TAX ME IF YOU CAN! OPTIMAL NONLINEAR INCOME TAX BETWEEN COMPETING GOVERNMENTS*

Etienne Lehmann
Laurent Simula
Alain Trannoy

We investigate how potential tax-driven migrations modify the Mirrlees income tax schedule when two countries play Nash. The social objective is the maximin and preferences are quasi-linear in consumption. Individuals differ both in skills and migration costs, which are continuously distributed. We derive the optimal marginal income tax rates at the equilibrium, extending the Diamond-Saez formula. We show that the level and the slope of the semi-elasticity of migration (on which we lack empirical evidence) are crucial to derive the shape of optimal marginal income tax. *JEL* Codes: D82, H21, H87, F22.

I. Introduction

The globalization process has not just made the mobility of capital easier. The transmission of ideas, meanings, and values across national borders associated with the decrease in transportation costs has also reduced the barriers to international labor mobility. In this context, individuals are more likely to vote with their feet in response to high income taxes. This is particularly the case for highly skilled workers, as recently emphasized by Liebig et al. (2007), Kleven et al. (2013), and Kleven et al. (2014). Consequently, the possibility of tax-driven migrations appears as an important policy issue and must be taken into account as a salient constraint when thinking about the design of taxes and benefits affecting households.

*We thank the editor, the anonymous referees, Spencer Bastani, Craig Brett, Sören Blomquist, Tomer Blumkin, Pierre Boyer, Giacomo Corneo, Peter Egger, Nathalie Ethchart-Vincent, John Hassler, Andreas Haufler, Laurence Jacquet, Eckhart Janeba, Hubert Kempf, Marko Koethenbuerger, Raphael Lalive, Jean-Marie Lozachmeur, Jean-Baptiste Michau, Fabien Moizeau, Olle Folke, Sebastian Koehne, Per Krusell, Ali Sina Onder, Torsten Persson, Thomas Piketty, Jukka Pirttilä, Régis Renault, Emmanuel Saez, Håkan Selin, Yotam Shem-Tov, Emmanuelle Taugourdeau, Farid Toubal, Bruno Van der Linden, John D. Wilson, Owen Zidar, and numerous seminar participants for helpful comments and discussions. A part of this research was realized while Etienne Lehmann was visiting UC Berkeley. Financial support from CEG/UC Berkeley, Fulbright, and IUF is gratefully acknowledged.

The Quarterly Journal of Economics (2014), 1995–2030. doi:10.1093/qje/qju027.
Advance Access publication on September 24, 2014.

The goal of this article is to cast light on this issue from the viewpoint of optimal tax theory. We investigate in what respects potential migrations affect the nonlinear income tax schedules that competing governments find optimal to implement in a Nash equilibrium. For this purpose, we consider the archetypal case of two countries between which individuals are free to move. We extend the model of Mirrlees (1971) to this setting and highlight the impact of potential migrations. By assumption, taxes can only be conditioned on income and are levied according to the residence principle.

The migration margin differs from the "usual" extensive margin because it intrinsically is associated with competition. In contrast, many papers have investigated the extensive margin where agents decide whether to work, either in isolation as in Laroque (2005) or in combination with an intensive choice as in Saez (2002), Kleven et al. (2009), or Jacquet et al. (2013). The possibility that individuals can move between countries shares some similarities with the mobility between economic sectors, which is at the core of the recent analysis by Scheuer and Rothschild (2013). However, in the latter article, agents interact with only one policy maker. Moreover, the agents necessarily remain productive in their home economy, so that there is no specific conflict from the policy maker's viewpoint between the desire to maintain national income per capita and redistribution.

To represent migration responses to taxation in a realistic way, we introduce a distribution of migration costs at each skill level. Hence, every individual is characterized by three characteristics: her birthplace, her skill, and the cost she would incur in case of migration, the last two being private information. As emphasized by Borjas (1999), "the migration costs probably vary among persons [but] the sign of the correlation between costs and (skills) is ambiguous." This is why we do not make any assumption on the correlation between skills and migration costs. Individuals make decisions along two margins. The choice of taxable income operates on the intensive margin, whereas the location choice operates on the extensive margin. In accordance with Hicks's idea, an individual decides to move abroad if her indirect utility in her home country is lower than her utility abroad net of her migration costs. To make the analysis more transparent, we assume away income effects on labor supply as in Diamond (1998) and consider the most redistributive social objective (maximin). Absent mobility, the optimal marginal tax rates under the maximin are the largest implementable ones.[1] We therefore expect that the effect of migration will be maximum under this criterion.

Because of the combination of asymmetric information and potential migration, each government has to solve a self-selection problem with random participation à la Rochet and Stole (2002). Intuitively, each government faces a trade-off between three conflicting objectives: (i) redistributing incomes to achieve a fairer allocation of resources; (ii) limiting the variations of the tax liability with income to reduce marginal tax rates, thereby prevent distortions along the intensive margin; and (iii) minimizing the distortions along the extensive margin to avoid a too large leakage of taxpayers. An additional term appears in the optimal marginal tax rate formula to take the third objective into account. This term depends on the semi-elasticity of migration, defined as the percentage change in the mass of taxpayers of a given skill level when their consumption is increased by one unit. Our main message is that the shape of the tax function depends on the slope of the semi-elasticity, which cannot be deduced from the slope of the elasticity. The theoretical analysis calls for a change of focus in the empirical analysis: in an open economy, if one wants to say something about the shape of tax function, one needs to estimate the profile of the semi-elasticity of migration with respect to earning capacities. We articulate this message with the main findings of the article.

We characterize the best response of each policy maker and obtain a simple formula for the optimal marginal tax rates. The usual optimal tax formula obtained by Piketty (1997), Diamond (1998), and Saez (2001) for a closed economy is augmented by a "migration effect." When the marginal tax rates are slightly increased on some income interval, everyone with larger income faces a lump-sum increase in taxes. This reduces the number of taxpayers in the given country. The magnitude of this new effect is proportional to the semi-elasticity of migration.

Second, we provide a full characterization of the overall shape of the tax function. When the semi-elasticity of migration is constant along the skill distribution, the tax function is increasing. This situation is for example obtained in a symmetric equilibrium when skills and migration costs are independently

1. See Boadway and Jacquet (2008) for a study of the optimal tax scheme under the maximin in the absence of individual mobility.

distributed, as assumed by Morelli et al. (2012) and Blumkin et al. (2014). A similar profile is obtained when the semi-elasticity of migration is decreasing in skills, for example, because of a constant elasticity of migration. When the semi-elasticity is increasing, the tax function may be either increasing, with positive marginal tax rates, or hump-shaped, with negative marginal tax rates in the upper part of the income distribution. A sufficient condition for the hump-shaped pattern is that the semi-elasticity becomes arbitrarily large in the upper part of the skill distribution. If this is the case, progressivity of the optimal tax schedule does not only collapse because of tax competition; the tax liability itself becomes strictly decreasing. There are then “middle-skilled” individuals who pay higher taxes than top-income earners, a situation that can be seen as a “curse of the middle-skilled” (Simula and Trannoy 2010).

Third, we numerically illustrate that the slope is as important as the level of the semi-elasticity, even when one focuses on the upper part of the income distribution. To make this point, we consider three economies, with an income distribution based on that of the United States, which only differ by the profile of the migration responses. More specifically, the average elasticity of migration within the top percentile is the same in all of them. We take this number from the study by Kleven et al. (2014). However, we consider different plausible scenarios for the slope of the semi-elasticity. We obtain dramatically different optimal tax schedules. Obtaining an estimate of the profile of the semi-elasticity is therefore essential to make public policy recommendations.

The article is organized as follows. Section II reviews the literature related to this article. Section III sets up the model. Section IV derives the optimal tax formula in the Nash equilibrium. Section V shows how to sign the optimal marginal tax rates and provides some further analytical characterization of the whole tax function. Section VI numerically investigates the sensitivity of the tax function to the slope of the semi-elasticity of migration. Section VII concludes.

II. Related Literature

We can distinguish two phases in the literature devoted to optimal income taxation in an open economy. In Mirrlees (1971) seminal paper, migrations are supposed to be impossible.

However, Mirrlees emphasizes that this is an assumption one would rather not make because the threat of migration has probably a major influence on the degree of progressivity of actual tax systems. Mirrlees (1982) and Wilson (1980, 1982a) are the first to relax this assumption. Mirrlees (1982) assumes that incomes are exogenously given and derives a tax formula à la Ramsey, the optimal average tax being inversely proportional to the elasticity of migration. Leite-Monteiro (1997) considers the same framework, with differentiated lump-sum taxes and two countries, and shows that tax competition may result in more redistribution in one of the countries. Wilson (1980, 1982a) considers the case of a linear tax. Osmundsen (1999) is the first to apply contract theory with type-dependent outside options to the issue of optimal income taxation in an open economy. He studies how highly skilled individuals distribute their working time between two countries. However, there is no individual trade-off between consumption and effort along the intensive margin.

A second generation of articles investigates optimal nonlinear income tax models in an open economy with the main ingredients that matter, that is, asymmetric information, intensive choice of effort, migration costs, and location choice. Among them, Hamilton and Pestieau (2005), Piaser (2007), and Lipatov and Weichenrieder (2012) consider tax competition on nonlinear income tax schedules in the two-type model of Stiglitz (1982). However, in a two-type setting, the possibility of countervailing incentives is ruled out by assumption. This is one of the reasons Morelli et al. (2012) and Bierbrauer et al. (2013) consider more than two types. Brewer et al. (2010), Simula and Trannoy (2010, 2011), and Blumkin et al. (2014) consider tax competition over nonlinear income tax schedules in a model with a continuous skill distribution. Thanks to the continuous population, it is possible to have insights into the marginal tax rates over the whole income range. Brewer et al. (2010) find that top marginal tax rates should be strictly positive under a Pareto unbounded skill distribution and derive a simple formula to compute them. In contrast, Blumkin et al. (2014) find that top marginal tax rates should be zero. Our article makes clear that this discrepancy arises because Brewer et al. (2010) assume that the elasticity of migration is constant in the upper part of the income distribution. This implies that the semi-elasticity is decreasing. Blumkin et al. (2014) conversely assume that the skills and migration costs are independently distributed. This implies that the semi-elasticity of migration is constant and, thus, that the asymptotic elasticity of migration is infinite. So, the asymptotic marginal tax rate is zero. This is also the case in the framework considered by Bierbrauer et al. (2013). Two utilitarian governments compete when labor is perfectly mobile whatever the skill level. They show that there does not exist equilibria in which individuals with the highest skill pay positive taxes to either country. In our model, there will be some perfectly mobile agents at each skill level. This feature makes our symmetric Nash equilibrium different from the autarkic solution. However, there will also be agents with strictly positive migration costs. Finally, Simula and Trannoy (2010, 2011) assume a single level of migration cost per skill level. There is thus a skill level below which the semi-elasticity of migration is zero and above which it is infinite. This is the reason Simula and Trannoy (2010) find that marginal tax rates may be negative in the upper part of the income distribution. The present article proposes a general framework that encompasses all previous studies.

III. Model

We consider an economy consisting of two countries, indexed by $i = A, B$. The same constant return to scales technology is available in both countries. Each worker is characterized by three characteristics: her native country $i \in \{A, B\}$, her productivity (or skill) $w \in [w_0, w_1]$, and the migration cost $m \in \mathbb{R}^+$ she supports if she decides to live abroad. Note that w_1 may be either finite or infinite and w_0 is nonnegative. In addition, the empirical evidence that some people are immobile is captured by the possibility of infinitely large migration costs. This in particular implies that there will always be a mass of natives of skills w in each country.[2] The migration cost corresponds to a loss in utility, due to various material and psychic costs of moving: application fees, transportation of persons and household's goods, forgone earnings, costs of speaking a different language and adapting to another culture, costs of leaving one's family and friends, and so

2. We could instead assume that $m \in [0, \overline{m}]$, but this would only complicate the analysis. In particular, we might have to deal with the possibilities of "exclusion" of consumer types (namely, a government trying to make its poor emigrate), as typical in the nonlinear price competition literature. In our optimal tax setting, this possibility of exclusion would raise difficult ethical issues, which we prefer to avoid.

on.[3] We do not make any restriction on the correlation between skills and migration costs. We simply consider that there is a distribution of migration costs for each possible skill level.

We denote by $h_i(w)$ the continuous skill density in country $i = A, B$, by $H_i(w) \equiv \int_{w_0}^{w} h_i(x)dx$ the corresponding cumulative distribution function (CDF) and by N_i the size of the population. For each skill w, $g_i(m|w)$ denotes the conditional density of the migration cost and $G_i(m|w) \equiv \int_0^m g_i(x|w)dx$ the conditional CDF. The initial joint density of (m, w) is thus $g_i(m|w)h_i(w)$ while $G_i(m|w)h_i(w)$ is the mass of individuals of skill w with migration costs lower than m.

Following Mirrlees (1971), the government does not observe individual types (w, m). Moreover, it is constrained to treat native and immigrant workers in the same way.[4] Therefore, it can only condition transfers on earnings y through an income tax function $T_i(\cdot)$. It is unable to base the tax on an individual's skill level w, migration cost m, or native country.

III.A. Individual Choices

Every worker derives utility from consumption c, and disutility from effort and migration, if any. Effort captures the quantity as well as the intensity of labor supply. The choice of effort corresponds to an intensive margin and the migration choice to an extensive margin. Let $v(y;w)$ be the disutility of a worker of skill w to obtain pretax earnings $y \geq 0$ with $v'_y > 0 > v'_w$ and $v''_{yy} > 0 > v''_{yw}$. Let $\mathbb{1}$ be equal to 1 if she decides to migrate and to 0 otherwise. Individual preferences are described by the quasi-linear utility function:

$$c - v(y;w) - \mathbb{1} \cdot \mathrm{m}. \tag{1}$$

Note that the Spence-Mirrlees single-crossing condition holds because $v''_{yw} < 0$. The quasi-linearity in consumption implies that there is no income effect on taxable income and appears as a reasonable approximation. For example, Gruber and Saez (2002)

3. Alternatively, the cost of migration can be regarded as the costs incurred by cross-border commuters, who still reside in their home country but work across the border.

4. In several countries, highly skilled foreigners are eligible to specific tax cuts for a limited time duration. This is the case in Sweden and Denmark. These exemptions are temporary.

estimate both income and substitution effects in the case of reported incomes and find small and insignificant income effects. The cost of migration is introduced in the model as a monetary loss.

1. Intensive Margin. We focus on income tax competition under the residence principle. Everyone living in country i is liable to an income tax $T_i(\cdot)$, which is solely based on earnings $y \geq 0$, and thus in particular independent of the native country. Because of the separability of the migration costs, two individuals living in the same country and having the same skill level choose the same gross income/consumption bundle, irrespective of their native country. Hence, a worker of skill w, who has chosen to work in country i, solves:

$$U_i(w) \equiv \max_y \; y - T_i(y) - v(y;w). \tag{2}$$

We call $U_i(w)$ the *gross* utility of a worker of skill w in country i. It is the net utility level for a native and the utility level absent migration cost for an immigrant. We call $Y_i(w)$ the solution to program (2) and $C_i(w) = Y_i(w) - T(Y_i(w))$ the consumption level of a worker of skill w in country i.[5] The first-order condition can be written as:

$$1 - T_i'(Y_i(w)) = v_y'(Y_i(w);w). \tag{3}$$

Differentiating equation (3), we obtain the elasticity of gross earnings with respect to the retention rate $1 - T_i'$,

$$\varepsilon_i(w) \equiv \frac{1 - T_i'(Y_i(w))}{Y_i(w)} \frac{\partial Y_i(w)}{\partial\left(1 - T_i'(Y_i(w))\right)} = \frac{v_y'(Y_i(w);w)}{Y_i(w)\, v_{yy}''(Y_i(w);w)}, \tag{4}$$

and the elasticity of gross earnings with respect to productivity w:

$$\alpha_i(w) \equiv \frac{w}{Y_i(w)} \frac{\partial Y_i(w)}{\partial w} = -\frac{w\, v_{yw}''(Y_i(w);w)}{Y_i(w)\, v_{yy}''(Y_i(w);w)}. \tag{5}$$

2. Migration Decisions. A native of country A of type (w, m) gets utility $U_A(w)$ if she stays in A and utility $U_B(w) - m$ if she relocates to B. She therefore emigrates if and only

5. If equation (2) admits more than one solution, we make the tie-breaking assumption that individuals choose the one preferred by the government.

if: $m < U_B(w) - U_A(w)$. Hence, among individuals of skill w born in country A, the mass of emigrants is given by $G_A(U_B(w) - U_A(w)\,|w)h_A(w)N_A$ and the mass of agents staying in their native country by $(1 - G_A(U_B(w) - U_A(w)|w)h_A(w)N_A$. Natives of country B behave in a symmetric way.

Combining the migration decisions made by agents born in the two countries, we see that the mass of residents of skill w in country A, denoted $\varphi_A(U_A(w) - U_B(w);w)$, depends on the difference in the gross utility levels $\Delta = U_A(w) - U_B(w)$, with:

$$\varphi_i(\Delta;w) \equiv \begin{cases} h_i(w)N_i + G_{-i}(\Delta|w)\,h_{-i}(w)N_{-i} & \text{when } \Delta \geq 0, \\ (1 - G_i(-\Delta|w))\,h_i(w)N_i & \text{when } \Delta \leq 0. \end{cases} \tag{6}$$

We impose the technical restriction that $g_A(0|w)h_A(w)N_A = g_B(0|w)h_B(w)N_B$ to ensure that $\varphi_i(\cdot;w)$ is differentiable. This restriction is automatically verified when A and B are symmetric or when there is a fixed cost of migration, implying $g_i(0|w) = 0$. We have:

$$\frac{\partial\varphi_i(\cdot;w)}{\partial\Delta} = \begin{cases} g_{-i}(\Delta|w)\,h_{-i}(w)N_{-i} & \text{when } \Delta \geq 0, \\ g_i(-\Delta|w)\,h_i(w)N_i & \text{when } \Delta \leq 0. \end{cases}$$

Hence, $\varphi_i(\cdot;w)$ is increasing in the difference Δ in the gross utility levels. By symmetry, the mass of residents of skill w in country B is given by $\varphi_B(U_B(w) - U_A(w);w)$.

All the responses along the extensive margin can be summarized in terms of elasticity concepts. We define the semi-elasticity of migration in country i as:

$$\eta_i(\Delta_i(w);w) \equiv \frac{\partial\varphi(\Delta_i(w);w)}{\partial\Delta}\frac{1}{\varphi(\Delta_i(w);w)} \text{ with } \Delta_i(w) = U_i(w) - U_{-i}(w). \tag{7}$$

Because of quasi-linearity in consumption, this semi-elasticity corresponds to the percentage change in the density of taxpayers with skill w when their consumption $C_i(w)$ is increased at the margin. The elasticity of migration is defined as:

$$\nu_i(\Delta_i(w);w) \equiv C_i(w) \times \eta(\Delta_i(w), w). \tag{8}$$

In words, equation (8) means that if the consumption of the agents of skill w is increased by 1% in country i, the mass of taxpayers with this skill level in country i will change by $\nu_i(\Delta_i(w);w)\%$. Defining the elasticity by multiplying by $C_i(w)$

instead of $\Delta_i(w)$ will pay dividends in terms of ease of exposition later.

III.B. *Governments*

In country $i = A, B$, a benevolent policy maker designs the tax system to maximize the welfare of the worst-off individuals. We chose a maximin criterion for several reasons. The maximin tax policy is the most redistributive, as it corresponds to an infinite aversion to income inequality. A first motivation is therefore to explore the domain of potential redistribution in the presence of tax competition. A second motivation is that in an open economy, there is no obvious way of specifying the set of agents whose welfare is to count (Blackorby et al. 2005). The policy maker may care for the well-being of the natives, irrespective of their country of residence. Alternatively, it may only account for the well-being of the native taxpayers, or for that of all taxpayers irrespective of native country. As an economist, there is no reason to favor one of these criteria (Mirrlees 1982). In our framework and in a second-best setting, these criteria are equivalent. This provides an additional reason for considering maximin governments. The budget constraint faced by country i's government is:

$$\int_{w_0}^{w_1} T_i(Y(w))\,\varphi_i(U_i(w) - U_{-i}(w); w)\,dw \geq E, \tag{9}$$

where $E \geq 0$ is an exogenous amount of public expenditures to finance.[6]

IV. OPTIMAL TAX FORMULA

Following Mirrlees (1971), the standard optimal income tax formula provides the optimal marginal tax rates that should be implemented in a closed economy (e.g., Atkinson and Stiglitz 1980; Diamond 1998; Saez 2001). From another perspective,

6. The dual problem is to maximize tax revenues, subject to a minimum utility requirement for the worst-off individuals, $U_i(w_0) \geq \underline{U}_i(w_0)$. In a closed economy, the dual problem gives rise to the same marginal tax rates as the leviathan (maximization of tax revenues without a minimum utility requirement). Indeed, a variation in the minimum utility requirement $\underline{U}_i(w_0)$ corresponds to a lump-sum transfer and does not alter the profile of marginal tax rates. This is no longer the case in an open economy because a variation in $\underline{U}_i(w_0)$ alters each $U_i(w)$, and thus each $\Delta_i(w)$, thereby modifying the density of taxpayers.

these rates can also be seen as those that should be implemented by a supranational organization ("world welfare point of view," Wilson 1982b) or in the presence of tax cooperation. In this section, we derive the optimal marginal tax rates when policy makers compete on a common pool of taxpayers. We investigate in which way this formula differs from the standard one.

IV.A. *Best Responses*

We start with the characterization of each policy maker's best response. Because a taxpayer interacts with only one policy maker at the same time, it is easy to show that the standard taxation principle holds. Hence, it is equivalent to choose a nonlinear income tax, taking individual choices into account, or to directly select an allocation satisfying the usual incentive-compatible constraints $C_i(w) - v(Y_i(w);w) \geq C_i(x) - v(Y_i(x);w)$ for every $(w,x) \in [w_0, w_1]^2$. Due to the single-crossing condition, these constraints are equivalent to:

$$U_i'(w) = -v_w'(Y_i(w);w), \tag{10}$$

$$Y_i(\cdot)\ \text{nondecreasing}. \tag{11}$$

The best-response allocation of government i to government $-i$ is therefore solution to:

$$\max_{U_i(w),Y_i(w)} \quad U_i(w_0) \quad \text{s.t.} \quad U_i'(w) = -v_w'(Y_i(w);w) \quad \text{and}$$

$$\int_{w_0}^{w_1} (Y_i(w) - v(Y_i(w);w) - U_i(w))\,\varphi_i(U_i(w) - U_{-i}(w);w)\,dw \geq E, \tag{12}$$

in which $U_{-i}(.)$ is given.[7] To save on notations, we from now on drop the i subscripts and denote the skill density of taxpayers and the semi-elasticity in the Nash equilibrium by $f^*(w) = \varphi_i(U_i(w) - U_{-i}(w);w)$ and $\eta^*(w) = \eta_i(U_i(w) - U_{-i}(w);w)$, respectively.

IV.B. *Nash Equilibria*

In Appendix 1, we derive the first-order conditions for equation (12) and rearrange them to obtain a characterization of the

7. The government solves a similar problem as in a closed economy in which agents would also respond to taxation along their participation margin, except that in our setting the reservation utility is exogenous to the government.

optimal marginal tax rates in a Nash equilibrium.[8] We provide an intuitive derivation based on the analysis of the effects of a small tax reform perturbation around the equilibrium.

PROPOSITION 1. In a Nash equilibrium, the optimal marginal tax rates are:

$$\frac{T'(Y(w))}{1-T'(Y(w))} = \frac{\alpha(w)}{\varepsilon(w)} \frac{X(w)}{w f^*(w)}, \tag{13}$$

with

$$X(w) = \int_w^{w_1} [1 - \eta^*(x)\, T(Y(x))] f^*(x)\, dx. \tag{14}$$

Our optimal tax formula (13) differs from the one derived by Piketty (1997), Diamond (1998), and Saez (2001) for a closed economy in two ways: on the one hand, the mass of taxpayers $f^*(\cdot)$ naturally replaces the initial density of skills and, on the other hand, $\eta^*(\cdot)T(Y(\cdot))$ appears in the expectation term $X(w)$. The starred terms capture the competitive nature of Nash equilibrium.[9]

Proposition 1—and all other results—hold in the absence of symmetry. The symmetric case where the two countries are identical ($N_A = N_B$, $h_A(\cdot) = h_B(\cdot) = h(\cdot)$, and $g_A(\cdot|w) = g_B(\cdot|w) = g(\cdot|w)$) is particularly interesting. Indeed, both countries then implement the same policy, which implies $U_A(w) = U_B(w)$. Then, in the equilibrium, no one actually moves but the tax policies differ from the closed-economy ones because of the *threat* of

8. If the solution to the relaxed program that ignores the monotonicity constraint is characterized by incomes that are nondecreasing in skills, then this solution is also the solution to the full program that also includes the monotonicity constraint. In a closed economy and with preferences that are concave in effort, bunching arises when there is a mass point in the skill distribution (Hellwig 2010). In our model, mass points are ruled out by assumption. Moreover, in our simulations, bunching was never optimal.

9. An alternative benchmark would be to look at the country-specific tax schedules that a unique tax authority would implement, taking into account the possibilities of international migration. In such an institutional environment, the well-being of the population would obviously be larger. However, we believe that this benchmark is—for the moment—very idealistic and we therefore prefer to contrast our results to autarky.

migration. The skill density of taxpayers $f^*(\cdot)$ is therefore equal to the exogenous skill density $h(\cdot)$ while equation (7) implies that the semi-elasticity of migration reduces to the structural parameter $g(0|\cdot)$. Obviously, if $g(0|w) \equiv 0$ for all skill levels, the optimal fiscal policy coincides with the optimal tax policy in a closed economy. For instance, this is the case when migration costs include a fixed-cost component. However, in practice, countries are asymmetric and the semi-elasticity is positive as long as the difference in utility in the two countries is larger than the lower bound of the support of the distribution of migration costs. The main difference is that for asymmetric countries the mass of taxpayers $f^*(\cdot)$ and the semi-elasticity of migration $\eta^*(\cdot)$ are both endogenous.

IV.C. Interpretation

We now give an intuitive proof which in particular clarifies the economic interpretation of $X(w)$. To this aim, we investigate the effects of a small tax reform in a unilaterally deviating country: the marginal tax rate $T'(Y(w))$ is uniformly increased by a small amount Δ on a small interval $[Y(w) - \delta, Y(w)]$ as shown in Figure I. Hence, tax liabilities above $Y(w)$ are uniformly increased by $\Delta\,\delta$. This gives rise to the following effects.

First, an agent with earnings in $[Y_i(w) - \delta, Y_i(w)]$ responds to the rise in the marginal tax rate by a substitution effect. From equation (4), the latter reduces her taxable income by:

$$dY(w) = \frac{Y(w)}{1 - T'(Y(w))}\,\varepsilon(w)\,\Delta.$$

This decreases the taxes she pays by an amount:

$$dT(Y(w)) = T'(Y(w))\,dY(w) = \frac{T'(Y(w))}{1 - T'(Y(w))}\,Y(w)\,\varepsilon(w)\,\Delta.$$

Taxpayers with income in $[Y_i(w) - \delta, Y_i(w)]$ have a skill level within the interval $[w - \delta_w, w]$ of the skill distribution. From equation (5), the widths δ and δ_w of the two intervals are related through:

$$\delta_w = \frac{w}{Y(w)}\frac{1}{\alpha(w)}\,\delta.$$

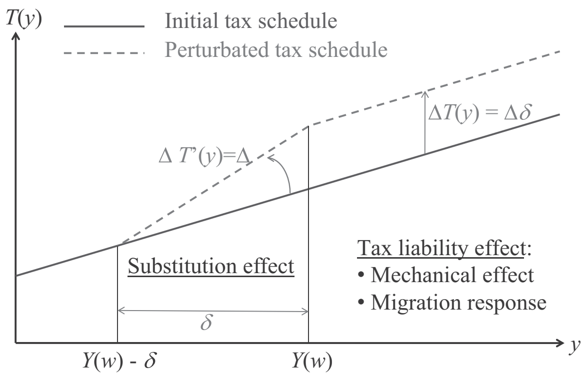


FIGURE I

Small Tax Reform Perturbation

The mass of taxpayers whose earnings are in the interval $[Y_i(w) - \delta, Y_i(w)]$ being $\delta_w f^*(w)$, the total substitution effect is equal to:

$$(15) \qquad dT(Y(w))\,\delta_w\, f^*(w) = \frac{T'(Y(w))}{1 - T'(Y(w))}\frac{\varepsilon(w)}{\alpha(w)}\, w\, f^*(w)\,\Delta\,\delta.$$

Second, every individual with skill x above w faces a lump-sum increase $\Delta\delta$ in her tax liability. In the absence of migration responses, this *mechanically* increases collected taxes from those x-individuals by $f^*(x)\Delta\,\delta$. This is referred to as the "mechanical" effect in the literature. However, an additional effect takes place in the present open-economy setting. The reason is that the unilateral rise in tax liability reduces the gross utility in the deviating country, compared to its competitor. Consequently, the number of emigrants increases or the number of immigrants decreases. From equation (7), the number of taxpayers with skill x decreases by $\eta^*(x) f^*(x)\,\Delta\,\delta$, and thus collected taxes are reduced by:

$$(16) \qquad \eta^*(x)\, T(Y(x)) f^*(x)\,\delta\,\Delta.$$

We define the tax liability effect $X(w)\,\delta\,\Delta$ as the sum of the mechanical and migration effects for all skill levels x above w, where $X(w)$—defined in equation (14)—is the intensity of the tax liability effects for all skill levels above w.

The unilateral deviation we consider cannot induce any first-order effect on the tax revenues of the deviating country; otherwise the policy in the deviating country would not be a best response. This implies that the substitution effect equation (15) must be offset by the tax liability effect $X(w)\,\delta\,\Delta$. We thus obtain Proposition 1's formula.

An alternative way of writing formula (13) illuminates the relationship between the marginal and the average optimal tax rates and captures the long-held intuition that migration is a response to average tax rates. Using the definition of the elasticity of migration, we obtain:

$$\frac{T'(Y(w))}{1-T'(Y(w))}=\frac{\alpha(w)}{\varepsilon(w)}\frac{1-F^*(w)}{wf^*(w)}\left[1-\mathbb{E}_f^*\left(\frac{T(Y(x))}{Y(x)-T(Y(x))}\nu^*(x)|x\geq w\right)\right]. \tag{17}$$

We see that the new "migration factor" makes the link between the marginal tax rate at a given w and the mean of the average tax rates above this w. More precisely, it corresponds to the weighted mean of the average tax rates $\frac{T(Y(x))}{Y(x)-T(Y(x))}$ weighted by the elasticity of migration $\nu^*(x)$, for everyone with productivity x above w. The reason is that migration choices are basically driven by average tax rates, instead of the marginal tax rates.

V. The Profile of the Optimal Marginal Tax Rates

It is trivial to show that the optimal marginal tax rate is equal to zero at the top if skills are bounded from above. It directly follows from equation (17) computed at the upper bound. We also find that the optimal marginal tax rate at the bottom is nonnegative.[10] Our contribution is to characterize the overall shape of the tax function, and thus of the entire profile of the optimal marginal tax rates.

The second-best solution is potentially complicated because it takes into account both the intensive labor supply decisions and the location choices. To derive qualitative properties, we follow the method 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 (migration costs m remain private

10. Indeed, in equation (13), the effect of a lump-sum increase in the tax liability of the least skilled agents is given by $X(w_0)$.

information). We call this benchmark the Tiebout best, as a tribute to Tiebout's seminal introduction of migration issues in the field of public finance.

V.A. The Tiebout Best as a Useful Benchmark

In the Tiebout best, each government faces the same program as in the second best but without the incentive-compatibility constraint (10):

$$\max_{U_i(w),Y_i(w)} \quad U_i(w_0)$$

$$\text{s.t.} \int_{w_0}^{w_1} (Y_i(w) - v(Y_i(w);w) - U_i(w))\,\varphi(U_i(w) - U_{-i}(w);w)\,dw \geq E. \tag{18}$$

The first-order condition with respect to gross earnings $v'(Y(w);w) = 1$ highlights the fact that there is no need to implement distortionary taxes given that skills w are observable. Therefore, a set of skill-specific lump-sum transfers $\tilde{T}_i(w)$ decentralizes the Tiebout best. We now consider the optimality condition with respect to $U(w)$. Because preferences are quasi-linear in consumption, increasing utility $U(w)$ by one unit for a given $Y(w)$ amounts to giving one extra unit of consumption, that is, to decreasing $\tilde{T}_i(w)$ by one unit. In the policy maker's program, the only effect of such a change is to tighten the budget constraint. In the Tiebout best, the $f^*(w)$ workers' taxes are reduced by one unit. However, the number of taxpayers with skill w increases by $\eta^*(w)f^*(w)$ according to equation (7), each of these paying $\tilde{T}_i(w)$. In the Tiebout best, the negative migration effect of an increase in tax liability fully offsets the positive mechanical effect, implying:

$$\tilde{T}_i(w) = \frac{1}{\eta^*(w)}. \tag{19}$$

The tax liability $\tilde{T}_i(w)$ required from the residents with skill $w > w_0$ is equal to the inverse of their semi-elasticity of migration $\eta_i^*(w)$. The least productive individuals receive a transfer determined by the government's budget constraint. Therefore, the optimal tax function is discontinuous at $w = w_0$, as illustrated shortly. We can alternatively express the best response of country i's policy maker using the elasticity of migration instead of the semi-elasticity. We recover the formula derived by Mirrlees (1982):

$$\frac{\tilde{T}_i(w)}{Y_i(w)-\tilde{T}_i(w)}=\frac{1}{\nu(\Delta_i;w)}. \tag{20}$$

Combining best responses, we easily obtain the following characterization for the Nash equilibrium in the Tiebout best. We state it as a proposition because it provides a benchmark to sign second-best optimal marginal tax rates.

PROPOSITION 2. In a Nash equilibrium equilibrium, the Tiebout-best tax liabilities are given by equation (19) for every $w > w_0$, with an upward jump discontinuity at w_0.

V.B. *Signing Optimal Marginal Tax Rates*

The Tiebout-best tax schedule provides insights into the second-best solution, where both skills and migration costs are private information. Using equation (19), equation (14) can be rewritten as:

$$X(w)=\int_w^{w_1}\left[\tilde{T}(x)-T(Y(x))\right]\eta^*(x)f^*(x)\,dx. \tag{21}$$

We see that the tax level effect $X(w)$ is a weighted sum of the difference between the Tiebout-best tax liabilities and second-best tax liabilities for all skill levels x above w. The weights are given by the product of the semi-elasticity of migration and the skill density, that is, by the mass of pivotal individuals of skill w, who are indifferent between migrating or not. In the Tiebout best, the mechanical and migration effects of a change in tax liabilities cancel out. Therefore, the Tiebout-best tax schedule defines a target for the policy maker in the second best, where distortions along the intensive margin have also to be minimized. The second-best solution thus proceeds from the reconciliation of three underlying forces: (i) maximizing the welfare of the worst-off, (ii) being as close as possible to the Tiebout-best tax liability to limit the distortions stemming from the migration responses, and (iii) being as flat at possible to mitigate the distortions coming from the intensive margin. These three goals cannot be pursued independently because of the incentive constraints (10). The following proposition is established in Appendix 2, but we provide graphs that cast light on the main intuitions. We consider the case of purely redistributive tax policies ($E = 0$).

PROPOSITION 3. Let $E = 0$. In a Nash equilibrium:

(i) if $\eta^{*\prime}(\cdot) = 0$, then $T'(Y(w)) > 0$ and $T(Y(w)) < \frac{1}{\eta^*}$ for all $w \in (w_0, w_1)$;
(ii) if $\eta^{*\prime}(\cdot) < 0$, then $T'(Y(w)) > 0$ for all $w \in (w_0, w_1)$;
(iii) if $\eta^{*\prime}(\cdot) > 0$, then either
   (a) $T'(Y(w)) \geq 0$ for all $w \in (w_0, w_1)$;
   (b) or there exists a threshold $\breve{w} \in [w_0, w_1)$ such that $T'(Y(w)) \geq 0$ for all $w \in (w_0, \breve{w})$ and $T'(Y(w)) < 0$ for all $w \in (\breve{w}, w_1)$.
(iv) if $\eta^{*\prime}(\cdot) > 0$ and $\lim_{w \to \infty} \eta^*(w) = +\infty$, then there exists a threshold $\breve{w} \in (w_0, w_1)$ such that $T'(Y(w)) \geq 0$ for all $w \in (w_0, \breve{w})$ and $T'(Y(w)) < 0$ for all $w \in (\breve{w}, w_1)$.

This proposition casts light on the part played by the slope of the semi-elasticity of migration. It considers the three natural benchmarks that come to mind when thinking about it. First, the costs of migration may be independent of w as in Morelli et al. (2012) and Blumkin et al. (2014), implying a constant semi-elasticity in a symmetric equilibrium. This makes sense, in particular, if most relocation costs are material (moving costs, flight tickets, etc.).[11] Second, one might want to consider a constant elasticity of migration, as in Brewer et al. (2010) and Piketty and Saez (2012). In this case, the semi-elasticity must be decreasing: if everyone receives one extra unit of consumption in country i, then the relative increase in the number of taxpayers becomes smaller for more skilled individuals. Third, the costs of migration may be decreasing in w. This seems to be supported by the empirical evidence that highly skilled workers are more likely to emigrate than low-skilled ones (Docquier and Marfouk 2007). This suggests that the semi-elasticity of migration may be increasing in skills. A special case is investigated in Simula and

11. Morelli et al. (2012) compare a unified nonlinear optimal taxation with the equilibrium taxation that would be chosen by two competing tax authorities if the same economy were divided into two states. In their conclusion, they discuss the possible implications of modifying this independence assumption and consider that allowing for a negative correlation might be more reasonable.

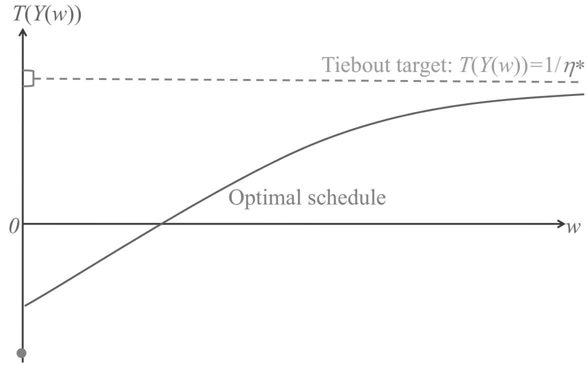


FIGURE II
Constant Semi-Elasticity of Migration

Trannoy (2010, 2011), with a semi-elasticity equal to 0 up to a threshold and infinite above.

The case of a constant semi-elasticity of migration is illustrated in Figure II. The dashed line represents the Tiebout target given by equation (19). It consists of a constant tax level, equal to at $\frac{1}{\eta^*} > 0$ for all $w > w_0$ and redistributes the obtained collected taxes to workers of skill w_0. It is therefore negative at w_0 and then jumps upward to a positive value $\frac{1}{\eta^*} > 0$ for every $w > w_0$. The solid line corresponds to the Nash equilibrium tax schedule in the second best. A flat tax schedule, with $T(Y(w)) \equiv \frac{1}{\eta^*(w)}$, would maximize tax revenues and avoid distortions along the intensive margin. However, it would not benefit to workers of skill w_0. Actually, the laissez-faire policy with $T(Y) \equiv 0$, which is feasible because $E = 0$, would provide workers of skill w_0 with a higher utility level. Consequently, the best compromise is achieved by a tax schedule that is continuously increasing over the whole skill distribution, from a negative value—so that workers of skill w_0 receive a net transfer—to positive values that converge to the Tiebout target $\frac{1}{\eta^*}$ from below. In particular, implementing a negative marginal tax rate at a given w would just make the tax liabilities of the less skilled individuals further away from the Tiebout target, thereby reducing the transfer to the w_0 individuals.

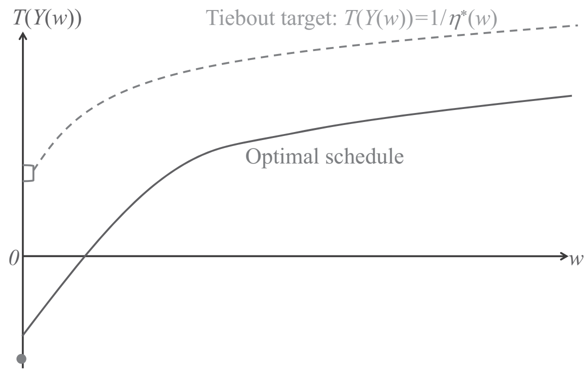


FIGURE III
Decreasing Semi-Elasticity of Migration

The case of a decreasing semi-elasticity of migration is illustrated in Figure III. The Tiebout target is thus increasing above w_0. This reinforces the rationale for having an increasing tax schedule over the whole skill distribution in the second best.

The case of an increasing semi-elasticity of migration is illustrated in Figure IV. The Tiebout target is now decreasing for $w > w_0$. To provide the workers of skill w_0 with a net transfer, the tax schedule must be negative at w_0. It then increases to get closer to the Tiebout target. This is why marginal tax rates must be positive in the lower part of the skill distribution. As shown in Figure IV, two cases are possible for larger w. In case a, the tax schedule is always slowly increasing, to get closer to the Tiebout target, as skill increases. The optimal marginal tax rates are therefore always positive. In case b, the Tiebout target is so decreasing that once the optimal tax schedule becomes close enough to the Tiebout target, it becomes decreasing in skills so as to remain close enough to the target. When the semi-elasticity of migration tends to infinity, the target converges to 0 as skill goes up. Consequently, the optimal tax schedule cannot remain below the target and only case b can occur, as illustrated in Figure V.

V.C. Asymptotic Properties

First, the studies by Brewer et al. (2010) and Piketty and Saez (2012) can be recovered as special cases of our analysis.

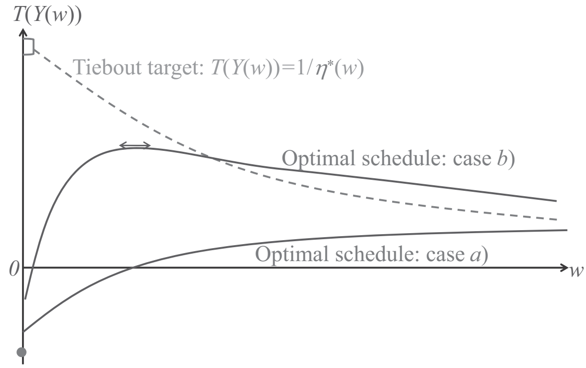


FIGURE IV

Increasing Semi-Elasticity of Migration

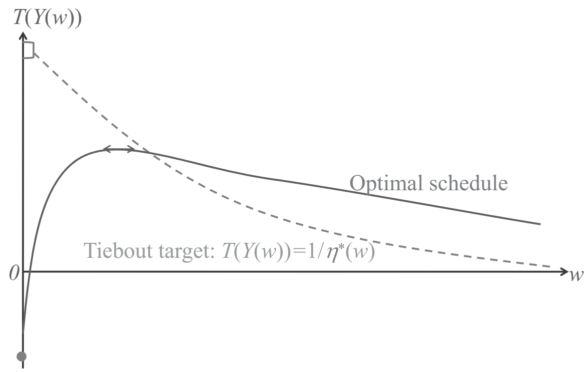


FIGURE V

Increasing Semi-Elasticity of Migration, Converging to Infinity

The latter look at the asymptotic marginal tax rate given potential migration. They assume that the elasticity of migration is constant, equal to ν. From equation (8), a constant elasticity of migration is a special case of a decreasing semi-elasticity, because $C(w)$ must be nondecreasing in the second best. They also assume that the elasticities $\varepsilon(w)$, $\alpha(w)$ converge asymptotically to ε and α, respectively. They finally assume that the distribution of skills is Pareto in its upper part, so that $\frac{wf^*(w)}{\alpha(w)(1-F^*(w))}$ asymptotically converges to k. Making skill w tends to infinity in the optimal tax formula (17), we retrieve their formula for the optimal asymptotic marginal tax rate:[12]

$$T'(Y(\infty)) = \frac{1}{1 + k\varepsilon + \nu}. \tag{22}$$

We see that the asymptotic marginal tax rate is then strictly positive. For example, if $k = 1.5$, $\varepsilon = 0.25$, and $\nu = 0.25$, we obtain $T'(Y(\infty)) = 61.5\%$ instead of 72.7% in the absence of migration responses. Note that when migration costs and skills are independently distributed and the skill distribution is unbounded, as assumed by Blumkin et al. (2014), the elasticity of migration tends to infinity according to equation (8). In this case, the asymptotic optimal marginal tax rate is equal to zero. The result of a zero asymptotic marginal tax obtained by Blumkin et al. (2014) is thus a limiting case of Piketty and Saez (2012).

Second, one may wonder whether the optimal tax schedule must converge asymptotically to the Tiebout target, as suggested in Figure II for the case of a constant elasticity of migration.[13] We can however provide counterexamples where this is not the case. For instance, when the skill distribution is unbounded and approximated by a Pareto distribution, and when the elasticity of migration converges asymptotically to a constant value ν_0, the optimal tax schedule converges to an asymptote that increases at a slope given by the optimal asymptotic marginal tax rate provided by Piketty's and Saez's (2012) formula. Conversely, the Tiebout target is given by equation (20). The Tiebout target therefore converges to an asymptote that increases at a pace $\frac{1}{1+\nu_0}$, which is larger than the asymptotic optimal marginal tax rate. The two schedules must therefore diverge when the skill level tend to infinity.

V.D. Discussion

Proposition 3 shows that the slope of the semi-elasticity of migration is crucial to derive the shape of optimal income tax. According to equation (8), even under the plausible case where the elasticity of migration is increasing over the skill distribution, the semi-elasticity may be either decreasing or increasing,

12. By L'Hôpital's rule, $\lim_{w \mapsto w_1} \frac{T(Y(w))}{Y(w) - T(Y(w))} = \lim_{w \mapsto w_1} \frac{T'(Y(w))}{1 - T'(Y(w))}$.

13. In this case, when the skill distribution is unbounded, Blumkin et al. (2014) show that the tax liability converges to the Tiebout target (which they call the Laffer tax) when the skill increases to infinity.

depending on whether the elasticity of migration is increasing at a lower or higher pace than consumption. In the former case, the optimal tax schedule is increasing and the optimal marginal tax rates are positive everywhere. In the latter case, the optimal tax schedule may be hump-shaped and optimal marginal tax rates may be negative in the upper part of the skill distribution. Therefore, the qualitative features of the optimal tax schedule may be very different, even with a similar elasticity of migration in the upper part of the skill distribution. This point will be emphasized by the numerical simulations of the next section.

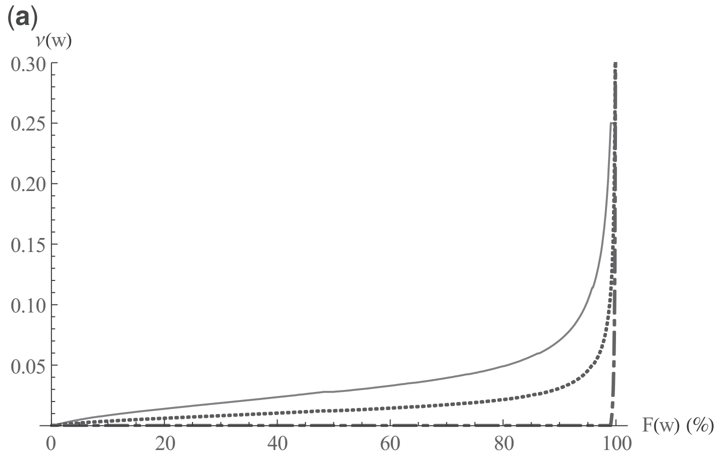
One may wonder why this is the slope of the semi-elasticity of migration and not that of the elasticity that matters in Proposition 3. This is because the distortions along the intensive margin depend on whether marginal tax rates are positive or negative, that is, whether the optimal tax liability is increasing or decreasing. Consequently, the second-best optimal tax schedule inherits the qualitative properties of the Tiebout-best solution, in which tax liabilities are equal to the inverse of the semi-elasticity of migration. We see that to clarify how migrations affect the optimal tax schedule, it is not sufficient to use an empirical strategy that only estimates the level of the migration response, as estimated by Liebig et al. (2007), Kleven et al. (2013), or Kleven et al. (2014). Our theoretical analysis thus calls for a change of focus in the empirical analysis: in an open economy, one needs to also estimate the profile of the semi-elasticity of migration with respect to earning capacities.

VI. Numerical Illustration

This section numerically implements the equilibrium optimal tax formula, so as to emphasize the part played by the slope of the semi-elasticity of migration. In particular, we illustrate the fact that the marginal tax rates faced by rich individuals may be highly sensitive to the overall shape of this semi-elasticity.

For simplicity, we consider that the world consists of two symmetric countries. The distribution of the skill levels is based on the Current Population Survey (CPS) data (2007) extended by a Pareto tail, so that the top 1% of the population gets 18% of total income, as in the United States. The disutility of effort is given by $v(y;w) = (\frac{y}{w})^{1+\frac{1}{\epsilon}}$. This specification implies a constant elasticity of gross earnings with respect to the retention rate ε, as in Diamond (1998) and Saez (2001). We choose $\epsilon = 0.25$, which is a reasonable value based on the survey by Saez et al. (2012).

Even though the potential impact of income taxation on migration choices has been extensively discussed in the theoretical literature, there are still few empirical studies estimating the migration responses to taxation. A first set of studies consider the determinants of migration across U.S. states (see Barro and Sala-i Martin 1991, 1992; Ganong and Shoag 2013; Suarez Serrato and Zidar 2013). They find that per capita income has a positive effect on net migration rates into a state. This conclusion is entirely compatible with an explanation based on tax differences between U.S. states but may also be due to other differences (e.g., in productivities, housing rents, amenities, or public goods). Strong structural assumptions are therefore required to disentangle the pure tax component. A second set of studies focuses exclusively on migration responses to taxation. Liebig et al. (2007) use differences across Swiss cantons and compute migration elasticities for different subpopulations, in particular for different groups in terms of education. Young and Varner (2011) use a millionaire tax specific to New Jersey. Because the salience of this millionaire tax is limited, their estimates of the causal effect of taxation on migration are not statistically significant, except for extremely specific subpopulations. Still, their results suggest that the elasticity of migration is increasing in the upper part of the income distribution. Only two studies are devoted to the estimation of migration elasticities between countries. Kleven et al. (2013) examine tax-induced mobility of football players in Europe and find substantial mobility elasticities. More specifically, the mobility of domestic players with respect to domestic tax rate is rather small around 0.15, but the mobility of foreign players is much larger, around 1. Kleven et al. (2014) confirm that these large estimates apply to the broader market of highly skilled foreign workers and not just to football players. They find an elasticity above 1 in Denmark. In a given country, the number of foreigners at the top is however relatively small. Hence, these findings would translate into a global elasticity at the top of about 0.25 (see Piketty and Saez 2012). Our model pertains to international migrations; based on our survey of the empirical literature, we believe that the best we can do is to use an average elasticity of 0.25 for the top 1%. Moreover, there is no empirical evidence regarding the slope of the semi-elasticity of migration.



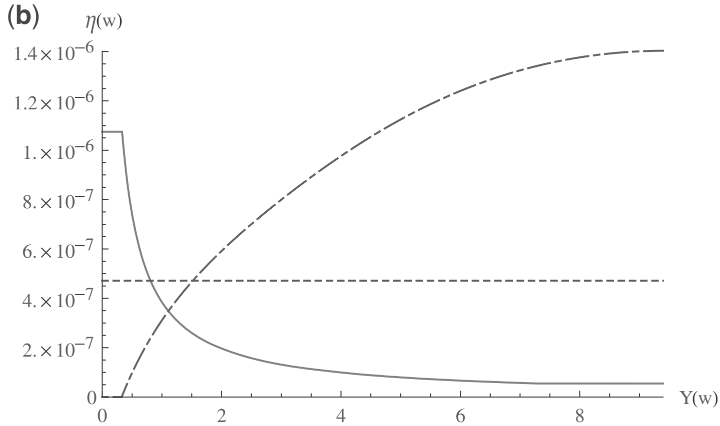


FIGURE VI

Elasticity (a) and Semi-Elasticity (b) of Migration; Case 1 (Plain), Case 2 (Dotted), and Case 3 (Dashed) (Y in Millions of USD)

We therefore investigate three possible scenarios, as shown in Figure VI. In each of them, the average elasticity in the actual economy top 1% of the population is equal to 0.25. In the first scenario, the semi-elasticity is constant up to the top centile and then decreasing in such a way that the elasticity of migration is constant within the top centile. In the second scenario, the semi-elasticity is constant throughout the whole skill distribution. In the third scenario, the semi-elasticity is 0 up to the top centile and then increasing. Note that in the three scenarios, the elasticity of migration is nondecreasing along the skill distribution and remains finite, whereas the semi-elasticity of migration is constant across the bottom 99% of the skill distribution. The average elasticity in the population is higher in the first scenario (0.028) than in the second (0.013) and third (0.003).

The optimal tax liabilities and optimal marginal tax rates in the equilibrium are shown on the the left and right panels of Figure VII, respectively. The x-axis represents annual gross earnings in millions of U.S. dollars. In addition to the three scenarios presented above, we added the tax schedule that would be chosen in a closed economy or in the presence of tax coordination (black curves). Even though the average elasticity of migration is the same for the top 1% of income earners in the three scenarios, we observe significant differences due to variations in the shape of the semi-elasticity of migration. Moreover, the threat of migration implies a nonnegligible decrease in the total taxes paid by top income earners whereas differences in the slope of the semi-elasticity may translate into large differences in marginal tax rates for high-income earners. These numerical results put the stress on the need for empirical studies on the slope of the semi-elasticity of migration, in addition to its level.

In the first case, the tax function is close to being linear for high-income earners and remains close to the closed-economy benchmark. In the second case, the tax function is more concave for large incomes but remains increasing. In the third case, the tax function becomes decreasing around $Y = \$2.9$ million. In particular, the richest people are not those paying the largest taxes. It is very striking that the largest difference in tax liabilities is observed in the third case which exhibits the lowest average elasticity of migration over the total population. This illustrates the fact that the profile of the semi-elasticity of migration within the top centile has a much stronger impact on the optimal tax schedule than the average elasticity of migration within the bottom 99% of the population.

VII. Concluding Comments

This article characterizes the nonlinear income tax schedules that competing Rawlsian governments should implement when individuals with private information on skills and migration costs

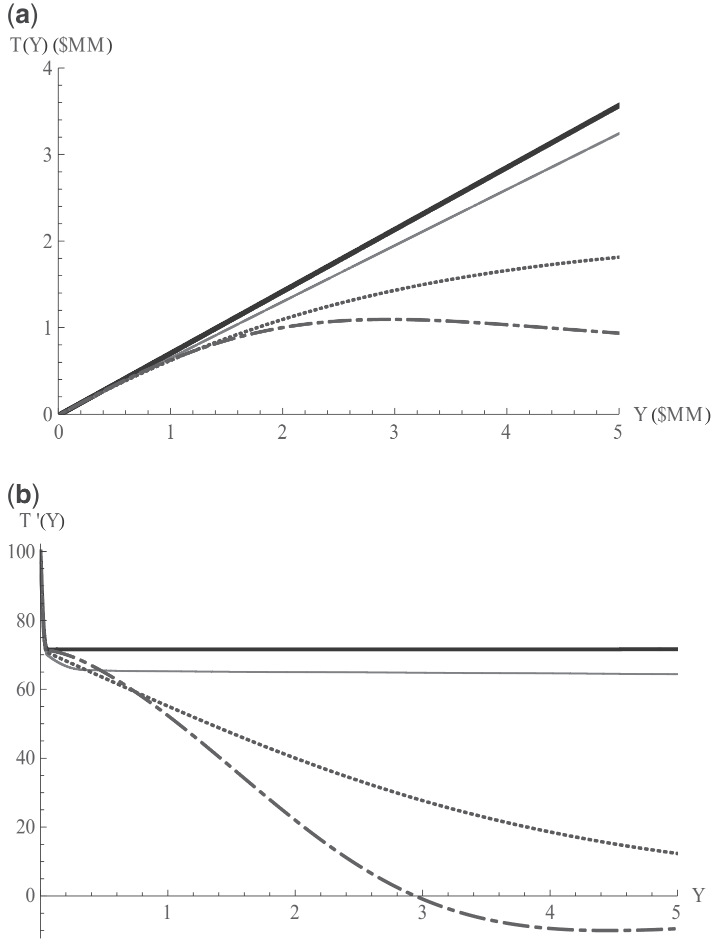


FIGURE VII

Optimal Tax Liabilities (a) and Optimal Marginal Tax Rates (b); Autarky (Bold), Case 1 (Plain), Case 2 (Dotted), and Case 3 (Dashed) (Y in Millions of USD)

decide where to live and how much to work. First, we obtain an optimality rule in which a migration term comes in addition to the standard formula obtained by Diamond (1998) for a closed economy. Second, we show that the optimal tax schedule for top income earners not only depends on the intensity of the migration response of this population, which has been estimated by Liebig et al. (2007), Kleven et al. (2013), and Kleven et al. (2014), but also on the way the semi-elasticity of migration varies along the skill distribution. If the latter is constant or decreasing, optimal marginal tax rates are positive. Conversely, marginal tax rates may be negative if the semi-elasticity of migration is increasing along the skill distribution. To illustrate the sensitivity of marginal tax rates to the slope, we numerically compare three economies that are identical in all aspects, including the average elasticity of migration among the top percentile of the distribution, except that they differ in term of the slope of the semi-elasticity of migration along the skill distribution. We obtain significantly different tax schedules.

Therefore, it is not sufficient to estimate the elasticity of migration. The level as well as the slope of the semi-elasticity of migration are crucial to derive the shape of optimal marginal income tax, even for high income earners. The empirical specification (2) of Kleven et al. (2014) does not allow one to recover the slope of the semi-elasticity. Another specification with additional terms should be estimated.

Different conclusions can be drawn from our results. From a first perspective, the uncertainty about the profile of the semi-elasticity of migration may justify very low, maybe even negative, marginal tax rates for the top 1% of the income earners. This may partly explain why Organisation for Economic Co-operation and Development (OECD) countries were reducing their top marginal tax rates before the financial crisis of 2007. From a second perspective, the potential consequences of mobility might be so substantial in terms of redistribution that governments might want to hinder migration. For example, departure taxes have recently been implemented in Australia, Bangladesh, Canada, the Netherlands, and South Africa. Finally, from a third viewpoint, the problem is not globalization per se but the lack of cooperation between national tax authorities.

A first possibility is an agreement among national policy makers resulting in the implementation of supranational taxes, for example at the EU level. A second possibility relies on the exchange of information between nation-states. Thanks to this exchange, the policy maker of a given country would be able to levy taxes on its citizens living abroad, as implemented by the United States. Indeed, in a citizenship-based income tax system, moving abroad would not change the tax schedule an individual faces, so that the distortions due to tax competition would vanish. There has been some advances in the direction of a better exchange of information between tax authorities. For example, the OECD Global Forum Working Group on Effective Exchange of Information was created in 2002 and contains two models of agreements against harmful tax practices. However, these agreements remain for the moment nonbinding and are extremely incomplete.

Among the various potential extensions, a particularly promising one concerns the possibility for governments to offer nonlinear preferential tax treatments to foreign workers, as in Denmark (Kleven et al. 2014). In particular, it would be interesting to compare our results with the social outcome arising when discrimination based on citizenship is allowed.

Appendix 1: Proof of Proposition 1

We use the dual problem to characterize best response allocations:

$$\max_{U_i(w),Y_i(w)} \int_{w_0}^{w_1} (Y_i(w) - v(Y_i(w);w) - U_i(w))\, \varphi_i(U_i(w) - U_{-i}(w);w)\, dw$$

$$\text{s.t.} \quad U_i'(w) = -v_w'(Y_i(w);w) \quad \text{and} \quad U_i(w_0) \geq \underline{U}_i(w_0), \tag{23}$$

in which $U_i(w_0)$ is given. We adopt a first-order approach by assuming that the monotonicity constraint is slack. We further assume that $Y(\cdot)$ is differentiable. Denoting $q(\cdot)$ the co-state variable, the Hamiltonian associated to equation (23) is:

$$\mathcal{H}(U_i, Y_i, q;w) \equiv [Y_i - v(Y_i;w) - U_i]\, \varphi_i(U_i - U_{-i};w) - q(w)\, v_w'(Y_i;w).$$

Using Pontryagin's principle, the first-order conditions for a maximum are:

$$1 - v_y'(Y_i(w);w) = \frac{q(w)}{\varphi_i(\Delta_i(w);w)}\, v_{yw}''(Y_i(w);w), \tag{24}$$

$$q'(w) = \left\{1 - [Y_i(w) - v(Y_i(w);w) - U_i(w)]\, \eta_i(\Delta_i(w);w)\right\} \varphi_i(\Delta_i(w);w), \tag{25}$$

$$q(w_1) = 0 \text{ when } w_1 < \infty \text{ and } q(w_1) \to 0 \text{ when } w_1 \to \infty, \tag{26}$$

$$q(w_0) \leq 0. \tag{27}$$

Integrating equation (25) between w and w_1 and using the transversality condition (26), we obtain:

$$q(w) = -\int_{w}^{w_1} [1 - \eta^*(x)\,T(Y(x))] f^*(x)\,dx. \tag{28}$$

Defining $X(w) = -q(w)$ leads to equation (14). Equation (24) can be rewritten as:

$$1 - v'_y(Y(w);w) = -\frac{X(w)}{f^*(w)}\, v''_{yw}(Y(w);w). \tag{29}$$

Dividing equation (5) by equation (4) and making use of equation (3), we get

$$v''_{yw}(Y(w);w) = -\frac{\alpha(w)}{\varepsilon(w)}\,\frac{1 - T'(Y(w))}{w}.$$

Plugging equation (3) and the latter equation into equation (29) leads to equation(13).

Appendix 2: Proof of Proposition 3

From equation (13), $T'(Y(w))$ has the same sign as the tax level effect $X(w)$. The transversality condition (27) is equivalent to $X(w_0) \geq 0$, whereas equation (26) is equivalent to $\lim_{w \mapsto w_1} X(w) = 0$. From equation (14), the derivative of $X(w)$ is

$$X'(w) = \left[T(Y(w)) - \frac{1}{\eta^*(w)} \right] \eta^*(w) f^*(w). \tag{30}$$

We now turn to the proofs of the different parts of Proposition 3.

i. $\eta^*(w)$ Is Constant and Equal to η^*

We successively show that any configuration but $T'(Y(w)) > 0$ for all $w \in (w_0, w_1)$ contradicts at least one of the transversality conditions (26) or (27). We start by establishing the following lemmas, which will also be useful for the case where $\eta^*(w)$ is decreasing.

LEMMA 1. Assume that for any $w \in [w_0, w_1]$, $\eta^{*\prime}(w) \leq 0$ and assume there exists a skill level $\hat{w} \in (w_0, w_1)$ such that $T'(Y(\hat{w})) \leq 0$ and $T(Y(\hat{w})) > \frac{1}{\eta^*(\hat{w})}$. Then $X(w_0) < 0$, so the transversality condition (27) is violated.

Proof. As $T(Y(w)) = Y(w) - C(w)$ and $\eta(w)$ are continuous functions of w, there exists by continuity an open interval around $\hat{w}$ where $T(Y(w)) > \frac{1}{\eta^*(w)}$. Let $w^* \in [w_0, \hat{w})$ be the lowest bound of this interval. Then either $w^* = w_0$ or $T(Y(w^*)) = \frac{1}{\eta^*(w^*)}$. Moreover, for all $w \in (w^*, \hat{w}]$, one has that $T(Y(w)) > \frac{1}{\eta^*(w)}$, thereby $X'(w) > 0$ according to equation (30). Hence, one has that $X(w) < X(\hat{w}) \leq 0$, thereby $T'(Y(w)) < 0$ for all $w \in [w^*, \hat{w})$. Consequently, $T(Y(w^*)) > T(Y(\hat{w})) > \frac{1}{\eta^*(\hat{w})} \geq \frac{1}{\eta^*(w^*)}$. So, one must have $w^* = w_0$. Finally, we get $X(w^*) = X(w_0) < 0$, which contradicts the transversality condition (27). **QED**

LEMMA 2. Assume that for any $w \in [w_0, w_1]$, $\eta^{*\prime}(w) \leq 0$ and assume there exists a skill level $\hat{w} \in (w_0, w_1)$ such that $T'(Y(\hat{w})) \leq 0$ and $T(Y(\hat{w})) < \frac{1}{\eta^*(\hat{w})}$. Then $X(w_1) < 0$, so the transversality condition (26) is violated.

Proof. As $T(Y(w)) = Y(w) - C(w)$ and $\eta(w)$ are continuous functions of w, there exists by continuity an open interval around $\hat{w}$ where $T(Y(w)) < \frac{1}{\eta^*(w)}$. Let $w^* \in (\hat{w}, w_1]$ be the highest bound of this interval. Then either $w^* = w_1$ or $T(Y(w^*)) = \frac{1}{\eta^*(w^*)}$. Moreover, for all $w \in [\hat{w}, w^*)$, one has that $T(Y(w)) < \frac{1}{\eta^*(w)}$, thereby $X'(w) < 0$, according to equation (30). Hence, one has that $X(w) < X(\hat{w}) \leq 0$, thereby $T'(Y(w)) < 0$ for all $w \in (\hat{w}, w^*]$. Consequently, $T(Y(w^*)) < T(Y(\hat{w})) < \frac{1}{\eta^*(\hat{w})} \leq \frac{1}{\eta^*(w^*)}$. So, one must have $w^* = w_1$. Finally, we get $X(w^*) = X(w_1) < 0$, which contradicts the transversality condition (26). **QED**

From Lemmas 1 and 2, it is not possible to have $T'(Y(\hat{w})) \leq 0$ and $T(Y(\hat{w})) \neq \frac{1}{\eta(\hat{w})}$, otherwise one of the transversality conditions is violated. Assume there exists a skill level $\hat{w} \in (w_0, w_1)$ such that $T'(Y(\hat{w})) < 0$ and $T(Y(\hat{w})) = \frac{1}{\eta_0(\hat{w})}$. By continuity, there exists $\varepsilon > 0$ such that $T'(Y(\hat{w} - \varepsilon)) < 0$ and $T(Y(\hat{w} - \varepsilon)) > \frac{1}{\eta^*}$, in which case, Lemma 1 applies.

Last, assume there exists a skill level $\hat{w} \in (w_0, w_1)$ such that $T'(Y(\hat{w})) = 0$ and $T(Y(\hat{w})) = \frac{1}{\eta^*(\hat{w})}$. According to the Cauchy-Lipschitz theorem (equivalently, the Picard–Lindelöf theorem), the differential system of equations in $U(w)$ and $X(w)$ defined by equations (10) and (30) (and including equation (13) to express

$Y(w)$ as a function of $X(w)$) with initial conditions that correspond to $T'(Y(\hat{w})) = X(\hat{w}) = 0$ and $T(Y(\hat{w})) = \frac{1}{\eta^*(\hat{w})}$ admits a unique solution where $X(w) \equiv 0$ and $T(\cdot) = \frac{1}{\eta^*}$ for all w. From equation (9), such a solution provides excess budget resources when E is assumed nil and provides less utility level than the *laissez faire* policy where $T(\cdot) = 0$.

Consequently, any case where $T'(Y(\hat{w})) \leq 0$ for $w \in (w_0, w_1)$ leads to the violation of at least one of the transversality conditions.

We finally show that $T(Y(w)) < \frac{1}{\eta^*(w)}$ for all $w \in (w_0, w_1)$. Assume by contradiction that there exists a skill level $\hat{w} \in (w_0, w_1)$ such that $T(Y(\hat{w})) \geq \frac{1}{\eta^*(\hat{w})}$. Because $T'(Y(w)) > 0$ and $\eta^{*\prime}(w) = 0$ for all $w \in (w_0, w_1)$, we have $T(Y(w)) > \frac{1}{\eta^*(w)}$ for all $w \in (\hat{w}, w_1]$. Equation (30) thus implies $X'(w) > 0$ for all $w \in (\hat{w}, w_1]$. Moreover, as we know from above that $T'(Y(w)) > 0$ for all $w \in (w_0, w_1)$, we have in particular $X(\hat{w}) > 0$. Combined with $X'(w) > 0$ for all $w \in (\hat{w}, w_1]$, this implies that $X(w)$ does not tend to zero as w tends to w_1, which contradicts the transversality condition (26).

ii. *$\eta^*(w)$ Is Decreasing*

If there exists a skill level $\hat{w} \in (w_0, w_1)$ such that $T'(Y(\hat{w})) \leq 0$ and $T(Y(\hat{w})) > \frac{1}{\eta^*(\hat{w})}$, Lemma 1 applies. If there exists a skill level $\hat{w} \in (w_0, w_1)$ such that $T'(Y(\hat{w})) \leq 0$ and $T(Y(\hat{w})) < \frac{1}{\eta^*(\hat{w})}$, Lemma 2 applies. Finally, if there exists a skill level $\hat{w} \in (w_0, w_1)$ such that $T'(Y(\hat{w})) \leq 0$ and $T(Y(\hat{w})) = \frac{1}{\eta^*(\hat{w})}$, then the function $w \mapsto T(Y(w)) - \frac{1}{\eta^*(w)}$ is nonpositive and admits a negative derivative at $\hat{w}$, as $\eta^{*\prime}(\cdot) < 0$. Hence, there exists $\overline{w} > \tilde{w}$ such that $T(Y(w)) < \frac{1}{\eta^*(w)}$, thereby $X'(w) < 0$ for all $w \in (\tilde{w}, \overline{w}]$. Consequently, $X'(\overline{w}) < 0$ (equivalently $T(Y(\overline{w})) < \frac{1}{\eta^*(\overline{w})}$) and $X(\overline{w}) < X(\tilde{w}) = 0$ (equivalently $T'(Y(\overline{w})) < 0$), in which case Lemma 2 applies at $\overline{w}$. Consequently, any case where $T'(Y(\hat{w})) \leq 0$ for $w \in (w_0, w_1)$ leads to the violation of at least one of the transversality conditions, which ends the proof of part ii of Proposition 3.

iii. *$\eta^*(w)$ Is Increasing*

We first show two useful lemmas.

LEMMA 3. Assume that for any $w \in [w_0, w_1]$, $\eta^{*\prime}(w) > 0$ and assume there exists a skill level $\hat{w} \in (w_0, w_1)$ such that $T'(Y(\hat{w})) \geq 0$ and $T(Y(\hat{w})) \geq \frac{1}{\eta^*(\hat{w})}$. Then, $X(w_1) > 0$, so the transversality condition (26) is violated.

Proof. We first show that we can assume that $T(Y(\hat{w})) > \frac{1}{\eta^*(\hat{w})}$ without any loss of generality. Assume that $T(Y(\hat{w})) = \frac{1}{\eta^*(\hat{w})}$ and $T'(Y(\hat{w})) \geq 0$. As $\eta^{*\prime}(\cdot) > 0$, the function $w \mapsto T(Y(w)) - \frac{1}{\eta^*(w)}$ is nonnegative and admits a positive derivative at $\hat{w}$. Hence, there exists $\overline{w} > \tilde{w}$ such that $T(Y(w)) > \frac{1}{\eta^*(w)}$, thereby $X'(w) > 0$ for all $w \in (\tilde{w}, \overline{w}]$. Consequently, $X'(\overline{w}) > 0$ (equivalently $T(Y(\overline{w})) > \frac{1}{\eta^*(\overline{w})}$) and $X(\overline{w}) > X(\tilde{w}) = 0$ (equivalently $T'(Y(\overline{w})) > 0$).

Consider now that $T'(Y(\hat{w})) \geq 0$ and $T(Y(\hat{w})) > \frac{1}{\eta^*(\hat{w})}$. As $T(Y(w)) = Y(w) - C(w)$ and $\eta(w)$ are continuous functions of w, there exists by continuity an open interval around $\hat{w}$ where $T(Y(w)) > \frac{1}{\eta^*(w)}$. Let $w^* \in (\hat{w}, w_1]$ be the highest bound of this interval. Then, either $w^* = w_1$ or $T(Y(w^*)) = \frac{1}{\eta^*(w^*)}$. Moreover, for all $w \in [\hat{w}, w^*)$, we have $T(Y(w)) > \frac{1}{\eta^*(w)}$, and thereby $X'(w) > 0$ according to equation (30). Hence, we have $X(w) > X(\hat{w}) \geq 0$, thereby $T'(Y(w)) > 0$ for all $w \in (\hat{w}, w^*]$. Consequently, $T(Y(w^*)) > T(Y(\hat{w})) > \frac{1}{\eta^*(\hat{w})} > \frac{1}{\eta^*(w^*)}$. So, $w^* = w_1$. Finally, we get $X(w^*) = X(w_1) > 0$, which contradicts the transversality condition (26). **QED**

LEMMA 4. Assume that for any $w \in [w_0, w_1]$, $\eta^{*\prime}(w) > 0$ and assume there exists a skill level $\hat{w} \in (w_0, w_1)$ such that $T'(Y(\hat{w})) \geq 0$ and $T(Y(\hat{w})) < \frac{1}{\eta^*(\hat{w})}$. Then, $X(w) > 0$ for all $w < \hat{w}$.

Proof. As $T(Y(w)) = Y(w) - C(w)$ and $\eta(w)$ are continuous functions of w, there exists by continuity an open interval around $\hat{w}$ where $T(Y(w)) < \frac{1}{\eta^*(w)}$. Let $w^* \in [w_0, \hat{w})$ be the lowest bound of this interval. Then, either $w^* = w_0$ or $T(Y(w^*)) = \frac{1}{\eta^*(w^*)}$. Moreover, for all $w \in (w^*, \hat{w}]$, we have $T(Y(w)) < \frac{1}{\eta^*(w)}$, and thereby $X'(w) < 0$ according to equation (30). Hence, we have $X(w) > X(\hat{w}) \geq 0$, thereby $T'(Y(w)) > 0$ for all $w \in (w^*, \hat{w}]$. Consequently, $T(Y(w^*)) < T(Y(\hat{w})) < \frac{1}{\eta^*(\hat{w})} < \frac{1}{\eta^*(w^*)}$. So, $w^* = w_0$ and $X(w) > 0$ for all $w \in [w_0, \hat{w})$. **QED**

According to the transversality condition (27), either $X(w_0) > 0$ or $X(w_0) = 0$. However, in the case where $X(w_0) = 0$, we must have $T'(Y(w)) < 0$ for all $w \in (w_0, w_1)$. Otherwise, either there exists $\hat{w} \in (w_0, w_1)$ such that $T'(Y(\hat{w})) \geq 0$ and $T(Y(\hat{w})) \geq \frac{1}{\eta^*(\hat{w})}$, in which case Lemma 3 implies that the transversality condition (26) is violated, or there exists $\hat{w} \in (w_0, w_1)$ such that $T'(Y(\hat{w})) \geq 0$ and $T(Y(\hat{w})) < \frac{1}{\eta^*(\hat{w})}$, in which case Lemma 4 implies that $X(w_0) > 0$. Consequently, if $X(w_0) = 0$, we must have $T'(Y(w)) < 0$ for all $w \in (w_0, w_1)$. Using equation (9) and the assumption that $E = 0$, this implies that $T(Y(w_0)) > 0 > T(Y(w_1))$. Hence, this policy provides the workers of skill w_0 with less utility than the laissez-faire policy $T(\cdot) = 0$, which contradicts $X(w_0) = 0$. We therefore have established that $X(w_0) > 0$.

By continuity of function $X(\cdot)$, either $X(w) > 0$ for all $w \in [w_0, w_1)$ (equivalently $T'(Y(w)) \geq 0$ for all $w < w_1$) which corresponds to case (a) of part iii of Proposition 3, or there exists $\breve{w} \in (w_0, w_1)$ such that $X(w) > 0$ for all $w < \breve{w}$ and $X(\breve{w}) = 0$. We now show that in the latter case, we must have $T'(Y(w)) < 0$, or equivalently $X(w) < 0$, for all $w \in (\breve{w}, w_1)$. Otherwise, either there exist $\hat{w} \in (\breve{w}, w_1)$ such that $T'(Y(\hat{w})) \geq 0$ and $T(Y(\hat{w})) \geq \frac{1}{\eta^*(\hat{w})}$, in which case Lemma 3 implies that the transversality condition (26) is violated, or there must exist $\hat{w} \in (\breve{w}, w_1)$ such that $T'(Y(\hat{w})) \geq 0$ and $T(Y(\hat{w})) < \frac{1}{\eta^*(\hat{w})}$, in which case Lemma 4 implies that $X(\breve{w}) > 0$), which contradicts $X(\breve{w}) = 0$. Therefore, if there exists $\breve{w} \in (w_0, w_1)$ such $X(\breve{w}) = 0$, then we must have $X(w) > 0$, thereby $T'(Y(w)) > 0$, for all $w < \breve{w}$ and $X(w) < 0$, thereby $T'(Y(w)) < 0$, for all $w \in (\breve{w}, w_1)$, which corresponds to case (b) of part iii of Proposition 3.

iv. $\eta^*(w)$ Is Increasing and Tends to Infinity

From case iii, we know that the marginal tax rates are either positive or there exists a threshold above which they are negative. Assume by contradiction that the marginal tax rates are positive. Then, the tax function is increasing. In addition, it must be positive for some individuals so as to clear the budget constraint. As the semi-elasticity of migration increases to infinity, there exists a skill level $\hat{w}$ at which $T'(Y(\hat{w})) \geq 0$ and $T(Y(\hat{w})) > \frac{1}{\eta^*(\hat{w})}$. Then, the transversality condition (26) is violated according to Lemma 3, which leads to the desired contradiction.

CRED—UNIVERSITÉ PARIS II PANTHÉON ASSAS AND CREST
UPPSALA CENTER FOR FISCAL STUDIES AND UPPSALA UNIVERSITY
AIX-MARSEILLE UNIVERSITÉ (AIX-MARSEILLE SCHOOL OF ECONOMICS), CNRS, AND EHESS

REFERENCES

Atkinson, Anthony B., and Joseph E. Stiglitz, *Lectures on Public Economics* (New York: McGraw-Hill, 1980).

Barro, Robert J., and Xavier Sala-I-Martin, "Convergence across States and Regions," *Brookings Papers on Economic Activity*, 1 (1991), 107–182.

———, "Convergence," *Journal of Political Economy*, 100 (1992), 223–251.

Bierbrauer, Felix, Craig Brett, and John A. Weymark, "Strategic Nonlinear Income Tax Competition with Perfect Labor Mobility," *Games and Economic Behavior*, 82 (2013), 292–311.

Blackorby, Charles, Walter Bossert, and David Donaldson, *Population Issues In Social Choice Theory, Welfare Economics and Ethics* (Cambridge: Cambridge University Press, 2005).

Blumkin, Tomer, Efraim Sadka, and Yotam Shem-Tov, "International Tax Competition: Zero Tax Rate at the Top Re-established," *International Tax and Public Finance*, 2014. doi: 10.1007/s10797-014-9335-y.

Boadway, Robin, and Laurence Jacquet, "Optimal Marginal and Average Income Taxation under Maximin," *Journal of Economic Theory*, 143 (2008), 425–441.

Borjas, George J., "The Economic Analysis of Immigration," in *Handbook of Labor Economics*, O. Ashenfelter, and D. Card, eds. (Amsterdam: North-Holland, 1999), chapter 28, pp. 1697–1760.

Brewer, Mike, Emmanuel Saez, and Andrew Shephard, "Means-Testing and Tax Rates on Earnings," in *Dimensions of Tax Design: The Mirrlees Review*, J. Mirrlees, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles, and J. Poterba, eds. (Oxford: Oxford University Press, 2010), chapter 3, pp. 90–173.

Diamond, Peter A., "Optimal Income Taxation: An Example with U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 88 (1998), 83–95.

Docquier, Fréderic, and Abdeslam Marfouk, "International Migration by Educational Attainment (1990–2000)—Release 1.1," in *International Migration, Remittances and Development*, Calglar Ozden, and Maurice Schiff, eds. (New York: Palgrave Macmillan, 2007), chapter 5.

Ganong, Peter, and Daniel Shoag, "Why Has Regional Income Convergence in the U.S. Declined?," Technical report, HKS WP RWP12-028, 2013.

Gruber, Jonathan H., and Emmanuel Saez, "The Elasticity of Taxable Income: Evidence and Implications," *Journal of Public Economics*, 84 (2002), 1–32.

Hamilton, Jonathan C., and Pierre Pestieau, "Optimal Income Taxation and the Ability Distribution: Implications for Migration Equilibria," *International Tax and Public Finance*, 12 (2005), 29–45.

Hellwig, Martin F., "Incentive Problems with Unidimensional Hidden Characteristics: A Unified Approach," *Econometrica*, 78 (2010), 1201–1237.

Jacquet, Laurence, Etienne Lehmann, and Bruno Van der Linden, "Optimal Redistributive Taxation with Both Extensive and Intensive Responses," *Journal of Economic Theory*, 148 (2013), 1770–1805.

Kleven, Henrik, Claus T. Kreiner, and Emmanuel Saez, "The Optimal Income Taxation of Couples," *Econometrica*, 77 (2009), 537–560.

Kleven, Henrik J., Camille Landais, and Emmanuel Saez, "Taxation and International Migration of Superstars: Evidence from the European Football Market," *American Economic Review*, 103 (2013), 1892–1924.

Kleven, Henrik J., Camille Landais, Emmanuel Saez, and Esben A. Schultz, "Migration and Wage Effects of Taxing Top Earners: Evidence from the Foreigners' Tax Scheme in Denmark," *Quarterly Journal of Economics*, 129 (2014), 333–378.

Laroque, Guy, “Income Maintenance and Labor Force Participation,” *Econometrica*, 73 (2005), 341–376.

Leite-Monteiro, Manuel, “Redistributive Policy with Labour Mobility across Countries,” *Journal of Public Economics*, 65 (1997), 229–244.

Liebig, Thomas, Patrick A. Puhani, and Alfonso Sousa-Poza, “Taxation and Internal Migration—Evidence from the Swiss Census Using Community-Level Variation in Income Tax Rates,” *Journal of Regional Science*, 47 (2007), 807–836.

Lipatov, Vilen, and Alfons Weichenrieder, “Optimal Income Taxation with Tax Competition,” Working Paper 1207, Oxford University Centre for Business Taxation, 2012.

Mirrlees, James A., “An Exploration in the Theory of Optimum Income Taxation,” *Review of Economic Studies*, 38 (1971), 175–208.

———, “Migration and Optimal Income Taxes,” *Journal of Public Economics*, 18 (1982), 319–341.

Morelli, Massimo, Huanxing Yang, and Lixin Ye, “Competitive Nonlinear Taxation and Constitutional Choice,” *American Economic Journal: Microeconomics*, 4 (2012), 142–175.

Osmundsen, Petter, “Taxing Internationally Mobile Individuals—A Case of Countervailing Incentives,” *International Tax and Public Finance*, 6 (1999), 149–164.

Piaser, Gwenaël, “Labor Mobility and Income Tax Competition,” in *International Taxation Handbook*, Greg. N. Gregoriou, and Reads, eds. (Amsterdam: Elsevier, 2007), chapter 4, pp. 73–94.

Piketty, Thomas, “La redistribution fiscale face au chômage,” *Revue Française d’Economie*, 12 (1997), 157–201.

Piketty, Thomas, and Emmanuel Saez, “Optimal Labor Income Taxation,” Technical Report WP No. 18521, NBER 2012.

Rochet, Jean-Charles, and Lars A. Stole, “Nonlinear Pricing with Random Participation,” *Review of Economic Studies*, 69 (2002), 277–311.

Saez, Emmanuel, “Using Elasticities to Derive Optimal Income Tax Rates,” *Review of Economic Studies*, 68 (2001), 205–229.

———, “Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses,” *Quarterly Journal of Economics*, 117 (2002), 1039–1073.

Saez, Emmanuel, Joel Slemrod, and Seth H. Giertz, “The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review,” *Journal of Economic Literature*, 50 (2012), 3–50.

Scheuer, Florian, and Casey Rothschild, “Redistributive Taxation in the Roy Model,” *Quarterly Journal of Economics*, 128 (2013), 623–668.

Simula, Laurent, and Alain Trannoy, “Optimal Income Tax under the Threat of Migration by Top-Income Earners,” *Journal of Public Economics*, 94 (2010), 163–173.

———, “Shall We Keep the Highly Skilled at Home? The Optimal Income Tax Perspective,” *Social Choice and Welfare* (2011), 1–32.

Stiglitz, Joseph E., “Self-Selection and Pareto Efficient Taxation,” *Journal of Public Economics*, 17 (1982), 213–240.

Suarez Serrato, Juan C., and Owen Zidar, “Who Benefits from State Corporate Tax Cuts? A Local Labor Markets Approach with Heterogeneous Firms,” Technical report, UC Berkeley and Stanford University, 2013.

Wilson, John D., “The Effect of Potential Emigration on the Optimal Linear Income Tax,” *Journal of Public Economics*, 14 (1980), 339–353.

———, “Optimal Linear Income Taxation in the Presence of Emigration,” *Journal of Public Economics*, 18 (1982a), 363–379.

———, “Optimal Income Taxation and Migration: A World Welfare Point of View,” *Journal of Public Economics*, 18 (1982b), 381–397.

Young, Cristobal, and Charles Varner, “Millionaire Migration and State Taxation of Top Incomes: Evidence from a Natural Experiment,” *National Tax Journal*, 64 (2011), 255–284.