Relativity, Mobility, and Optimal Nonlinear Income Taxation in an Open Economy*

Darong Dai†  Wenzheng Gao‡  Guoqiang Tian§

February, 2019

Abstract

Recent evidence suggests that globalization has reduced the barriers to international labor mobility and induced more cross-country social comparisons. In an open economy with tax-driven migrations and consumption externalities (relativity), 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 rate for all skills. 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 rate. We find reasonable combinations of relativity, mobility and inequality such that tax competition results in equilibrium top tax rates higher than proposed in autarky. Under both Nash and Stackelberg tax competition: If the migration probability of top-income workers is around 50%, numerical calculation using parameter estimates from empirical studies shows that the country with labor inflow (outflow) implements over 10% lower (higher) marginal tax rates than suggested by the autarky equilibrium of Kanbur and Tuomala (2013).

Keywords: Relative consumption; International labor mobility; Maximin; Optimal income taxation; Nonlinear taxation; Tax competition.

JEL classification codes: D63; H21; H23; J61.

1 Introduction

Since Veblen (1899), economists recognize that the well-being of economic agents depends on relative consumption (“relativity”) in addition to absolute consumption, so taxing consumption externalities seems to be welfare enhancing as any other Pigouvian tax.1 Also, evidences reported by Senik (2009) and Clark and Senik (2010) show that social comparisons increase the demand for income redistribution, so the proper tax policy should play both externality-correcting and income-redistributing roles. Kanbur and Tuomala (2013) (K&T hereafter) was the first paper to examine whether nonlinear income tax is an effective tool for reducing inequalities and attenuating possible externalities arising from relative income concerns. They,

---

1 Helpful comments from Pedro Bento, Ian Dew-Becker, Li Gan, Amy Glass, Dennis Jansen, Silvana Krasteva, Quan Li, Fernando Luco, Jonathan Meer, Tatevik Sekhposyan, Manuelita Ureta, Yuzhe Zhang and Sarah Zubairy are gratefully acknowledged. The usual disclaimer certainly applies.

† Institute for Advanced Research (IAR), Shanghai University of Finance and Economics, Shanghai 200433, P. R. China. E-mail: dai.darong@mail.shufe.edu.cn. Financial support from the Bradley Fellowship of Texas A&M University is gratefully acknowledged.

‡ School of Economics, Nankai University, Tianjin 300071, P. R. China. E-mail: wenzhenggao@outlook.com.

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

nevertheless, focus on a closed economy. Recent evidence suggests that globalization has reduced the barriers to international labor mobility. For example, Kleven et al. (2013), Kleven et al. (2014) and Akcigit et al. (2016) estimate large migration elasticities with respect to tax rate for highly skilled workers. Besides, as is mentioned in Piketty and Saez (2013), the migration elasticity for all top earners is likely to increase over time as labor markets become better integrated and the fraction of foreign workers grows. The mobility of taxpayers thus induces tax competition between countries.²

The goal of the current study is, therefore, to address the following questions. How do the externality-correcting and income-redistributing roles of income taxation policy interact in an open economy with competing governments? How does the tax competition induced by cross-border labor mobility affect the interplay of relativity and inequality in determining the optimal structure of income taxes? And, how might the effect change across alternative forms of tax competition? To the best of our knowledge, none of the papers in the literature answers these questions in the optimal income tax framework inspired by Mirrlees (1971).

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 (2015), Nash competition is not necessarily the most realistic one since the ability to commit to public policy may differ across countries. Thus, we further 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 skill and migration cost. The distributions of skills and costs are continuously differentiable and assumed to be common knowledge, while the values of them are 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, maximin³, in the spirit of Rawls (1971). 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, endogenous location choice allows workers to have a reservation utility that depends on the tax policy of the opponent country. Throughout, taxes can only be conditioned on the observable 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 for all skill levels. The optimal tax formula obtained by Oswald (1983) and K&T for closed economies is augmented by a migration effect that changes the Pigouvian-tax term and the Mirrleesian-tax term, leading to a much more comprehensive formula. In addition, as in Lehmann et al. (2014), 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. 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 for all skill levels. Interpreting the Tiebout-best

²A recent example is the “Tax Cuts and Jobs Act” signed into law by President Trump.

³As is demonstrated by Boadway and Jacquet (2008), focusing on the maximin objective significantly simplifies the analytical analysis of the optimal income tax structure.
as the usual first-best, the constrained optimum exhibits a version of “no distortion at the top” for the highest skills, but with a downward distortion relative to the Tiebout best for the lowest skills.

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., Alvarez-Cuadrado and Long, 2012; K&T; Aronsson and Johansson-Stenman, 2014) or focus on how labor mobility and income inequality together shape the optimal tax schedules with two competing governments (e.g., Simula and Trannoy, 2012; Lehmann et al., 2014; Lipatov and Weichenrieder, 2015). Even though these studies have provided some useful insights, 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. The second strand uses the social optimality of income tax schedules that is biased and hence misleading for policy suggestions as people do care about social status in reality. See, for example, Fong (2001), and Heffetz and Frank (2011), for evidences. As such, the social welfare function without taking into account between-individual externalities may not be well-defined, especially in terms of the optimal design of redistributive taxation.

To analyze how relativity and inequality together shape the marginal tax rate facing 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. Under both Nash and Stackelberg competition, if the migration elasticity is no greater than one, then the higher is inequality, the lower is the effect of relativity in raising the top tax rate, and vice versa. Thus, relativity and inequality play a substitutive role under such migration intensities. Our result subsumes the prediction of K&T regarding top-income workers as a special case with a zero migration elasticity. If the migration elasticity is greater than one, then the higher is inequality, the higher is the effect of relativity in raising the top tax rate, and vice versa, as long as the ex post mass of top-income workers is below some threshold; otherwise, relativity and inequality play a substitutive role. In fact, Kleven et al. (2013), Kleven et al. (2014) and Akcigit et al. (2016) find empirical evidence that the migration elasticities for highly paid foreigners with respect to the tax rate can be larger than one, so we establish under reasonable migration possibilities a complementary relationship that is exactly the opposite of the prediction of K&T. In addition, the form of strategic tax competition turns out to be qualitatively irrelevant.

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, our calculation shows that the country with large labor inflow imposes much lower (less redistributive) tax rates than suggested by K&T, while the country with large labor outflow imposes much higher (more redistributive) ones. By and large, it is thus a less redistributive top tax rate that attracts the inflow of residents of the highest skill level. 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 on the competition effect. As an implication for open economies, normative public policy design on income taxes must take between-country tax competition, tax-driven migrations and relative consumption concerns seriously; otherwise, workers are likely to face welfare loss or the economy could face efficiency loss. In fact, these concerns seem to be the major motive behind Trump’s Massive Tax Cuts, although the tax reform per se may not be an optimal solution.

Our study is related to the literature studying optimal nonlinear income taxation in an open economy, such as Mirrlees (1982), Simula and Trannoy (2010), Bierbrauer et al. (2013),
Lehmann et al. (2014), Aronsson and Johansson-Stenman (2015), and Blumkin et al. (2015). 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 all but Aronsson and Johansson-Stenman (2015) completely ignore the effect of relative consumption concern placed on the design of Mirrlees income taxes. As numerically illustrated in Section 5, relative consumption concern does result in quantitatively significant effects on the optimal marginal tax rates and hence should not be ignored. Though Aronsson and Johansson-Stenman (2015) 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. Therefore, 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 sets up the model. Section 3 derives the optimal tax formula in Nash equilibrium and establishes some qualitative properties. Section 4 derives the optimal tax formula in Stackelberg equilibrium and establishes some qualitative properties. Section 5 provides some numerical examples regarding the optimal asymptotic marginal tax rates and compares our results with those calculated with K&T-formula. Section 6 concludes. Proofs are relegated to Appendix.

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 $n_{-i}$, with $0 < n_{-i} \leq 1$. Each worker is characterized by three characteristics: her native country $i \in \{A, B\}$, her skill $w \in [\underline{w}, \overline{w}]$ with $0 < w < \overline{w} \leq \infty$, and the migration cost $m \in \mathbb{R}^+$ she supports if deciding to live abroad. If a worker faces an infinitely large migration cost, then she is immobile. Following Lehmann et al. (2014), we do not make any restriction on the correlation between skills and migration costs.

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 $m$.

Following Mirrlees (1971), 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 1).

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 $i$:

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

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, $\mu_i$ is a domestic comparison consumption level, and $\mu_{-i}$ is a cross-country
Figure 1: Agents and Relationships

comparison consumption level,$^4$ with $v' > 0 \geq v''$, $h' > 0$ and $h'' > 0$. Following common practice,$^5$ comparison consumption levels are constructed as follows:

$$\mu_i = \int_{w}^{w'} c_i(w) f_i(w) dw,$$

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

**Assumption 2.1 (Bounded Jealousy)** For $\psi_i(\mu_i, \mu_{-i}) \equiv \partial \psi / \partial \mu_i < 0$, $\psi_{-i}(\mu_i, \mu_{-i}) \equiv \partial \psi / \partial \mu_{-i} < 0$, we have $\max\{|\psi_i(\mu_i, \mu_{-i})|, |\psi_{-i}(\mu_i, \mu_{-i})|\} < v'(c_i(w))$ for $i \in \{A, B\}$ and $w \in [w, w']$.

Assumption 2.1 states that the utility contribution of relative consumption is strictly smaller than that of absolute consumption. The assumption is consistent with general intuition as well as real data (see Clark et al., 2008).

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

$$c_i(w) = y_i(w) - T_i(y_i(w)).$$

Each worker is assumed to be small relative to the whole economy, and hence she takes $\mu_i$ and $\mu_{-i}$ as exogenously given. If she stays in country $i$, she maximizes (1) subject to $I = 0$ and (3), yielding the first-order condition:

$$\frac{h'(l_i(w))}{w v'(c_i(w))} = 1 - T'_i(y_i(w)).$$

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

We now proceed to her migration decision. We assume that migration occurs if and only if $m < U_{-i}(w) - U_i(w)$. As in Lehmann et al. (2014), 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:

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

---

$^4$ Piketty (2014) 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. Using survey-data for countries in Western Europe, Becchetti et al. (2013) find that the contribution of cross-country comparisons to well-being increased from the early 1970s to 2002.

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:

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

and

$$\theta_i(\Delta_i(w); w) \equiv c_i(w)\eta_i(\Delta_i(w); w).$$

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 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 (1) and (4), it is easy to show that $U_i(w) = \min\{U_i(w) : w \in [\underline{w}, \overline{w}]\}$. That is, the worst-off are exactly those workers with wage rate $w$ 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 (Gereffi and Luo, 2014). Second, they have the lowest migration (or foot-voting) ability, as migration rates increase in skill (Docquier and Marfouk, 2006). Third, especially for those in developed countries, they are worse off 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 (1971).

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

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

where $R \geq 0$ is an exogenous revenue requirement. As $\nu_c(\cdot) > 0$, (8) must be binding. In particular, here the participation constraint has been incorporated into the fiscal budget constraint (8) through the ex post skill density $\phi_i$ that endogenizes the tax base. As $\phi_i$ is a function of the difference of gross utilities, $U_i(w) - U_{-i}(w)$, that depends on the taxation policies of both jurisdictions, the strategic interaction plays a role through the fiscal budget constraint of each government.

The second is the set of incentive-compatibility constraints:

$$\nu(c_i(w)) - h(y_i(w)/w) \geq \nu(c_i(w')) - h(y_i(w')/w) \quad \forall w, w' \in [\underline{w}, \overline{w}].$$

The necessary conditions for (9) to be satisfied are:

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

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

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

$$\max_{\{U_i(w), l_i(w), \nu_i\}} U_i(w)$$

subject to (2), (8), (10) and $\dot{y}_i(w) \geq 0$ for all $w$. 


3 Nash Equilibrium

3.1 Optimal Tax Formula

We state the solution to the maximization problem above under Nash competition in the following theorem.

**Theorem 3.1** In a Nash equilibrium with \( \dot{y}_i(w) > 0 \) for all \( w \), the second-best marginal tax rates verify:

\[
\frac{T_i'(y_i(w))}{1 - T_i'(y_i(w))} = \frac{\gamma_i}{\lambda_i} \frac{\tilde{f}_i(w)}{\tilde{f}_i(w)} + \frac{\tilde{A}_i(w)B_i(w)C_i(w)}{\tilde{A}_i(w)B_i(w)C_i(w)}
\]

(11)

where: \( A_i(w) \equiv 1 + \{l_i(w)h''(l_i)/h'(l_i)\} \), \( B_i(w) \equiv \left[\tilde{F}_i(w) - \tilde{F}_i(w)\right] / w \tilde{f}_i(w) \),

\[
C_i(w) = \frac{\nu'(c_i(w))}{\nu'(c_i(w))} \int_w^{\bar{w}} \left\{ \frac{1}{\nu'(c_i(t))} \left[ 1 + \frac{\gamma_i}{\lambda_i} \frac{f_i(t)}{f_i(t)} \right] - T_i(y_i(t))\tilde{y}_i(t) \right\} \frac{\tilde{f}_i(t)dt}{\tilde{F}_i(w) - \tilde{F}_i(w)}
\]

(12)

and

\[
\frac{\gamma_i}{\lambda_i} = \frac{-\int_w^{\bar{w}} \frac{\psi_i(\mu_i, \mu_i)}{\nu'(c_i(w))} \tilde{f}_i(w)dw}{1 + \int_w^{\bar{w}} \psi_i(\mu_i, \mu_i) \frac{f_i(w)}{\nu'(c_i(w))} \tilde{f}_i(w)dw}
\]

(13)

with \( \tilde{F}_i(w) \equiv \int_w^w \tilde{f}_i(t)dt \) 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 \( \dot{y}_i(w) > 0 \) for all \( w \) holds.

**Proof.** See Appendix. ■

Our optimal tax formula (11) differs from the classic one derived by Diamond (1998) and Saez (2001) 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 \( C_i(w) \) as a tax level effect, and (iii) there is a Pigovian tax used to correct consumption externalities. Also, (i) and (ii) constitute new features relative to K&T.

To intuitively interpret the optimal tax formula (11), we investigate the effects of a small tax reform, as shown in Figure 2, in a unilaterally deviating country i: the second-best marginal 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., Saez, 2001). 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 Lehmann et al. (2014), 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 (12) that income effect will in turn reduce the positive mechanical effect. Since the optimal tax formula (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: A Small Tax Reform Perturbation

To see how relativity changes the average tax rate (ATR)\(^6\) and marginal tax rate (MTR), we also numerically solve the optimal tax formula (11) under the following assumptions (see Figure 3). First, these two countries are assumed to be symmetric. Second, following Jacquet et al. (2013), we put the mode \(w_m = $19,800\) and the highest skill level \(\bar{w} = $40,748\), and assume that workers within this income interval have a Pareto income distribution, with density function \(f(w) = aw^a / w^{a+1}\) for \(w_m \leq w \leq \bar{w}\). Third, we use the quasilinear-in-consumption preference with a constant elasticity of labor supply, formally \(u_i = c_i - l_i^{1+\frac{1}{\epsilon}} / (1 + \frac{1}{\epsilon}) + \sigma_D \mu_i + \sigma_F \mu_{-i}\) with \(\sigma_D, \sigma_F \in (-1, 0)\). And fourth, also similar to the distribution assumption used by Jacquet et al. (2013), we let the conditional distribution of migration costs be logistic:

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

Parameter values for simulation are given by \(\epsilon = 0.25, \ a = 2, \ \chi = 0.5, \ l = 0.33\) and \(\sigma_D \in (-1, 0)\). It follows from Figure 3 that both ATR and MTR increase as the degree of relative consumption concern \(|\sigma_D|\) increases, for any \(w \in [w_m, \bar{w}]\).

### 3.2 Qualitative Properties

To derive the qualitative properties of the optimal tax formula established in Theorem 3.1, we follow the approach developed by Jacquet et al. (2013) 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. Using the same terminology as Lehmann et al. (2014), we call this benchmark the Tiebout best.

**Lemma 3.1** In a Nash equilibrium, we have the following predictions:

\(^6\)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.

\(^7\)See, e.g., Diamond and Saez (2011), Kanbur and Tuomala (2013) and Jacquet et al. (2013).
(i) The Tiebout-best tax liabilities are given by

$$T^*_i(y_i(w)) = \frac{1}{v'(c_i(w))\tilde{\eta}_i(w)} \left[ 1 + \frac{\gamma_i f_i(w)}{\lambda_i \tilde{f}_i(w)} \right]$$

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

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

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

with $\gamma_i/\lambda_i$ given in Theorem 3.1.

Proof. See Appendix. ■

Under jealousy-type consumption comparison, it follows from (13) that $\gamma_i/\lambda_i > 0$. So for all but the bottom skill, the Tiebout-best tax liabilities are strictly decreasing in the semi-elasticity of migration, as shown in part (i). The intuition for this result is straightforward. In addition, if the revenue requirement $R$ is sufficiently small, then it follows from the fiscal-budget constraint (8) that the worst-off workers receive net transfers in the Tiebout-best economy. 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. Also, the ex ante to ex post density ratio $f_i(w)/\tilde{f}_i(w)$ and jealousy comparison impose a complementary effect on the Tiebout-best tax liabilities and tax rates.

Assuming quasilinear-in-consumption preferences, the formula of Tiebout-best tax liability obtained by Lehmann et al. (2014), namely the tax liability required from the residents with skill levels above $\underline{w}$ is equal to the inverse of their semi-elasticity of migration $\tilde{\eta}_i(w)$, is augmented by the positive externality-corrective tax component $\gamma_i f_i(w)/(\lambda_i \tilde{f}_i(w))$. In particular, if attention is restricted to the pure redistributive taxation under $R = 0$, then the worst-off residents receive more transfers than in the economy considered by Lehmann et al. (2014). As such, the relativity concern leads the Tiebout-best tax schedule to be a more redistributive one. This is somehow consistent with the empirical finding that social comparisons increase the demand for income redistribution (e.g., Senik, 2009; Clark and Senik, 2010).

To characterize the second-best tax schedules, we give the following:

Proposition 3.1 Suppose Assumption 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 all $w \in (w, \bar{w})$, then $T_i'(y_i(w)) > 0$ for all $w \in (w, \bar{w})$.

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

(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 all $w \in [w, \bar{w}]$, 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 $w f_i(w)$ is non-decreasing in $w$.

(iv) If $f_i(w)/\tilde{f}_i(w)$ is non-increasing in $w$, and $\frac{\gamma_i f_i(w)}{\lambda_i f_i(\tilde{w})} \leq \frac{\tilde{h}_i(w)}{\tilde{h}_i(w)}$, then there exists a $\tilde{w} \in (w, \bar{w})$ such that $T_i'(y_i(\tilde{w})) > 0$, then

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

Proof. See Appendix. ■

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 Nash equilibrium tax schedule is thus featured as follows. 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. However, for workers of the highest skill, marginal tax rate is the same under the second-best and the Tiebout-best, which may be regarded as a version of “no distortion at the top” when the Tiebout-best is interpreted as the usual first-best.

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

**Proposition 3.2** 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 $\bar{w} = \infty$ and Pareto index $a_i > 1$.

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

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

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

Proof. See Appendix. ■

Given that the optimal tax formula (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 $\gamma_i/\lambda_i$, 

10
the measure of labor flow $\alpha_i(\infty)$, the elasticity of labor supply $\varepsilon$, the degree of income inequality $1/a_i$, and the elasticity of migration $\hat{\theta}_i(\infty)$. In particular, AMTR is strictly decreasing in the elasticity of migration.

The following proposition characterizes the composition effect of consumption relativity and income inequality on the optimal AMTR.

**Proposition 3.3** Suppose $\psi(\mu_i, \mu_{-i}) = \sigma_D \mu_i + \hat{\psi}(\mu_{-i})$ for a constant $\sigma_D \in (-1, 0)$, $F_i(w) = F_{-i}(w)$ and $\partial \hat{F}_i(\infty)/\partial a_i = 0$, then we have the following predictions.

(i) If $\hat{\theta}_i(\infty) \leq 1$, then

$$\frac{\partial^2 T_i'(y_i(\infty))}{\partial (-\sigma_D) \partial (1/a_i)} < 0.$$ 

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

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

**Proof.** See Appendix.

If migration elasticity is no 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 some threshold; otherwise, relativity and inequality play a complementary role in shaping the optimal AMTR.

The substitution pattern implies that the demand for income redistribution driven by social comparisons is weakened by ability inequality, or the demand for redistribution driven by ability inequality is weakened by social comparisons. This holds in two cases: (1) the migration elasticity is small or the intensity of tax competition is relativity low, and (2) the migration elasticity is large while the ex post mass of residents of the top skill level is large. In the former case, as the tax base is relatively immobile, if relativity and inequality enhance each other in motivating redistribution, the equity concern is overemphasized and efficiency losses arise due to the weakened incentive of labor supply, thus violating the best balance between equity and efficiency. In the latter case, as the high-skill tax base is sufficiently mobile, if the redistributive motive is too strong, then the ex post mass of high-skill workers would not be large. The case for the complementary pattern can be analogously analyzed. In a word, here the composition effect of consumption relativity and income inequality on optimal AMTRs reflects the fundamental equity-efficiency tradeoff in an open economy.

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 3.3 subsumes the corresponding prediction of K&T as a special case with $\hat{\theta}_i(\infty) = 0$ and $\alpha_i(\infty) = 1$.

---

8Kleven et al. (2013), Kleven et al. (2014) and Akcigit et al. (2016), indeed, find empirical evidence that the migration elasticities for highly paid foreigners with respect to the tax rate can be larger than one.
4 Stackelberg Equilibrium

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 solution to the optimal tax design problem under Stackelberg competition in the following theorem.

**Theorem 4.1** In a Stackelberg equilibrium, the optimal tax formula is the same as that in the Nash equilibrium, except that:

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

with

$$
\frac{\partial \mu_{-i}}{\partial \mu_i} = \frac{\int_w \frac{\partial c_{-i}(w)}{\partial \mu_i} f_{-i}(w)dw}{1 - \int_w \frac{\partial c_{-i}(w)}{\partial \mu_i} f_{-i}(w)dw}
$$

for the leader country $i$.

**Proof.** See Appendix.

Theorems 4.1 and 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.

4.2 Qualitative Properties

Using Theorem 4.1, the following corollary is immediate.

**Corollary 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 3.1-3.2 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 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 $\sigma_D, \sigma_F \in (-1, 0)$ 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 Proposition 3.3 carry over to the current equilibrium.
Proof. See Appendix. ■

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 conveyed by Proposition 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 4.2 If Assumption 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. ■

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.

5 Numerical Illustration

In this section we provide some numerical examples on the optimal AMTR established in Proposition 3.2. Although these exercises are very coarse, they 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.

For simplicity, we use the linear utility function of relative consumptions shown in condition (a) of Proposition 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 \( \sigma_D = -0.5 \) and 0, 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 Saez et al., 2012), 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 Diamond and Saez, 2011), while \( a_i = 3 \) is chosen to be larger than this realistic number to represent an experimental scenario with more equal income distribution.

We consider two relativity scenarios. \( \sigma_D = 0 \) denotes the benchmark case without relative consumption, whereas \( \sigma_D = -0.5 \) measures the degree of jealousy. 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 in absolute value and of the opposite sign (see, e.g., Luttmer, 2005). Given the assumption of quasi-linear preferences, \( \sigma_D = -0.5 \) seems to be reasonable. In fact, it is consistent with the finding of Alpizar et al. (2005) who use survey-experimental methods to see how much we care about absolute versus relative income and consumption. By assumption, the degree of cross-country social comparisons is smaller than that of domestic social comparisons, so we let \( \sigma^2 = 0.04 \) in what follows.\(^9\) Following Piketty and Saez (2013), 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 1.

If \( \Delta_i(\infty) \geq 0 \), namely top-income workers get an indirect utility in country \( i \) which is no 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

\(^9\)In fact, we have not found any realistic estimates of this parameter in empirical literature.
Table 1: Parameter Values

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Description</th>
<th>Source/Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_i(\infty)$</td>
<td>0.25</td>
<td>Migration elasticity</td>
<td>Piketty &amp; Saez (2013)</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>(0.1,0.4)</td>
<td>Labor-supply elasticity</td>
<td>Saez et al. (2012)</td>
</tr>
<tr>
<td>$\alpha_i$</td>
<td>2</td>
<td>Pareto index</td>
<td>Diamond &amp; Saez (2011)</td>
</tr>
<tr>
<td>$\sigma_D$</td>
<td>-0.5</td>
<td>Domestic relativity</td>
<td>Clark et al. (2008)</td>
</tr>
<tr>
<td>$\sigma_F$</td>
<td>-0.2</td>
<td>Cross-country relativity</td>
<td>$</td>
</tr>
</tbody>
</table>

Table 2: Economic Mechanism Governing Quantitative Findings

<table>
<thead>
<tr>
<th>Relativity effect</th>
<th>Nash</th>
<th>Nash</th>
<th>Stack</th>
<th>Stack</th>
</tr>
</thead>
<tbody>
<tr>
<td>Migration effect</td>
<td>Small</td>
<td>Large</td>
<td>Small</td>
<td>Large</td>
</tr>
<tr>
<td>MTR$^O$ Labor inflow</td>
<td>&lt;</td>
<td>&lt;&lt;&lt;</td>
<td>&gt;</td>
<td>&lt;&lt;&lt;</td>
</tr>
<tr>
<td>MTR$^O$ Labor outflow</td>
<td>&lt;</td>
<td>&gt;&gt;</td>
<td>&gt;</td>
<td>&gt;&gt;</td>
</tr>
</tbody>
</table>

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.

5.1 A Comparison with K&T

In the following tables, we use red numbers to denote the optimal AMTRs calculated using the formula of K&T. In both types of equilibrium, we obtain three main findings under different values of $\alpha_i(\infty)$.

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. Third, for each given elasticity and given degree of inequality, AMTR significantly increases under jealousy, compared to the benchmark case without relative consumption concerns.

We summarize the economic mechanism in Table 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, relativity and migration are determinant factors in the comparison.

Since no one internalizes the cross-country consumption externality under Nash competition, the relativity effects on MTR are 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, regardless of whether 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
cases with labor inflow and 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 cases with labor inflow and 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 analogous between Nash and Stackelberg tax competition.

Figure 4: \(\alpha_i(\infty) = 0.5, \varepsilon = \tilde{\theta}_i(\infty) = 0.25, a_i > 1, \) and \(\sigma_D \in (-1, 0)\).

Figure 5: \(\alpha_i(\infty) = 2, \varepsilon = \tilde{\theta}_i(\infty) = 0.25, a_i > 1, \) and \(\sigma_D \in (-1, 0)\).

5.1.1 Nash vs. K&T

Tables 3-6 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 4). However, if $\Delta_i(\infty) \leq 0$, they are greater than those in K&T. Moreover, the migration effect is magnified by consumption comparison.

5.1.2 Stackelberg vs. K&T

Table 7: AMTR (%) under $\Delta_i(\infty) \geq 0$ with $\alpha_i(\infty) = 0.95$

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>$\alpha_i = 2$</th>
<th>$\alpha_i = 3$</th>
<th>$\alpha_i = 2$</th>
<th>$\alpha_i = 3$</th>
<th>$\alpha_i = 2$</th>
<th>$\alpha_i = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_D = -0.5$</td>
<td>65.2,69.2</td>
<td>61.5,64.7</td>
<td>65.8,70.0</td>
<td>62.0,65.4</td>
<td>67.0,71.4</td>
<td>63.1,66.7</td>
</tr>
<tr>
<td>$\sigma_D = 0$</td>
<td>35.1,38.5</td>
<td>27.4,29.4</td>
<td>36.3,39.9</td>
<td>28.5,30.7</td>
<td>38.7,42.9</td>
<td>30.8,33.3</td>
</tr>
</tbody>
</table>

Tables 7-10 compare optimal AMTRs in Stackelberg equilibrium, denoted by black 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 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 6).
Figure 6: $\alpha_i(\infty) = 0.5$, $\varepsilon = \tilde{\theta}_i(\infty) = 0.25$, $a_i > 1$, $\sigma_D < 0$, and $\sigma_F^2 = 0.04$.

Figure 7: $\alpha_i(\infty) = 2$, $\varepsilon = \tilde{\theta}_i(\infty) = 0.25$, $a_i > 1$, $\sigma_D < 0$, and $\sigma_F^2 = 0.04$.

Table 8: AMTR (%) under $\Delta_i(\infty) \geq 0$ with $\alpha_i(\infty) = 0.55$

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>$\alpha_i = 2$</th>
<th>$\alpha_i = 3$</th>
<th>$\alpha_i = 2$</th>
<th>$\alpha_i = 3$</th>
<th>$\alpha_i = 2$</th>
<th>$\alpha_i = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_D = -0.5$</td>
<td>61.7,69.2</td>
<td>57.5,64.7</td>
<td>62.3,70.0</td>
<td>58.1,65.4</td>
<td>63.6,71.4</td>
<td>59.3,66.7</td>
</tr>
<tr>
<td>$\sigma_D = 0$</td>
<td>36.4,38.5</td>
<td>28.9,29.4</td>
<td>38.5,39.9</td>
<td>31.1,30.7</td>
<td>40.8,42.9</td>
<td>33.2,33.3</td>
</tr>
</tbody>
</table>

Table 9: AMTR (%) under $\Delta_i(\infty) \leq 0$ with $\alpha_i(\infty) = 1.05$

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>$\alpha_i = 2$</th>
<th>$\alpha_i = 3$</th>
<th>$\alpha_i = 2$</th>
<th>$\alpha_i = 3$</th>
<th>$\alpha_i = 2$</th>
<th>$\alpha_i = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_D = -0.5$</td>
<td>72.1,69.2</td>
<td>69.1,64.7</td>
<td>72.5,70.0</td>
<td>69.5,65.4</td>
<td>73.5,71.4</td>
<td>70.4,66.7</td>
</tr>
<tr>
<td>$\sigma_D = 0$</td>
<td>37.6,38.5</td>
<td>30.2,29.4</td>
<td>38.8,39.9</td>
<td>31.3,30.7</td>
<td>41.0,42.9</td>
<td>33.5,33.3</td>
</tr>
</tbody>
</table>

5.2 Nash vs. Stackelberg: the Leader Country

Tables 11-12 illustrate Proposition 4.2 by comparing AMTRs under these two types of equilibrium. As is obvious, regardless of whether $\Delta_i(\infty) \geq 0$ or $\Delta_i(\infty) \leq 0$, these AMTRs in Nash
Table 10: AMTR (%) under $\Delta_i(\infty) \leq 0$ with $\alpha_i(\infty) = 1.55$

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>$\alpha_i = 2$</th>
<th>$\alpha_i = 3$</th>
<th>$\alpha_i = 2$</th>
<th>$\alpha_i = 3$</th>
<th>$\alpha_i = 2$</th>
<th>$\alpha_i = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_D = -0.5$</td>
<td>78.0, 69.2</td>
<td>75.7, 64.7</td>
<td>78.4, 70.0</td>
<td>76.1, 65.4</td>
<td>79.1, 71.4</td>
<td>76.7, 66.7</td>
</tr>
<tr>
<td>$\sigma_D = 0$</td>
<td>38.7, 38.5</td>
<td>31.5, 29.4</td>
<td>38.8, 39.9</td>
<td>31.3, 30.7</td>
<td>41.0, 42.9</td>
<td>33.5, 33.3</td>
</tr>
</tbody>
</table>

Figure 8: $\alpha_i(\infty) = 0.5$, $\varepsilon = \tilde{\theta}_l(\infty) = 0.25$, $a_i > 1$, $\sigma_D < 0$, and $\sigma_F^2 = 0.04$.

Figure 9: $\alpha_i(\infty) = 2$, $\varepsilon = \tilde{\theta}_l(\infty) = 0.25$, $a_i > 1$, $\sigma_D < 0$, and $\sigma_F^2 = 0.04$.

equilibrium are in general smaller than those in Stackelberg equilibrium (see also Figures 8-9), denoted by blue numbers in these two tables. The differences of marginal tax rates under any given degree of relativity concerns may not be that large just because we let $\alpha_i(\infty) = 0.95$ and $\alpha_i(\infty) = 1.05$. Both are close to 1 and this implies that the migration effect is small. That is, as $\alpha_i(\infty)$ deviates from 1 further, regardless of whether from the below or the above, the differences of AMTRs would increase.
Table 11: AMTR (%) under $\Delta_i(\infty) \geq 0$ with $\alpha_i(\infty) = 0.95$

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>$\alpha_i(\infty)$</th>
<th>$a_i$</th>
<th>$\sigma_D$</th>
<th>AMTR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.25</td>
<td>2</td>
<td>-0.5</td>
<td>65.2,70.5</td>
</tr>
<tr>
<td>0.25</td>
<td>0.33</td>
<td>3</td>
<td>0.25</td>
<td>61.5,67.3</td>
</tr>
<tr>
<td>0.33</td>
<td>0.33</td>
<td>2</td>
<td>0.33</td>
<td>65.8,71.0</td>
</tr>
<tr>
<td>0.33</td>
<td>0.5</td>
<td>3</td>
<td>0.25</td>
<td>62.0,67.8</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>2</td>
<td>0.33</td>
<td>67.0,71.9</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>3</td>
<td>0.25</td>
<td>63.1,68.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\sigma_D$</th>
<th>AMTR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>35.1,37.4</td>
</tr>
<tr>
<td>0</td>
<td>27.4,30.0</td>
</tr>
<tr>
<td>0</td>
<td>36.3,38.5</td>
</tr>
<tr>
<td>0</td>
<td>28.5,31.1</td>
</tr>
<tr>
<td>0</td>
<td>38.7,40.8</td>
</tr>
<tr>
<td>0</td>
<td>30.8,33.2</td>
</tr>
</tbody>
</table>

Table 12: AMTR (%) under $\Delta_i(\infty) \leq 0$ with $\alpha_i(\infty) = 1.05$

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>$\alpha_i(\infty)$</th>
<th>$a_i$</th>
<th>$\sigma_D$</th>
<th>AMTR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.25</td>
<td>2</td>
<td>-0.5</td>
<td>66.8,72.1</td>
</tr>
<tr>
<td>0.25</td>
<td>0.33</td>
<td>3</td>
<td>0.25</td>
<td>63.3,69.1</td>
</tr>
<tr>
<td>0.33</td>
<td>0.33</td>
<td>2</td>
<td>0.33</td>
<td>67.4,72.5</td>
</tr>
<tr>
<td>0.33</td>
<td>0.5</td>
<td>3</td>
<td>0.25</td>
<td>63.8,69.5</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>2</td>
<td>0.33</td>
<td>68.5,73.5</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>3</td>
<td>0.25</td>
<td>64.8,70.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\sigma_D$</th>
<th>AMTR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>35.1,37.6</td>
</tr>
<tr>
<td>0</td>
<td>27.4,30.2</td>
</tr>
<tr>
<td>0</td>
<td>36.3,38.8</td>
</tr>
<tr>
<td>0</td>
<td>28.5,31.3</td>
</tr>
<tr>
<td>0</td>
<td>38.7,41.0</td>
</tr>
<tr>
<td>0</td>
<td>30.8,33.5</td>
</tr>
</tbody>
</table>

6 Conclusion

In this paper 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 with private information on skills and migration costs decide where to live and how much to work. In addition to the standard Nash, we also examine the scenario wherein governments play Stackelberg.

Firstly, we obtain an optimal tax formula that can be interpreted as a nontrivial generalization of those obtained by Diamond (1998), Saez (2001), K&T, Lehmann et al. (2014) and Aronsson and Johansson-Stenman (2015). 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 various combinations of parameters measuring relative consumption, labor mobility and income inequality. Thirdly, we show that the leader country imposes a higher marginal tax rate in Stackelberg equilibrium than that in Nash equilibrium. And fourthly, we provide a complete characterization on how relativity and inequality jointly determine the optimal AMTR under both Nash and Stackelberg tax competition, and find that both the elasticity and level of migration are determinant for predicting when relativity and inequality are complementar or substitutive in shaping the optimal tax rates placed on top-income workers. In consequence, this finding shows the qualitative irrelevance of tax competition form in terms of shaping the interplay of relativity and inequality in determining top tax rates.

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 do generate heterogenous degrees of impact on optimal tax rates, the identification of the form of tax competition should be of practical relevance, which however awaits future research.
References


Appendix: Proofs

**Proof of Theorem 3.1.** We shall complete the proof in 3 steps.

Step 1. Given the FOC (4) of individual choice, the indirect utility of a type-$w$ worker in country $i \in \{A, B\}$ can be written as

$$U_i(w) = v(\varphi_i(U_i(w), l_i(w), \mu_i, \mu_{-i})) - h(l_i(w)) + \psi(\mu_i, \mu_{-i}), \quad (14)$$

where we treat individual consumption $c_i(w)$ as an implicit function of $U_i(w)$, $l_i(w)$, $\mu_i$, $\mu_{-i}$, and equivalently rewrite it as $\varphi_i(\cdot)$. By applying the Implicit Function Theorem, we get from (14) that

$$\frac{\partial \varphi_i}{\partial l_i} = \frac{h'(l_i(w))}{v'(c_i(w))}, \quad \frac{\partial \varphi_i}{\partial U_i} = \frac{1}{v'(c_i(w))}, \quad \frac{\partial \varphi_i}{\partial \mu_i} = -\frac{\psi_i(\mu_i, \mu_{-i})}{v'(c_i(w))} \quad \text{and} \quad \frac{\partial \varphi_i}{\partial \mu_{-i}} = -\frac{\psi_{-i}(\mu_i, \mu_{-i})}{v'(c_i(w))}. \quad (15)$$

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:

$$L_i(\{U_i(w), l_i(w)\}_{w \in [\underline{w}, \overline{w}]}, \mu_i; \lambda_i, \gamma_i, \{\varsigma_i(w)\}_{w \in [\underline{w}, \overline{w}]})$$

$$= U_i(w) + \lambda_i \int_{\underline{w}}^{\overline{w}} \left\{ [w l_i(w) - \varphi_i(U_i(w), l_i(w), \mu_i, \mu_{-i})] \phi_i(U_i(w) - U_{-i}(w); w) - R \frac{w}{\overline{w} - w} \right\} dw$$

$$+ \int_{\underline{w}}^{\overline{w}} \varsigma_i(w) \left[ h'(l_i(w)) \frac{l_i(w)}{w} - \dot{U}_i(w) \right] dw$$

$$+ \gamma_i \left[ \mu_i - \int_{\underline{w}}^{\overline{w}} \varphi_i(U_i(w), l_i(w), \mu_i, \mu_{-i}) f_i(w) dw \right]$$

where $\lambda_i > 0$ is the multiplier associated with the binding budget constraint (8), $\varsigma_i(w)$ is the multiplier associated with the FOIC conditions (10), and $\gamma_i$ is the multiplier associated with the comparison consumption constraint (2). Integrating by parts, we obtain

$$\int_{\underline{w}}^{\overline{w}} \varsigma_i(w) \dot{U}_i(w) dw = \varsigma_i(\overline{w})U_i(\overline{w}) - \varsigma_i(\underline{w})U_i(\underline{w}) - \int_{\underline{w}}^{\overline{w}} \varsigma_i(w)U_i(w) dw. \quad (17)$$

Plugging (17) in (16) gives rise to

$$L_i(\{U_i(w), l_i(w)\}_{w \in [\underline{w}, \overline{w}]}, \mu_i; \lambda_i, \gamma_i, \{\varsigma_i(w)\}_{w \in [\underline{w}, \overline{w}]})$$

$$= U_i(w) + \lambda_i \int_{\underline{w}}^{\overline{w}} \left\{ [w l_i(w) - \varphi_i(U_i(w), l_i(w), \mu_i, \mu_{-i})] \phi_i(U_i(w) - U_{-i}(w); w) - R \frac{w}{\overline{w} - w} \right\} dw$$

$$+ \varsigma_i(w)U_i(w) - \varsigma_i(\overline{w})U_i(\overline{w}) + \int_{\underline{w}}^{\overline{w}} \left[ \varsigma_i(w)h'(l_i(w)) \frac{l_i(w)}{w} + \varsigma_i(w)U_i(w) \right] dw$$

$$+ \gamma_i \left[ \mu_i - \int_{\underline{w}}^{\overline{w}} \varphi_i(U_i(w), l_i(w), \mu_i, \mu_{-i}) f_i(w) dw \right].$$

Assuming that there is no bunching of workers of different skills and the existence of an interior solution, applying (15) shows that the necessary conditions can be written as follows:

$$\frac{\partial L_i}{\partial l_i(w)} = \lambda_i \left[ w - \frac{h'(l_i(w))}{v'(c_i(w))} \right] \tilde{f}_i(w) - \gamma_i \frac{h'(l_i(w))}{v'(c_i(w))} f_i(w)$$

$$+ \frac{\varsigma_i(w)h'(l_i(w))}{w} \left[ 1 + \frac{l_i(w)h''(l_i(w))}{h'(l_i(w))} \right] = 0 \quad \forall w \in [\underline{w}, \overline{w}], \quad (18)$$
\[
\frac{\partial L_i}{\partial U_i(w)} = -\frac{\lambda_i f_i(w)}{v'(c_i(w))} + \lambda_i T_i(y_i(w))\hat{\eta}_i(w)\hat{f}_i(w) - \frac{\gamma_i f_i(w)}{v'(c_i(w))} + \dot{s}_i(w) = 0 \quad \forall w \in (\underline{w}, \overline{w}),
\]

(19)

\[
\frac{\partial L_i}{\partial U_i(w)} = 1 + s_i(w) = 0,
\]

(20)

\[
\frac{\partial L_i}{\partial \mu_i} = -s_i(w) = 0,
\]

(21)

\[
\frac{\partial L_i}{\partial \mu_i} = \lambda_i \int_{\underline{w}}^{\overline{w}} \psi_i(\mu_i, \mu_i - i) \hat{f}_i(w) dw + \gamma_i \left[ 1 + \int_{\underline{w}}^{\overline{w}} \psi_i(\mu_i, \mu_i - i) f_i(w) dw \right] = 0.
\]

(22)

Using (19), we get

\[
\frac{\dot{s}_i(w)}{\lambda_i} = \frac{\gamma_i}{\lambda_i} \frac{f_i(w)}{v'(c_i(w))} + \frac{\hat{f}_i(w)}{v'(c_i(w))} - T_i(y_i(w))\hat{\eta}_i(w)\hat{f}_i(w).
\]

Integrating on both sides of this equation and using the transversality condition (21), we obtain

\[
\frac{-s_i(w)}{\lambda_i} = \frac{\gamma_i}{\lambda_i} \int_{\underline{w}}^{\overline{w}} \frac{f_i(t)}{v'(c_i(t))} dt + \int_{\underline{w}}^{\overline{w}} \left[ \frac{1}{v'(c_i(t))} - T_i(y_i(t))\hat{\eta}_i(t) \right] \hat{f}_i(t) dt.
\]

(23)

Rearranging (18) via using FOC (4), we have

\[
\frac{T_i'(y_i(w))}{1 - T_i'(y_i(w))} = \frac{l_i(w)h''(l_i(w))}{w\hat{f}_i'(w)} \left[ 1 + \frac{l_i(w)h''(l_i(w))}{h'(l_i(w))} \right] \forall w \in [\underline{w}, \overline{w}].
\]

(24)

Substituting (23) into (24) gives the first-order conditions characterizing the optimum marginal tax rates, with \(\gamma_i/\lambda_i\) determined by solving (22).

Step 3. To derive a sufficient condition for the optimal marginal tax profile to satisfy the SOIC conditions, we rewrite the FOC (4) as

\[
\frac{v'(c_i(w))}{h'(y_i(w)/w)} = \frac{1}{w[1 - T_i'(y_i(w))]},
\]

(25)

Noting that

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

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

\[
\frac{d\text{LHS}}{dw} < 0 \implies \hat{y}_i(w) > 0.
\]

(26)

Also, noting that

\[
\frac{d\text{RHS}}{dw} = - \{ w \left[ 1 - T_i'(y_i(w)) \right] \}^2 \left[ 1 - T_i'(y_i(w)) - w \frac{dT_i'(y_i(w))}{dw} \right],
\]

we thus arrive at

\[
\frac{dT_i'(y_i(w))}{dw} \leq 0 \implies \frac{d\text{RHS}}{dw} < 0.
\]

(27)
Therefore, (25) combined with (26) and (27) implies that
\[
\frac{dT^i(y_i(w))}{dw} \leq 0 \implies \dot{y}_i(w) > 0,
\]
as desired. ■

**Proof of Lemma 3.1.** The Lagrangian of government \( i \)'s problem can be expressed as
\[
\mathcal{L}_i({\{U_i(w), l_i(w)\}}_{w \in [w, \bar{w}]}, \mu_i; \lambda_i, \gamma_i) = U_i(w) + \lambda_i \int_w^{\bar{w}} \left\{ [w l_i(w) - \varphi_i(U_i(w), l_i(w), \mu_i, \mu_{-i})] \phi_i(U_i(w) - U_{-i}(w); w) - \frac{R}{w - w} \right\} dw + \gamma_i \left[ \mu_i - \int_w^{\bar{w}} \phi_i(U_i(w), l_i(w), \mu_i, \mu_{-i}) f_i(w) dw \right].
\]
where \( \lambda_i > 0 \) is the multiplier associated with the binding budget constraint (8) and \( \gamma_i \) is the multiplier associated with the comparison consumption constraint (2). Assuming that there is no bunching of workers of different skills and the existence of an interior solution, applying (15) gives these necessary conditions:
\[
\frac{\partial \mathcal{L}_i}{\partial U_i(w)} = \lambda_i \left[ T_i(y_i(w)) \bar{y}_i(w) - \frac{1}{v'(c_i(w))} \right] \tilde{f}_i(w) - \gamma_i \frac{f_i(w)}{v'(c_i(w))} = 0 \quad \forall w \in (w, \bar{w}),
\]
\[
\frac{\partial \mathcal{L}_i}{\partial l_i(w)} = \lambda_i \left( w - \frac{h'(l_i(w))}{v'(c_i(w))} \right) \tilde{f}_i(w) - \gamma_i \frac{h'(l_i(w)) f_i(w)}{v'(c_i(w))} = 0 \quad \forall w \in [w, \bar{w}],
\]
and
\[
\frac{\partial \mathcal{L}_i}{\partial \mu_i} = \lambda_i \int_w^{\bar{w}} \frac{\psi_i(\mu_i, \mu_{-i})}{v'(c_i(w))} \tilde{f}_i(w) dw + \gamma_i \left[ 1 + \int_w^{\bar{w}} \frac{\psi_i(\mu_i, \mu_{-i})}{v'(c_i(w))} f_i(w) dw \right] = 0.
\]
By using (28), we obtain the Tiebout-best tax liabilities. By using (29) and the FOC (4), we obtain the Tiebout-best marginal tax rates. The ratio \( \gamma_i / \lambda_i \) is determined by (30). 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 \( w = \bar{w} \). ■

**Proof of Proposition 3.1.** We shall complete the proof in 6 steps.

**Step 1.** By Assumption 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 \( \forall w \in (w, \bar{w}) \). Therefore, by directly comparing the formulas of marginal tax rates established in Theorem 3.1 and Lemma 3.1, claim (i) is immediate.

**Step 2.** By applying the transversality condition (20) to equation (24), it is easy to see that
\[
\frac{T^i(y_i(w))}{1 - T^i(y_i(w))} > \frac{\gamma_i f_i(w)}{\lambda_i f_i(w)} > 0
\]
under Assumption 2.1. Similarly, for a bounded skill distribution with \( \bar{w} < \infty \), applying the transversality condition (21) to equation (24) gives
\[
\frac{T^i(y_i(\bar{w}))}{1 - T^i(y_i(\bar{w}))} = \frac{\gamma_i f_i(\bar{w})}{\lambda_i f_i(\bar{w})} > 0
\]
under Assumption 2.1. By using Lemma 3.1 again, the required assertion (ii) follows.

**Step 3.** Suppose \( h(\cdot) \) takes the iselastic 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, \( 1/w f_i(w) \) 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 \( w f_i(w) \) are non-decreasing beyond the mode \( w_m \). Also, noting from term \( C_i(w) \) that

\[
\frac{d}{dw} \int_{\omega} \left\{ \frac{1}{v'(c_i(t))} \left[ 1 + \frac{\gamma_i f_i(t)}{\lambda_i f_i(t)} \right] - T_i(y_i(t)) \right\} \tilde{f}_i(t) dt
= \left\{ T_i(y_i(w)) - \frac{1}{v'(c_i(w))\tilde{\eta}_i(w)} \left[ 1 + \frac{\gamma_i f_i(w)}{\lambda_i f_i(w)} \right] \right\} \tilde{\eta}_i(w) \tilde{f}_i(w),
\]

hence an application of Lemma 3.1 completes the proof of claim (iii).

Step 4. Define

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

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

\[
\Sigma'_i(w) = \left\{ T_i(y_i(w)) - \frac{1}{v'(c_i(w))\tilde{\eta}_i(w)} \left[ 1 + \frac{\gamma_i f_i(w)}{\lambda_i f_i(w)} \right] \right\} \tilde{\eta}_i(w) \tilde{f}_i(w) = \xi_i(w)\tilde{\eta}_i(w)\tilde{f}_i(w), \tag{31}
\]

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

\[
\frac{d}{dw} \left\{ \frac{1}{v'(c_i(w))\tilde{\eta}_i(w)} \left[ 1 + \frac{\gamma_i f_i(w)}{\lambda_i f_i(w)} \right] \right\} > 0
\]

under Assumption 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

\[
\frac{\gamma_i}{\lambda_i} \frac{d[f_i(w)/\tilde{f}_i(w)]}{dw} \leq \frac{\dot{\eta}_i(w)}{\tilde{\eta}_i(w)}
\]

is sufficient for

\[
\frac{d}{dw} \left\{ \frac{1}{v'(c_i(w))\tilde{\eta}_i(w)} \left[ 1 + \frac{\gamma_i f_i(w)}{\lambda_i f_i(w)} \right] \right\} \leq 0. \tag{32}
\]

If we assume that \( \Sigma_i(w) \geq 0 \), then we get from the optimal tax formula in Theorem 3.1 that \( T'_i(y_i(w)) > 0 \). Then applying (31) and (32) shows that \( \xi'_i(w) > 0 \) given \( \Sigma_i(w) \geq 0 \).

Step 5. Assume that there exists a \( \hat{w} \in (\omega, \bar{w}) \) such that \( \Sigma_i(\hat{w}) \geq 0 \). Then we have two cases to consider in what follows, namely either \( \Sigma'_i(\hat{w}) \geq 0 \) or \( \Sigma'_i(\hat{w}) < 0 \). If \( \Sigma'_i(\hat{w}) \geq 0 \), then we have both \( \xi_i(\hat{w}) \geq 0 \) and \( \xi'_i(\hat{w}) > 0 \). So the continuity of \( \xi_i(w) \) with respect to \( w \) implies that there is an open interval with lower bound \( \hat{w} \) 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 \( (\hat{\tilde{w}}, \tilde{w}) \) be a maximal interval on which \( \Sigma_i(w) > 0 \) with \( \hat{w} < \tilde{w} \leq \bar{w} \). As a consequence, \( 0 \leq \Sigma_i(\hat{w}) < \Sigma(\tilde{w}) \), which implies that \( \xi'_i(w) > 0 \) for \( \forall w \in [\hat{w}, \tilde{w}] \). As a result, \( 0 \leq \xi_i(\hat{w}) < \xi_i(\tilde{w}) \), which leads us to \( \Sigma'_i(\hat{w}) > 0 \) by using (31). Therefore, \( \Sigma_i(\cdot) \) is increasing on \([\hat{w}, \bar{w}]\) given that
\(\Sigma_i'(\tilde{w}) \geq 0\). We know from the transversality condition (21) that \(\Sigma_i(\tilde{w}) = -\zeta_i(\tilde{w})/\lambda_i = 0\). As we have already shown that \(0 \leq \Sigma_i(\tilde{w}) < \Sigma_i(\bar{w})\), 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 (31) and \(\bar{\zeta}_i'(\tilde{w}) > 0\). Similarly, the continuity of \(\xi_i(\tilde{w})\) with respect to \(w\) implies that there is an open interval with upper bound \(\tilde{w}\) such that \(\xi_i(\tilde{w}) < 0\), and hence \(\Sigma_i'(\tilde{w}) < 0\), on this interval. \(\Sigma_i(\tilde{w})\) 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 \(w \leq w^* < \tilde{w}\). In consequence, \(0 \leq \Sigma_i(w) < \Sigma(w^*)\), 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 (31). Therefore, \(\Sigma_i(\tilde{w})\) will not stop decreasing until reaching the lower bound \(w^*\), namely \(\Sigma_i(w)\) will be decreasing on \([\tilde{w}, w^*]\). We know from the transversality condition (20) that \(\Sigma_i(w) = -\zeta_i(w)/\lambda_i > 0\). Since we have already shown that \(0 \leq \Sigma_i(\tilde{w}) < \Sigma_i(w)\), thus the transversality condition is fulfilled in this case. By using (31) and Lemma 3.1 again, the required assertion (iv) follows.

**Proof of Proposition 3.2.** It follows from condition (b) that \(A_i(w) = 1 + \varepsilon\), a fixed positive constant. The ex post skill distribution term \(B_i(w)\) can be decomposed through

\[
B_i(w) = \frac{1-F_i(w)}{w f_i(w)} \cdot \frac{\bar{F}_i(\infty) - \bar{F}_i(w)}{1-F_i(w)}.
\]

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

\[
\lim_{w \uparrow \infty} \frac{\bar{F}_i(\infty) - \bar{F}_i(w)}{1-F_i(w)} = \lim_{w \uparrow \infty} \frac{\bar{f}_i(w)}{f_i(w)}.
\]

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

\[
C_i(w) = \int_{\infty}^w \left[ 1 + \frac{\gamma_i f_i(t)}{\lambda_i f_i(t)} - \frac{T_i(y_i(t))}{y_i(t) - T_i(y_i(t))} \bar{T}_i(t) \right] \bar{f}_i(t) dt.
\]

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

\[
\lim_{w \uparrow \infty} C_i(w) = 1 + \frac{\gamma_i}{\lambda_i} \alpha_i(\infty) - \frac{T_i(y_i(\infty))}{1 - T_i'(y_i(\infty))} \bar{T}_i(\infty).
\]

So, we get from the optimal tax formula derived in Theorem 3.1 that

\[
\frac{T_i'(y_i(\infty))}{1 - T_i'(y_i(\infty))} = \frac{\gamma_i}{\lambda_i} \alpha_i(\infty) + (1 + \varepsilon) \frac{1}{a_i} \left[ 1 + \frac{\gamma_i}{\lambda_i} \alpha_i(\infty) - \frac{T_i'(y_i(\infty))}{1 - T_i'(y_i(\infty))} \bar{T}_i(\infty) \right],
\]

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

**Proof of Proposition 3.3.** We shall complete the proof in 3 steps. Step 1. By applying condition (a) assumed in Proposition 3.2 and the assumption \(\psi_i(\mu_i, \mu_{-i}) = \sigma_D \in (-1, 0)\) to equation (13) produces

\[
\frac{\gamma_i}{\lambda_i} = \frac{-\sigma_D}{1 + \sigma_D} \bar{F}_i(\infty) > 0,
\]

in which it is unnecessary that \(\bar{F}_i(\infty) = 1\). Also, if \(F_i(w) = F_{-i}(w)\), then by using the definition of \(\bar{f}_i(w)\) we obtain \(\partial \alpha_i(\infty)/\partial(1/a_i) = 0\). Therefore, as long as \(\partial \bar{F}_i(\infty)/\partial(1/a_i) = 0\), we
must have \( \partial \left( \frac{\alpha_i(\infty)}{\lambda_i} \right) / \partial (1/a_i) = 0 \). 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

\[
\frac{\partial T'_i(y_i(\infty))}{\partial \left( \frac{\alpha_i(\infty)}{\lambda_i} \right)} = \left\{ 1 + \frac{\gamma_i}{\lambda_i} \alpha_i(\infty) + (1 + \varepsilon)(1/a_i) \left[ 1 + \frac{\gamma_i}{\lambda_i} \alpha_i(\infty) + \tilde{\theta}_i(\infty) \right] \right\}^2 
\]

by which we hence obtain

\[
\frac{\partial^2 T'_i(y_i(\infty))}{\partial \left( \frac{\alpha_i(\infty)}{\lambda_i} \right) \partial (1/a_i)} = \frac{1 + \varepsilon}{\left\{ 1 + \frac{\gamma_i}{\lambda_i} \alpha_i(\infty) + (1 + \varepsilon)(1/a_i) \left[ 1 + \frac{\gamma_i}{\lambda_i} \alpha_i(\infty) + \tilde{\theta}_i(\infty) \right] \right\}^3} 
\]

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

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

\[
1 + \frac{\gamma_i}{\lambda_i} \alpha_i(\infty) < \frac{\tilde{\theta}_i(\infty) \left\{ 2 + \frac{1 + \varepsilon}{a_i} \left[ 1 + \tilde{\theta}_i(\infty) \right] \right\}}{\left( 1 + \frac{1 + \varepsilon}{a_i} \right) \left[ \tilde{\theta}_i(\infty) - 1 \right]}, \quad \text{if } \frac{\tilde{\theta}_i(\infty) \left\{ 2 + \frac{1 + \varepsilon}{a_i} \left[ 1 + \tilde{\theta}_i(\infty) \right] \right\}}{\left( 1 + \frac{1 + \varepsilon}{a_i} \right) \left[ \tilde{\theta}_i(\infty) - 1 \right]} > 0
\]

Also, noting that

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

the desired assertion in part (ii) follows. ■

**Proof of Theorem 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
We thus obtain
\[ \mathcal{L}_i(\{U_i(w), l_i(w)\}_{w \in [\underline{w}, \overline{w}]}; \mu_i; \lambda_i, \gamma_i, \{\varsigma_i(w)\}_{w \in [\underline{w}, \overline{w}]}) \]
\[ = U_i(w) + \lambda_i \int_{\underline{w}}^{\overline{w}} \left\{ [w l_i(w) - \varphi_i(U_i(w), l_i(w), \mu_i, \mu_{-i}(\mu_i))] \phi_i(U_i(w) - U_{-i}(w); w) - \frac{R}{w - \overline{w}} \right\} dw \]
\[ + \varsigma_i(w) U_i(w) - \varsigma_i(\overline{w}) U_i(\overline{w}) + \int_{\underline{w}}^{\overline{w}} \left[ \varsigma_i(w) h'(l_i(w)) \frac{l_i(w)}{w} + \varsigma_i(w) U_i(w) \right] dw \]
\[ + \gamma_i \left[ \mu_i - \int_{\underline{w}}^{\overline{w}} \varphi_i(U_i(w), l_i(w), \mu_i, \mu_{-i}(\mu_i)) f_i(w) dw \right]. \]

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
\[ \frac{\partial \mu_{-i}}{\partial \mu_i} = \frac{\int_{\underline{w}}^{\overline{w}} \varphi_{-i}(U_{-i}(w), l_{-i}(w), \mu_{-i}, \mu_i) f_{-i}(w) dw}{1 - \int_{\underline{w}}^{\overline{w}} \varphi_{-i}(U_{-i}(w), l_{-i}(w), \mu_{-i}, \mu_i) f_{-i}(w) dw}. \tag{33} \]

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 3.1 but
\[ \frac{\partial \mathcal{L}_i}{\partial \mu_i} = -\lambda_i \int_{\underline{w}}^{\overline{w}} \left( \frac{\partial \varphi_i}{\partial \mu_i} + \frac{\partial \varphi_i}{\partial \mu_{-i}} \frac{\partial \mu_{-i}}{\partial \mu_i} \right) f_i(w) dw \]
\[ + \gamma_i \left[ 1 - \int_{\underline{w}}^{\overline{w}} \left( \frac{\partial \varphi_i}{\partial \mu_i} + \frac{\partial \varphi_i}{\partial \mu_{-i}} \frac{\partial \mu_{-i}}{\partial \mu_i} \right) f_i(w) dw \right] = 0, \]
where \( \frac{\partial \mu_{-i}}{\partial \mu_i} \) is given by equation (33). The proof is thus complete. \( \blacksquare \)

**Proof of Proposition 4.1.** Applying condition (a) and (15) to equation (33) shows that
\[ \frac{\partial \mu_{-i}}{\partial \mu_i} = \frac{-\sigma_F}{1 + \sigma_D}, \]
substituting which into the formula of \( \gamma_i/\lambda_i \) shown in Theorem 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
\[ \frac{\partial (\gamma_i/\lambda_i)}{\partial \sigma_D} = \frac{-\sigma_F^2 + (1 + \sigma_D)^2}{[(1 + \sigma_D)^2 - \sigma_F^2]^2} \tilde{F}_i(\overline{w}) < 0 \]
and
\[ \frac{\partial (\gamma_i/\lambda_i)}{\partial \sigma_F^2} = \frac{1 + \sigma_D}{[(1 + \sigma_D)^2 - \sigma_F^2]^2} \tilde{F}_i(\overline{w}) > 0. \]
As a result, using chain rule, Corollary 4.1 and Proposition 3.2 gives rise to
\[ \frac{\partial^2 T_i(y_i(\infty))}{\partial \sigma_D \partial (1/a_i)} = \frac{\partial^2 T_i(y_i(\infty))}{\partial (\frac{\gamma_i}{\lambda_i})} \cdot \alpha_i(\infty) \frac{\partial (\gamma_i/\lambda_i)}{\partial \sigma_D} > 0 \]
and
\[ \frac{\partial^2 T_i'(y_i(\infty))}{\partial \sigma_i^2 \partial (1/a_i)} = \frac{\partial^2 T_i'(y_i(\infty))}{\partial (\frac{\alpha_i(\infty)}{\lambda_i}) \partial (1/a_i)} \cdot \frac{\alpha_i(\infty)}{\partial \sigma_i^2} < 0 \]

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.

**Proof of Proposition 4.2.** It follows from Theorem 4.1 that
\[ \frac{\partial (\gamma_i/\lambda_i)}{\partial (\partial \mu_i/\partial \mu_i)} = \frac{\Lambda}{\left[ 1 - \int_0^w \left( \frac{\partial c_i(w)}{\partial \mu_i} + \frac{\partial c_i(w)}{\partial \mu_{-i}} \right) \tilde{f}_i(w) dw \right]^2} \]

where
\[ \Lambda \equiv \left[ 1 - \int_0^w \frac{\partial c_i(w)}{\partial \mu_i} f_i(w) dw \right] \int_0^w f_i(w) dw + \int_0^w \frac{\partial c_i(w)}{\partial \mu_{-i}} \tilde{f}_i(w) dw \int_0^w \frac{\partial c_i(w)}{\partial \mu_i} \tilde{f}_i(w) dw. \]

If Assumption 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 4.1 and Assumption 2.1 again gives rise to \( \partial \mu_{-i}/\partial \mu_i > 0 \). Since we get from the optimal tax formula in Theorem 3.1 that optimal marginal tax rates strictly increase in \( \gamma_i/\lambda_i \) and \( \partial \mu_{-i}/\partial \mu_i = 0 \) in the Nash equilibrium, the required assertion accordingly follows.