The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

CESifo Working Papers

www.cesifo.org/wp

Tax Me If You Can! Optimal Nonlinear Income Tax between Competing Governments

Etienne Lehmann
Laurent Simula
Alain Trannoy

CESIFO WORKING PAPER NO. 4351
CATEGORY 1: PUBLIC FINANCE
JULY 2013

Presented at CESifo Area Conference on Public Sector Economics, April 2013

CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 4351*

Tax Me If You Can! Optimal Nonlinear Income Tax between Competing Governments

Abstract

We investigate how the optimal nonlinear income tax schedule is modified when taxpayers can evade taxation by emigrating. We consider two symmetric countries with Maximin governments. Workers choose their labor supply along the intensive margin. The skill distribution is continuous, and, for each skill level, the distribution of migration cost is also continuous. We show that optimal marginal tax rates are nonnegative at the symmetric Nash equilibrium when the semi-elasticity of migration is decreasing in the skill level. When the semi-elasticity of migration is increasing in the skill level, either optimal marginal tax rates are positive everywhere or they are positive for the lower part of the skill distribution and then negative. Numerical simulations are calibrated using plausible values of the semi-elasticity of migration for top income earners. We show that the shape of optimal tax schedule varies significantly, depending on the profile of the semi-elasticity of migration over the entire skill distribution - a profile over which we lack empirical evidence.

JEL-Code: D820, H210, H870.

Keywords: optimal income tax, income tax competition, migration, labor mobility, Nash-equilibrium tax schedules.

Etienne Lehmann
CREST-INSEE
Timbre J360, 15 boulevard Gabriel Péri
France – 92245 Malakoff Cedex
etienne.lehmann@ensae.fr

Laurent Simula
Uppsala Center for Fiscal Studies (UCFS) &
Department of Economics
Uppsala University
Box 513
Sweden – 75120 Uppsala
laurent.simula@nek.uu.se

Alain Trannoy
Aix-Marseille University (Aix-Marseille
School of Economics), CNRS & EHESS
2, rue de la Charité
France – 13002 Marseille
alain.trannoy@univmed.fr

July 24, 2013
We are grateful to Spencer Bastani, Sören Blomquist, Tomer Blumkin, Pierre Boyer, Nathalie Ethchart-Vincent, John Hassler, Laurence Jacquet, Eckhart Janeba, Hubert Kempf, Jean-Baptiste Michau, Olle Folke, Sebastian Koehne, Per Krusell, Ali Sina Onder, Torsten Persson, Régis Renault, Emmanuel Saez, Håkan Selin, Emmanuelle Taugourdeau, Farid Toubal, Bruno Van der Linden and John D. Wilson, as well as participants to seminars at CREST, Ecole Polytechnique, UCFS workshop, the University of Mannheim, the University of Rennes (CREM), the University of Panthéon Assas Paris 2, the ENS-Cachan/Paris-Sud Hotelling seminar, the University of Cergy (THEMA) and the IIES (Stockholm). Etienne Lehmann is also research fellow at IRES-Université Catholique de Louvain, IDEP, IZA and CESifo and junior research fellow of Institut Universitaire de France. Laurent Simula is also research fellow at CESifo and a research associate at Centre d'Economie de la Sorbonne - Ecole Normale Supérieure de Cachan.

I Introduction

The globalization process does not only make the mobility of capital easier. The transmission of ideas, meanings and values across national borders associated with the decrease in transportation costs has reduced 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 in particular the case for highly skilled workers, as recently emphasized by Liebig, Puhani, and Sousa-Poza (2007) across Swiss cantons as well as by Kleven, Landais, and Saez (2010) and Kleven, Landais, Saez, and Schultz (2013) across European countries. Tax-driven migrations appear has an important constraint for the design of redistributive income tax policies. This raises two important public policy issues. On the one hand, what is the social cost of the lack of coordination between tax autorities? On the other hand, given the lack of coordination, what is the best policy in a particular country?

In this article, we address these issues from the viewpoint of optimal income tax theory. For this purpose, we extend the model of Mirrlees (1971) to a setting with two symmetric countries between which individuals may migrate. The world population consists of individuals differing in skills as well as in 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. An individual decides to move abroad if her indirect utility in her home country is lower than her reservation utility, equal to her utility abroad net of her migration costs.[1] As emphasized by Borjas (1999), the migration costs "probably vary among persons [but] the sign of the correlation between costs and wages is ambiguous". For this reason, we do not make any assumption on the joint distribution of skills (herein equal to wages) and migration costs. In each country, a benevolent policymaker aims at redistributing incomes across individuals to achieve a fairer allocation. In doing so, the latter knows the joint distribution of skills and migration costs, but is unable to observe the skill and the migration cost of a particular individual. Because of the combination of asymmetric information and potential migrations, each government faces a self-selection problem with random participation *à la* Rochet and Stole (2002). In order to highlight the main economic effects, we place ourselves in the situation that would lead, in each country, to the most progressive tax scheme in the absence of mobility and examine to which extent the latter is modified due to tax competition. We therefore assume that each policymaker maximizes the well-being of its worst-off citizens (maximin).[2] To make the analysis more transparent, we assume away income effects on taxable income.

We first characterize the best-response of each policymaker and obtain a simple formula for the optimal marginal tax rates. The usual optimal tax formula obtained by Diamond

[1] This is in accordance with Hicks's idea that migration decisions are based on the comparison of earnings opportunities across countries, net of moving costs, which is the cornerstone of practically all modern economic studies of migration (Sjaastad, 1962; Borjas, 1999).

[2] See Boadway and Jacquet (2008) for a recent study of the optimal tax scheme under the Maximin in the absence of individual mobility.

(1998) for a closed economy is augmented by a "migration effect". Indeed, 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, defined as the percentage change in the density of taxpayers of a given skill level when their consumption is increased by one unit.[3]

Second, we provide a full characterization of the overall shape of the tax function. When the semi-elasticity of migration is constant, the tax function is increasing. This situation is for example obtained when skills and migration costs are independently distributed, as assumed by Morelli, Yang, and Ye (2012) and Blumkin, Sadka, and Shem-Tov (2012). A similar profile is obtained when the semi-elasticity of migration is decreasing in skills, because e.g. of a constant elasticity of migration. When the semi-elasticity is increasing, the tax function may be either increasing, with positive top 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 for top income earners. 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 cast light on the limitations of the empirical literature which tries to estimate the sufficient statistics to compute top optimal taxes in the presence of migration. The latter provides local estimates of the elasticity of migration, but only for high income earners. It is easy to recover the semi-elasticity for this population. However, as far as we know, there is no empirical study providing insights into the slope of the semi-elasticity. We show that the slope is equally important as the level of the semi-elasticity, even when one focuses on the top of the income distribution. The level and the slope are the two sufficient statistics to estimate. To make this point, we consider three economies which only differ by the profile of their migration responses. We calibrate the three of them in such a way that the average elasticity of migration within the top percentile is the same. We take this number from the study by Kleven, Landais, Saez, and Schultz (2013). However, we consider three different plausible scenarios for the slope of the semi-elasticity. We obtain dramatically different optimal tax schedules. For example, an agent earning 2 millions of USD per year faces an average optimal tax rate of about 64% in the scenario with a decreasing semi-elasticity, 53% in the scenario with a constant semi-elasticty and 48% in the scenario with an increasing elasticity. In this latter scenario, the marginal tax rates become negative above 3 millions of annual income, so that the richest people do not pay the highest taxes, in absolute levels.

Intuitively, each government faces a trade-off between three conflicting objectives: redistributing incomes to achieve a fairer allocation of resources, i.e. maximizing the transfer

[3]The elasticity of migration corresponds to the product of the semi-elasticity and consumption level.

to the least productive agents; avoiding a too large leakage of taxpayers, i.e. minimizing the distortions along the extensive margin; limiting the variations of the tax liability with income, to reduce marginal tax rates, and thereby prevent distortions along the intensive margin. If individual skills were public information, the government would be able to satisfy these three objectives at the same time, thanks to a set of differentiated lump-sum transfers. We refer to these transfers as the "Tiebout" tax liabilities. They are equal to the inverse of the semi-elasticity of migration, except for the lowest skilled, who receive the sum of all collected taxes. The Tiebout tax schedule is therefore discontinuous at the lowest skill level. When skills – as well as migration costs – are private information, the Tiebout tax schedule is no longer implementable. The government wishes to implement a policy with tax liabilities as close as possible to the Tiebout target, while keeping marginal tax rate sufficiently low to limit the distortions along the intensive margin. When the semi-elasticity of migration is constant, the Tiebout target has an upward jump discontinuity at the bottom of the skill distribution and then remains constant, with positive value. The optimal second-best tax schedule is therefore increasing, located in between the lower and upper bounds defined by the Tiebout target. When the semi-elasticity is decreasing in skills, the Tiebout target is increasing in skills above the initial discontinuity. The second-best tax function has therefore the same shape as with a constant semi-elasticity. In contrast, when the semi-elasticity is increasing, the Tiebout target is positive but decreasing above the initial discontinuity. There are then two possibilities. The optimal schedule may be slowly increasing, getting closer to the Tiebout target as skills go up. In this case, the optimal marginal tax rates are positive. Alternatively, the tax schedule may be initially increasing and then becomes decreasing to remain as close as possible to the Tiebout target, in which case marginal tax rates are negative in the upper part of the skill distribution.

We investigate a world consisting of two countries, which are symmetric in terms of initial population, individual preferences, and social maximin objectives. In this symmetric setting, the semi-elasticity of migration at a given skill level is simply equal to the density of the migration costs at that skill level evaluated at zero. It thus corresponds to an exogenous structural parameter, and we can unambiguously distinguish the circumstances under which it is constant, decreasing or increasing along the skill distribution. In the more realistic case of tax competition between two slightly different economies, the level and the slope of the semi-elasticity of migration remain the relevant statistics to estimate. However, the semi-elasticity of migration has to be evaluated at the difference between the endogenous utility levels in the two countries. It is then endogenous and it is difficult to find exogenous conditions under which it is constant, decreasing or increasing in skills. However, our results on the sign of the optimal marginal tax rates remain valid under these restrictions on the endogenous semi-elasticity of migration.

Our paper is related to different literatures. Mirrlees (1982) considers the case where individuals choose to live in either of two regions and their income is exogenous. As far as we know, Osmundsen (1999) is one of the first to examine income taxation with type-dependent outside options. He studies how highly skilled individuals distribute their working time between two countries. However, there is no individual trade-off between consumption and effort. Leite-Monteiro (1997) consider the case with personalized income taxes. Huber (1999), 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, the two-type assumption rules out by assumption the possibility of countervailing incentives. This is one of the reason why Bierbrauer, Brett, and Weymark (2011) and Morelli, Yang, and Ye (2012) consider many types. Brewer, Saez, and Shephard (2010), Simula and Trannoy (2010), Simula and Trannoy (2011) and Blumkin, Sadka, and Shem-Tov (2012) 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 characterize and quantify the full income tax schedule. Brewer, Saez, and Shephard (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, Sadka, and Shem-Tov (2012) find that top marginal tax rates should be zero. This is because the first paper assumes that the elasticity of migration is constant. This implies that the semi-elasticity is decreasing and, thus, that the Tiebout target is increasing. Blumkin, Sadka, and Shem-Tov (2012) 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. Finally, Simula and Trannoy (2010) and Simula and Trannoy (2011) assume that there is 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 why Simula and Trannoy (2010) find that marginal tax rates may be negative in the upper part of the income distribution. The present paper proposes a general framework that encompasses all of these studies. It clarifies why apparently contradictory results were derived in these articles.

The article is organized as follows. Section II sets up the model. Section III derives the optimal tax formula in the symmetric Nash equilibrium. Section IV shows how to sign the optimal marginal tax rates and provides some further characterization of the whole tax function. Section V uses numerical simulations to investigate the sensitivity of the tax function to the slope of the semi-elasticity of migration. Section VI concludes.

II Model

We consider an economy consisting of two symmetric countries, indexed by $i = A, B$. There is a mass 2 of workers. The same technology is available in both countries. It exhibits constant returns to scales. Hence, workers are paid up to their productivity, which is independent of location. 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 non-negative. In addition, the empirical evidence that some people are immobile is captured by the possibility of infinitely large migration costs.[4] 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, etc.[5] We do not make any restriction on the relationship between skills and migration costs. We simply consider that there is a distribution of migration costs for each possible skill level.

The joint distribution of skills w and migration costs m is initially identical in the two countries. We denote by $f(w)$ the continuously-differentiable skill density, and by $F(w) \equiv \int_{w_0}^{w} f(x)\, dx$ the corresponding cumulative distribution function (CDF). For each skill w, $g(m\,|w)$ denotes the conditional density of the migration cost and $G(m\,|w) \equiv \int_0^m g(x\,|w)\, dx$ the conditional CDF. In each country, the initial joint density of (m, w) is thus $g(m|w)\, f(w)$; note that $G(m\,|w)\, f(w)$ is the density of individuals of skill w whose migration cost is 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.[6] Therefore, it can only condition transfers on earnings y through an income tax function $T_i(.)$. It is unable to base the tax on an individual's skill level w, migration cost m, or native country.

II.A Individual Choices

Every worker derives utility from consumption c, and disutility from effort and migration, if any. In the original article by Mirrlees (1971), effort is synonymous of labour supply. Note that effort is a more general concept than working hours, and can encompass choices made by self-workers and entrepreneurs. Let $v(y; w)$ be the disutility of a worker of skill w to obtain pre-tax earnings $y \geq 0$. Let $\mathbb{1}$ be equal to 1 if she decides to migrate, and to zero otherwise. Individual preferences are described by the quasi-linear utility function:

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

The quasi-linearity in consumption implies that there is no income effect on taxable income. Even though there is much less evidence on the magnitude of the income effects in the literature estimating the effect of taxation on reported income than in the labour supply literature, the quasi-linear specification seems to be a reasonable approximation. For example, Gruber and Saez (2002) estimate both income and substitution effects in the case of reported

[4]Alternatively, we could assume that $m \in [0, \overline{m}]$ but this would only complexify the analysis without changing the main insights.

[5]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.

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

incomes, and find small and insignificant income effects. The cost of migration is additively separable. It is introduced in the model as a monetary loss, which might be due, as previously emphasized, to material or psychological costs. Because of additive separability, 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. Also, if the two countries implement the same tax schedule, a given individual chooses the same bundle at home and abroad.

The choice of effort corresponds to an intensive margin and the migration choice to an extensive margin.

II.A.1 Intensive Margin

The disutility $v(.;.)$ of effort is a twice continuously differentiable function. It is increasing and convex in effort, thereby in pre-tax earnings y. Moreover, it is decreasing in w because it is easier for a more productive individual to earn a given pre-tax income y. Finally, the marginal cost of increasing pre-tax income is larger for more productive agents. Because indifference curves have equation $c = v(y;w) + u$, this assumption implies that the Spence-Mirrlees strict single-crossing condition holds.

Every individual living in country i is liable to an income tax $T_i(.)$, which is solely based on earnings $y \geq 0$, and thus in particular independent of the native country. We focus on income tax competition under the residence principle.[7] Therefore, a worker of skill w, who has chosen to work in country i, solves:

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

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 programme (2) and $C_i(w) = Y_i(w) - T(Y_i(w))$ the consumption level of a worker of skill w in country i.[8] The first-order condition can be written as:

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

Increasing effort to get one extra unit of pre-tax income increases consumption by $1 - T_i'(Y_i(w))$ units, but reduces utility by $v_y'(Y_i(w);w)$ units. Differentiating (3), we obtain the elasticity of gross earnings with respect to the retention rate $1 - T_i'$ and skill level w:

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

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

[7] US citizens, though, are liable to the US income tax on their world incomes; however, US citizens living abroad benefit from a general tax exclusion of \$92,900 in 2011.

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

II.A.2 Extensive Margin

Migration decisions correspond to a choice along the extensive margin. We start with the migration decisions of individuals born in country A. An individual 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 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 $G\left(U_B(w) - U_A(w)\,|w\right) f(w)$ and the mass of agents staying in their native country by $\left(1 - G\left(U_B(w) - U_A(w)\,|w\right)\right) f(w)$. Individuals born in country B behave in a symmetric way.

It is important to note that, at a given skill level, migration flows are going in only one direction. 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 depends on the difference in the gross utility levels $\Delta = U_A(w) - U_B(w)$, with:

$$\varphi\left(\Delta; w\right) \equiv \begin{cases} \left(1 + G(\Delta|w)\right)\, f(w) & \text{when } \Delta \geq 0, \\ \left(1 - G(-\Delta|w)\right)\, f(w) & \text{when } \Delta \leq 0. \end{cases} \tag{6}$$

The function $\varphi(.; w)$ is continuously differentiable, with derivative $\partial\varphi(.; w)/\partial\Delta = g\left(|\Delta|\,\middle|w\right)\, f(w)$. It 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 $\varphi\left(U_B(w) - U_A(w); w\right)$.

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:

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

It corresponds to the percentage change in the density of taxpayers with skill w when their consumption $C_i\left(w\right)$ is increased at the margin. The elasticity of migration is defined as $\nu_i\left(\Delta_i(w); w\right) \equiv C_i\left(w\right) \times \eta\left(\Delta_i(w), w\right)$.

II.B Governments

In each country $i = A, B$, a benevolent policy-maker designs the tax system so as 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 one, 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, Bossert, and Donaldson, 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 favour one of these criteria (Mirrlees, 1982). We focus on the maximin because the set of agents whose welfare is accounted for is then independent of the tax policy. So all these criteria are equivalent.[9] The budget constraint faced by country i's government is:

$$\int_{w_0}^{w_1} T_i\left(Y\left(w\right)\right)\ \varphi\left(U_i\left(w\right)-U_{-i}\left(w\right);w\right)\ dw \geq E \tag{8}$$

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

III 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, these rates can also be seen as those that should be implemented by a supranational organization ("world welfare point of view" in Wilson (1982)) 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. We start with the characterization of the best response allocations, before focusing on the symmetric Nash equilibria. We provide a formal as well as an intuitive derivation based on the analysis of the effects of a small tax reform perturbation around the equilibrium (Piketty, 1997; Saez, 2001).

Each government is unable to condition taxes on skill levels w, migration costs m and native country. It thus faces a multidimensional screening problem. However, because migration costs enter separably in the individual utility function (1), two individuals with the same skill level make the same intensive choice irrespective of their other personal characteristics. The fact that migration costs and native country are unobservable thus only matters for the migration decision and not for the intensive one. The government's problem thus belongs to the class of multidimensional screening problems with random participation (Rochet and Stole, 2002; Jacquet, Lehmann, and Van der Linden, 2013).

The standard taxation principle holds. The tax policy can thus be decentralized by an allocation satisfying the usual incentive-compatible constraints. Due to the single-crossing condition $v'' < 0$, these constraints are equivalent to:

$$U_i'\left(w\right) = -v_w'\left(Y_i\left(w\right);w\right), \tag{9}$$

$$Y_i(\cdot) \text{ non-decreasing.} \tag{10}$$

[9]This equivalence holds because, at each skill level, the conditional distribution of migration costs m is unbounded from above. Hence, there are always individuals for whom migration is not a valuable option. See Simula and Trannoy (2011).

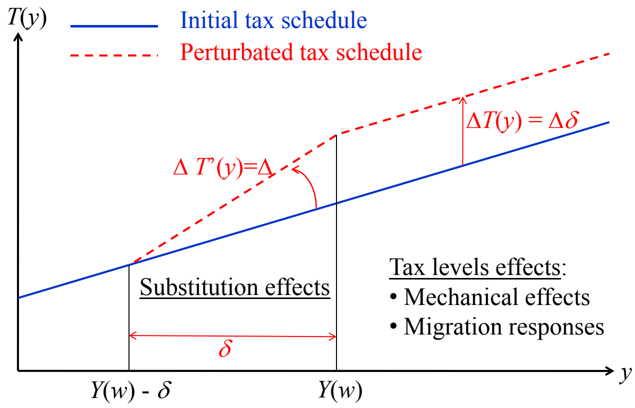

Figure 1: Small Tax Reform Perturbation

The best-response allocation of government i to government $-i$ is solution to:[10]

$$\max_{U_i(w),Y_i(w)} \quad U_i(w_0) \quad \text{s.t.} \quad U_i'(w) = -v_w'(Y_i(w); w) \quad \text{and}$$
$$\int_{w_0}^{w_1} (Y_i(w) - v(Y_i(w); w) - U_i(w))\, \varphi\left(U_i(w) - \underline{U}_{-i}(w); w\right) dw \geq E,$$

in which $\underline{U}_{-i}(.)$ is given. We use the dual problem to characterize best response allocations:

$$\max_{U_i(w),Y_i(w)} \quad \int_{w_0}^{w_1} (Y_i(w) - v(Y_i(w); w) - U_i(w)) \;\; \varphi\left(U_i(w) - \underline{U}_{-i}(w); w\right) dw$$
$$\text{s.t.} \quad U_i'(w) = -v_w'(Y_i(w); w) \quad \text{and} \quad U_i(w_0) \geq \underline{U}_i(w_0), \tag{11}$$

in which $\underline{U}_i(w_0)$ is given.

We focus on symmetric Nash equilibria for two reasons. First, they provide insights into the asymmetric world, as will be shown below. Second, by investigating symmetric equilibria, we illustrate the impact of the *threat* of migration because nobody actually moves. In this case, the gross utility levels of an agent of skill w are the same in A and in B, and we drop the A and B subscripts which are no longer necessary. This implies that $\varphi(\Delta_i; w) = f(w)$ and $\eta(\Delta_i; w) = g(0|w)$, using (6) and (7). Note that the semi-elasticity of migration, denoted $\eta_0(w)$, is equal to the *structural* parameter $g(0|w)$. The optimality conditions (22a)-(22d) can be simplified to obtain the following characterization.

Proposition 1. *In a symmetric Nash equilibrium, the marginal tax rates are:*

$$\frac{T'(Y(w))}{1 - T'(Y(w))} = \frac{\alpha(w)}{\varepsilon(w)} \; \frac{X(w)}{w \; f(w)}, \tag{12}$$

with

$$X(w) = \int_w^{w_1} [1 - \eta_0(x) \; T(Y(x))] \, f(x) \; dx. \tag{13}$$

[10]We are adopt the first-order approach that consists in considering only the first-order incentive constraint (9) and not the second-order one (10). If the solution to this "relaxed" program is non-decreasing, which is the case in all of our simulations, then the solution to this relaxed program is also the solution to the true program that also considers (10).

We provide a proof based on optimal control theory in Appendix I.A. We now give an intuitive proof which in particular clarifies the economic interpretation of $X(w)$. 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 Δ on the interval $[Y(w)-\delta, Y(w)]$ as shown in Figure 1.[11] Hence tax liabilities above $Y(w)$ are uniformly increased by $\Delta\ \delta$. This gives rise to the following effects.

First, everyone with earnings in $[Y_i(w)-\delta, Y_i(w)]$ responds to the rise in the marginal tax rate by a *substitution effect*. Each of them reduces her taxable income by:

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

according to (4). This reduces the tax she pays by:

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

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 (5), their widths δ and δ_w are related through:

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

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

$$dT\left(Y(w)\right)\ \delta_w\ f(w) = \frac{T'\left(Y(w)\right)}{1-T'\left(Y(w)\right)}\frac{\varepsilon(w)}{\alpha(w)} w\ f(w)\ \Delta\ \delta. \tag{14}$$

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 (7), the number of taxpayers with skill x decreases by $\eta_0(x)\ f(x)\ \Delta\ \delta$, and thus collected taxes are reduced by

$$\eta_0(x)\ T\left(Y(x)\right)\ f(x)\ \Delta\ \delta \tag{15}$$

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 $X(w)$ is the intensity of the tax liability effects for all skill levels above w and is defined in (13).

The unilateral deviation we consider cannot induce any first-order effect on the tax revenues of the deviating country. This implies that the substitution effect (14) must be offset by the tax level $X(w)\ \Delta\ \delta$. We thus obtain the optimal income tax formula (12) of Proposition 1.

[11] Tax perturbation approaches have, in particular, been used by Piketty (1997) and Saez (2001) to characterize optimal income taxes in a closed economy.

Our optimal tax formula (12) is identical to the one provided by Diamond (1998) and Saez (2001), except for the presence of $\eta_0(w)\ T(Y(w))$ in the expectation term. This new term captures the migration responses associated to a unilateral deviation of one government in its fiscal policy. The intensity of these migration responses for workers of skill w is given by the semi-elasticity $\eta_0(w)$. At a symmetric equilibrium, this semi-elasticity is equal to the density of migration cost at zero, conditional on skill w, namely $g(0|w)$.

Obviously, if one had $g(0|w) \equiv 0$ for all skill levels, the optimal fiscal policy at the symmetric Nash equilibrium would then coincide with the optimal tax policy in a closed economy. This would be for instance the case when migration costs includes a fixed cost component. The latter observation might cast a doubt about the practical interest of Proposition 1. However, we can extend our reasonings to the case of asymmetric economies. Then, the relevant semi-elasticity to consider in the migration response (15) would be the semi-elasticity at $U_A(w)-U_B(w)$, which is endogenous and equal to $g\left(|U_A(w) - U_B(w)|\,|w\right)\ f(w)/\varphi\left(U_A(w) - U_B(w); w\right)$, instead of the structural parameter $g(0|w)$. The former semi-elasticity remains positive, which is empirically relevant, 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. Beyond the symmetric case, the semi-elasticity of migration computed at the difference in equlibrium utilities remains a key sufficient statistics to compute the optimal tax schedule in an open economy. For the sake of tractability, we focus on the symmetric case with positive $g(0|w)$.

An alternative way of writing the formula (12) given in Proposition 1 illuminates the relationship between the marginal and the average optimal tax rates. Using in Equation (12) that the elasticity of migration is

$$\nu_0(w) \equiv (Y(w) - T\left(Y(w)\right))\ \ \eta_0(w), \tag{16}$$

we obtain:

$$\frac{T'\left(Y\left(w\right)\right)}{1-T'\left(Y\left(w\right)\right)} = \frac{\alpha\left(w\right)}{\varepsilon\left(w\right)}\frac{1-F\left(w\right)}{wf\left(w\right)}\left[1-\mathbb{E}_f\left(\frac{T\left(Y\left(x\right)\right)}{Y\left(x\right)-T\left(Y\left(x\right)\right)}\nu_0\left(x\right)|x\geq w\right)\right]. \tag{17}$$

This alternative way of writing the optimal tax rate formula shows that the new "migration factor" makes the link between the marginal tax rate at a given w and the mean of the average tax liabilities 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 $\nu_0(x)$, 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.

IV 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 the distribution of skills is bounded from above. We also find that the optimal marginal tax rate at the bottom is nonnegative. 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 both the intensive labor supply decisions and the location choices into account. To derive qualitative properties, we follow the method developed by Jacquet, Lehmann, and Van der Linden (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 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.

IV.A A Useful Benchmark: The Tiebout Best

The Tiebout benchmark is first-best in the intensive labor supply dimension and second-best in the migration dimension. It thus corresponds to the "first-and-a-half" best of (Jacquet, Lehmann, and Van der Linden, 2013).

In the second best setting, the first-order incentive constraint (9) implies that the utility level in non-decreasing in skills. Therefore, the maximin objective is equivalent to the maximization of the utility $U(w_0)$ of the least skilled workers. This equivalence does no longer hold when skills are observable. There are consequently different possible ways of defining the Tiebout best: it may correspond to the maximization of the utility of the worst off or alternatively to the maximization of the utility of the least skilled worker. We chose the second definition, because the Tiebout best is then particularly helpful to cast light on the second best.

In the Tiebout best, each government faces the same program as in second-best but without the incentive-compatibility constraint (9):

$$\begin{aligned} &\max_{U_i(w),Y_i(w)} \quad U_i(w_0) \\ &\text{s.t.} \quad \int_{w_0}^{w_1} \left(Y_i\left(w\right) - v\left(Y_i\left(w\right);w\right) - U_i\left(w\right)\right)\ \varphi\left(U_i\left(w\right) - \underline{U}_{-i}\left(w\right);w\right)\ dw \geq E, \end{aligned}$$

We first consider the condition with respect to gross earnings $Y(w)$, which is $v'\left(Y(w);w\right) = 1$ in the symmetric Tiebout equilibrium. Because w is observable, there is no need to implement distortionary taxes. A set of lump-sum transfers $\tilde{T}_i(w)$, differentiated by skill levels, is sufficient to decentralize the Tiebout solution.

We now consider the optimality condition with respect to $U(w)$. Because preferences are quasilinear in consumption, increasing utility $U(w)$ by one unit for a given $Y(w)$ amounts to giving one extra unit of consumption, i.e. to decreasing by one unit the skill-specific lump-sum transfer $\tilde{T}_i(w)$ for $w > 0$. Regarding the policymaker's programme, the only effect of such a change is to tighten the budget constraint. At the symmetric equilibrium, taxes are reduced by one unit for $f(w)$ workers. However, the number of taxpayers increases by $\eta_0(w)\ f(w)$ according to (7). As a result,

$$\tilde{T}(w) = \frac{1}{\eta_0(w)}. \tag{18}$$

.

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 $\eta_i(\Delta_i(w), w)$. The least productive individuals receive a transfer determined by the budget constraint. Therefore, the optimal tax function is discontinuous at $w = w_0$, as illustrated in Figures 3 – 4. Using (16), we can alternatively express the best response of country i's policy-maker using the elasticity of migration instead of the semi-elasticity. We obtain the formula derived by Mirrlees (1982):

$$\frac{\tilde{T}_i(w)}{Y_i(w) - \tilde{T}_i(w)} = \frac{1}{\nu(\Delta_i; w)}. \tag{19}$$

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

Proposition 2. *In the Tiebout symmetric equilibrium, the tax liabilities are given by* (18) *for every $w > w_0$, with an upwards jump discontinuity at w_0.*

In the Tiebout best, the optimal tax liability is increasing in skill when the semi-elasticity of migration $\eta_0(.)$ is decreasing. Symmetrically, the tax liability is decreasing when $\eta_0(.)$ is increasing. Knowing how the semi-elasticity of migration varies with skills is therefore key to determine the profile of the optimal tax schedule.

Three natural benchmarks come to mind when thinking about the slope of the semi-elasticity. First, the costs of migration may be decreasing in w. This seems to be supported by the empirical evidence that highly skilled are more likely to emigrate than low skilled (Docquier and Marfouk, 2006). This suggests that the semi-elasticity of migration may be increasing in skills. A special case is investigated in Simula and Trannoy (2010) and Simula and Trannoy (2011), with a semi elasticity eqaul to zero up to a threshold and infinite above. Second, the costs of migration may be independent of w as in Blumkin, Sadka, and Shem-Tov (2012) and Morelli, Yang, and Ye (2012). This makes sense, in particular, if most relocation costs are material (moving costs, flight tickets, etc.).[12] Third, one might want to consider a constant elasticity of migration, as in Brewer, Saez, and Shephard (2010) and Piketty and Saez (2012). In this case, the semi-elasticity must be non-increasing: 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.

The next subsections will show that the profile of the semi-elasticity of migration will also play an essential part in the second best.

IV.B Signing Optimal Marginal Tax Rates

The Tiebout-best tax schedule provides insight into the second-best solution, where individuals skills and migration costs are private information. Using (18), Equation (13) can

[12] Morelli, Yang, and Ye (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.

be rewritten as:

$$X(w) = \int_{w}^{w_1} \left[\tilde{T}(x) - T\left(Y(x)\right)\right] \eta_0(x) f(x)\ dx. \tag{20}$$

We see that the tax level effect at w is the weighted sum of the difference between the Tiebout optimal 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, i.e. by the mass of pivotal individuals of skill w, who are indifferent between migrating or not. Intuitively, in the Tiebout best, the mechanical and migration effects of a change in tax liabilities cancel out. The tax schedule in the Tiebout best thus 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; *iii*) being as flat at possible to mitigate the distortions coming from the intensive margin.

In the Tiebout best, these three goals can be pursued independently. This is however no longer the case in the second best because of the incentive constraints (9). The following proposition is established in Appendix I.B.

Proposition 3. *Let $E = 0$. In the symmetric Nash equilibrium,*

i) if $\eta_0' = 0$, marginal tax rates are positive $T'(Y(w)) > 0$ for $w \in (w_0, w_1)$;

ii) if $\eta_0' < 0$, marginal tax rates are positive $T'(Y(w)) > 0$ for $w \in (w_0, w_1)$;

iii) if $\eta_0' > 0$, then

- *(a) the marginal tax rates are either positive $T'(Y(w)) \geq 0$ for $w \in (w_0, w_1)$;*
- *(b) or there exists a threshold $\breve{w} \in [w_0, w_1)$ such that $T'(Y(w)) \geq 0$ for $w \in (w_0, \breve{w})$ and $T'(Y(w)) < 0$ for $w \in (\breve{w}, w_1)$.*

The assumption that the exogenous amount of public good $E = 0$ is nil implies that the *laissez faire* policy with $T(Y) \equiv 0$ satisfies the budget constraint. So the best response of each government must be such that the individuals of skill w_0 are at least as well as in the *laissez faire*. The above Proposition is proved in Appendix I.B but the following graphs illustrates the intuitions.

The case where the semi-elasticity of migration is constant is illustrated in Figure 2. The dashed line represents the "Tiebout target" given by Equation (18). It consists of a constant tax level, equal to at $1/\eta_0(w) > 0$ for all $w > w_0$ and redistribute the obtained collected taxes to workers of skill w_0. It is therefore negative at w_0 and then jumps upwards to a positive value $1/\eta_0(w) > 0$ for any $w > w_0$. The solid line corresponds to the Nash-equilibrium tax schedule in the second best. A flat tax schedule, with $T(Y(w)) \equiv 1/\eta_0(w)$, would maximize tax revenues and avoid distortions along the intensive margin. It would however not benefit to workers of skill w_0. Actually, the *laissez faire* policy where $T(Y(w)) \equiv 0$ would provide workers of skill w_0 with a higher utility level. Consequently, the best compromise is achieved by a tax schedule that is copntinuously 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 remain below $1/\eta_0(w)$.

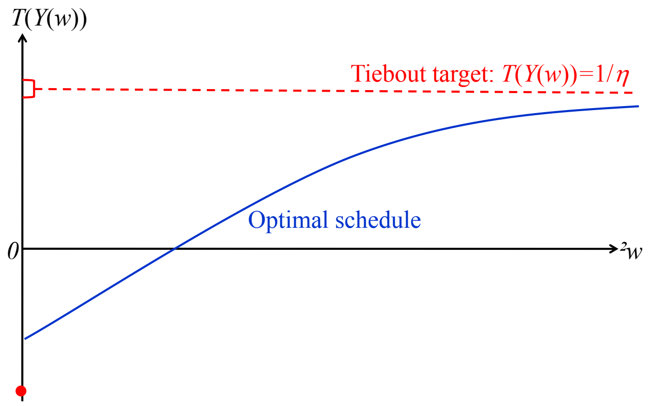

Figure 2: Constant Semi-Elasticity of Migration

The case where the semi-elasticity of migration is decreasing is illustrated in Figure 3. From (18), the Tiebout target is now increasing above w_0. This reinforces the rationale for having an increasing tax schedule over the whole skill distribution.

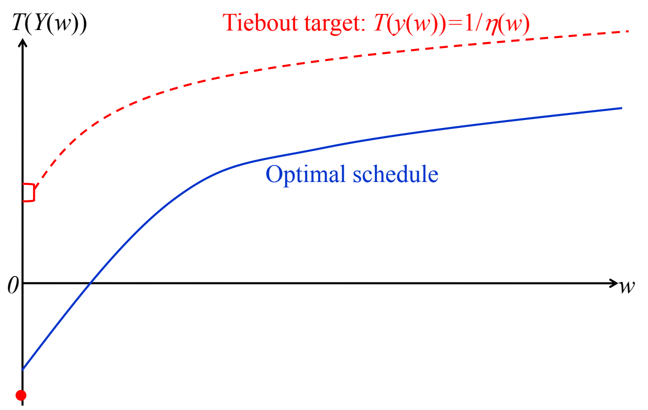

Figure 3: Decreasing Semi-Elasticity of Migration

The case where the semi-elasticity of migration is increasing is illustrated in Figure 4. From (18), 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. Two cases are possible for larger w. In the first case, 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 the second case, the Tiebout target is so decreasing than once the optimal tax schedule becomes close enough to the target, it becomes decreasing in skills so as to remain close enough to the Tiebout target (cf. Figure 4.a)). Note that when the semi-elasticity of migration tends to infinity, the Tiebout target converges to 0 as skill goes up. Consequently, the optimal tax schedule cannot remain below the Tiebout target, and only the second case can occur. This leads to the following Proposition, which is Proved in Appendix I.C.

Proposition 4. *Assume that* $\eta_0'(w) > 0$ *and* $\lim_{w_1 \to \infty} \eta_0(w) = \infty$*, then there exists a threshold* $\hat{w} \in (w_0, w_1)$ *under which* $T'(Y(w)) > 0$ *and above which* $T'(Y(w)) < 0$*.*

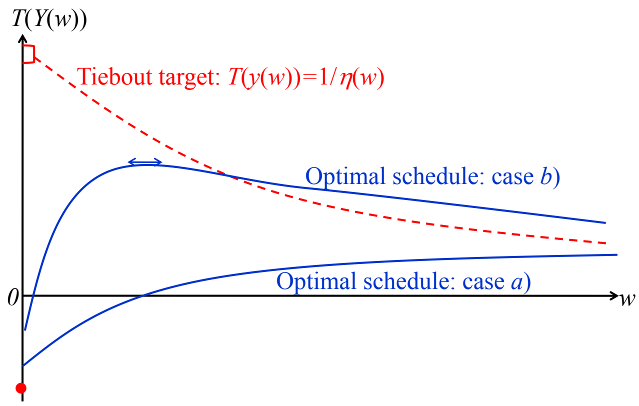

Figure 4: Increasing Semi-Elasticity of Migration

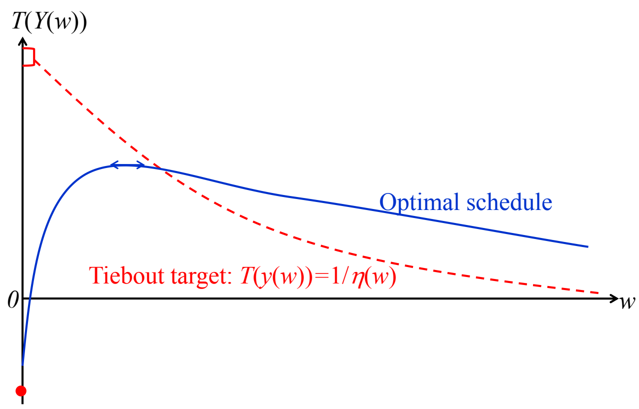

Figure 5: Increasing Semi-Elasticity of Migration

According to Propositions 3 and 4 the elasticity of migration is not a sufficient statistics to characterize the optimal tax schedule in an open economy. According to (16), 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, 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 at the top of the skill distribution. This point will be emphasized by the numerical simulations provided in the next sections.

One may wonder why it is the slope of the semi-elasticity of migration and not that of the elasticity that matters in Proposition 3. This is because the distorsions along the intensive margin depend on whether marginal tax rates are positive or negative, i.e. on 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 in order to clarify how the threat of migration affects 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, Puhani, and Sousa-Poza (2007), Kleven, Landais, and Saez (2010) or Kleven, Landais, and Saez (2010)). One should add in the empirical specification a term that interacts tax liabilities with income levels, so as to also estimate η'_0.

Brewer, Saez, and Shephard (2010) and Piketty and Saez (2012) look at the asymptotic marginal tax rate. They assume that the elasticity of migration is constant ($\nu_0(x) = \nu_0$). From Equation (16), a constant elasticity of migration is a special case of a decreasing semi-elasticity, because $C(w)$ must be non-decreasing in the second best. They also assume that the elasticities $\varepsilon(w)$, $\alpha(w)$ and $\nu(w)$ converge asymptotically to $\varepsilon(\infty)$, $\alpha(\infty)$ and $\nu(\infty)$ respectively. They finally assume that the distribution of skills is Pareto in its upper part, so that $k = (wf(w))/(\alpha(\infty)(1 - F(w)))$. Making skill w tends to infinity in the optimal tax formula (17), we retrieve their formula for the optimal asymptotic marginal tax rate:[13]

$$T'(Y(\infty)) = \frac{1}{1 + k\varepsilon + \nu_0}. \tag{21}$$

We see that the asymptotic marginal tax rate is then strictly positive. For example, if $k = 1.5$, $\varepsilon = 0.25$ and $\nu_\infty = 0.25$, 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, Sadka, and Shem-Tov (2012), the elasticity of migration tends to infinity according to (16). In this case, the asymptotic optimal marginal tax rate is equal to zero. The result of a zero asymptotic marginal tax obtained by Blumkin, Sadka, and Shem-Tov (2012) is thus a limiting case of Piketty and Saez (2012).

Finally, one may wonder whether the optimal tax schedule must converge asymptotically to the Tiebout target, as suggested in Figure 2 for the case of a constant elasticity of migration.[14] We can however provide counter-examples where this cannot be the case. For

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

[14]In this case, when the skill distribution is unbounded, Blumkin, Sadka, and Shem-Tov (2012) show that the tax liability converges to the Tiebout target (that they call the "Laffer tax") when the skill increases to infinity.

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 (19). The Tiebout target therefore converges to an asymptote that increases at a pace $1/(1+\nu_0)$, which is larger than the asymptotic optimal marginal tax rate. The two schedules must therefore diverge when the skill level tend to infinity.

V Simulations

This section provides numerical simulations of the equilibrium optimal tax schedule that competing policy-makers should implement. One of our objectives is to emphasize the part played by the slope of the semi-elasticity of migration. In particular, we will show that the marginal tax rates faced by rich individuals may be quite sensitive to the overall shape of the semi-elasticity.

We use the distribution of weakly earnings for singles without children in 2007 (CPS data) to recover the skill distribution $f(w)$, using the workers' first-order condition (3). We compute annual earnings Y and then proceed by inversion to find the value of w, assuming an approximation of the federal (Table 1) and local (Table 2) income tax in 2007 (See Appendix B). Following Diamond (1998) and Saez (2001), we correct for top coding by extending the obtained estimation with a Pareto distribution of coefficient 1.59. The disutility of effort is given by $v(y;w) = (y/w)^{1+1/\epsilon}$. This specification implies a constant elasticity of gross earnings with respect to the retention rate ϵ, as in Diamond (1998) and Saez (2001). In a recent survey, Saez, Slemrod, and Giertz (2012) conclude that "the best available estimates range from 0.12 to 0.4" in the United States. We use a central value, $\epsilon = 0.25$. Public expenditures E are kept at their initial level \$18, 157, which corresponds to 33.2% of the total gross earnings of single without children. Our calibration provides a very good approximation of the top of the income distribution as described by Alvaredo, Atkinson, Piketty, and Saez (2013). In the absence of migration responses, we find that the top 0.1%, top 1%, top 5% and top 10% of the population respectively get 6.5%, 18.2%, 34.7% and 45.4% of total income. The corresponding numbers in the World Top Income Database are 8.2%, 18.3%, 33.8% and 45.7%.

The semi-elasticity of migration is a key parameter in our computations. Even though the potential impact of income taxation on migration choices has been extensively discussed in the theoretical literature since Tiebout's (1956) seminal contribution, there are still few empirical results. Kleven, Landais, and Saez (2010) study 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, Landais, Saez, and Schultz

(2013) confirm that these results apply to the broader market of highly skilled foreign workers and not only to football players. They find an elasticity above 1 in Denmark. In a given country, the number of foreigners at the stop is relatively small. Hence, these findings would translate into a global elasticity at the top of at most 0.25 for most countries (see Piketty and Saez (2012)).

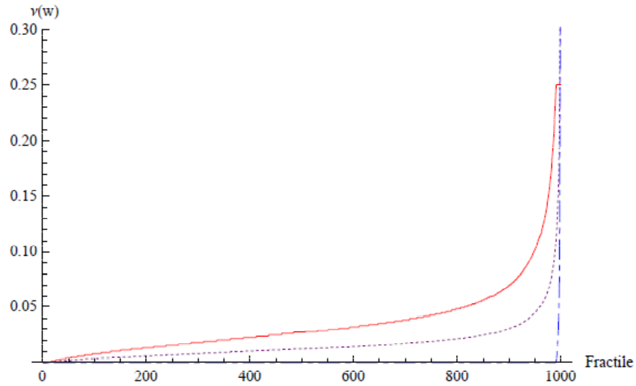

Figure 6: Elasticity of Migration by Fractile of the Actual Earnings Distribution. Case 1 (Red), Case 2 (Purple - dotted) and Case 3 (Blue - dashed)

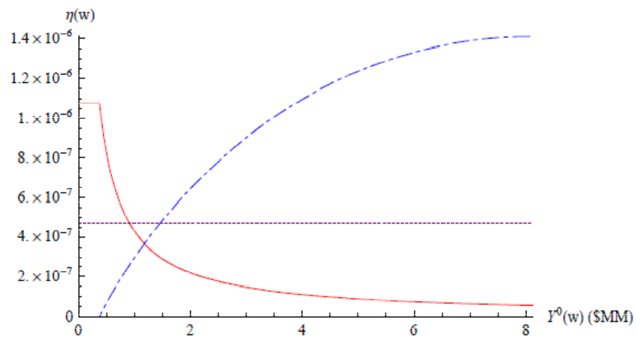

Figure 7: Semi-Elasticity of Migration as a Function of Actual Gross Earnings in Millions of US$. Case 1 (Red), Case 2 (Purple - dotted) and Case 3 (Blue - dashed)

As far as we know, there are no empirical studies regarding the possible shape of the elasticity or semi-elasticity of migration. We therefore investigate three possible scenarios. In each of them, the average elasticity in the actual economy top 1% of the population is equal to 0.25, as shown in Figure 6, where the population is divided into 1000 fractiles, based on individual earnings $Y^0(w)$ in the actual economy. The average elasticity in the population is much lower: 0.025 in the first one, 0.01 in the second one and 0.003 in the third one. 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. This is shown in Figure 7. In the second scenario, the semi-elasticity is constant throughout the whole skill distribution. In the third scenario, the semi-elasticity is zero up to the top centile and then increasing.

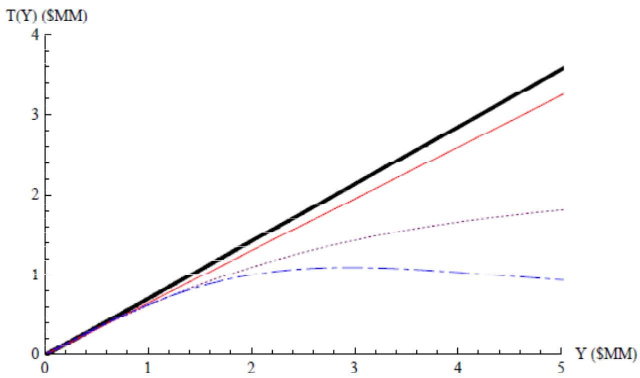

Figure 8: Optimal Tax Liabilities. Autarky (Bold), Case 1 (Red), Case 2 (Purple - dotted) and Case 3 (Blue - dashed).

The optimal equilibrium tax liabilities are shown in Figure 8. The x-axis represents gross earnings and the y-axis the total tax paid, both expressed in millions of US dollars. In addition to the three scenarios presented above, we added the tax liabilities that would be chosen in a closed economy or in the presence of tax coordination (cf. black curve). We observe that the threat of migration implies a non-negligible decrease in the total taxes paid by top income earners. Even though the average elasticity of migration is the same for the top 1% of income earners in the three cases, we observe significant differences due to variations in the shape of the semi-elasticity of migration. In the first case, the tax function is close to 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 $Y = \$3.2$ millions. The richest people are not those paying the largest taxes.

The effect of fiscal competition on tax progressivity is emphasized in Figure 9, which shows the average tax rate. The tax policy is progressive in case 1, but strongly regressive in the two other cases. The average tax rate for rich people ($5 millions of annual earnings) is about 65% in case 1, 39% in case 2 and 21% in case 3.

Figure 10 casts light on the differences in the optimal marginal tax rates. What we see is that differences in the slope of the semi-elasticity of migration may translate into large differences in marginal tax rates for high-income earners. Consequently, our 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.

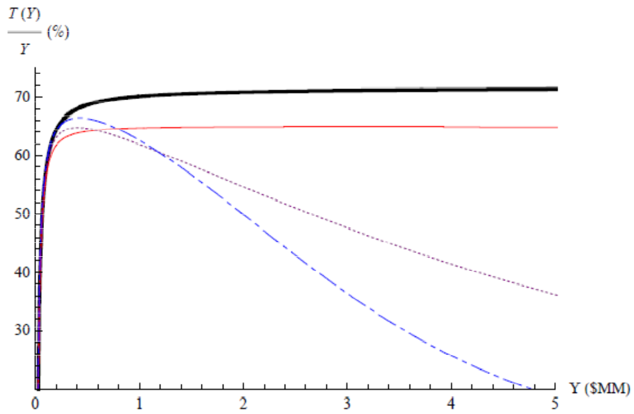

Figure 9: Optimal Average Tax Rates. Autarky (Bold), Case 1 (Red), Case 2 (Purple - dotted) and Case 3 (Blue - dashed).

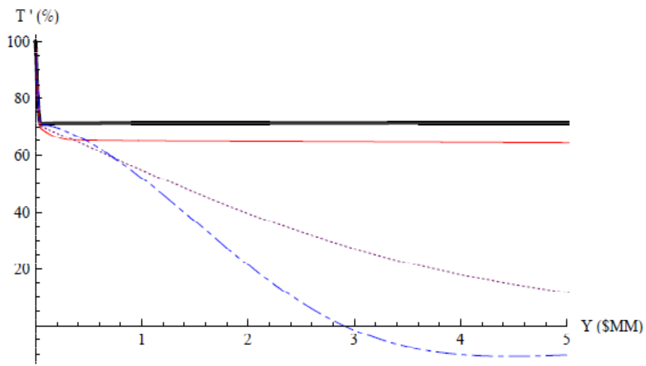

Figure 10: Optimal Marginal Tax Rates. Autarky (Bold), Case 1 (Red), Case 2 (Purple - dotted) and Case 3 (Blue - dashed).

VI Conclusion

This paper characterizes the nonlinear income tax schedules that competing Rawlsian governments should implement, when the countries are ex ante identical and individuals with private information on skills and migration costs 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 popluation, which has been estimated by Liebig, Puhani, and Sousa-Poza (2007), Kleven, Landais, and Saez (2010) and Kleven, Landais, Saez, and Schultz (2013), but also on the way in which 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. Given our calibration, optimal top marginal tax rates are positive and around 65% when the semi-elasticity is increasing, but are negative over $3,000,000 in the scenario with an increasing semi-elasticity of migration.

Therefore, it is not sufficient to only estimate the elasticity of migration. The slope of the semi-elasticity with respect to skills is also required. A first step in that direction would be to include an interaction between the disposable income and the pre-tax income in the empirical specification (2) of Kleven, Landais, Saez, and Schultz (2013). Thanks to this interaction, it would be possible to estimate how the elasticity of migration – and thereby the semi-elasticity – varies along the skill distribution. We argue that one cannot conclude about the optimal tax schedule in the presence of migration, even at the top, without further empirical evidence on the slope of the semi-elasticity.

Three different policies can be drawn from our results. From a *conservative* perspective, the uncertainty about the profile of the semi-elasticity of migration may justify very low, and maybe even negative, marginal tax rates for the top 1% of the income earners. This may partly explain why OECD countries have been reducing their top marginal tax rates over the past decades. From a more *nationalistic* perspective, the potential consequences of mobility might be so substantial in terms of redistribution that governments might want to hinder emigration, for example by the implementation of a departure tax as in Australie, Bengladesh, Canada, Netehrlands, South Africa[15], or to append tagging schemes to the generic income tax schedule so as to attract highly skilled foreigners at a low cost. Finally, from a *federalist* viewpoint, with which we sympathize the most, the problem is not globalization *per se* but

[15] cf. http://en.wikipedia.org/wiki/Departure_tax

the lack of cooperation between national tax authorities. In order to circumvent the negative impact of tax competition, it might make sense for policymakers to levy taxes on citizens living abroad, as implemented by the United States. Let us assume that policymakers levy taxes on their citizens living abroad abroad. For simplicity, let us assume that expatriates are not taxed abroad. If individuals do not move abroad so as to get higher wages, there will be no migration and the closed-economy tax schedules can be implemented. Therefore a tax system based on citizenship instead of residence may be preferable when taxpaywers can vote with their feet. Some additional residence-based taxes may be added to finance local public goods for example. The stability of such a tax system is however conditioned on the policymakers' willingness to exchange information. There has been some advances in this direction. 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 non-binding and very incomplete.

An limitation of our model is its inability to disentangle these three points of views. A reason is the assumption that the skill distribution is the same whether or not migration is allowed. In doing so, we do not allow for gains that may arise due to globalization and increasing returns to scale (Krugman (1991)). Another limitation is that we focus on competition between symmetric countries. The actual lack of fiscal coordination and the competition between Nation States are certainly inseparable from the asymmetries in size, production technology or preferences for redistribution (see Alesina and Spolaore (1997) and Bolton and Roland (1997)). The cost of cooperation is then substantially larger, as emphasized by the fiscal competition literature (see e.g. the recent survey of Keen and Konrad (2013)). Introducing such asymmetries within our framework is an extension that belongs to our research agenda.

A Proofs

I.A Proposition 1

We adopt a first-order approach by assuming that the monotonicity constraint is slack. We further assume that $Y(.)$ is differentiable. Denoting $q(.)$ the co-state variable, the Hamiltonian associated to Problem (11) is:

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

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

$$1 - v'_y(Y_i(w); w) = \frac{q(w)}{\varphi(\Delta_i(w); w)} \; v''_{yw}(Y_i(w); w)\,, \tag{22a}$$

$$q'(w) = \{1 - [Y_i(w) - v(Y_i(w); w) - U_i(w)] \;\; \eta_i(\Delta_i(w); w)\} \;\; \varphi(\Delta_i(w); w)\,, \tag{22b}$$

$$q(w_1) = 0 \text{ when } w_1 < \infty \text{ and } q(w_1) \to 0 \text{ when } w_1 \to \infty, \tag{22c}$$

$$q(w_0) \leq 0. \tag{22d}$$

Integrating Equation (22b) between w and w_1 at the symmetric equilibrium and using the transversality condition (22c), we obtain:

$$q(w) = -\int_{w}^{w_1} [1 - \eta_0(x)\ T(Y(x))]\ f(x)\ dx. \tag{23}$$

Defining $X(w) = -q(w)$ leads to (13). Equation (22a) can be rewritten at the symmetric equilibrium as:

$$1 - v'_y\,(Y(w); w) = -\frac{X\,(w)}{f(w)}\ v''_{yw}\,(Y(w); w) \tag{24}$$

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

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

Plugging (3) and the latter equation into (24) leads to (12).

I.B Proposition 3

From (12), $T'(Y(w))$ has the same sign as the tax level effect $X(w)$. The transversality condition (22d) is equivalent to $X(w_0) \geq 0$, while (22c) is equivalent to $\lim_{w \mapsto w_1} X(w) = 0$. From (13), the derivative of $X(w)$ is

$$X'(w) = \left[T(Y(w)) - \frac{1}{\eta_0(w)}\right]\ \eta_0(w)\ f(w) \tag{25}$$

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

i) $\eta_0(w)$ is constant and equal to η_0

We successively show that any configuration but $T(Y(w)) < 1/\eta_0$ for all $w \in (w_0, w_1)$ contradicts at least one transversality conditions (22c) or (22d). We start by establishing the following Lemmas.

Lemma 1. *Assume that for any $w \in [w_0, w_1]$, $\eta'_0(w) \leq 0$ and assume there exists a skill level $\hat{w} \in (w_0, w_1)$ such that $T'(Y(\hat{w})) \leq 0$ and $T(Y(\hat{w})) > 1/\eta_0(\hat{w})$. Then $X(w_0) < 0$, so the transversality condition (22d) 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 $T(Y(w)) > 1/\eta_0(w)$. Let $w^* \in [w_0, \tilde{w})$ be the lowest bound of this interval. Then either $w^* = w_0$ or $T(Y(w^*)) = 1/\eta_0(w^*)$. Moreover, for all $w \in (w^*, \hat{w}]$, one has that $T(Y(w)) > 1/\eta_0(w)$, thereby $X'(w) > 0$, according to (25). Hence, one has that $X(w) < X(\hat{w}) \leq 0$, thereby $T'\,(Y(w)) < 0$ for all $w \in [w^*, \hat{w})$. Consequently, $T(Y(w^*)) > T(Y(\hat{w})) \geq 1/\eta_0(\hat{w}) \geq 1/\eta_0(w^*)$. So, one must have $w^* = w_0$. Finally, we get $X(w^*) = X(w_0) < 0$, which contradicts the transversality condition (22d). $\square$

Lemma 2. *Assume that for any $w \in [w_0, w_1]$, $\eta'_0(w) \leq 0$ and assume there exists a skill level $\hat{w} \in (w_0, w_1)$ such that $T'(Y(\hat{w})) \leq 0$ and $T(Y(\hat{w})) < 1/\eta_0(\hat{w})$. Then $X(w_1) < 1$, so the transversality condition (22d) 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 $T(Y(w)) < 1/\eta_0(w)$. Let $w^* \in (\hat{w}, w_1]$ be the highest bound of this interval. Then either $w^* = w_1$ or $T(Y(w^*)) = 1/\eta_0(w^*)$. Moreover, for all $w \in [\hat{w}, w^*)$, one has that $T(Y(w)) < 1/\eta_0(w)$, thereby $X'(w) < 0$, according to (25). Hence, one has that $X(w) < X(\hat{w}) \leq 0$, thereby $T'\,(Y(w)) < 0$ for all $w \in (\hat{w}, w^*]$. Consequently, $T(Y(w^*)) < T(Y(\hat{w})) \leq 1/\eta_0(\hat{w}) \leq 1/\eta_0(w^*)$. So, one must have $w^* = w_1$. Finally, we get $X(w^*) = X(w_1) < 0$, which contradicts the transversality condition (22c). $\square$

From Lemmas 1 and 2, it is not possible to have $T'(Y(\hat{w})) \leq 0$ and $T(Y(\hat{w})) \neq 1/\eta(\hat{w})$, 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 $T(Y(\hat{w})) = 1/\eta_0(\hat{w})$. By continuity there exists $\varepsilon > 0$ such that $T'(Y(\hat{w} - \varepsilon)) < 0$ and $T(Y(\hat{w} - \varepsilon)) > 1/\eta_0$, 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 $T(Y(\hat{w})) = 1/\eta_0(\hat{w})$. According to the Cauchy-Lipschitz theorem (equivalently, the Picard–Lindelöf theorem), the differential system of equations in $U(w)$ and $X(w)$ defined by (9) and (25) (and including (12) to express $Y(w)$ as a function of $X(w)$) with initial condition that corresponds to $T'(Y(\hat{w})) = X(\hat{w}) = 0$ and $T(Y(\hat{w})) = 1/\eta_0(\hat{w})$ admits a single solution where for all w $X(w) \equiv 0$ and $T(.) = 1/\eta_0$. From (8), such a solution provides excess budget resources when E is assumed nil and provides less utility level than the *laissez faire* policy where $T(.) = 0$.

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, which ends the proof of Part *i*) of Proposition 3.

***ii*) $\eta_0(w)$ is decreasing**

If there exists a skill level $\hat{w} \in (w_0, w_1)$ such that $T'(Y(\hat{w})) \leq 0$ and $T(Y(\hat{w})) > 1/\eta_0(\hat{w})$, 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 $T(Y(\hat{w})) < 1/\eta_0(\hat{w})$, 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 $T(Y(\hat{w})) = 1/\eta_0(\hat{w})$, then function $w \mapsto T(Y(w)) - 1/\eta_0(w)$ is non-positive and admits a negative derivative at $\hat{w}$, as $\eta_0'(.) < 0$. Hence, there exists $\overline{w} > \tilde{w}$ such that $T(Y(w)) < 1/\eta_0(w)$, thereby $X'(w) < 0$ for all $w \in (\tilde{w}, \overline{w}]$. Consequently, $X'(\overline{w}) < 0$ (equivalently $T(Y(\overline{w})) < 1/\eta_0(\overline{w})$) 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*) $\eta_0(w)$ is increasing**

We first show the following Lemma.

Lemma 3. *Assume that for any $w \in [w_0, w_1]$, $\eta_0'(w) > 0$ and assume there exists a skill level $\hat{w} \in (w_0, w_1)$ such that $T'(Y(\hat{w})) \geq 0$ and $T(Y(\hat{w})) \geq 1/\eta_0(\hat{w})$. Then, $X(w_1) > 0$, so the transversality condition* (22c) *is violated.*

Proof We first show that we can assume that $T(Y(\hat{w})) > 1/\eta_0(\hat{w})$ without any loss of generality. Assume that $T(Y(\hat{w})) = 1/\eta_0(\hat{w})$ and $T'(Y(\hat{w})) \geq 0$. As $\eta_0'(.) > 0$, function $w \mapsto T(Y(w)) - 1/\eta_0(w)$ is non-negative and admits a positive derivative at $\hat{w}$. Hence, there exists $\overline{w} > \tilde{w}$ such that $T(Y(w)) > 1/\eta_0(w)$, thereby $X'(w) > 0$ for all $w \in (\tilde{w}, \overline{w}]$. Consequently, $X'(\overline{w}) > 0$ (equivalently $T(Y(\overline{w})) > 1/\eta_0(\overline{w})$) and $X(\overline{w}) > X(\tilde{w}) = 0$ (equivalently $T'(Y(\overline{w})) > 0$.

Consider now that $T'(Y(\hat{w})) \geq 0$ and $T(Y(\hat{w})) > 1/\eta_0(\hat{w})$. As $T(Y(w)) = Y(w) - C(w)$ and $\eta(w)$ are continuous functions of w, there exists by continuity an open set around $\hat{w}$ where $T(Y(w)) > 1/\eta_0(w)$. Let $w^* \in (\hat{w}, w_1]$ be the highest bound of this interval. Then either $w^* = w_1$ or $T(Y(w^*)) = 1/\eta_0(w^*)$. Moreover, for all $w \in [\hat{w}, w^*)$, one has that $T(Y(w)) > 1/\eta_0(w)$, thereby $X'(w) > 0$, according to (25). Hence, one has that $X(w) > X(\hat{w}) \geq 0$, thereby $T'(Y(w)) > 0$ for all $w \in (\hat{w}, w^*]$. Consequently, $T(Y(w^*)) > T(Y(\hat{w})) \geq 1/\eta_0(\hat{w}) > 1/\eta_0(w^*)$. So, one must have $w^* = w_1$. Finally, we get $X(w^*) = X(w_1) > 0$, which contradicts the transversality condition (22c). $\square$

Assume by contradiction that $X(w_0) = 0$ and there exists a skill level $\hat{w} \in (w_0, w_1)$ such that $T'(Y(\hat{w})) \geq 0$ and $T(Y(\hat{w})) < 1/\eta_0(\hat{w})$. 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 $T(Y(w)) < 1/\eta_0(w)$. Let $w^* \in [w_0, \tilde{w})$ be the lowest bound of this interval. Then either $w^* = w_0$ or $T(Y(w^*)) = 1/\eta_0(w^*)$. Moreover, for all $w \in (w^*, \hat{w}]$, one has that $T(Y(w)) < 1/\eta_0(w)$, thereby $X'(w) < 0$, according to (25). Hence, one has that $X(w) > X(\hat{w}) \geq 0$, thereby $T'(Y(w)) > 0$ for all $w \in [w^*, \hat{w})$. Consequently, $T(Y(w^*)) < T(Y(\hat{w})) < 1/\eta_0(\hat{w}) < 1/\eta_0(w^*)$. So, one must have $w^* = w_0$. Finally, we get $X(w^*) = X(w_0) > 0$, which contradicts the presumption that $X(w_0)$.

Consequently, if $X(w_0) = 0$, we must have $T'(Y(w)) < 0$ for all $w \in (w_0, w_1)$. Using (8) and the assumption that $E = 0$, this implies that $T(Y(w_0)) > 0 > T(Y(w_1))$. Hence such policy provides less utility to workers of skill w_0 than the laissez faire policy $T(.) = 0$, which contradicts the presumption that $X(w_0) = 0$.

We consider hereafter the case where $X(w_0) > 0$. There thus exists $\breve{w} \in (w_0, w_1]$ such that (equivalently $(X(w) \geq 0$ $T'(Y(w)) \geq 0)$ for $w \leq \breve{w}$ and either $\breve{w} = w_1$ or there exists $w_2 \in (\breve{w}, w_1]$ such that $X(w) < 0$ (equivalently $T'(Y(w) < 0)$ for all $w \in (\breve{w}, w_2)$ and $X(w) > 0$ in the neighborhood to the right of w_2.

If $\breve{w} < w_1$, either we have $w_2 = w_1$ or we must have $T'(Y(w_2)) = 0$ by continuity of $T'(.)$. Moreover, according to Lemma 3, one must have $T(Y(w)) < 1/\eta_0(w)$ for all $w \in (\breve{w}, w_2)$, otherwise the transversality condition $X(w_1) = 0$ would be violated. Consequently, one has that $X'(w) < 0$ for all $w \in (\breve{w}, w_2)$, so one has that $0 = X(\breve{w}) < X(w_2)$. Hence we have that $T'(Y(w_2)) < 0$, implying that $w_2 = w_1$, which ends the proof of Part *iii)* of Proposition 3

I.C Proposition 4

From Proposition 3, either marginal tax rates are positive, or there exists a threshold above which marginal tax rate is negative. Assume by contradiction that marginal tax rates are positive. Then, the tax schedule is increasing. It must also be positive to clear the budget constraint. As the semi-elasticity of migration increases to infinity, there thus exists a skill level $\hat{w}$ at which $T'(Y(\hat{w})) \geq 0$ and $T(Y(\hat{w})) > 1/\eta_0(\hat{w})$. Then the transversality condition (22c) is violated according to Lemma 3, which leads to the desired contradiction. So, marginal tax rate must be negative above some skill level.

B Numerical Simulations

The simulation program consists in solving the differential system of Equations (9) and (25) in $U(w)$ and $X(w)$, using $T(Y(w)) = Y(w) - v(Y(w); w) - U(w)$ and (24), with the terminal condition $X(w_1) = 0$. The remaining terminal condition $U(w_1)$ is selected to clear the budget constraint (8). The algorithm actually solves this system by the Newton-Raphson method[16] using a discrete grid over $[w_0, w_1]$.

We calibrate the skill distribution using the distribution of weekly earnings among singles without dependent extracted from CPS 2008. We multiply this weakly earnings by 52 to get annual earnings. Given the specified utility function $c - (y/w)^{1+(1/\varepsilon)}$, we recover the skill level for each earnings observation from (3), using an approximation of the federal income tax schedule for singles described in Table 1 and an approximation of the local income tax in California described in Table 2.

\$0	\$7,550	\$30,650	\$74,200	\$154,800	\$336,500
10%	15%	25%	28%	33%	35%

Table 1: Approximation of the Federal Income tax

\$0	\$13,251	\$31,397	\$40,473	\$50,090	\$59,166
1%	2%	4%	6%	8%	9.3%

Table 2: Approximation of the local Income tax (California)

We use a Gaussian Kernel of bandwidth \$1,157.2. We expand this estimated density by a Pareto density of the form $k\ w^{-(p+1)}$. The skill level at which the expansion occurs and the scale parameter p are selected to insure that the density $f(.)$ remains continuously differentiable. The truncation at w_1 implies that the ratio $(1 - F(w)) / (w\ f(w))$ is not constant at $1/p$ and instead tends to zero at w_1, despite the Pareto expansion. This the reason why we add at the highest point w_1 of the grid of skills a mass point whose weight is such that $(1 - F(w)) / (w\ f(w))$ is constant at $1/p$ in the upper part of the skill distribution.

[16] i.e. it approximates the solution of $f'(w) = \Phi(f(w), w)$ by $f(w_{i-1}) \simeq f(w_i) + (w_{i-1} - w_i)\ \Phi(w_i, f(w_i))$.

This lead us with an approximation of the current economy. Parameter p is selected to get plausible values for the shares of total income earned by the top 1% of the population.

In each scenario, we calibrate the semi elasticity of migration $g(0|w)$ such that the average of the elasticity of migration $(Y(w) - T^{\text{actual}}(Y(w)))g(0|w)$ in this approximation of the current economy is 0.25 for the top 1% of the income distribution. In the scenario with a constant elasticity of migration, this is done such that for each skill level in the top 1%, the elasticity of migration is equal to 0.25. In the scenario with an increasing semi-elasticity of migration, we choose a quadratic - concave specification for the function $w \mapsto g(0|w)$ such that the semi elasticity of migration is nil at the 99% percentile.

References

Alesina, A., and E. Spolaore (1997): "On the Number and Size of Nations," *The Quarterly Journal of Economics*, 112(4), 1027–56.

Alvaredo, F., A. B. Atkinson, T. Piketty, and E. Saez (2013): "The World Top Incomes Database, http://topincomes.g-mond.parisschoolofeconomics.eu/," .

Atkinson, A. (1990): "Public Economics and the Economic Public," *European Economic Review*, 34(2-3), 225–248.

Atkinson, A., and J. Stiglitz (1980): *Lectures on Public Economics*. McGraw-Hill Inc.

Bierbrauer, F., C. Brett, and J. A. Weymark (2011): "Strategic Nonlinear Income Tax Competition with Perfect Labor Mobility," Discussion Paper 3329, CESifo WP.

Blackorby, C., W. Bossert, and D. Donaldson (2005): *Population Issues In Social Choice Theory, Welfare Economics and Ethics*. Cambridge University Press.

Blumkin, T., E. Sadka, and Y. Shem-Tov (2012): "International Tax Competition: Zero Tax Rate at the Top Re-established," Discussion Paper 3820, CESifo WP.

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

Bolton, P., and G. Roland (1997): "The Breakup of Nations: A Political Economy Analysis," *The Quarterly Journal of Economics*, 112(4), 1057–90.

Borjas, G. (1999): "The Economic Analysis of Immigration," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, chap. 28, pp. 1697–1760. North-Holland.

Brewer, M., E. Saez, and A. Shephard (2010): "Means-testing and Tax Rates on Earnings," in *Dimensions of Tax Design: the Mirrlees Review*, ed. by J. Mirrlees, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles, and J. Poterba, chap. 3, pp. 90–173. Oxford University Press.

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

Docquier, F., and A. Marfouk (2006): "International migration by educational attainment (1990-2000) - Release 1.1," in *International Migration, Remittances and Development*, ed. by C. Ozden, and M. Schiff, chap. 5. Palgrave Macmillan, New-York.

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

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

Huber, B. (1999): "Tax Competition and Tax Coordination in an Optimal Income Tax Model," *Journal of Public Economics*, 71, 441–458.

Jacquet, L., E. Lehmann, and B. Van der Linden (2013): "Optimal Redistributive Taxation with both Extensive and Intensive Responses," *Journal of Economic Theory*, Forthcoming.

Keen, M., and K. A. Konrad (2013): "The Theory of International Tax Competition and Coordination," Discussion Paper 2012–06, Working Paper of the Max Planck Institute for Tax Law and Public Finance, Forthcoming *Handbook of Public Economics*, vol. 5.

Kleven, H., C. Landais, and E. Saez (2010): "Taxation and International Migration of Superstars: Evidence from the European Football Market," NBER Working Papers 16545, NBER, Forthcoming in American Economic Review.

Kleven, H. J., C. Landais, E. Saez, and E. A. Schultz (2013): "Migration and Wage Effects of Taxing Top Earners: Evidence from the Foreigners' Tax Scheme in Denmark," NBER Working Papers 18885, National Bureau of Economic Research, Inc.

Krugman, P. (1991): "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99(3), 483–99.

Leite-Monteiro, M. (1997): "Redistributive Policy with Labour Mobility Accross Countries," *Journal of Public Economics*, 65, 229–244.

Liebig, T., P. Puhani, and A. Sousa-Poza (2007): "Taxation And Internal Migration-Evidence From The Swiss Census Using Community-Level Variation In Income Tax Rates," *Journal of Regional Science*, 47(4), 807–836.

Lipatov, V., and A. Weichenrieder (2012): "Optimal income taxation with tax competition," (1207).

Maggi, G., and A. Rodriguez-Clare (1995): "On Countervailing Incentives," *Journal of Economic Theory*, 66, 238–263.

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

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

Morelli, M., H. Yang, and L. Ye (2012): "Competitive Nonlinear Taxation and Constitutional Choice," *American Economic Journal: Microeconomics*, 4(1), 142–75.

Osmundsen, P. (1999): "Taxing Internationally Mobile Individuals– A Case of Countervailing Incentives," *International Tax and Public Finance*, 6, 149–164.

Piaser, G. (2007): "Labor Mobility and Income Tax Competition," in *International Taxation Handbook*, ed. by G. N. Gregoriou, and R. Colin, chap. 4, pp. 73–94. Elsevier.

Piketty, T. (1997): "La redistribution fiscale face au chômage," *Revue Française d'Economie*, 12, 157–201.

Piketty, T., and E. Saez (2012): "Optimal Labor Income Taxation," Discussion Paper WP No, 18521, NBER, Forthcoming *Handbook of Public Economics*, vol. 5.

Rochet, J.-C., and A. Stole, L (2002): "Nonlinear Pricing with Random Participation," *The Review of Economic Studies*, 69(1), pp. 277–311.

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

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

Saez, E., J. Slemrod, and S. H. Giertz (2012): “The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review,” *Journal of Economic Literature*, 50(1), 3–50.

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

Simula, L., and A. Trannoy (2011): “Shall we keep the highly skilled at home? The optimal income tax perspective,” *Social Choice and Welfare*, pp. 1–32.

Sjaastad, L. (1962): “The Costs and Returns of Human Migration,” *Journal of Political Economy*, 70, 80–93.

Stiglitz, J. (1982): “Self-Selection and Pareto Efficient Taxation,” *Journal of Public Economics*, 17(2), 213–240.

Tiebout, C. (1956): “A Pure Theory of Local Expenditures,” *Journal of Political Economy*, 64, 416–424.

Wilson, J. D. (1982): “Optimal income taxation and migration: A world welfare point of view,” *Journal of Public Economics*, 18(3), 381 – 397.