alpha) & \textit{for } w \in [w_\alpha, w_\beta] \textit{ if } w_\alpha > \underline{w}, \\ y^{R*}(w_\beta) & \textit{for } w \in [w_\alpha, w_\beta] \textit{ if } w_\beta < \overline{w}, \\ y^{R*}(w) & \textit{for } w \in (w_\beta, w_\gamma), \\ \overline{y}^{R*}(\overline{w}) & \textit{for } w \in [w_\gamma, \overline{w}] \textit{ if } w_\gamma \geq w_\beta, \end{cases} \tag{B.62}$$

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

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

$$y^*(w) = \begin{cases} \underline{y}^{M*}(\underline{w}) & \textit{for } w = \underline{w}, \\ y^{M*}(w) & \textit{for } w \in (\underline{w}, w_\alpha), \\ y^{M*}(w_\alpha) & \textit{for } w \in [w_\alpha, w_\beta] \textit{ if } w_\alpha > \underline{w}, \\ y^{R*}(w_\beta) & \textit{for } w \in [w_\alpha, w_\beta] \textit{ if } w_\beta < \overline{w}, \\ y^{R*}(w) & \textit{for } w \in (w_\beta, w_\gamma), \\ \overline{y}^{R*}(\overline{w}) & \textit{for } w \in [w_\gamma, \overline{w}] \textit{ if } w_\gamma \geq w_\beta, \end{cases} \tag{B.63}$$

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

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

$$y^*(w) = \begin{cases} \underline{y}^{M*}(\underline{w}) & \text{for } w \in [\underline{w}, w_\eta] \text{ if } w_\eta \leq w_\alpha, \\ y^{M*}(w) & \text{for } w \in (w_\eta, w_\alpha), \\ y^{M*}(w_\alpha) & \text{for } w \in [w_\alpha, w_\beta] \text{ if } w_\alpha > \underline{w}, \\ y^{R*}(w_\beta) & \text{for } w \in [w_\alpha, w_\beta] \text{ if } w_\beta < \overline{w}, \\ y^{R*}(w) & \text{for } w \in (w_\beta, \overline{w}), \\ \overline{y}^{R*}(\overline{w}) & \text{for } w = \overline{w}, \end{cases} \tag{B.64}$$

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

Proof. We shall complete the proof in 5 steps.

Step 1. By Proposition 3.4.3 (ii), there always exists a downward discontinuity at $w = k$, which hence requires the surgery of ironing in the current complete solution. By Proposition 3.4.3 (i) and (iii), there would be two additional downward discontinuities at, respectively, $w = \underline{w}$ and $w = \overline{w}$ whenever $-\overline{w}U'(\overline{w}) < U(\underline{w}) < 0$ holds. This case corresponds to part (i) of Theorem B.2.1. Similarly, based on Proposition 3.4.3, part (ii) of Theorem B.2.1 considers the case with two discontinuities at $w = k$ and $w = \overline{w}$, and part (iii) of Theorem B.2.1 considers the case with two discontinuities at $w = k$ and $w = \underline{w}$. Importantly, it follows from Proposition 3.4.3 that at least one of these two endpoints $w = \underline{w}$ and $w = \overline{w}$ exhibits a downward discontinuity in the schedule derived under the first-order approach. Due to the similarity between these cases, here we just prove part (i) of Theorem B.2.1 to economize on the space.

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

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

We now equivalently rewrite the maximand of problem (B.1) as follows

$$\begin{aligned} U(k) = &\int_{\underline{w}}^{k} \left\{ \left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \frac{\tilde{f}(w)}{\Gamma(\underline{w}, \overline{w})} + \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \left[1 - \frac{\Gamma(w, \overline{w})}{\Gamma(\underline{w}, \overline{w})} \right] \right\} dw \\ &+ \int_{k}^{\overline{w}} \left\{ \left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \frac{\tilde{f}(w)}{\Gamma(\underline{w}, \overline{w})} - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \frac{\Gamma(w, \overline{w})}{\Gamma(\underline{w}, \overline{w})} \right\} dw. \end{aligned} \tag{B.65}$$

Taking into account the bunching possibility, (B.65) should be modified as follows:

$$\begin{aligned} U^*(k) = &\int_{\underline{w}}^{k} \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w|w \notin \mathcal{B}^M\}} dw \\ &+ \int_{\underline{w}}^{k} \tilde{\Phi}^{M*}(w, y(w), \Gamma(w_-, w_+), \Gamma(w_-, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w|w \in \mathcal{B}^M\}} dw \\ &+ \int_{k}^{\overline{w}} \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w|w \notin \mathcal{B}^R\}} dw \\ &+ \int_{k}^{\overline{w}} \tilde{\Phi}^{R*}(w, y(w), \Gamma(w_-, w_+), \Gamma(w_+, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w|w \in \mathcal{B}^R\}} dw, \end{aligned} \tag{B.66}$$

in which

$$\begin{aligned} &\tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \\ \equiv &\left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \frac{\tilde{f}(w)}{\Gamma(\underline{w}, \overline{w})} + \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \left[1 - \frac{\Gamma(w, \overline{w})}{\Gamma(\underline{w}, \overline{w})} \right]; \\ &\tilde{\Phi}^{M*}(w, y(w), \Gamma(w_-, w_+), \Gamma(w_-, \overline{w}), \Gamma(\underline{w}, \overline{w})) \\ \equiv &\left[y(w) - h\left(\frac{y(w)}{w}\right) \right] \frac{\Gamma(w_-, w_+)}{\Gamma(\underline{w}, \overline{w})} + \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \left[1 - \frac{\Gamma(w_-, \overline{w})}{\Gamma(\underline{w}, \overline{w})} \right] \end{aligned} \tag{B.67}$$

and

$$
\begin{aligned}
&\tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \\
&\equiv \left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \frac{\tilde{f}(w)}{\Gamma(\underline{w}, \overline{w})} - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \frac{\Gamma(w, \overline{w})}{\Gamma(\underline{w}, \overline{w})}; \\
&\tilde{\Phi}^{R*}(w, y(w), \Gamma(w_-, w_+), \Gamma(w_+, \overline{w}), \Gamma(\underline{w}, \overline{w})) \\
&\equiv \left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \frac{\Gamma(w_-, w_+)}{\Gamma(\underline{w}, \overline{w})} - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \frac{\Gamma(w_+, \overline{w})}{\Gamma(\underline{w}, \overline{w})}
\end{aligned}
\tag{B.68}
$$

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 (B.66) are needed for types that are not bunched. Departing from the first-order approach, if an extra unit of consumption is given to type-w workers, it must be given to all workers who are bunched with them, whose mass is $\Gamma(w_-, w_+)$. Also, if w is bunched, in the maximax case, some of this extra consumption can be reclaimed from workers of lower types than those bunched with w, whose mass is $\Gamma(\underline{w}, \overline{w}) - \Gamma(w_-, \overline{w})$. The corresponding workers in the maximin case are those workers of higher types than those bunched with w, whose mass is $\Gamma(w_+, \overline{w})$.

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

$$
\begin{aligned}
\max_{y(\cdot)} \Bigg\{ &\int_{\underline{w}}^{w_\eta} \tilde{\Phi}^{M*}(w, y(w), \Gamma(w_-, w_+), \Gamma(w_-, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta \in (\underline{w}, w_\alpha)\} \cap \{w_- = \underline{w}\} \cap \{w_+ = w_\eta\}} dw \\
&+ \int_{w_\eta}^{w_\alpha} \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta \leq w_\alpha\}} dw \\
&+ \int_{w_\alpha}^{k} \tilde{\Phi}^{M*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\alpha, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}} dw \\
&+ \int_{k}^{w_\beta} \tilde{\Phi}^{R*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\beta < \overline{w}\}} dw \\
&+ \int_{w_\beta}^{w_\gamma} \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\gamma \geq w_\beta\}} dw \\
&+ \int_{w_\gamma}^{\overline{w}} \tilde{\Phi}^{R*}(w, y(w), \Gamma(w_-, w_+), \Gamma(w_+, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\gamma \in (w_\beta, \overline{w})\} \cap \{w_- = w_\gamma\} \cap \{w_+ = \overline{w}\}} dw \Bigg\},
\end{aligned}
\tag{B.69}
$$

subject to

$$y(w) = \begin{cases} y^*(\underline{w}, w_\eta) & \text{for } w \in [\underline{w}, w_\eta], \\ y^*(w_\alpha, w_\beta) & \text{for } w \in [w_\alpha, w_\beta], \\ y^*(w_\gamma, \overline{w}) & \text{for } w \in [w_\gamma, \overline{w}]. \end{cases} \tag{B.70}$$

In consequence, for the current purpose, problem (B.69) can be simplified as the following unconstrained maximization problem:

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

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

$$\begin{aligned} & \frac{\partial \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial y(w)} \\ & + \frac{\partial \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial \tilde{f}(w)} \frac{\partial \tilde{f}(w)}{\partial y(w)} \\ & + \frac{\partial \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial \Gamma(w, \overline{w})} \frac{\partial \Gamma(w, \overline{w})}{\partial y(w)} = 0 \ \text{ for } \forall w \in (w_\eta, w_\alpha), \end{aligned} \tag{B.71}$$

and

$$\begin{aligned} & \frac{\partial \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial y(w)} \\ & + \frac{\partial \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial \tilde{f}(w)} \frac{\partial \tilde{f}(w)}{\partial y(w)} \\ & + \frac{\partial \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial \Gamma(w, \overline{w})} \frac{\partial \Gamma(w, \overline{w})}{\partial y(w)} = 0 \ \text{ for } \forall w \in (w_\beta, w_\gamma). \end{aligned} \tag{B.72}$$

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

Step 5. We now show that $y^*(w_\alpha, w_\beta) = y^{M*}(w_\alpha)$ if $w_\alpha > \underline{w}$ and $y^*(w_\alpha, w_\beta) = y^{R*}(w_\beta)$ if $w_\beta < \overline{w}$. Based on the above ironing procedure, we can apply the same reasoning used to prove the Proposition 3 of [14] 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 (3.7) must therefore bind at k', which implies that the slope of $y^*(\cdot)$ is zero at k'. Since $y^*(\cdot)$ is continuous, we obtain that there exists a $w_\beta > k'$ such that $y^*(\cdot)$ is constant on $[k, w_\beta]$ and coincides with the maximin income schedule $y^{R*}(\cdot)$ on $[w_\beta, w_\gamma]$. Similarly, if there exists a type $k' < k$ for which $y^*(k')$ is not the maximax income, formally $y^*(k') \neq y^{M*}(k')$, we can use the same argument to show that there exists a $w_\alpha < k'$ such that $y^*(\cdot)$ is constant on $[w_\alpha, k]$ and coincides with the maximax income schedule $y^{M*}(\cdot)$ on $[w_\eta, w_\alpha]$.

By further setting $y^*(\underline{w}, w_\eta) \equiv \underline{y}^{M*}(\underline{w})$ and $y^*(w_\gamma, \overline{w}) \equiv \overline{y}^{R*}(\overline{w})$ in (B.70), then the desired income schedule given by (B.62) is established. □

Theorem B.2.2. *For any proposer of type $k \in (\underline{w}, \overline{w})$, the bridge endpoints are chosen as follows.*

(i) The optimal values of the bridge endpoints w_α and w_β are determined by the first-order condition (B.75) for $w_\beta < \overline{w}$ and by the first-order condition (B.91) for $w_\alpha > \underline{w}$.

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

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

Proof. We shall complete the proof in 4 steps.

Step 1. We first determine the optimal endpoints w_α and w_β of the bridge connecting the maximax income schedule and the maximin income schedule. Our proof employs the procedure developed by [14].

Suppose $w_\beta < \overline{w}$ holds. By continuity of income schedule $y^*(\cdot)$, we get from Theorem B.2.1 that $y^*(w_\beta) = y^{R*}(w_\beta)$. Also, $y^*(w_\beta) = y^*(w_\alpha)$ because income is a constant on the bridge. If we also have $w_\alpha > \underline{w}$, then by continuity again, $y^*(w_\alpha) = y^{M*}(w_\alpha)$. Define

$$\psi(w_\beta) \equiv \begin{cases} (y^{M*})^{-1}(y^{R*}(w_\beta)) & \text{if } w_\alpha > \underline{w}, \\ w_\alpha & \text{if } w_\alpha = \underline{w}. \end{cases} \tag{B.73}$$

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

$$\begin{aligned}\Xi(w_\beta; k) \equiv & \int_{w_\eta}^{\psi(w_\beta)} \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta \leq w_\alpha\} \cap \{y(w) = y^{M*}(w)\}} dw \\ & + \int_{\psi(w_\beta)}^{k} \tilde{\Phi}^{M*}(w, y(w), \Gamma(\psi(w_\beta), w_\beta), \Gamma(\psi(w_\beta), \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\} \cap \left\{y(w) = y^{R*}(w_\beta)\right\}} dw \\ & + \int_{k}^{w_\beta} \tilde{\Phi}^{R*}(w, y(w), \Gamma(\psi(w_\beta), w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\} \cap \left\{y(w) = y^{R*}(w_\beta)\right\}} dw \\ & + \int_{w_\beta}^{w_\gamma} \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{w_\gamma \geq w_\beta\right\} \cap \{y(w) = y^{R*}(w)\}} dw. \end{aligned} \tag{B.74}$$

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

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

Using (B.74), the first-order condition with respect to w_β can be derived as

$$\Psi_1 + \Psi_2(k) + \Psi_3(k) + \Psi_4 = 0, \tag{B.75}$$

in which

$$\begin{aligned} & \Psi_1 = \\ & \frac{\mathrm{d}\psi(w_\beta)}{\mathrm{d}w_\beta} \Big\{ \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta \leq w_\alpha\} \cap \{y(w) = y^{M*}(w)\} \cap \left\{w = \psi(w_\beta)\right\}} \\ & - \tilde{\Phi}^{M*}(w, y(w), \Gamma(\psi(w_\beta), w_\beta), \Gamma(\psi(w_\beta), \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\} \cap \left\{y(w) = y^{R*}(w_\beta)\right\} \cap \left\{w = \psi(w_\beta)\right\}} \Big\}, \end{aligned} \tag{B.76}$$

$$
\begin{aligned}
&\Psi_2(k) = \\
&\int_{\psi(w_\beta)}^{k} \frac{\partial \tilde{\Phi}^{M*}(w, y^{R*}(w_\beta), \Gamma(\psi(w_\beta), w_\beta), \Gamma(\psi(w_\beta), \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial y^{R*}(w_\beta)} \frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta} \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}} dw \\
&+ \int_{\psi(w_\beta)}^{k} \frac{\partial \tilde{\Phi}^{M*}(w, y^{R*}(w_\beta), \Gamma(\psi(w_\beta), w_\beta), \Gamma(\psi(w_\beta), \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial \Gamma(\psi(w_\beta), w_\beta)} \frac{\partial \Gamma(\psi(w_\beta), w_\beta)}{\partial w_\beta} \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}} dw \\
&+ \int_{\psi(w_\beta)}^{k} \frac{\partial \tilde{\Phi}^{M*}(w, y^{R*}(w_\beta), \Gamma(\psi(w_\beta), w_\beta), \Gamma(\psi(w_\beta), \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial \Gamma(\psi(w_\beta), \overline{w})} \frac{\partial \Gamma(\psi(w_\beta), \overline{w})}{\partial w_\beta} \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}} dw \\
&\equiv \int_{\psi(w_\beta)}^{k} [\Psi_{21}(w) + \Psi_{22}(w) + \Psi_{23}(w)] \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}} dw,
\end{aligned}
\tag{B.77}
$$

$$
\begin{aligned}
&\Psi_3(k) = \\
&\int_{k}^{w_\beta} \frac{\partial \tilde{\Phi}^{R*}(w, y^{R*}(w_\beta), \Gamma(\psi(w_\beta), w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial y^{R*}(w_\beta)} \frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta} \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\}} dw \\
&+ \int_{k}^{w_\beta} \frac{\partial \tilde{\Phi}^{R*}(w, y^{R*}(w_\beta), \Gamma(\psi(w_\beta), w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial \Gamma(\psi(w_\beta), w_\beta)} \frac{\partial \Gamma(\psi(w_\beta), w_\beta)}{\partial w_\beta} \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\}} dw \\
&+ \int_{k}^{w_\beta} \frac{\partial \tilde{\Phi}^{R*}(w, y^{R*}(w_\beta), \Gamma(\psi(w_\beta), w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial \Gamma(w_\beta, \overline{w})} \frac{\partial \Gamma(w_\beta, \overline{w})}{\partial w_\beta} \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\}} dw \\
&\equiv \int_{k}^{w_\beta} [\Psi_{31}(w) + \Psi_{32}(w) + \Psi_{33}(w)] \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\}} dw,
\end{aligned}
\tag{B.78}
$$

and

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

Step 2. By using (B.67) and (B.68), we can have

$$\begin{aligned}
&\tilde{\Phi}^{M*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\alpha, \overline{w}), \Gamma(\underline{w}, \overline{w})) \\
&\quad \equiv \left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \frac{\Gamma(w_\alpha, w_\beta)}{\Gamma(\underline{w}, \overline{w})} + \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right)\left[1 - \frac{\Gamma(w_\alpha, \overline{w})}{\Gamma(\underline{w}, \overline{w})}\right], \\
&\tilde{\Phi}^{R*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w})) \\
&\quad \equiv \left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \frac{\Gamma(w_\alpha, w_\beta)}{\Gamma(\underline{w}, \overline{w})} - \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \frac{\Gamma(w_\beta, \overline{w})}{\Gamma(\underline{w}, \overline{w})}
\end{aligned} \tag{B.80}$$

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

By using (B.73), (B.80), (B.67) and (B.68), we can rewrite (B.76) as

$$\begin{aligned}
\Psi_1 =& \frac{\mathrm{d}\psi(w_\beta)}{\mathrm{d}w_\beta}\Big\{\tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta \le w_\alpha\} \cap \{y(w) = y^{R*}(w_\beta)\} \cap \{w = \psi(w_\beta)\}} \\
&- \tilde{\Phi}^{M*}(w, y(w), \Gamma(w, w_\beta), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\} \cap \{y(w) = y^{R*}(w_\beta)\} \cap \{w = \psi(w_\beta)\}}\Big\} \\
=& \frac{\mathrm{d}\psi(w_\beta)}{\mathrm{d}w_\beta}\Big\{\tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \\
&- \tilde{\Phi}^{M*}(w, y(w), \Gamma(w, w_\beta), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w}))\Big\} \cdot \mathbb{I}_{\{w_\eta \le w_\alpha\} \cap \{w_\alpha > \underline{w}\} \cap \{y(w) = y^{R*}(w_\beta)\} \cap \{w = \psi(w_\beta)\}} \\
=& \frac{\mathrm{d}\psi(w_\beta)}{\mathrm{d}w_\beta}\left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \frac{\tilde{f}(w) - \Gamma(w, w_\beta)}{\Gamma(\underline{w}, \overline{w})} \cdot \mathbb{I}_{\{w_\eta \le w_\alpha\} \cap \{w_\alpha > \underline{w}\} \cap \{y(w) = y^{R*}(w_\beta)\} \cap \{w = \psi(w_\beta)\}} \\
=& \frac{\mathrm{d}\psi(w_\beta)}{\mathrm{d}w_\beta}\left[y^{R*}(w_\beta) - h\left(\frac{y^{R*}(w_\beta)}{\psi(w_\beta)}\right)\right] \frac{\tilde{f}(\psi(w_\beta)) - \Gamma(\psi(w_\beta), w_\beta)}{\Gamma(\underline{w}, \overline{w})} \cdot \mathbb{I}_{\{w_\eta \le w_\alpha\} \cap \{w_\alpha > \underline{w}\}}.
\end{aligned} \tag{B.81}$$

By using (B.77) and (B.80), we have

$$\begin{aligned}
\Psi_{21}(w) = &\Big\{\left[1 - \frac{1}{w} h'\left(\frac{y^{R*}(w_\beta)}{w}\right)\right] \frac{\Gamma(\psi(w_\beta), w_\beta)}{\Gamma(\underline{w}, \overline{w})} \\
&+ \left[\frac{1}{w^2} h'\left(\frac{y^{R*}(w_\beta)}{w}\right) + \frac{y^{R*}(w_\beta)}{w^3} h''\left(\frac{y^{R*}(w_\beta)}{w}\right)\right]\left[1 - \frac{\Gamma(\psi(w_\beta), \overline{w})}{\Gamma(\underline{w}, \overline{w})}\right]\Big\} \frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta},
\end{aligned} \tag{B.82}$$

$$\Psi_{22}(w) = \left[y^{R*}(w_\beta) - h\left(\frac{y^{R*}(w_\beta)}{w}\right)\right]\frac{1}{\Gamma(\underline{w},\overline{w})} \times \left\{\tilde{f}(w_\beta) - \tilde{f}\left(\psi(w_\beta)\right)\frac{\mathrm{d}\psi(w_\beta)}{\mathrm{d}w_\beta} + \left[\int_{\psi(w_\beta)}^{w_\beta}\frac{\partial\tilde{f}(w)}{\partial y^{R*}(w_\beta)}\frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta}dw\right]\right\}, \tag{B.83}$$

and

$$\Psi_{23}(w) = \frac{y^{R*}(w_\beta)}{w^2}h'\left(\frac{y^{R*}(w_\beta)}{w}\right)\frac{\tilde{f}\left(\psi(w_\beta)\right)}{\Gamma(\underline{w},\overline{w})}\frac{\mathrm{d}\psi(w_\beta)}{\mathrm{d}w_\beta}. \tag{B.84}$$

Similarly, By using (B.78) and (B.80), we have

$$\Psi_{31}(w) = \left\{\left[1 - \frac{1}{w}h'\left(\frac{y^{R*}(w_\beta)}{w}\right)\right]\frac{\Gamma(\psi\left(w_\beta\right), w_\beta)}{\Gamma(\underline{w},\overline{w})} - \left[\frac{1}{w^2}h'\left(\frac{y^{R*}(w_\beta)}{w}\right) + \frac{y^{R*}(w_\beta)}{w^3}h''\left(\frac{y^{R*}(w_\beta)}{w}\right)\right]\frac{\Gamma(w_\beta,\overline{w})}{\Gamma(\underline{w},\overline{w})}\right\}\frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta}, \tag{B.85}$$

$$\Psi_{32}(w) = \left[y^{R*}(w_\beta) - h\left(\frac{y^{R*}(w_\beta)}{w}\right)\right]\frac{1}{\Gamma(\underline{w},\overline{w})} \times \left\{\tilde{f}(w_\beta) - \tilde{f}\left(\psi(w_\beta)\right)\frac{\mathrm{d}\psi(w_\beta)}{\mathrm{d}w_\beta} + \left[\int_{\psi(w_\beta)}^{w_\beta}\frac{\partial\tilde{f}(w)}{\partial y^{R*}(w_\beta)}\frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta}dw\right]\right\}, \tag{B.86}$$

and

$$\Psi_{33}(w) = \frac{y^{R*}(w_\beta)}{w^2}h'\left(\frac{y^{R*}(w_\beta)}{w}\right)\frac{\tilde{f}\left(w_\beta\right)}{\Gamma(\underline{w},\overline{w})}. \tag{B.87}$$

Finally, by using (B.80), we can rewrite (B.79) as

$$\begin{aligned}
\Psi_4 &= \tilde{\Phi}^{R*}(w, y(w), \Gamma(\psi(w), w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\} \cap \left\{y(w) = y^{R*}(w_\beta)\right\} \cap \left\{w = w_\beta\right\}} \\
&\quad - \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{w_\gamma \geq w_\beta\right\} \cap \{y(w) = y^{R*}(w)\} \cap \left\{w = w_\beta\right\}} \\
&= \Big[\tilde{\Phi}^{R*}(w, y(w), \Gamma(\psi(w), w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) - \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w}))\Big] \\
&\quad \times \mathbb{I}_{\left\{w_\gamma \geq w_\beta\right\} \cap \left\{w_\beta < \overline{w}\right\} \cap \{y(w) = y^{R*}(w)\} \cap \left\{w = w_\beta\right\}} \\
&= \left\{\left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \frac{\Gamma\left(\psi(w), w\right) - \tilde{f}(w)}{\Gamma\left(\underline{w}, \overline{w}\right)}\right\} \mathbb{I}_{\left\{w_\gamma \geq w_\beta\right\} \cap \left\{w_\beta < \overline{w}\right\} \cap \{y(w) = y^{R*}(w)\} \cap \left\{w = w_\beta\right\}}.
\end{aligned} \tag{B.88}$$

<u>*Step 3.*</u> Suppose $w_\alpha > \underline{w}$ holds. By continuity of income schedule $y^*(\cdot)$, we get from Theorem B.2.1 that $y^*(w_\alpha) = y^{M*}(w_\alpha)$. Also, $y^*(w_\beta) = y^*(w_\alpha)$ because income is a constant on the bridge. If we also have $w_\beta < \overline{w}$, then by continuity again, $y^*(w_\beta) = y^{R*}(w_\beta)$. Define

$$\varphi(w_\alpha) \equiv \begin{cases} (y^{R*})^{-1}(y^{M*}(w_\alpha)) & \text{if } w_\beta < \overline{w}, \\ w_\beta & \text{if } w_\beta = \overline{w}. \end{cases} \tag{B.89}$$

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

$$\begin{aligned}
\Xi(w_\alpha; k) \equiv &\int_{w_\eta}^{w_\alpha} \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta \leq w_\alpha\} \cap \{y(w) = y^{M*}(w)\}} dw \\
&+ \int_{w_\alpha}^{k} \tilde{\Phi}^{M*}(w, y(w), \Gamma(w_\alpha, \varphi(w_\alpha)), \Gamma(w_\alpha, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\} \cap \{y(w) = y^{M*}(w_\alpha)\}} dw \\
&+ \int_{k}^{\varphi(w_\alpha)} \tilde{\Phi}^{R*}(w, y(w), \Gamma(w_\alpha, \varphi(w_\alpha)), \Gamma(\varphi(w_\alpha), \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\} \cap \{y(w) = y^{M*}(w_\alpha)\}} dw \\
&+ \int_{\varphi(w_\alpha)}^{w_\gamma} \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{w_\gamma \geq w_\beta\right\} \cap \{y(w) = y^{R*}(w)\}} dw.
\end{aligned} \tag{B.90}$$

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

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

Using (B.90), the first-order condition with respect to w_α can be derived as

$$\Lambda_1 + \Lambda_2(k) + \Lambda_3(k) + \Lambda_4 = 0, \tag{B.91}$$

in which

$$\begin{aligned}\Lambda_1 = & \left[\tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) - \tilde{\Phi}^{M*}(w, y(w), \Gamma(w, \varphi(w)), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w}))\right] \\ & \times \mathbb{I}_{\{w_\eta \le w_\alpha\}\cap\{w_\alpha > \underline{w}\}\cap\{y(w)=y^{M*}(w)\}\cap\{w=w_\alpha\}} \\ = & \left\{\left[y(w) - h\left(\frac{y(w)}{w}\right)\right]\frac{\tilde{f}(w) - \Gamma(w, \varphi(w))}{\Gamma(\underline{w}, \overline{w})}\right\}\mathbb{I}_{\{w_\eta \le w_\alpha\}\cap\{w_\alpha > \underline{w}\}\cap\{y(w)=y^{M*}(w)\}\cap\{w=w_\alpha\}},\end{aligned} \tag{B.92}$$

$$\Lambda_2(k) = \int_{w_\alpha}^{k} \left[\Lambda_{21}(w) + \Lambda_{22}(w) + \Lambda_{23}(w)\right] \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}} dw, \tag{B.93}$$

with

$$\begin{aligned}\Lambda_{21}(w) = & \left\{\left[1 - \frac{1}{w}h'\left(\frac{y^{M*}(w_\alpha)}{w}\right)\right]\frac{\Gamma(w_\alpha, \varphi(w_\alpha))}{\Gamma(\underline{w}, \overline{w})}\right. \\ & \left. + \left[\frac{1}{w^2}h'\left(\frac{y^{M*}(w_\alpha)}{w}\right) + \frac{y^{M*}(w_\alpha)}{w^3}h''\left(\frac{y^{M*}(w_\alpha)}{w}\right)\right]\left[1 - \frac{\Gamma(w_\alpha, \overline{w})}{\Gamma(\underline{w}, \overline{w})}\right]\right\}\frac{\mathrm{d}y^{M*}(w_\alpha)}{\mathrm{d}w_\alpha},\end{aligned} \tag{B.94}$$

$$\begin{aligned}\Lambda_{22}(w) = & \left[y^{M*}(w_\alpha) - h\left(\frac{y^{M*}(w_\alpha)}{w}\right)\right]\frac{1}{\Gamma(\underline{w}, \overline{w})} \\ & \times \left\{\tilde{f}\left(\varphi(w_\alpha)\right)\frac{\mathrm{d}\varphi(w_\alpha)}{\mathrm{d}w_\alpha} - \tilde{f}(w_\alpha) + \left[\int_{w_\alpha}^{\varphi(w_\alpha)} \frac{\partial \tilde{f}(w)}{\partial y^{M*}(w_\alpha)}\frac{\mathrm{d}y^{M*}(w_\alpha)}{\mathrm{d}w_\alpha}dw\right]\right\},\end{aligned} \tag{B.95}$$

and

$$\Lambda_{23}(w) = \frac{y^{M*}(w_\alpha)}{w^2}h'\left(\frac{y^{M*}(w_\alpha)}{w}\right)\frac{\tilde{f}\left(w_\alpha\right)}{\Gamma(\underline{w}, \overline{w})}; \tag{B.96}$$

$$\Lambda_3(k) = \int_{k}^{\varphi(w_\alpha)} \left[\Lambda_{31}(w) + \Lambda_{32}(w) + \Lambda_{33}(w)\right] \cdot \mathbb{I}_{\{w_\beta < \overline{w}\}} dw, \tag{B.97}$$

with

$$\Lambda_{31}(w) = \left\{ \left[1 - \frac{1}{w}h'\left(\frac{y^{M*}(w_\alpha)}{w}\right)\right] \frac{\Gamma(w_\alpha, \varphi(w_\alpha))}{\Gamma(\underline{w}, \overline{w})} - \left[\frac{1}{w^2}h'\left(\frac{y^{M*}(w_\alpha)}{w}\right) + \frac{y^{M*}(w_\alpha)}{w^3}h''\left(\frac{y^{M*}(w_\alpha)}{w}\right)\right] \frac{\Gamma(\varphi(w_\alpha), \overline{w})}{\Gamma(\underline{w}, \overline{w})} \right\} \frac{\mathrm{d}y^{M*}(w_\alpha)}{\mathrm{d}w_\alpha}, \tag{B.98}$$

$$\Lambda_{32}(w) = \left[y^{M*}(w_\alpha) - h\left(\frac{y^{M*}(w_\alpha)}{w}\right)\right] \frac{1}{\Gamma(\underline{w}, \overline{w})} \times \left\{ \tilde{f}\left(\varphi(w_\alpha)\right) \frac{\mathrm{d}\varphi(w_\alpha)}{\mathrm{d}w_\alpha} - \tilde{f}(w_\alpha) + \left[\int_{w_\alpha}^{\varphi(w_\alpha)} \frac{\partial \tilde{f}(w)}{\partial y^{M*}(w_\alpha)} \frac{\mathrm{d}y^{M*}(w_\alpha)}{\mathrm{d}w_\alpha} dw\right] \right\}, \tag{B.99}$$

and

$$\Lambda_{33}(w) = \frac{y^{M*}(w_\alpha)}{w^2}h'\left(\frac{y^{M*}(w_\alpha)}{w}\right) \frac{\tilde{f}\left(\varphi(w_\alpha)\right)}{\Gamma(\underline{w}, \overline{w})} \frac{\mathrm{d}\varphi(w_\alpha)}{\mathrm{d}w_\alpha}; \tag{B.100}$$

and finally

$$\begin{aligned}
\Lambda_4 =& \frac{\mathrm{d}\varphi(w_\alpha)}{\mathrm{d}w_\alpha} \times \\
& [\tilde{\Phi}^{R*}(w, y(w), \Gamma(w_\alpha, w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\} \cap \{y(w) = y^{M*}(w_\alpha)\} \cap \{w = \varphi(w_\alpha)\}} \\
& - \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{w_\gamma \geq w_\beta\right\} \cap \{y(w) = y^{M*}(w_\alpha)\} \cap \{w = \varphi(w_\alpha)\}}] \\
=& \left[\tilde{\Phi}^{R*}(w, y(w), \Gamma(w_\alpha, w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) - \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w}))\right] \\
& \times \mathbb{I}_{\left\{w_\gamma \geq w_\beta\right\} \cap \left\{w_\beta < \overline{w}\right\} \cap \{y(w) = y^{M*}(w_\alpha)\} \cap \{w = \varphi(w_\alpha)\}} \frac{\mathrm{d}\varphi(w_\alpha)}{\mathrm{d}w_\alpha} \\
=& \left\{ \left[y^{M*}(w_\alpha) - h\left(\frac{y^{M*}(w_\alpha)}{\varphi(w_\alpha)}\right)\right] \frac{\Gamma\left(w_\alpha, \varphi(w_\alpha)\right) - \tilde{f}(\varphi(w_\alpha))}{\Gamma\left(\underline{w}, \overline{w}\right)} \right\} \frac{\mathrm{d}\varphi(w_\alpha)}{\mathrm{d}w_\alpha} \mathbb{I}_{\left\{w_\gamma \geq w_\beta\right\} \cap \left\{w_\beta < \overline{w}\right\}}.
\end{aligned} \tag{B.101}$$

Step 4. By using (B.69), we can establish the first-order conditions implicitly solving for $y(\underline{w})$ and $y(\overline{w})$, and we denote the solutions as $\underline{y}^{M*}(\underline{w})$ and $\overline{y}^{R*}(\overline{w})$, respectively. Since for each of these two bridges, one of the endpoints is already fixed, the other endpoint is actually fixed given the established optimal income schedule. That is, w_η is determined by setting $y^{M*}(w_\eta) = \underline{y}^{M*}(\underline{w})$ and w_γ is determined by setting $y^{R*}(w_\gamma) = \overline{y}^{R*}(\overline{w})$. □

Proposition B.2.1. *If the following condition*

$$\begin{cases} \frac{\mathrm{d}w_\alpha}{\mathrm{d}w_\beta} \geq \frac{\tilde{f}(w_\beta)}{\tilde{f}(w_\alpha)} & \textit{for } w_\beta < \overline{w}, \\ \frac{\tilde{f}(w_\alpha)}{\tilde{f}(w_\beta)} \geq \frac{\mathrm{d}w_\beta}{\mathrm{d}w_\alpha} & \textit{for } w_\alpha > \underline{w} \end{cases} \tag{B.102}$$

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

Proof. We shall complete the proof in 2 steps.

Step 1. We first consider the case with $w_\beta < \overline{w}$. It follows from (B.74) and (B.75) that

$$\frac{\partial^2 \Xi(w_\beta; k)}{\partial w_\beta \partial k} = \frac{\mathrm{d}\Psi_2(k)}{\mathrm{d}k} + \frac{\mathrm{d}\Psi_3(k)}{\mathrm{d}k}. \tag{B.103}$$

By using equations (B.82)-(B.87), (B.103) can be explicitly expressed as

$$\begin{aligned} & \frac{\mathrm{d}\Psi_2(k)}{\mathrm{d}k} + \frac{\mathrm{d}\Psi_3(k)}{\mathrm{d}k} \\ = {} & [\Psi_{21}(k) + \Psi_{22}(k) + \Psi_{23}(k)] - [\Psi_{31}(k) + \Psi_{32}(k) + \Psi_{33}(k)] \\ = {} & [\Psi_{21}(k) - \Psi_{31}(k)] + \underbrace{[\Psi_{22}(k) - \Psi_{32}(k)]}_{=0} + [\Psi_{23}(k) - \Psi_{33}(k)] \\ = {} & \left[\frac{1}{k^2} h'\left(\frac{y^{R*}(w_\beta)}{k}\right) + \frac{y^{R*}(w_\beta)}{k^3} h''\left(\frac{y^{R*}(w_\beta)}{k}\right)\right] \left[1 - \frac{\Gamma(\psi(w_\beta), \overline{w})}{\Gamma(\underline{w}, \overline{w})} + \frac{\Gamma(w_\beta, \overline{w})}{\Gamma(\underline{w}, \overline{w})}\right] \frac{\mathrm{d}y^{R*}(w_\beta)}{\mathrm{d}w_\beta} \\ & + \frac{y^{R*}(w_\beta)}{k^2} h'\left(\frac{y^{R*}(w_\beta)}{k}\right) \frac{1}{\Gamma(\underline{w}, \overline{w})} \left[\tilde{f}(\psi(w_\beta)) \frac{\mathrm{d}\psi(w_\beta)}{\mathrm{d}w_\beta} - \tilde{f}(w_\beta)\right]. \end{aligned} \tag{B.104}$$

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

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

As such

$$\frac{\partial^2 \Xi(w_\beta; k)}{\partial w_\beta \partial k} = \frac{\mathrm{d}\Psi_2(k)}{\mathrm{d}k} + \frac{\mathrm{d}\Psi_3(k)}{\mathrm{d}k} > 0 \tag{B.105}$$

whenever

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

as desired. In particular, $\mathrm{d}\psi(w_\beta)/\mathrm{d}w_\beta > 0$ based on the construction of $\psi(\cdot)$ given in the proof of Theorem B.2.2 as well as the monotonicity of the income schedule. In the case of (B.105), $\Xi(w_\beta; k)$ is a supermodular function, and an application of Topkis Theorem (see Topkis [67], Theorem 6.1) implies that $w_\beta(k)$ is nondecreasing in k.

Step 2. We now consider the case with $w_\alpha > \underline{w}$. It follows from (B.90) and (B.91) that

$$\frac{\partial^2 \Xi(w_\alpha; k)}{\partial w_\alpha \partial k} = \frac{\mathrm{d}\Lambda_2(k)}{\mathrm{d}k} + \frac{\mathrm{d}\Lambda_3(k)}{\mathrm{d}k}. \tag{B.106}$$

By using equations (B.93)-(B.100), (B.106) can be explicitly expressed as

$$\begin{aligned}
&\frac{\mathrm{d}\Lambda_2(k)}{\mathrm{d}k} + \frac{\mathrm{d}\Lambda_3(k)}{\mathrm{d}k} \\
=\ &[\Lambda_{21}(k) + \Lambda_{22}(k) + \Lambda_{23}(k)] - [\Lambda_{31}(k) + \Lambda_{32}(k) + \Lambda_{33}(k)] \\
=\ &[\Lambda_{21}(k) - \Lambda_{31}(k)] + \underbrace{[\Lambda_{22}(k) - \Lambda_{32}(k)]}_{=0} + [\Lambda_{23}(k) - \Lambda_{33}(k)] \\
=\ &\left[\frac{1}{k^2}h'\left(\frac{y^{M*}(w_\alpha)}{k}\right) + \frac{y^{M*}(w_\alpha)}{k^3}h''\left(\frac{y^{M*}(w_\alpha)}{k}\right)\right]\left[1 - \frac{\Gamma(w_\alpha, \overline{w})}{\Gamma(\underline{w}, \overline{w})} + \frac{\Gamma(\varphi(w_\alpha), \overline{w})}{\Gamma(\underline{w}, \overline{w})}\right]\frac{\mathrm{d}y^{M*}(w_\alpha)}{\mathrm{d}w_\alpha} \\
&+ \frac{y^{M*}(w_\alpha)}{k^2}h'\left(\frac{y^{M*}(w_\alpha)}{k}\right)\frac{1}{\Gamma(\underline{w}, \overline{w})}\left[\tilde{f}(w_\alpha) - \tilde{f}(\varphi(w_\alpha))\frac{\mathrm{d}\varphi(w_\alpha)}{\mathrm{d}w_\alpha}\right].
\end{aligned} \tag{B.107}$$

Since $\mathrm{d}y^{M*}(w_\alpha)/\mathrm{d}w_\alpha > 0$ by assumption, we have by (B.107) that

$$\frac{\partial^2 \Xi(w_\alpha; k)}{\partial w_\alpha \partial k} = \frac{\mathrm{d}\Lambda_2(k)}{\mathrm{d}k} + \frac{\mathrm{d}\Lambda_3(k)}{\mathrm{d}k} > 0 \tag{B.108}$$

whenever

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

as desired in (B.102). In the case of (B.108), $\Xi(w_\alpha; k)$ is a supermodular function, and an application of Topkis Theorem (see [67]) implies that $w_\alpha(k)$ is nondecreasing in k. □

Proposition B.2.2. *The bridge endpoint $w_\eta(k)$ is decreasing whereas the bridge endpoint $w_\gamma(k)$ is increasing in the type k of the proposer.*

Proof. We shall complete the proofs in 3 steps.

Step 1. By using (B.67), we have

$$\begin{aligned}&\tilde{\Phi}^{M*}(w, y(w), \Gamma(w_-, w_+), \Gamma(w_-, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta \in (\underline{w}, w_\alpha)\} \cap \{w_- = \underline{w}\} \cap \{w_+ = w_\eta\}} \\ &= \left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \frac{\Gamma(\underline{w}, w_\eta)}{\Gamma(\underline{w}, \overline{w})} \cdot \mathbb{I}_{\{w_\eta \in (\underline{w}, w_\alpha)\}}.\end{aligned} \tag{B.109}$$

By using (B.68), we have

$$\begin{aligned}&\tilde{\Phi}^{R*}(w, y(w), \Gamma(w_-, w_+), \Gamma(w_+, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{w_\gamma \in (w_\beta, \overline{w})\right\} \cap \{w_- = w_\gamma\} \cap \{w_+ = \overline{w}\}} \\ &= \left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \frac{\Gamma(w_\gamma, \overline{w})}{\Gamma(\underline{w}, \overline{w})} \cdot \mathbb{I}_{\left\{w_\gamma \in (w_\beta, \overline{w})\right\}}.\end{aligned} \tag{B.110}$$

Substituting (B.109) and (B.110) into (B.69) results in

$$\begin{aligned}U^*(k) = &\int_{\underline{w}}^{w_\eta} \left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \frac{\Gamma(\underline{w}, w_\eta)}{\Gamma(\underline{w}, \overline{w})} \cdot \mathbb{I}_{\{w_\eta \in (\underline{w}, w_\alpha)\}} dw \\ &+ \int_{w_\eta}^{w_\alpha} \tilde{\Phi}^{M*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\eta \leq w_\alpha\}} dw \\ &+ \int_{w_\alpha}^{k} \tilde{\Phi}^{M*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\alpha, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}} dw \\ &+ \int_{k}^{w_\beta} \tilde{\Phi}^{R*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\}} dw \\ &+ \int_{w_\beta}^{w_\gamma} \tilde{\Phi}^{R*}(w, y(w), \tilde{f}(w), \Gamma(w, \overline{w}), \Gamma(\underline{w}, \overline{w})) \cdot \mathbb{I}_{\left\{w_\gamma \geq w_\beta\right\}} dw \\ &+ \int_{w_\gamma}^{\overline{w}} \left[y(w) - h\left(\frac{y(w)}{w}\right)\right] \frac{\Gamma(w_\gamma, \overline{w})}{\Gamma(\underline{w}, \overline{w})} \cdot \mathbb{I}_{\left\{w_\gamma \in (w_\beta, \overline{w})\right\}} dw.\end{aligned} \tag{B.111}$$

Step 2. With respect to $y(\underline{w})$, we get from (B.111) and (B.80) that

$$\begin{aligned}\frac{\partial^2 U^*(k)}{\partial y(\underline{w})\partial k} &= \frac{\partial\tilde{\Phi}^{M*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\alpha, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial\Gamma(\underline{w}, \overline{w})} \frac{\partial\Gamma(\underline{w}, \overline{w})}{\partial y(\underline{w})} \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}\cap\{w=k\}} \\ &\quad - \frac{\partial\tilde{\Phi}^{R*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial\Gamma(\underline{w}, \overline{w})} \frac{\partial\Gamma(\underline{w}, \overline{w})}{\partial y(\underline{w})} \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\}\cap\{w=k\}} \\ &= \underbrace{\frac{\partial\Gamma(\underline{w}, \overline{w})}{\partial y(\underline{w})}}_{<0} \cdot \frac{y(k)}{k^2} h'\left(\frac{y(k)}{k}\right) \frac{1}{\Gamma(\underline{w}, \overline{w})^2} \cdot \underbrace{[\Gamma(w_\alpha, \overline{w}) - \Gamma(w_\beta, \overline{w})]}_{>0} \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}\cap\left\{w_\beta < \overline{w}\right\}} \\ &< 0.\end{aligned} \tag{B.112}$$

Hence, (B.112) implies that $U^*(k)$ is a submodular function, and an application of Topkis Theorem (see [67]) implies that $y(\underline{w})$ is decreasing in k. Since the other endpoint $w_\eta(k)$ is completely determined by the value of $y(\underline{w})$, we get that $w_\eta(k)$ is decreasing in k.

Step 3. With respect to $y(\overline{w})$, we get from (B.111) and (B.80) that

$$\begin{aligned}\frac{\partial^2 U^*(k)}{\partial y(\overline{w})\partial k} &= \frac{\partial\tilde{\Phi}^{M*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\alpha, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial\Gamma(w_\alpha, \overline{w})} \frac{\partial\Gamma(w_\alpha, \overline{w})}{\partial y(\overline{w})} \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}\cap\{w=k\}} \\ &\quad + \frac{\partial\tilde{\Phi}^{M*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\alpha, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial\Gamma(\underline{w}, \overline{w})} \frac{\partial\Gamma(\underline{w}, \overline{w})}{\partial y(\underline{w})} \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}\cap\{w=k\}} \\ &\quad - \frac{\partial\tilde{\Phi}^{R*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial\Gamma(w_\beta, \overline{w})} \frac{\partial\Gamma(w_\beta, \overline{w})}{\partial y(\overline{w})} \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\}\cap\{w=k\}} \\ &\quad - \frac{\partial\tilde{\Phi}^{R*}(w, y(w), \Gamma(w_\alpha, w_\beta), \Gamma(w_\beta, \overline{w}), \Gamma(\underline{w}, \overline{w}))}{\partial\Gamma(\underline{w}, \overline{w})} \frac{\partial\Gamma(\underline{w}, \overline{w})}{\partial y(\underline{w})} \cdot \mathbb{I}_{\left\{w_\beta < \overline{w}\right\}\cap\{w=k\}} \\ &= \underbrace{\frac{\partial\Gamma(\underline{w}, \overline{w})}{\partial y(\overline{w})}}_{>0} \cdot \frac{y(k)}{k^2} h'\left(\frac{y(k)}{k}\right) \frac{1}{\Gamma(\underline{w}, \overline{w})^2} \cdot \underbrace{[\Gamma(w_\alpha, \overline{w}) - \Gamma(w_\beta, \overline{w})]}_{>0} \cdot \mathbb{I}_{\{w_\alpha > \underline{w}\}\cap\left\{w_\beta < \overline{w}\right\}} \\ &> 0.\end{aligned} \tag{B.113}$$

Hence, (B.113) implies that $U^*(k)$ is a supermodular function, and an application of Topkis Theorem (see [67]) implies that $y(\overline{w})$ is increasing in k. Since the other endpoint $w_\gamma(k)$ is completely determined by the value of $y(\overline{w})$, we accordingly have that $w_\gamma(k)$ is increasing in k. □

Theorem B.2.3. *If the following condition*

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

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

Proof. Given Propositions B.2.1 and B.2.2, the proof is the same as that of Theorem 3.5.1. □

We, therefore, have established the voting equilibrium under the complete solution of the tax design problem.

B.3 The Relation between Ex Ante and Ex Post Median Skill Levels

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

$$\Gamma(\underline{w}, \overline{w}) = \int_{\underline{w}}^{\overline{w}} \tilde{f}(w)dw \ = \ \int_{\underline{w}}^{w_m} \tilde{f}(w)dw + \int_{w_m}^{\overline{w}} \tilde{f}(w)dw. \tag{B.114}$$

By using (3.8), we get the right-hand terms of (B.114) as

$$\int_{\underline{w}}^{w_m} \tilde{f}(w)dw = \frac{1}{2} \ + \ \underbrace{L^I([\underline{w}, w_m]) - L^O([\underline{w}, w_m])}_{L^{NI}([\underline{w}, w_m]) \ = \ \text{net labor inflow}} \tag{B.115}$$

and

$$\int_{w_m}^{\overline{w}} \tilde{f}(w)dw = \frac{1}{2} \ + \ \underbrace{L^I([w_m, \overline{w}]) - L^O([w_m, \overline{w}])}_{L^{NI}([w_m, \overline{w}]) \ = \ \text{net labor inflow}}, \tag{B.116}$$

in which the measures of labor inflows are defined as

$$\begin{aligned} L^I([\underline{w}, w_m]) &\equiv \int_{\{w \in [\underline{w}, w_m] | \Delta(w) \geq 0\}} G_-(\Delta(w)|w) f_-(w) n_- dw, \\ L^I([w_m, \overline{w}]) &\equiv \int_{\{w \in [w_m, \overline{w}] | \Delta(w) \geq 0\}} G_-(\Delta(w)|w) f_-(w) n_- dw, \end{aligned} \tag{B.117}$$

and the measures of labor outflows are defined as

$$\begin{aligned} L^O([\underline{w}, w_m]) &\equiv \int_{\{w\in[\underline{w},w_m]|\Delta(w)\le 0\}} G(-\Delta(w)|w)f(w)dw, \\ L^O([w_m, \overline{w}]) &\equiv \int_{\{w\in[w_m,\overline{w}]|\Delta(w)\le 0\}} G(-\Delta(w)|w)f(w)dw. \end{aligned} \tag{B.118}$$

By using (B.114)-(B.118), 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 B.3.1. *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 B.3.2. *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 B.3.3. *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 B.3.1-B.3.3 consider three possible cases corresponding to three possible ex post measures of workers of all skill levels.

Proposition B.3.1 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 B.3.2-B.3.3 in the similar way.