Optimal Non-Linear Income Tax When Agents Vote With Their Feet

Laurent Simula and Alain Trannoy*
EHESS, IDEP and GREQAM†

June 26, 2006

Abstract

Individuals living in a Mirrleesian economy have outside options consisting in moving to a laissez-faire country while paying migration costs. Three social points of view are distinguished according to the agents whose welfare is to count and type-dependent participation constraints taken into account. The first-best allocations are characterized by a curse of the middle-skilled workers. In the second-best, we extend Mirrlees's and Diamond's formulae when emigration of the nationals is prevented and implement the latter numerically using French data. We then examine for which individuals the participation constraints should be introduced and show that it depends on countervailing incentives.

Keywords: Optimal Income Taxation, Type-Dependent Participation Constraints, Individual Mobility

JEL classification: H21; H31; D82; F22

1 Introduction

In Mirrlees's (1971) seminal article, "migration is supposed to be impossible". However, "since the threat of migration is a major influence on the degree of progression in actual tax systems, at any rate outside the United States, this is [an] assumption one would rather not make". The impact of this threat is certainly even more topical after 25 years of increasing labour mobility, as testified by the political concern, in some EU countries but also in Canada, about emigration or potential emigration of highly skilled individuals because of tax purposes. There are for

*We are grateful to Thomas Aronsson, David Bevan, Philippe Choné, David de la Croix, Jeremy Edwards, Jean Hindriks, Mathias Hungerbühler, Laurence Jacquet, Jean-Marie Lozachmeur, Hamish Low, François Maniquet, James Mirrlees, Gareth Myles, Frank Page, Pierre Pestieau, Panu Poutvaara, Emanuela Sciubba, John Weymark, David Wildasin for their suggestions. Our work has benefited from comments of seminar participants at GREQAM, the University of Cambridge, PET 2005, ESWC 2005, ASSET 2005, HECER Workshop on Fiscal Federalism, Séminaire Economique de Louvain, Doctorales ADRES 2006, Fifth Journées Louis-André Gérard-Varet. The usual caveats apply.

†Ecole des Hautes Etudes en Sciences Sociales, Research Group in Quantitative Economics of Aix-Marseilles and Institute for Public Economics, Centre de la Vieille Charité, 2 rue de la Charité, 13236 Marseille cedex 02, France. Emails: Laurent.Simula@ehess.fr and Alain.Trannoy@ehess.univ-mrs.fr

instance about 34 000 income taxpayers who leave France each year since 2000. These individuals paid three times more taxes than the average taxpayer and seven out of ten of them chose to relocate to another EU country or to North America, mostly to countries where income taxes are lower (DGI, 2005). If emigration for tax purposes remains difficult to quantify precisely, these figures suggest that international differences in income taxes should be regarded as *one* of the determinants of the migration decision. Such a determinant agrees with John 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).

This paper studies to which extent the mobility of highly skilled individuals for tax purposes alters the optimal non-linear income tax a Mirrleesian economy should implement. This mobility is a specific phenomenon, which induces both losses in taxes and in productive capacity in the left country. First, contrary to the brain drain, the key parameter is not the cross-country differences in productivity. Second, the set of instruments at the government' disposal is more limited than when it faces tax evasion (e.g. Chander and Wilde (1998)). It has indeed few alternatives but to lower taxes, that is to reduce redistribution, to prevent the departure of its highly skilled individuals: it thus has "carrots" but no "sticks". That results in a specific conflict between its desire to maintain the national income per capita in keeping taxes down and its desire to sustain the redistribution programme. Third and consequently, this mobility of individuals appears as a new constraint on the design of the optimal income tax scheme.

To address this issue, we characterize the optimal non-linear income tax in a Mirrleesian economy A whose individuals have type-dependent outside options consisting in emigrating to a *less* redistributive country B. A's government wants to redistribute incomes from the more to the less productive individuals as in Mirrlees's (1971) article, but has also to take individual rationality into account via the introduction of participation constraints for the individuals it wants to keep in A.

Most of the literature on optimal income taxation has focused on closed economies. The papers taking individual mobility into account have used models with no leisure-consumption choice (Mirrlees, 1982, Hindriks, 1999, Osmundsen, 1999), considered a world with two classes of individuals and lump-sum taxes (Leite-Monteiro, 1997), focused on linear taxes (Wilson, 1980, 1982, Simula and Trannoy, 2006a), or adopted Stiglitz's (1982) self-selection approach (Huber, 1999, Hamilton and Pestieau, 2005, Piaser, 2003). Among them, Leite-Monteiro (1997), Hindriks (1999), Huber (1999) and Piaser (2003) have adopted the point of view of tax competition. Hamilton and Pestieau (2005) have concentrated on migration equilibria.

This paper considers optimal non-linear income taxation when there is a continuum of individuals differing in productivity as well as migration costs and facing consumption-leisure choices in the absence of unemployment. To keep the model as simple and pure as possible, one considers that B is a laissez-faire country and that the individuals are initially living in A. B may thus be viewed as the limit case of a less redistributive country whose income tax policy is given. There is therefore no strategic competition on taxes between the countries. The problem would become much more difficult if B's tax policy were endogenous; for instance the usual revelation principle in A and B would generally vanish (Page and Monteiro, 2003). The focus on one-way migrations is justified by our interest in the threat of migration and should be regarded as a way of modelling an economy whose migration balance is in deficit.

The social objective is more complex to specify when the individuals can vote with their feet. It does not only depend on the government's aversion to income inequality, as in a closed economy, but also on the set of agents whose welfare is to count. We distinguish three different social points of view. From the Mercantilist point of view, A's government is interested in the welfare of its nationals whilst ensuring that every national lives in A. From the National point of view, A's government takes the welfare of its nationals into account, irrespective of their country of residence. From the Resident point of view, A's government cares about the welfare of its residents.

We introduce type-dependent participation constraints in the optimal non-linear income tax problem and examine for which productivity levels they should bind to obtain the highest social welfare per capita from the three different social points of view. Participation constraints appear as a natural way of modelling the possibility that the individuals vote with their feet. They express the fact that an individual will leave the country where he is living if the net utility he obtains there is less than his reservation utility, equal to the maximum utility he can obtain abroad. Participation constraints are however not used in the previous models of optimal income taxation dealing with individual mobility. This is rather surprising if one considers that optimal income taxation is a principal-agent model with "false moral hazard" (Laffont and Martimort, 2002). The specific issues raised by such constraints have been studied in models based on contract theory (see Lewis and Sappington (1989), Maggi and Rodriguez-Clare (1995), Jullien (2000)). We have shown in Simula and Trannoy (2006a) that linear taxes lack degrees of freedom when many constraints have to be taken into account. This rigidity of linear taxes, in which only two instruments are at the governments' disposal, explains why we choose to focus on non-linear taxes herein.

The basic assumptions we make are the following. First, the individual productivity, equal to the gross wage, does not change with the country of residence. In other words, both countries have the same production function, which distinguishes our framework from the literature on the brain drain. Second, migration costs depend on productivity which is thus the sole parameter of heterogeneity within the population. Third, migration costs are monotonic and do not increase faster than the laissez-faire utility. If individuals where initially living in B, an additional assumption on the level of the migration costs should be introduced to prevent emigration of low-skilled workers from B to A, in which we are not interested herein.

When each individual's productivity is public information (first-best), we show that it is always socially optimal to prevent emigration of the highly-skilled individuals from both National and Resident points of view, which coincide therefore with the Mercantilist point of view at the optimum. We establish that a curse of the middle-skilled workers replaces the curse of the high-skilled individuals obtained in closed economy (Mirrlees, 1974). Indeed, when leisure is a normal good, the first-best optimum allocation is such that it is no longer possible to demand as much work as without mobility from the most talented individuals. It is thus from the middle-skilled workers, insufficiently talented to threaten to emigrate, that the productive rent is extracted to the maximum. However, these middle-skilled workers cannot be taxed at will because they would otherwise choose to emigrate. Consequently, the redistribution in favour of the low-skilled individuals is reduced.

When each individual's productivity is private information (second-best), the optimal income tax problem takes both incentive-compatibility and participation constraints into account. We first study the properties of the optimum tax schedule from the Mercantilist point of view. We extend Mirrlees's (1971) and Diamond's (1998) optimal income tax formulae to the case where agents vote with their feet. Using the same tax reform perturbation around the optimal income tax scheme as in Piketty (1997) and Saez (2001), we show that a marginal increase in taxes has an additional negative effect on social welfare in the presence of potential mobility. Because of this effect, the optimal marginal tax rates have to be adjusted very carefully even at productivity levels where the individuals do not threaten to emigrate. In addition, this new effect favours a decrease in the optimal average tax rates, which appear as important variables even in the absence of income effects on labour supply. We examine the National and Resident points of view to consider whether it is always socially optimal to prevent emigration of the highly skilled workers and address the issue of the optimal size of A's resident population. Because the type-dependent participation constraints interact with the incentive compatibility conditions, the presence in A of highly-skilled individuals with appealing outside options may give rise to an indirect social cost through countervailing incentives. We derive a sufficient condition under which implementing a tax schedule which induces their emigration increases social welfare.

Section 2 sets up the model. Section 3 examines the first-best optimal allocations. Section 4 studies the properties of the second-best optimal allocations. Section 5 provides numerical simulations on French data. Section 6 concludes. Proofs are relegated to the appendices.

2 The Model

The world consists of two countries, A and B. All individuals are initially living in A. A's government implements a redistributive tax policy and B is committed to being a laissez-faire country. Both countries have the same production function with constant returns to scale. Hence, productivity levels, and therefore wages in the absence of taxation, are independent of the country in which the individuals are working.

2.1 Population

Individuals differ in productivity θ. For simplicity, an individual with productivity θ is called a θ-individual. The distribution function of θ, denoted F, is defined on a closed interval $\left[\underline{\theta}, \overline{\theta}\right] \subseteq \mathbb{R}^+$. Its derivative f is continuous and strictly positive on $\left[\underline{\theta}, \overline{\theta}\right]$. This distribution is common knowledge.

2.2 Individual Behaviour

All individuals have the same preferences over consumption x and labour l. If $\bar{l}$ is the time endowment, these preferences are represented by a utility function $U : \mathscr{X} \to \mathbb{R}$, where $\mathscr{X} := \left\{(x,l) \in \mathbb{R}^+ \times [0,\bar{l})\right\}$.

Assumption 1. *U is a $\mathscr{C}^2$ strictly concave function s.t. $U_x > 0$, $U_l < 0$ and $U \to -\infty$ as $x \overset{>}{\to} 0$ or $l \overset{<}{\to} \bar{l}$.*

Assumption 2. *Leisure is a normal good.*

Any θ-individual who works l units of time has gross income $z := \theta l$. We call

$$u(x,z;\theta) := U(x,z/\theta) \tag{1}$$

the personalized utility function and note that $u'_x = U'_x$, $u'_z = U'_l/\theta$, $u''_{xx} = U''_{xx}$, $u''_{xz} = U''_{xl}/\theta$, $u''_{zz} = U''_{ll}/\theta^2$. The marginal rate of substitution of gross income for consumption is given by

$$s(x,z;\theta) := -\frac{u'_z(x,z;\theta)}{u'_x(x,z;\theta)}. \tag{2}$$

Each individual decides about the optimal amount of consumption and labour so as to maximize his utility subject to his budget constraint. Using a tax function T, A's government can arrange that a θ-individual supplying l units of labour in A has disposable income ξ_A. We assume that ξ_A is an upper semi-continuous function so that every θ-individual attains its supremum in $\mathscr{X}$. Consequently, the utility maximization programme in A, $\max_{(x,l)\in\mathscr{X}} \{U(x,l) \text{ s.t. } x = \xi_A\}$, defines implicitly the consumption and labour supply functions in A, $x_A(\theta)$ and $l_A(\theta)$ respectively. The indirect utility in A is

$$V_A(\theta) := U(x_A(\theta), l_A(\theta)). \tag{3}$$

The utility maximization programme in B, $\max_{(x,l)\in\mathscr{X}} \{U(x,l) \text{ s.t. } x = z\}$, defines implicitly the consumption and labour supply functions in B, $x_B(\theta)$ and $l_B(\theta)$ respectively. We call e^H and e^M the Hicksian and Marshallian elasticities of labour supply with respect to the net-of-tax wage rate. The indirect utility in B is

$$V_B(\theta) := U(x_B(\theta), l_B(\theta)), \tag{4}$$

which is strictly increasing in θ under Assumption 1.

2.3 Emigration and Participation Constraints

An individual who leaves A pays a strictly positive *migration cost c*. This cost is introduced in the model as a "time-equivalent" loss in utility, which corresponds to different 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 his family and friends, etc. "[Migration] costs probably vary among persons [but] the sign of the correlation between costs and wages is ambiguous" (Borjas, 1999, p. 12). We consider that they depend on productivity and that their distribution is known to A's government. In addition:

Assumption 3. *$c : [\underline{\theta}, \overline{\theta}] \to \mathbb{R}^{++}$ is a $\mathscr{C}^2$ monotonic function satisfying $c'(\theta) < V'_B(\theta)$.*

Monotonicity implies that A's government will know $c(\theta)$ if it knows θ. Migration costs are allowed to be either non-increasing or non-decreasing provided they do not increase faster than the laissez-faire utility V_B. No restrictions are placed on their level.

The *reservation utility* is defined as the maximum utility an individual staying in A can obtain abroad. It is thus equal to $V_B(\theta) - c(\theta)$. Assumption 3 amounts therefore to considering that the outside opportunities are increasing in productivity. An individual will leave A if and only if his utility in this country is less than his reservation utility. Hence, the *participation constraint* for the θ-individuals is defined as

$$V_A(\theta) \geq V_B(\theta) - c(\theta). \tag{PC}$$

Definition 1. θ^* *(when it exists) is the infimum of the set* $\left\{\theta \in \left[\underline{\theta}, \overline{\theta}\right] : (PC) \textit{ active}\right\}$.

2.4 Social Objective and Tax Policy

We define a *national* as an individual born in A. Hence, all individuals have A's nationality. These individuals are allowed to vote with their feet and settle down in B, but are committed to working in the country where they live. Since the focus is on emigration of highly skilled individuals, we consider that A's resident population corresponds to an interval of types and introduce the following definition.

Definition 2. $\widehat{\theta} \in \left[\underline{\theta}, \overline{\theta}\right]$ *is the supremum productivity in A's resident population*[1].

A's government intends to implement the purely redistributive tax policy $T(\theta, l) : \mathscr{T} \to \mathbb{R}$ corresponding to the best compromise between equity and efficiency. Its desire to redistribute income is captured through its aversion to income inequality $\rho \in \mathbb{R}^+$. A zero aversion corresponds to utilitarianism and an infinite one to the Rawlsian maximin. It seems sensible to consider that A's government is not able to levy taxes in B since the fiscal prerogative is closely linked to national sovereignty and that it is not willing to redistribute income to the individuals staying in B. Consequently, $\mathscr{T} = \left[\underline{\theta}, \widehat{\theta}\right] \times \left[0, \overline{l}\right]$. Since $\xi_A = x_A$, $T := z_A - x_A$ so that the tax revenue constraint of A's government reads

$$\int_{\underline{\theta}}^{\widehat{\theta}} (z_A - x_A)\, dF(\theta) \geq 0. \tag{TR}$$

In the rest of the paper, γ denotes the Lagrange multiplier associated with (TR).

The social objective is more difficult to specify when the agents can vote with their feet. Indeed, it does not only depend on ρ which is captured through an isoelastic function à la Atkinson defined by $\phi_\rho : \mathbb{R}^{++} \to \mathbb{R}$, $\phi_\rho(U) = U^{1-\rho}/(1-\rho)$ for $\rho \neq 1$ and $\phi_1(U) = \ln U$, but also on the answers to the following questions. First, should we maximize total or average social welfare? We consider herein that the government is interested in social welfare per capita because we want to be able to compare allocations differing in population size. Second, who are the agents whose welfare is to count? We distinguish three social points of view.

From the *Mercantilist* point of view, A's government maximizes the average social welfare of its nationals,

$$W^M_{A,\rho} := \int_{\underline{\theta}}^{\overline{\theta}} \phi_\rho(V_A(\theta))\, dF(\theta), \tag{5}$$

[1]The rather usual assumption that the productivity distribution is defined on a compact set is required herein to treat $\widehat{\theta}$ as a free upper bound which is optimally chosen to maximize social welfare.

while ensuring through participation constraints that every national is living in the country,

$$V_A(\theta) \geq V_B(\theta) - c(\theta) \text{ for all } \theta \leq \overline{\theta}. \tag{6}$$

This point of view corresponds to the mercantilist idea, formulated by Bodin (1578), that "the only source of welfare is mankind itself". Emigration should therefore be prevented to keep the state prosperous. By analogy, we call Mercantilist a tax policy which prevents emigration of all the A's nationals.

From the *National* point of view, A's government maximizes the average social welfare of its nationals, whether they are in A or B,

$$W_{A,\rho}^N\left(\widehat{\theta}\right) := \int_{\underline{\theta}}^{\widehat{\theta}} \phi_\rho\left(V_A(\theta)\right) dF(\theta) + \int_{\widehat{\theta}}^{\overline{\theta}} \phi_\rho\left(V_B(\theta) - c(\theta)\right) dF(\theta) \tag{7}$$

and takes participation constraints into account for the nationals living in A,

$$V_A(\theta) \geq V_B(\theta) - c(\theta) \text{ for all } \theta \leq \widehat{\theta}. \tag{PC'}$$

A's government faces therefore a *fixed* population problem consisting in allocating its nationals between A and B. A first interpretation of this point of view rests on the idea that the fiscal system finds its legitimacy in its democratic adoption. The government should therefore care for every individual who has the right to vote[2]. A second interpretation is based on the *jus sanguinis*: when the nation is not defined as a territorial entity, it is not the place of residence but the nationality which matters.

From the *Resident* point of view, the government is interested in the average social welfare of its residents,

$$W_{A,\rho}^R\left(\widehat{\theta}\right) := \frac{1}{F\left(\widehat{\theta}\right)} \int_{\underline{\theta}}^{\widehat{\theta}} \phi_\rho\left(V_A(\theta)\right) dF(\theta). \tag{8}$$

The participation constraints are the same as in (PC'). The basic idea is that a public policy should take the welfare of all taxpayers into account. Since the welfare of the nationals living in B does not count, A's government deals with a population problem consisting in "*different* number choices" (Parfit, 1984). $W_{A,\rho}^R\left(\widehat{\theta}\right)$ is based on average utilitarianism, which is known to face the Mere Addition Paradox: the addition of individuals whose utility is less than the average utility in the initial population is regarded as suboptimal even if this change in population size affect no one else and does not involve social injustice. However, since we are focusing on emigration of the highest skilled individuals initially living in A, whose utility will be shown to be greater than the average utility in A, this paradox is not a matter herein. Finally, it should be noted that a government adopting the National or Resident points of view can always choose to implement a Mercantilist tax policy. Consequently, the Mercantilist point of view provides a lower bound to social welfare and constitutes a benchmark.

[2]In France, the 14th Article of the Declaration of the Rights of Man and of the Citizen, which has constitutional value, provides that: "*All citizens* have the right to vote, by themselves or through their representatives, for the need for the public contribution, to agree to it voluntarily, to allow implementation of it, and to determine its appropriation, the amount of assessment, its collection and its duration". For example, twelve senators represent the French nationals living abroad.

3 First-Best Optimal Allocations: the Curse of the Middle-Skilled Workers

This section is devoted to the first-best allocations where each individual's productivity is public information. In consequence, the government knows everything about every individual and implements taxes depending on productivity, i.e. $T(\theta, l) = T(\theta)$.

The indirect utility and social welfare obtained in the *closed* economy A, denoted $V_A^c(\theta)$ and W_A^c respectively, will be used as benchmarks. Recall that a curse of the high-skilled workers characterizes the social optimum in a closed economy, provided Assumption 2 holds and ρ is finite (Mirrlees, 1974): the optimum indirect utility $V_A^c(\theta)$ is decreasing in θ. When ρ is infinite, all individuals receive the same utility level.

3.1 Mercantilist Point of View

The government chooses the tax paid by each individual or, equivalently, the consumption-income bundle intended for each individual.

Problem 1 (Mercantilist Criterion, First-Best)**.** *Find $(x, l) \in \mathscr{X}$ to maximize $W_{A,\rho}^M$ subject to (6) and (TR).*

Proposition 1 (The Curse of the Middle-Skilled Workers)**.** *Let $V_A^C(\overline{\theta}) < V_B(\overline{\theta}) - c(\overline{\theta})$. Under Assumptions 2–3 and when ρ is finite, the optimum indirect utility in A is V-shaped in θ; there is a productivity level θ^* from which the participation constraints are binding.*

Figure 1 illustrates the proposition. $V_A^C(\theta)$ is the utility a θ-individual would obtain in A if the population were immobile. $V_B(\theta) - c(\theta)$ is the reservation utility. On panel (a), the government's aversion to income inequality is finite. $V_A^C(\theta)$ is strictly decreasing in θ. On the contrary $V_A(\theta)$ is decreasing up to θ^* from which it is equal to the increasing reservation utility. The θ^*-individuals are thus the worse-off when potential mobility is taken into account. Panel (b) depicts the optimum situation when the government's aversion to income inequality is infinite. The optimal indirect utility remains constant up to θ^*, from which it is equal to the reservation utility. In addition, the individuals with productivity less than θ^* receive a utility level which is less than that they would obtain if A were a closed economy.

To interpret Proposition 1, let us consider that the individuals living in A are initially unable to vote with their feet, because of high migration costs or legal barriers. Now some of them become mobile and threaten to emigrate if the utility they obtain in A is insufficient compared to what they could have in B. Faced with this situation, A's Mercantilist government introduces participation constraints to prevent its nationals from emigrating. These constraints split the population in two intervals: they are indeed inactive up to the productivity level θ^* from which they become active. It is therefore no longer possible to require the most talented individuals to work as much as without mobility, that is to require them to keep working even though the labour disutility exceeds the gains from the increase in income. The possibility that the high-skilled individuals vote with their feet prevents the government from extracting from them as much as before. That results in less redistribution in A. In consequence, the situation of the individuals for whom the participation constraints are not binding ($\theta \leq \theta^*$) gets worse. Eventually,

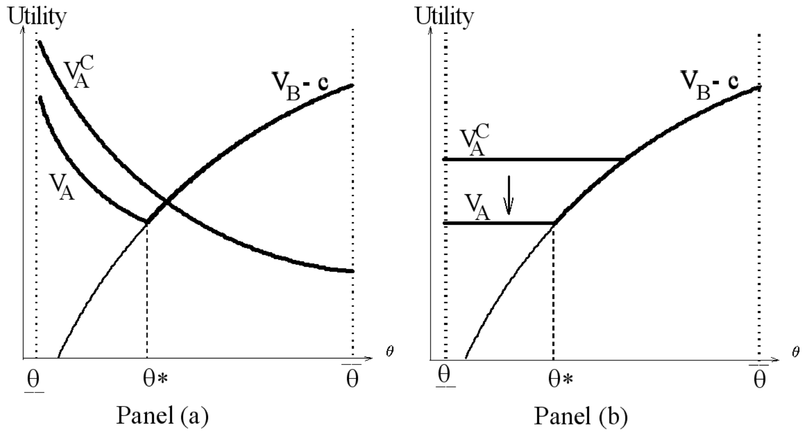

Figure 1: The curse of the middle-skilled workers

it is from the middle-skilled workers that the productive rent is extracted to the maximum, but within some limits since their utilities have to remain greater than their reservation utilities. The middle-skilled workers, insufficiently talented to threaten to leave the country, appear as the main victims of the openness. In other words, a curse of the middle-skilled workers replaces the curse of the high-skilled workers.

3.2 National and Resident Points of View

We examine if it is socially optimal to use participation constraints to prevent emigration of the high-skilled individuals from the National and Resident points of view.

Problem 2 (National and Resident Criteria, First-Best)**.** *Find* $(x,l) \in \mathscr{X}$ *and* $\widehat{\theta} \in [\underline{\theta}, \overline{\theta}]$ *solution to* $\max W^i_{A,\rho}\left(\widehat{\theta}\right)$ *s.t. (PC') and (TR),* $i = \{N,R\}$.

Proposition 2. *Under Assumptions 2–3, the optimal tax policy is Mercantilist from both National and Resident points of view.*

Hence, the National and Resident points of view coincide with the Mercantilist point of view at the optimum. Assume $\widehat{\theta} < \overline{\theta}$ is socially optimal. The individuals with productivity $\widehat{\theta}$ are indifferent between A and B, i.e. $V_A = V_B - c$ at $\widehat{\theta}$, and those with productivity greater than $\widehat{\theta}$ emigrate to B. Note that it is always *feasible* to make them relocate to A, without reducing the indirect utilities of A's residents, in giving them their laissez-faire utility V_B (or a bit more than their reservation utility). Since the laissez-faire utility of the individuals with productivity $\theta > \widehat{\theta}$ is always greater than the average utility within A's resident population, this feasible change results in an increase in the social objective. A contradiction. The social optimum corresponds therefore to the corner solution where the resident population is the national one.

4 Second-Best Optimal Allocations

The distribution of characteristics in the economy remains common knowledge, but the individual's productivity is now private information. A's government is thus restricted to setting taxes as a function of earnings, i.e. $T(\theta,l) = T(z)$. In addition, it has to ensure that the tax schedule is not only incentive compatible and budget balanced, but also individual rational.

4.1 Statement of the Problem

T is an *incentive compatible* tax schedule if and only if individuals living in A have an incentive to reveal their type truthfully when it is implemented. Formally, the incentive compatibility conditions read

$$u\left(x_A\left(\theta'\right), z_A\left(\theta'\right);\theta\right) \leq u\left(x_A(\theta), z_A(\theta);\theta\right) \text{ for all } \left(\theta,\theta'\right) \in \left[\underline{\theta},\widehat{\theta}\right]^2. \tag{IC}$$

To deal with this uncountable infinity of constraints, the Spence-Mirrlees single-crossing property is assumed to hold:

Assumption 4. $s'_\theta(x,z;\theta) < 0$.

Under this assumption, (IC) is equivalent to:

$$V'_A(\theta) = -\frac{z_A(\theta)}{\theta} u'_z\left(x_A(\theta), z_A(\theta);\theta\right) \text{ for } \theta \leq \widehat{\theta}, \tag{FOIC}$$

$$z_A(\theta) \text{ non-decreasing for } \theta \leq \widehat{\theta}. \tag{SOIC}$$

The proof of this equivalence is standard and is omitted. (FOIC) is an envelope condition which specifies how the indirect utility $V_A(\theta)$ must locally change. Since $V'_A(\theta) \geq 0$, $V_A(\theta)$ cannot be V-shaped and Proposition 1 does no longer apply. (SOIC) is a global monotonicity condition of gross income which implies that $z_A(\theta)$ is differentiable almost everywhere.

By the revelation principle, the most general class of direct revelation mechanisms (x_A, z_A) to consider is the class of almost everywhere differentiable functions. Here, we will restrict the analysis to the class of functions which are continuously differentiable, except at a finite number of points in $\left(\underline{\theta},\widehat{\theta}\right)$[3]. Hence, we will look for the optimal tax function among the admissible functions which are continuous but can exhibits kinks at a finite number of points corresponding to jumps of the marginal rate of tax. In addition, (SOIC) is equivalent to

$$z'_A(\theta) \geq 0 \text{ for } \theta \leq \widehat{\theta}. \tag{SOIC'}$$

Since A's government does not know for which productivity levels the participation constraints are binding, the incentive compatibility constraints and the participation conditions (PC') have to be taken simultaneously into account for all A's residents[4]. The optimization programmes we focus on read thus as follows.

[3] Similar restrictions are introduced in Ebert (1992) which provides a general setting to take the second-order condition for incentive compatibility explicitly into account.

[4] If the participation constraints (PC') were not type-dependent, it would be necessary and sufficient to check that they are satisfied at $\overline{\theta}$ since (FOIC) ensures that the optimal utility path is non-decreasing.

Problem 3 (Second-Best)**.** *Find $T(z_A)$ which maximizes $W^i_{A,\rho}$, $i = \{M,N,R\}$, subject to (i) (FOIC), (SOIC'), (PC'), (TR); (ii) $\widehat{\theta} = \overline{\theta}$ when $i = M$ and $\widehat{\theta} \in [\underline{\theta}, \overline{\theta}]$ otherwise.*

These problems are complex for four reasons. First, the participation constraints can a priori bind on any subset of the resident population, even at isolated points, because $V_A - V_B + c$ is not necessarily monotonic. Second, we take (SOIC) explicitly into account and allow bunching of individuals. Third, the participation constraints (PC') are pure state constraints in the optimization programme. The adjoint variables may thus have jump discontinuities. Lastly, from the National and Resident points of view, $\widehat{\theta}$ is free to vary between $\underline{\theta}$ and $\overline{\theta}$.

In solving Problem 3, we assume that the adjoint variables have a finite number of jump discontinuities and are $\mathscr{C}^1$ elsewhere. This ensures that the necessary conditions are equivalent to the sufficient ones, provided some concavity conditions are satisfied. There might be no solution satisfying this assumption, but it seems nevertheless sensible to make it because it would otherwise be very difficult to say anything about the optimal tax schemes. For later reference, we call ι the adjoint variable associated with (FOIC) and $\pi' \geq 0$ the Lagrange multiplier of (PC), which corresponds to the shadow price of a marginal increase in the reservation utility at θ. We also define

$$\pi(\theta) := \pi(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \pi'(\tau)\, d\tau \tag{9}$$

as the shadow price of a uniform marginal increase in the reservation utility for all $\theta' \leq \theta$ and note that $\pi(\theta)$ is a non-decreasing function, with derivative $\pi'(\theta)$ almost everywhere.

4.2 Tax Schedule for the Individuals Threatening to Emigrate

We derive some properties which are satisfied by all optimal tax schemes, irrespective of the social point of view which is considered. The first one deals with the sign of the tax function.

Property 1. *Let (PC) be active at θ. Then, $T(z_A(\theta)) > 0$.*

It is always possible to levy taxes on the individuals who threaten to emigrate. Indeed, let (PC) be active at some θ. Since $c(\theta) > 0$ under Assumption 3, $V_B(\theta) > V_A(\theta)$ and thus $T(z_A(\theta)) > 0$.

We now examine the case where the participation constraints are active on any non-empty open interval $I \subset [\underline{\theta}, \overline{\theta}]$. Hence,

$$u(x_A, z_A; \theta) \equiv V_B(\theta) - c(\theta) \text{ for } \theta \in I. \tag{10}$$

Differentiating and employing (FOIC) show that the rate of increase of the indirect utility the government has to give to the individuals so that they reveal their private information, is equal to the slope of the reservation utility on I. Indeed,

$$V_A'(\theta) = -z_A(\theta)\, u_z'(x_A, z_A; \theta)/\theta \equiv V_B'(\theta) - c'(\theta) \text{ for } \theta \in I. \tag{11}$$

Rearranging (11) yields

$$z_A(\theta) = -\theta \left[V_B'(\theta) - c'(\theta)\right]/u_z' \text{ for } \theta \in I, \tag{12}$$

which is differentiated and factorized by z'_A to get

$$z'_A(\theta) = \frac{\left[V'_B(\theta) - c'(\theta)\right]\left\{\theta\left(u''_{xz}x'_A + u''_{\theta z}\right) - \left(1 + \theta\frac{V''_B(\theta) - c''(\theta)}{V'_B(\theta) - c'(\theta)}\right)u'_z\right\}}{\left(u'_z\right)^2 - \theta\left(V'_B(\theta) - c'(\theta)\right)u''_{zz}} \text{ for } \theta \in I. \tag{13}$$

Under Assumptions 1 and 3, the sign of the RHS of the latter expression is given by the curly bracket. Therefore (SOIC') restricts the set of non-empty open intervals where (PC) can be binding.

Property 2. *(PC) can be active on a non-empty open interval only if*

$$\frac{\theta\left(u''_{xz}x'_A + u''_{\theta z}\right)}{u'_z} \leq 1 + \theta\frac{V''_B(\theta) - c''(\theta)}{V'_B(\theta) - c'(\theta)}. \tag{14}$$

The elasticity of the marginal reservation utility, evaluated at θ, appears on the RHS. The LHS captures the behavioural response of the θ-individuals to a slight change in their reservation utility. When u is quasilinear in consumption,

$$u(x_A, z_A; \theta) = x_A - v(z_A/\theta), \text{ with } v' > 0 \text{ and } v'' > 0, \tag{15}$$

we have $e^H = v'/(lv'')$, $u''_{xz} = 0$, $u''_{\theta z} = (1 + lv''/v')\,v'/\theta^2$ and $u'_z = -v'/\theta$ so that (14) reads

$$-1 - \frac{1}{e^H(\theta)} \leq 1 + \theta\frac{V''_B(\theta) - c''(\theta)}{V'_B(\theta) - c'(\theta)}, \tag{16}$$

the LHS of which is strictly negative.

Property 3. *Let preferences be quasilinear in consumption and consider any non-empty interval where (PC) is binding. Then, there is no bunching on this interval when $V_B(\theta) - c(\theta)$ is convex.*

By (15), the taxes levied on the individuals with gross income z_A are equal to

$$T(z_A) = z_A - V_A - v(z_A/\theta) \tag{17}$$

Employing (12), the average tax rate on I amounts to

$$\frac{T(z_A)}{z_A} = 1 - \frac{1 - T'}{\theta\left(V'_B - c'\right)}\left[V_B - c + v\left(\frac{V'_B - c'}{1 - T'}\right)\right] \text{ for } \theta \in I. \tag{18}$$

When the elasticity of labour supply is constant ($e^H(\theta) = e$), the disutility of labour can be described by

$$v(l) = l^{1+1/e}/(1 + 1/e) \tag{19}$$

and the following property is obtained.

Property 4. *Let preferences be quasilinear in consumption, e^H be constant, and I be a non-empty open interval where (PC) is active. Then, when migration costs are constant, equal to $\bar{c}$, we have $T(z_A) = \bar{c}$ and $T'(z_A) = 0$ for all $\theta \in I$.*

In this case, the individuals who threaten to emigrate are taxed as high as their migration costs. A's government extracts from them the maximum rent which is possible given the requirement of individual rationality of the tax schedule.

4.3 The Mercantilist Point of View

We study the impact of the threat of migration on the optimum tax scheme in A when A's government adopts the Mercantilist point of view. To this aim, Mirrlees's and Diamond's formulae are extended to the case where agents are allowed to vote with their feet. These formulae give the optimal marginal tax rates when bunching does not occur at the social optimum and are derived from necessary conditions.

4.3.1 A First Pass

We first look at a very simple situation to illuminate the basic economic relations behind the optimal marginal tax rate formula. A's government is Rawlsian and $\underline{\theta} = 0$. The social objective is thus to maximize the social benefit given to the worst-off individuals, i.e. to maximize tax revenue $\overline{G}$. Individual preferences are quasilinear in consumption so that there are no income effects on labour supply and the elasticity of labour supply is constant ($e^H(\theta) = e$). Migration costs are constant, equal to $c(\theta) = \overline{c}$. In addition, we restrict our attention to the cases where (PC) is only active on non-degenerate intervals. By Property 4, the taxes paid by the individuals threatening to emigrate will amount to $\overline{c}$.

We adopt the methodology described in Piketty (1997) and Saez (2001) to derive the optimal marginal tax rates without setting any formal optimization programme. We consider the effects of a small increase dT in the optimal marginal tax rates for income between z and $z + dz$. This tax perturbation has three effects on social welfare, captured through changes in tax revenue $\overline{G}$. The first two effects are the same as in a closed economy. The third one arises from potential emigration.

Mechanical effect: All individuals with income greater than z pay additional taxes $dTdz$. Since their proportion is given by $1 - F(\theta_z)$, the mechanical effect on tax revenue is

$$dG^+ = (1 - F(\theta_z)) \times dTdz. \tag{20}$$

Elasticity effect: The net-of-tax wage rate of the individuals with income between z_A and $z_A + dz_A$ decreases from $\theta_z(1 - T')$ to $\theta_z(1 - T' - dT)$, i.e. by $dT/(1 - T')\%$. The corresponding reduction in gross income z for the $fd\theta$ individuals is therefore $e \times dT/(1 - T') \times zfd\theta$. The resulting loss in taxes amounts to $dG_1^- = T' \times e \times dT/(1 - T') \times zfd\theta$. Since $d\theta = dz/[l(1+e)]$ by definition of e, the overall elasticity effect on tax revenue is

$$dG_1^- = \frac{T'}{1 - T'} \frac{e}{1+e} \times \theta f \times dTdz. \tag{21}$$

Participation effect: Individuals with income greater than z for whom (PC) is active have to be compensated for the increase in taxes. Since preferences are quasilinear in consumption and migration costs constant, an amount $dTdz$ has to be given to them. Let μ be a measure on $[\underline{\theta}, \overline{\theta}]$ and Θ^{PC} be the set of productivity where (PC) is active. Provided (PC) is only active on non-degenerate intervals, the participation effect on tax revenue is

$$dG_2^- = \mu\left(\left[\theta_z, \overline{\theta}\right] \cap \Theta^{PC}\right) \times dTdz. \tag{22}$$

At the social optimum, the small tax reform perturbation has no first-order effect. Consequently, it must be $dG^{+}+dG_1^{-}+dG_2^{-}=0$. Since θ_z has been chosen arbitrarily, the following result is obtained.

Proposition 3. *Assume (PC) is not active at isolated points. Then, when preferences are quasilinear in consumption, $e^H(\theta)=e$ and $c(\theta)=\overline{c}$, the Rawlsian optimal marginal tax rates are given by*

$$\frac{T'}{1-T'}=\left(1+\frac{1}{e}\right)\frac{1-F(\theta)}{\theta f(\theta)}\left(1-\frac{\mu\left(\left[\theta,\overline{\theta}\right]\cap\Theta^{PC}\right)}{1-F(\theta)}\right)\textit{ for all }\theta<\overline{\theta},\tag{23}$$

and $T'=0$ for $\theta=\overline{\theta}$, provided there is no bunching at the optimum.

When $\mu\left(\left[\theta,\overline{\theta}\right]\cap\Theta^{PC}\right)=0$, (23) reduces to the formula derived by Piketty (1997) in closed economy. Three points are worth noting. First, since e^H is constant, it can be rewritten as

$$\frac{T'}{1-T'}=\frac{T_{IC}'^{R}}{1-T_{IC}'^{R}}\left(1-\frac{\mu\left(\left[\theta,\overline{\theta}\right]\cap\Theta^{PC}\right)}{1-F(\theta)}\right),\tag{24}$$

where $T_{IC}'^{R}$ is the Rawlsian marginal tax rate the θ-individuals would face in A in the absence of individual mobility. As $\mu\left(\left[\theta,\overline{\theta}\right]\cap\Theta^{PC}\right)\leq 1-F(\theta)$, the openness of the economy results in a decrease in the marginal tax rates faced by all individuals, and not only those who threaten to emigrate. Second, $\mu\left(\left[\theta,\overline{\theta}\right]\cap\Theta^{PC}\right)/(1-F(\theta))$ is increasing in θ for all $\theta\leq\theta^{*}$. Hence, the closer θ to θ^{*}, the greater the reduction in marginal tax rates. Third, by Property 4 $T'=0$ on non-empty open intervals where (PC) is active. But, by (23), $T'=0$ if and only if $1-F(\theta)=\mu\left(\left[\theta,\overline{\theta}\right]\cap\Theta^{PC}\right)$. Therefore the participation constraints are active for all individuals with productivity $\theta>\theta^{*}$ when θ^{*} exists.

4.3.2 The General Case

We now extend the previous analysis by relaxing all simplifying assumptions, except the absence of bunching.

Proposition 4. *From the Mercantilist point of view and in the absence of bunching, the optimal marginal tax rates are given by*

$$\frac{T'}{1-T'}=A(\theta)B(\theta)C(\theta)\textit{ for all }\theta<\overline{\theta},\tag{25}$$

where $A(\theta):=\dfrac{1+e^M(\theta)}{e^H(\theta)}$, $C(\theta):=\dfrac{1-F(\theta)}{\theta f(\theta)}$ *and* $B(\theta):=B_1(\theta)-B_2(\theta)-B_3(\theta)$ *with*

$$B_1(\theta):=\frac{1}{1-F(\theta)}\int_{\theta}^{\overline{\theta}}\left[1-\frac{\phi_\rho'(V_A(\tau))}{\gamma}u_x'(x_A,z_A;\tau)\right]\Psi_{\theta\tau}dF(\tau),$$

$$B_2(\theta):=\frac{1}{1-F(\theta)}\int_{\theta}^{\overline{\theta}}\frac{\pi'(\tau)}{\gamma}u_x'(x_A,z_A;\tau)\Psi_{\theta\tau}d\tau,\quad \pi'\geq 0\ (=0\textit{ if }V_A>V_B-c),$$

$$B_3(\theta):=\frac{\iota\left(\overline{\theta}\right)}{\gamma}\frac{u_x'(x_A,z_A;\theta)}{1-F(\theta)},\quad \iota\left(\overline{\theta}\right)\geq 0\ \left(=0\textit{ if (PC) inactive at }\overline{\theta}\right),$$

where $\gamma > 0$ *and* $\Psi_{\theta\tau} = \exp\int_{\theta}^{\tau}\left(1 - \frac{e^{M}(\delta)}{e^{H}(\delta)}\right)\frac{z_A'(\delta)}{z_A(\delta)}d\delta$.

Proposition 4 extends Mirrlees's (1971) optimal income tax formula to take the threat of migration into account, using behavioural elasticities as in Saez (2001). (25) reflects the trade-off between efficiency and equity when the government has decided to maintain the national productive capacity to the maximum in preventing its nationals from leaving the country. $A(\theta)$ and $C(\theta)$ are the usual efficiency and demographic factors, respectively. We note however that the value of $A(\theta)$ is usually not the same whether the individuals can or cannot vote with their feet since it depends on gross income which is endogenous. The factor $B(\theta)$, which combines efficiency and equity, is the only factor which does not write as in Mirrlees's formula, in which the RHS of (25) reduces to $A(\theta)B_1(\theta)C(\theta)$.

Interpretation and Implications As previously, we consider a small tax reform perturbation around the optimal tax schedule to gain insight into the way the optimal marginal tax rates are determined. A small increase dT for gross income between z and $z+dz$ has four effects on social welfare. Three effects are already observed in closed economy, and have been thoroughly examined by Saez (2001). We start by mentioning them for completeness.

- *The three "usual" effects* allow us to grasp $A(\theta)$, $B_1(\theta)$ and $C(\theta)$.

First, the small increase in marginal tax rates mechanically results in individuals with gross income greater than z paying additional taxes. Second, the elasticity response from the taxpayers with gross income between z and $z+dz$ decreases their labour supply and reduces tax revenue. Third, under Assumption 2, the increase in taxes paid by these individuals has an income effect, leading them to work more, which is good for tax receipts.

- *The new participation effect* illuminates $B_2(\theta)$ and $B_3(\theta)$.

The tax reform perturbation *mechanically* results in an increase in taxes paid by all individuals with gross income above z. Consequently, those among them for whom the participation constraints were already active receive now a utility level below their reservation utility. Then the participation constraints (11) are no longer satisfied. Consequently, these individuals have to be compensated for the increase in taxes they face. The compensation proceeds as follows.

The *substitution effect* leads A's government to totally compensate them for staying in A. Each θ-individual is thus given $u_x(x_A, z_A; \theta) \times dTdz$ additional units of utility. Since $\pi'(\theta)$ is the shadow price of the participation constraint at θ and γ the Lagrange multiplier of the tax revenue constraint (TR), the cost in terms of social welfare of the compensation of all θ-individuals amounts to

$$\frac{\pi'(\theta)}{\gamma}u_x(x_A, z_A; \theta) \times dTdz. \tag{26}$$

The substitution effect combines with the usual *income effect*. Since leisure is a normal good under Assumption 2, the increase in the tax burden paid by all individuals with income greater than z induces them to work more. This allows A's government to increase the marginal taxes they face. As a result, it is not required to compensate the individuals who threaten to emigrate among them as high as the increase in taxes they face. The magnitude of the uncompensated behavioural response is summarized by $\Psi_{\theta\tau}$ which is larger than one, unless preferences are separable in consumption and leisure, in which case $\Psi_{\theta\tau} = 1$. Using (26) and $\Psi_{\theta\tau}$, one obtains the average compensation, including income effects, which has to be given to the individuals with productivity between θ_z and $\overline{\theta}$,

$$\frac{1}{1-F(\theta_z)} \int_{\theta_z}^{\overline{\theta}} \frac{\pi'(\tau)}{\gamma} u_x(x_A, z_A; \tau) \Psi_{\theta_z \tau} d\tau \times dT dz = B_2(\theta_z) \times dT dz, \tag{27}$$

$B_2(\theta)$ is positive for all θ below θ' when the participation constraints of the θ'-individuals are active, because increasing the marginal tax rates at θ makes the compensation of the θ'-individuals socially more expensive. In particular, $B_2(\theta)$ goes against progressivity on a range of gross income levels preceding that on which individuals hesitate to leave the country. In addition, as $\pi'(\theta) = 0$ for all $\theta < \theta^*$,

$$B_2(\theta) = \frac{1}{1-F(\theta)} \int_{\theta^*}^{\overline{\theta}} \frac{\pi'(\tau)}{\gamma} u_x(x_A, z_A; \tau) \Psi_{\theta^* \tau} d\tau \text{ for } \theta < \theta^*. \tag{28}$$

Differentiating, one obtains $B_2'(\theta) = f(\theta) / (1-F(\theta))^2 \times (1-F(\theta^*)) B_2(\theta^*)$ for $\theta < \theta^*$, which is strictly positive: the closer to θ^* the productivity level at which the small tax reform perturbation takes place, the higher the average compensation required to restore individual rationality of the tax schedule. When θ is greater than θ^*, it is not possible to determine the sign of $B_2(\theta)$ in the general case. In passing, we note that $B_1(\theta) - B_2(\theta)$ can be written as

$$B_1(\theta) - B_2(\theta) = \int_{\theta}^{\overline{\theta}} [1 - g(\tau)] \Psi_{\theta\tau} dF(\tau), \tag{29}$$

where $g(\theta) = \left[\frac{\phi_\rho'(V_A(\theta))}{\gamma} + \frac{\pi'(\theta)}{\gamma f(\theta)}\right] u_x'(x_A, z_A; \theta)$ is the social marginal weight of the θ-individuals within the population. Consequently, the participation constraints alter the redistributive tastes of the government to give higher social priority to the people threatening to emigrate. This change reflects the trade-off between sustaining redistribution and maintaining national productive capacity.

It remains to look at the compensation of the $\overline{\theta}$-individuals. The self-selection constraints (IC) imply that these individuals compare their utility level with those of the less productive individuals. The utility levels of individuals with income below z are unaltered. On the contrary, the compensation process described above increases the utility levels of some individuals with income above z. Consequently, the $\overline{\theta}$-individuals will have greater incentives to mimicking the $(1-F(\theta_z))\%$ individuals with productivity ranging between θ_z and $\overline{\theta}$. A's government has to compensate them in order to restore incentive compatibility of the tax schedule. Since $\iota(\overline{\theta})$ is the shadow price of adjusting the slope of V_A in order to induce individual truthtelling at $\overline{\theta}$, the average cost in terms of social welfare amounts to

$$\frac{1}{1-F(\theta_z)} \frac{\iota(\overline{\theta})}{\gamma} \times dT dz = B_3(\theta_z) \times dT dz \geq 0. \tag{30}$$

When the $\overline{\theta}$-individuals do not initially hesitate to leave A, $\iota\left(\overline{\theta}\right)=0$ and $B_3(\theta)$ vanishes. Since $B_3(\theta)$ is non-negative, it reinforces the decrease in marginal tax rates induced by $B_2(\theta)$. The participation effect results therefore in the adjustment of the optimal marginal tax rates to make the *average* tax rates compatible with the participation constraints. In consequence, A's government should be particularly cautious about increasing marginal tax rates even at productivity levels where individuals do not hesitate to leave the country.

We now give the formula to determine the optimal marginal rate of tax faced of the most productive individuals.

Property 5. *From the Mercantilist point of view and in the absence of bunching, the optimal marginal tax rate faced by the $\overline{\theta}$-individuals is*

$$\frac{T'\left(z_A\left(\overline{\theta}\right)\right)}{1-T'\left(z_A\left(\overline{\theta}\right)\right)}=-A\left(\overline{\theta}\right)\frac{\iota\left(\overline{\theta}\right)u'_x\left(x_A,z_A;\overline{\theta}\right)}{\gamma\overline{\theta}f\left(\overline{\theta}\right)}\leq 0\ \left(=0 \textit{ if (PC) inactive at } \overline{\theta}\right). \tag{31}$$

When individual mobility matters, the $\overline{\theta}$-individuals can be subsidized at the margin for staying in A. An example is provided in the simulation section. This property contrasts with the zero marginal tax rate at the top obtained in closed economy (Seade, 1977), in which case $\iota\left(\overline{\theta}\right)=0.<$

For completeness, we extends Diamond's (1998) formula to take the threat of migration into account. We call $D(\theta)$ the average marginal social welfare of the individuals with productivity above θ, $D(\theta):=\int_\theta^{\overline{\theta}}\phi'_\rho\left(V_A(\tau)\right)dF(\tau)/\left(1-F(\theta)\right)$, and $\Pi(\theta)$ the average shadow price of a marginal increase in the reservation utilities of all individuals with productivity greater than θ, $\Pi(\theta):=\int_\theta^{\overline{\theta}}\pi'(\tau)\,d\tau/\left(1-F(\theta)\right)$.

Corollary 1. *When preferences are quasilinear in consumption, Proposition 4 applies with $A(\theta):=1+1/e^H(\theta)$ and $B(\theta):=B_1(\theta)-B_2(\theta)-B_3(\theta)$,*

$$B_1(\theta):=1-\frac{D(\theta)}{\gamma};\ B_2(\theta):=\frac{\Pi(\theta)}{\gamma};\ B_3(\theta)=\frac{\iota\left(\overline{\theta}\right)}{\gamma(1-F(\theta))}.$$

4.4 National and Resident Points of View

From the Mercantilist point of view, the whole population is constrained to stay in A. We now relax this constraint to examine whether keeping everybody in the country is not too much expensive in terms of social welfare.

We proceed in two steps to determine which $\widehat{\theta}$ should be optimally chosen. We first consider the following problem in which $\widehat{\theta}$ is arbitrarily chosen by A's government.

Problem 4. *Given $\widehat{\theta}\in\left[\underline{\theta},\overline{\theta}\right]$,*

$$\begin{cases}\mathscr{W}^i_{A,\rho}\left(\widehat{\theta}\right)=\max_{x_A,z_A}W^i_{A,\rho}\left(\widehat{\theta}\right),\\ s.t.\ (FOIC),\ (SOIC'),\ (PC'),\ (TR).\end{cases}\qquad (i=\{N,R\})$$

We call $\iota^{i}_{\widehat{\theta}}(\theta)$, $i=\{N,R\}$, the adjoint function ι associated with the solution to Problem 4. The solution in $\widehat{\theta}$ to Problem 3, from the National and Resident points of view, is then obtained as

$$\widehat{\theta}^{i} \in \arg \max_{\widehat{\theta}\in[\underline{\theta},\overline{\theta}]} \mathscr{W}^{i}_{A,\rho}\left(\widehat{\theta}\right), \; i=\{N,R\}. \tag{32}$$

For all $\widehat{\theta} \in (\underline{\theta},\overline{\theta})$, the $\widehat{\theta}$-individuals are by construction indifferent between living in A and in B. Otherwise, the individuals with greater productivity would not be in B and $\widehat{\theta}$ would not be the supremum of A's resident population. The $\widehat{\theta}$-individuals receive therefore their reservation utility. Since $c>0$ under Assumption 3, we have[5]:

Property 6. *At the solution to Problem 4, for any $\widehat{\theta} \in \left(\underline{\theta},\overline{\theta}\right)$, the $\widehat{\theta}$-individuals pay strictly positive taxes.*

The National and Resident optimal tax schemes can now be characterized.

Property 7. *If $\iota^{i}_{\widehat{\theta}}$ is continuous at $\widehat{\theta}$ for all $\widehat{\theta} \in (\underline{\theta},\overline{\theta}]$ and if the $\overline{\theta}$-individuals pay strictly positive taxes at the second-best Mercantilist allocation, then $\widehat{\theta}^{i}=\overline{\theta}$, $i=\{N,R\}$.*

Property 8. *If there is a non-degenerate interval containing $\widehat{\theta}$ where (PC) is active at the solution to Problem 4 for all $\widehat{\theta} \in (\underline{\theta},\overline{\theta}]$, then $\widehat{\theta}^{i}=\overline{\theta}$, $i=\{N,R\}$.*

In both cases, implementing a tax scheme which induces emigration of the highly skilled individuals always results in a decrease in social welfare. It is therefore socially optimal to choose the greatest possible $\widehat{\theta}$ and to take participation constraints into account accordingly. The optimal income tax schedule is thus Mercantilist. When Property 7 or Property 8 hold from both National and Resident points of view, the three welfare criteria yield the same social optimum. This result parallels Proposition 2 obtained in the first-best setting. However, the explanation given in the first-best does no longer apply.

Indeed, let us assume that $\widehat{\theta}<\overline{\theta}$. Individuals with productivity above $\widehat{\theta}$ are living in B. Making them relocate has now three effects. (i.) It increases social welfare because these individuals have utility levels above the average in A's resident population while $\phi'>0$. (ii.) It increases tax receipts, income redistribution and thus social welfare because it is always possible to levy positive taxes on them since $c>0$. (iii.) It gives rise to a new mimicking behaviour. Indeed, the individuals who were already living in A can now have an incentive to mimicking the $\widehat{\theta}$-individuals who have the most appealing outside options. There is therefore an additional cost to pay to design an incentive-compatible tax scheme, which is bad for tax receipts and social welfare.

The third effect is crucial to understanding the possible interactions between the incentive-compatibility conditions and the type-dependent participation constraints. In closed economy, individuals have the usual incentive to understate θ to obtain greater social benefit whilst en-

[5]When $\widehat{\theta}=\underline{\theta}$, A's resident population is a set of mesure zero. All individuals have the same productivity and income redistribution is not a relevant issue. Hence, $T(z_A)=0$ at $\widehat{\theta}$.

joying more leisure[6]. The downward self-selection constraints are thus prevailing. When type-dependent participation constraints are taken into account, the individuals have not only the usual incentive, but may also be tempted to overstate θ. Indeed, since individuals with greater productivity have more attractive outside options, a less productive individual may choose to mimic them in working harder with a view to obtaining greater compensation for staying in A. This kind of behaviour is referred to as countervailing incentives. Countervailing incentives can give rise to an asymmetry in terms of informational constraints between the individuals with productivity below $\widehat{\theta}$ and the $\widehat{\theta}$-individuals. Indeed, contrary to the former, the latter can only have the usual incentive to mimicking less productive individuals. The cost of making the θ-individuals reveal their private information, represented by $\iota(\theta) \geq 0$, can thus have a *downward* jump discontinuity at $\widehat{\theta}$. This discontinuity corresponds to an *upward* jump discontinuity in π and is given by:

$$\iota\left(\widehat{\theta}^{-}\right) - \iota\left(\widehat{\theta}\right) = \pi\left(\widehat{\theta}\right) - \pi\left(\widehat{\theta}^{-}\right) \geq 0 \ \left(= 0 \text{ if (PC) inactive at } \widehat{\theta}\right), \tag{33}$$

where $\iota\left(\widehat{\theta}^{-}\right) := \lim_{\theta \overset{\leq}{\to} \widehat{\theta}} \iota(\theta)$.

Properties 7 and 8 hold when there is no sudden change in the structure of the individuals' mimicking behaviour when θ comes closer to $\widehat{\theta}$, that is when the usual downward mimicking behaviour is prevailing. Consequently, both positive effects of the presence in A of the highly skilled individuals are greater than the negative one so that the optimal tax scheme should be such that they stay in the home country. On the contrary, the trade-off between maintaining national capacity to the maximum and sustaining the redistribution programme can lead A's government to implement a tax schedule which induces emigration of the most productive individuals when both positive effects are insufficiently high to balance the negative one.

Property 9. *Consider the Mercantilist optimal allocation and the corresponding* $(\gamma, \iota, \pi, T, V_A)$. *If* $\overline{\theta}$ *is an isolated point where (PC) is active and* ι *has a jump discontinuity, then*
(i) $\widehat{\theta}^{N} < \overline{\theta}$ *if*

$$\gamma T(z_A) f + \iota\left(\overline{\theta}^{-}\right) V_A' < \left[\pi - \pi\left(\overline{\theta}^{-}\right)\right] \left[V_B' - c' - V_A'\right]; \tag{34}$$

(ii) $\widehat{\theta}^{R} < \overline{\theta}$ *if*

$$\left[\phi_\rho (V_B - c) - W_{A,\rho}^M + \gamma T(z_A)\right] f + \iota\left(\overline{\theta}^{-}\right) V_A' < \left[\pi - \pi\left(\overline{\theta}^{-}\right)\right] \left[V_B' - c' - V_A'\right], \tag{35}$$

where all functions are evaluated at $\overline{\theta}$ *except otherwise stated.*

(34) and (35) reflect the cost/benefit analysis of the presence in A of the most productive individuals. The positive LHS capture the direct effects of their presence. They have utility above the average and pay strictly positive taxes[7]. The non-negative RHS would vanish if the

[6] In the discrete population model of Guesnerie and Seade (1982), a sufficient condition for incentive-compatibility of the tax scheme is that the downward self-selection constraints are binding, which corresponds to a monotonic chain to the left (see also Weymark (1986, 1987)).

[7] $\iota\left(\overline{\theta}^{-}\right) \geq 0$ corresponds to $-\lim_{\theta \overset{\leq}{\to} \widehat{\theta}} \partial W_{A,\rho}^M / \partial V_A = -\lim_{\theta \overset{\leq}{\to} \widehat{\theta}} \partial W_{A,\rho}^M / \partial (V_B - c)$.

participation constraints were active on a non-degenerate interval at the top, since we would then have $V_A'\left(\overline{\theta}\right)=V_B'\left(\overline{\theta}\right)-c'\left(\overline{\theta}\right)$. Therefore, they correspond to the cost, in terms of social welfare, of not having binding participation constraints on the left of $\overline{\theta}$. This cost can be viewed as indirect because it is due to the mimicking behaviour of other individuals who aspire to benefiting from the appealing outside options of the $\overline{\theta}$-individuals in order to receive greater compensation for staying in A.

To interpret Property 9, let us consider the case depicted in Figure 2. The participation constraints (PC') are active on an interval (θ^*,θ^{**}), with $\theta^*<\theta^{**}<\overline{\theta}$, and at the isolated point $\overline{\theta}$. On (θ^*,θ^{**}), the slope of the reservation utility $V_B'-c'$ is compatible with individual truthtelling. However, on $\left(\theta^{**},\overline{\theta}\right)$, the incentive to mimicking the more skilled individuals becomes higher and incentive compatibility requires a steeper slope in V_A in order to reduce the gap between $V_A(\theta)$ and $V_A\left(\overline{\theta}\right)$: the participation constraints can no longer be active. Because of countervailing incentives, the individuals of $\left(\theta^{**},\overline{\theta}\right)$ have thus to be overcompensated to reveal their type. Countervailing incentives cease suddenly at $\overline{\theta}$, given that the highest skilled individuals can only have the usual downward incentives. At this point, it becomes feasible in terms of incentives to leave the individuals with no rent; hence, since the aim of A's government is to redistribute income, it is optimal to have binding participation constraints at the top. The interaction between the incentive compatibility conditions and the participation constraints is thus socially expensive because A's government would like to have binding participation constraints for the highly skilled individuals as in the first-best setting, in order to extract from them as much as possible and increase redistribution, but this is not feasible given the informational constraints it faces.

If A's government lets the $\overline{\theta}$-individuals emigrate to B in choosing a supremum productivity $\widehat{\theta}<\overline{\theta}$, it reduces the social cost of of countervailing incentives, since the individuals staying in A will no longer have the possibility to mimic them. The choice of $\widehat{\theta}$ by A's government can thus be regarded as a means of revealing private information, at least in the absence of bunching. Indeed, if A's government designs a tax policy such that the individuals with productivity greater than $\widehat{\theta}$ do not receive in A their reservation utility, it *knows* that $\widehat{\theta}$ is the supremum productivity in its resident population and consequently that the individuals with productivity greater than $\widehat{\theta}$ are in B. Property 9 tells us in which cases using this means improves social welfare.

5 Illustration on French Data

We provide optimal income tax schemes for the French economy.

5.1 Calibration

A's government is Rawlsian. The productivity distribution is described by a lognormal distribution obtained from the survey "Budget des familles", year 1995 (see Laslier, Trannoy, and Van Der Straeten (2003, Appendix C)). By kernel estimation, one obtains a mean of 0.2398 and a variance of 0.4403[8]. The productivity levels are rather ad hoc; we normalize them so that the median individuals have annual income equal to the median income in 1995, that is 13 320 euros. Since more than 99.99% of the population has a productivity which is less than 66 600 euros

[8]Since our model does not take the family size into account, the population is restricted to single individuals.

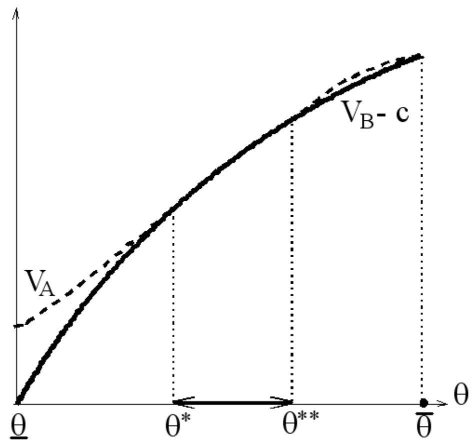

Figure 2: Case where $\overline{\theta}$ is an isolated point where (PC) is active

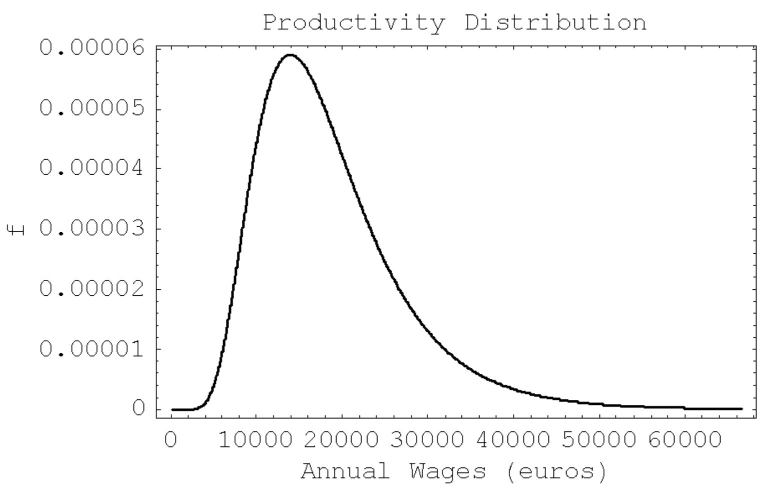

Figure 3: Productivity Distribution in France in 1995

per year, the upper productivity level $\overline{\theta}$ is set equal to this value. The lognormal distribution is completed by a uniform one to obtain the distribution depicted in Figure 3.

Following d'Autume (2000) who provides numerical results based on French data in the absence of individual mobility, we concentrate on the case where there are no income effects on labour supply and e^H is constant, equal to e. The utility function is thus given by (15) and (19). There is no clear consensus over the value of e in France. We choose $e = 0.2$ as a benchmark, which is not unreasonable.

The migration costs are the new ingredient of our model and play therefore a large part in the determination of the optimal income tax schemes. If these costs are used in the standard economic model of migration, there are however few empirical work concerning their level[9]. We provide numerical results for two different scenarios.

Constant migration costs amounts to focusing on material migration costs (transportation of persons and goods) which do not vary significantly from person to person. This case provides a benchmark.

Decreasing migration costs are justified by the idea that the highly productive individuals should have a better command of the language spoken in B as well as degrees or skills that should be more easily recognized in this country. We have rather accurate data concerning migration costs to Australia. The minimum costs of migration range from 7 500 to 11 200 euros for a single person. We take psychological costs into account in assuming that the less productive individuals would face migration costs of 11 200 euros and the most productive ones migration costs of 7500 euros. The migration costs are then adjusted linearly.

5.2 Numerical Results

The main lines of the methodology used to obtain the optimal income tax schemes is described in the Appendix. Further results are available in Simula and Trannoy (2006b).

5.2.1 Constant Migration Costs

Simulation results are presented in Table 1 and Figures 4–6. The optimal income tax schemes we have constructed have the following features. First, they present no bunching of individuals. Second, there is a gross income level z^* from which the participation constraints are active. Third, the individuals with gross income below z^* face strictly positive marginal tax rates; the other are not taxed at the margin.

In consequence, the optimum income tax-function is increasing up to z^* where it reaches a plateau. The theorem obtained by Seade (1982, p. 642) in a closed economy, stating that the optimal tax function is strictly increasing at *all* income levels, does no longer hold when the threat of migration is taken into account. Since gross income is strictly increasing in the absence of bunching, the average tax rates are decreasing above z^*. Hence, taking the threat of migration into account does not only make the tax scheme less progressive, it can also make it regressive.

[9] The IZA Database for Migration Literature (http://www.iza.org/iza/en/webcontent/links/migration) provides 34 matches for "moving costs". These references are mainly theoretical or estimate the macroeconomic costs of migration.

	Worse-Off				Median				Last Twentile				
$c(\theta)$	V	T	T'	T/z	V	T	T'	T/z	V	T	T'	T/z	%PC
closed	13089	-13089	100	$-\infty$	13414	-4471	84.4	-48.7	19873	13370	55.9	37.3	-
3000	11001	-11001	100	$-\infty$	11599	-1784	72.5	-17.3	32234	3000	0	7.1	36.3
6000	12086	-12086	100	$-\infty$	12512	-3183	80.1	-33.0	29231	6000	0	14.2	18.5
9000	12566	-12566	100	$-\infty$	12939	-3800	82.3	-40.4	26231	9000	0	21.3	9.9
12000	12809	-12809	100	$-\infty$	13159	-4111	83.4	-44.2	23228	12000	0	28.4	5.5
15000	12939	-12939	100	$-\infty$	13277	-4277	83.9	-46.3	21243	13410	33.1	34.4	3.1
21000	13049	-13049	100	$-\infty$	13378	-4417	84.3	-48.0	20221	13452	50.7	36.7	0.9
27000	13082	-13082	100	$-\infty$	13407	-4460	84.4	-48.5	19966	13393	54.6	37.1	0.3
33000	13088	-13088	100	$-\infty$	13413	-4468	84.5	-48.7	19906	13377	55.5	37.2	0.1
34000	as in a closed economy for all $c(\theta) \gtrsim 34000$												

Note: (closed) means closed economy; %PC: proportion of individuals for whom (PC) is active.
$c(\theta)$, V, T are expressed in euros; T', T/z are expressed in %.

Table 1: Optimum Allocations (Rawls, e=0.2, constant migration costs)

Figure 4 gives the value of $B(\theta)$, which would be identically equal to one if the economy were closed. When individual mobility matters, $B(\theta)$ is equal to one for the worse-off society's members and decrease at an increasing rate until z^*, from which it is zero. Figure 5 shows how the changes in $B(\theta)$ alter the optimal marginal tax rates. The quasi flat part of the curves, on the top LHS corner comes from the fact that $f(\theta)$ is very low and locally constant at these productivity levels. The optimal average tax rates are given in Figure 6.

5.2.2 Decreasing Migration Costs

The optimum tax scheme has the same features as the ones obtained for constant migration costs, except that there is a gross income level $z^0 < z^*$ from which marginal tax rates are strictly negative (Figure 7). Consequently, the optimum income tax-function is first strictly decreasing in gross income up to z^0, from which it is strictly decreasing. From the theoretical point of view, this example stresses the fact that taking the threat of migration into account can lead the government to subsidize highly skilled individuals, including the highest skilled ones, at the margin. The non-negativity property as well as the zero-at-the-top property of the optimal marginal rates of tax, obtained in closed economy, are thus lost.

6 Conclusion

The potential mobility of individuals in order to benefit from international differences in taxes, adds a conflict between the desire to maintain national capacity in keeping taxes down and the desire to sustain the redistribution programme to the fundamental trade-off between equity and efficiency formulated by Mirrlees (1971). This conflict is a matter of public debate, in the EU or in Canada for instance.

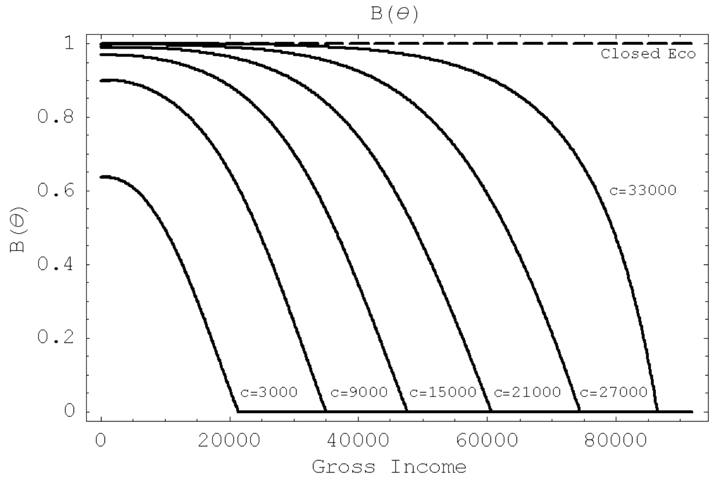

Figure 4: Value of $B(\theta)$ with Respect to Migration Costs (in euros). Rawls, $e = 0.2$, $c(\theta)$ constant.

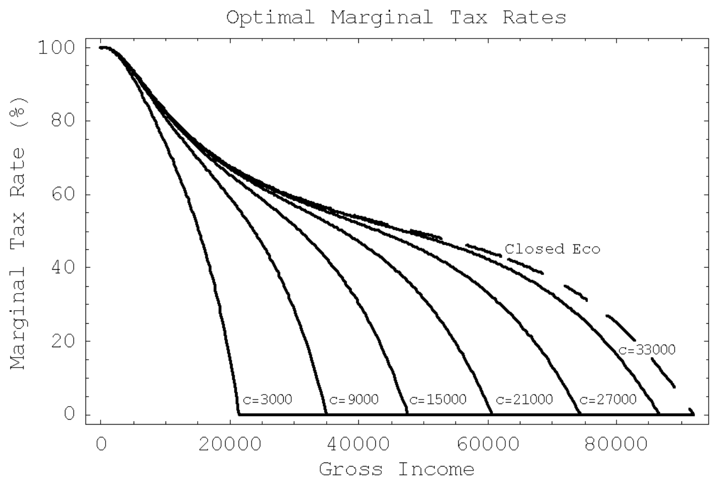

Figure 5: Optimal Marginal Tax Rates with Respect to Migration Costs (in euros). Rawls, $e = 0.2$, $c(\theta)$ constant.

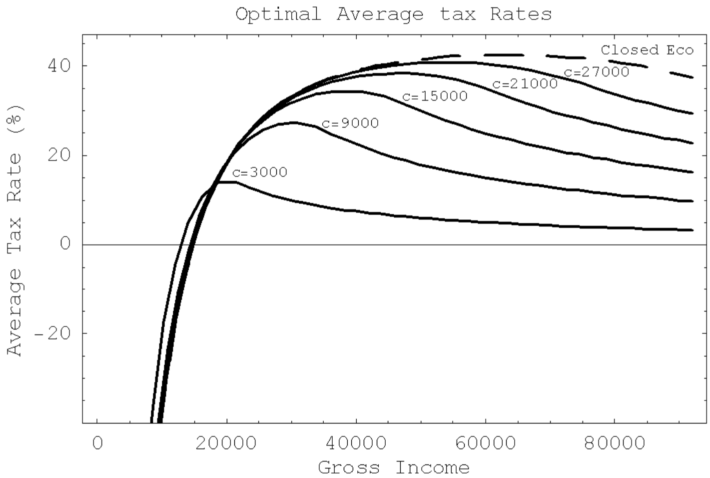

Figure 6: Optimal Average Tax Rates in A with Respect to Gross Income and Migration Costs (in euros). Rawls, $e = 0.2$, $c(\theta)$ constant.

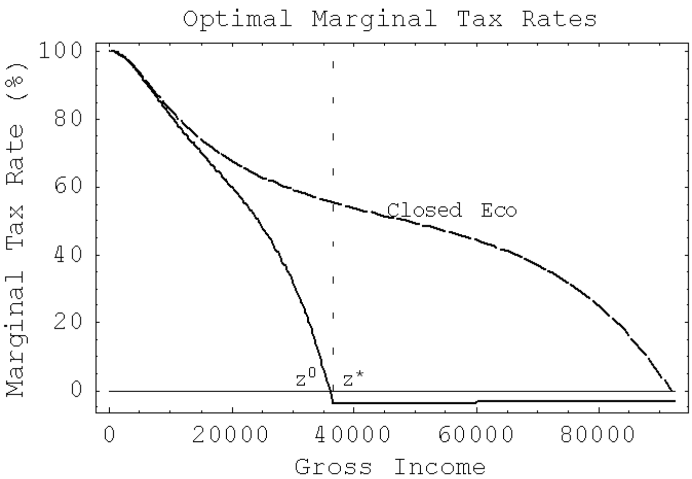

Figure 7: Optimal Marginal Tax Rates with respect to Gross Income (in euros). Rawls, $e = 0.2$, $c(\theta)$ decreasing.

We have introduced type-dependent participation constraints in the optimal income tax problem à la Mirrlees to address this issue. Three different social welfare criteria have been distinguished since it is not obvious to specify the set of agents whose welfare is to count when individuals can vote with their feet.

In the first-best, we have established that the same optimal allocation is obtained for the three social criteria. It is therefore socially optimal to implement a Mercantilist income tax scheme ensuring that the highly skilled individuals threatening to emigrate remain in A. A curse of the middle-skilled workers replaces the curse of the highly skilled workers when individual mobility matters. First, taxes levied on the highly skilled individuals are reduced. Second, the fiscal rent is extracted to the maximum from the most productive individuals among those who are not talented enough to threaten to emigrate. However, since it is not possible to tax the middle-skilled workers at will because of the participation constraints, there is less space for redistribution in favour of the low-skilled individuals.

In the second best, we have extended Mirrlees's and Diamond's formulae to take the threat of migration into account. The optimal marginal income tax rates have to be adjusted in order to obtain average tax rates compatible with the participation constraints. This goes against progressivity even before gross income levels from which individuals hesitate to leave the country. The importance of the average tax rates may be viewed as providing a theoretical basis for the idea of a "tax shield". The simulations on French data suggest that such a tax shield should depend on gross income. In addition, we have examined whether it is not too much expensive in terms of social welfare to prevent emigration of all highly skilled workers. The interaction between the incentive compatibility conditions and the participation constraints can give rise to countervailing incentives, which induce an indirect cost of the presence in A of the individuals with the highest outside options. It is socially optimal to implement a tax schedule such that the highest skilled individuals initially living in A choose to emigrate, when this indirect cost prevails.

A FIRST-BEST RESULTS

Proof of Proposition 1. Let π' and γ be the Lagrange multipliers of (6) and γ respectively. Under Assumption 1, the solution satisfies $x > 0$ and $l < \bar{l}$. The first-order conditions are

$$\left(\phi'_\rho + \pi'\right) U'_x = \gamma \text{ and } \left(\phi'_\rho + \pi'\right) U'_l = -\gamma\theta. \tag{36}$$

The complementarity slackness conditions corresponding to the participation constraints is

$$\pi' \geq 0;\ U(x,l) - V_B + c \geq 0;\ \pi' \left[U(x,l) - V_B + c\right] = 0 \text{ for all } \theta \in \left[\underline{\theta}, \overline{\theta}\right]. \tag{37}$$

Since $\phi'_\rho > 0$, it follows from (36) and (37) that $\gamma > 0$. The second-order conditions, equivalent to $U''_{xx} < 0$ and $U''_{xx}U''_{ll} - U'^2_{xl} > 0$ (since $\phi' > 0$ and $\pi' \geq 0$), are satisfied by strict concavity of U (Assumption 1).

Lemma 1. *Let Assumption 2 hold and J be a non-degenerate interval where $\pi' \equiv 0$. Then for all $\theta \in J$, (i.) $V'_A(\theta) < 0$ when $\rho < \infty$, (ii.) $V'_A(\theta) = 0$ when $\rho \to \infty$.*

The proof of the lemma is omitted (see Mirrlees (1974, Proposition, p. 251)). The proof is completed in two steps.

Step 1: Let Assumptions 2 and 3 hold. Then, $\exists\widetilde{\theta}: \left(\pi'\left(\widetilde{\theta}\right)>0\right) \Rightarrow \left(\pi'(\theta)>0,\ \forall\theta>\widetilde{\theta}\right)$. By (36), $\pi'(\theta)=\gamma/U_x'(x_A(\theta),l_A(\theta))-\phi_\rho'(U(x_A(\theta),l_A(\theta)))$, which is continuous in θ on $\left[\underline{\theta},\overline{\theta}\right]$. By continuity of π', there is an interval $[\widetilde{\theta},\theta'[$, with $\theta'>\widetilde{\theta}$, where $\pi'(\theta)>0$. Hence, by (37),

$$\forall\theta\in[\widetilde{\theta},\theta'[: V_A(\theta)=V_B(\theta)-c(\theta). \tag{38}$$

In addition, π' cannot differ from zero at an isolated θ since it is continuous; it is therefore sufficient to prove that there is no $J\subset[\widetilde{\theta},\overline{\theta}]$. We proceed by contradiction. Assume $J\subset[\widetilde{\theta},\overline{\theta}]$ exists. By Lemma 1, $V_A'\leq 0$ on J. In addition, under Assumption 3, $V_B'-c'>0$ on J. Hence V_A decreases and V_B-c strictly increases in θ on J. It thus follows from (38) that: $V_A(\theta)<V_B(\theta)-c(\theta)$ for all $\theta\in J$, which contradicts (6).

Step 2: Let Assumptions 2 and 3 hold. Then, θ^* exists if and only if $V_A^C\left(\overline{\theta}\right)<V_B\left(\overline{\theta}\right)-c\left(\overline{\theta}\right)$.

We have $\theta^*=\inf\left\{\theta\in\left[\underline{\theta},\overline{\theta}\right]:\pi'(\theta)>0\right\}$ and $\pi'\geq 0$ by (37).

(a) $\left[\theta^* \text{ exists} \Rightarrow V_A^C\left(\overline{\theta}\right)<V_B\left(\overline{\theta}\right)-c\left(\overline{\theta}\right)\right]$. We prove the contraposed statement $[V_A^C\left(\overline{\theta}\right)\geq V_B\left(\overline{\theta}\right)-c\left(\overline{\theta}\right)\Rightarrow\theta^*$ does not exist$]$. By Lemma 1, $V_A^C(\theta)$ is non-increasing in θ for any given $\rho\geq 0$. In addition, under Assumption 3, $V_B'-c'>0$. Hence, $V_A^C\left(\overline{\theta}\right)\geq V_B\left(\overline{\theta}\right)-c\left(\overline{\theta}\right)\Rightarrow V_A^C(\theta)\geq V_B\left(\overline{\theta}\right)-c\left(\overline{\theta}\right),\ \forall\theta\leq\overline{\theta}$, and thus $V_A(\theta)=V_A^C(\theta),\ \forall\theta\leq\overline{\theta}$. Consequently, $\left\{\theta\in\left[\underline{\theta},\overline{\theta}\right]:\pi'(\theta)>0\right\}=\emptyset$ so that θ^* does not exist.

(b) $\left[V_A^C\left(\overline{\theta}\right)<V_B\left(\overline{\theta}\right)-c\left(\overline{\theta}\right)\Rightarrow\theta^* \text{ exists}\right]$. We proceed by contradiction. Assume θ^* does not exist. Then, $\pi'(\theta)=0,\ \forall\theta\leq\overline{\theta}$ and $V_A(\theta)\equiv V_A^C(\theta)$. Consequently, $V_A^C\left(\overline{\theta}\right)<V_B\left(\overline{\theta}\right)-c\left(\overline{\theta}\right)\Leftrightarrow V_A\left(\overline{\theta}\right)<V_B\left(\overline{\theta}\right)-c\left(\overline{\theta}\right)$. A contradiction because of (6). □

Proof of Proposition 2. In this proof, we denote by $V_A(\theta;\rho)$ and $\pi'(\theta;\rho)$ the values of $V_A(\theta)$ and $\pi'(\theta)$ for a given ρ.

National Point of View. We proceed by contradiction. Assume $\widehat{\theta}<\overline{\theta}$ is socially optimal. It is possible to give the θ-individuals, with $\theta>\widehat{\theta}$, their laissez-faire utility V_B in A. Since $c>0$ and $\phi_\rho'>0$, we have $\phi_\rho(V_B-c)<\phi_\rho(V_B)$ and thus

$$\int_{\underline{\theta}}^{\widehat{\theta}}\phi_\rho(V_A(.;\rho))\,dF+\int_{\widehat{\theta}}^{\overline{\theta}}\phi_\rho(V_B)\,dF>\int_{\underline{\theta}}^{\widehat{\theta}}\phi_\rho(V_A(.;\rho))\,dF+\int_{\widehat{\theta}}^{\overline{\theta}}\phi_\rho(V_B-c)\,dF, \tag{39}$$

the RHS of which is $W_{A,\rho}^N\left(\widehat{\theta}\right)$. But the allocation on the LHS is also feasible. A contradiction.

Resident point of view. We proceed in two steps.

Step 1: $\phi_\rho\left(V_B\left(\widehat{\theta}\right)\right)>\phi_\rho\left(V_A(\underline{\theta};\rho)\right),\forall\rho$. Let $V_B\left(\widehat{\theta}\right)\leq V_A(\underline{\theta};\infty)$. Then, by (37), $\pi'(\theta;\infty)=0,\ \forall\theta\leq\widehat{\theta}$. By Lemma 1, $V_A(\underline{\theta};\infty)=V_A(\theta;\infty),\ \forall\theta\leq\widehat{\theta}$. Since $V_B'>0$, $V_B(\theta)<V_B\left(\widehat{\theta}\right)$, $\forall\theta\leq\widehat{\theta}$. Hence, $V_B(\theta)<V_A(\underline{\theta};\infty),\ \forall\theta<\widehat{\theta}$. This contradicts the fact that the laissez-faire is Pareto-efficient. Consequently,

$$V_B\left(\widehat{\theta}\right)>V_A(\underline{\theta};\infty). \tag{40}$$

In addition, A's government maximizes $V_A(\underline{\theta};\rho)$ when $\rho \to \infty$. Hence,

$$V_A(\underline{\theta};\infty) \geq V_A(\underline{\theta};\rho), \; \forall\rho. \tag{41}$$

By (40) and (41), $V_B\left(\widehat{\theta}\right) > V_A(\underline{\theta};\rho), \; \forall\rho$, which completes the step since $\phi'_\rho > 0$.

Step 2: $\widehat{\theta} = \overline{\theta}$ at the optimum. We proceed by contradiction. Assume $\widehat{\theta} < \overline{\theta}$ is optimal. Proposition 1 can be applied for the given $\widehat{\theta}$. Hence, $\phi_\rho(V_A(\underline{\theta};\rho)) \geq \phi_\rho(V_A(\theta;\rho))$ for all $\underline{\theta} \leq \theta < \theta^*$. Using Step 1,

$$\phi_\rho\left(V_B\left(\widehat{\theta}\right)\right) \geq \phi_\rho(V_A(\rho;\theta)), \; \forall\underline{\theta} \leq \theta < \theta^*. \tag{42}$$

In addition, $V_A = V_B - c$, $\forall\theta > \theta^*$. Since $\phi'_\rho > 0$ and $c > 0$,

$$\phi_\rho\left(V_B\left(\widehat{\theta}\right)\right) > \phi_\rho(V_B(\theta) - c(\theta)), \; \forall\theta^* < \theta \leq \widehat{\theta}. \tag{43}$$

Using (42) and (43),

$$\frac{1}{F\left(\widehat{\theta}\right)}\int_{\underline{\theta}}^{\widehat{\theta}} \phi_\rho(V_A(\rho;\theta))\,dF(\theta) < \phi_\rho\left(V_B\left(\widehat{\theta}\right)\right), \tag{44}$$

the LHS of which is $W^R_{A,\rho}\left(\widehat{\theta}\right)$. It is always feasible to give the θ-individuals with $\theta \geq \widehat{\theta}$ the utility $V_B(\theta)$ in A (or a bit less). By (44), this increases social welfare. A contradiction. □

B Second-Best Results

Proof of Property 4. $V_B(\theta) = \theta^{1+e}/(1+e)$. Let $\theta \in I$. By (11), $V'_A(\theta) = l_A^{1+1/e}/\theta = V'_B = \theta^e$ and thus $z_A(\theta) = \theta^e$. Hence, by (17),

$$T(z_A)|_I = c(\theta) + \left(1 - \frac{e}{1+e}\right)\theta^{1+e} - \frac{\theta^{1+e}}{1+e} = c(\theta). \tag{45}$$

□

B.1 Mercantilist Point of View

Proof of Proposition 3. We first state necessary conditions for the optimization problem. Since $\underline{\theta} = 0$ and $V_A = \theta l_A - T - v(l_A)$, the government chooses l_A and V_A to maximize

$$\overline{G} = \int_{\underline{\theta}}^{\overline{\theta}} [\theta l_A(\theta) - V_A(\theta) - v(l_A(\theta))]\,dF(\theta) \text{ s.t. (FOIC) and (6),}$$

l_A is control variable; V_A is state variable with adjoint variable ι. The Hamiltonian and Lagrangian are respectively

$$\begin{aligned} H^M &= \theta l_A - V_A - v(l_A) + \iota\frac{l_A}{\theta}v'(l_A), \\ L^M &= H^M + \pi'(V_A - V_B + c). \end{aligned}$$

Necessary conditions for a maximum include[10]:

$$\partial H^M/\partial l_A = 0 \Leftrightarrow \left(\theta - v'\right) f + \iota \left(\frac{v'}{\theta} + \frac{l_A}{\theta} v''\right) = 0, \tag{46}$$

$$\partial H^M/\partial V_A = -\iota'(\theta) \Leftrightarrow \iota'(\theta) = f - \pi', \tag{47}$$

$$\iota\left(\overline{\theta}\right) \geq 0 \ \left(= 0 \text{ when } V_A\left(\overline{\theta}\right) > V_B\left(\overline{\theta}\right) - c\left(\overline{\theta}\right)\right), \tag{48}$$

$$\iota\left(\underline{\theta}\right) \leq 0 \ \left(= 0 \text{ when } V_A\left(\underline{\theta}\right) > V_B\left(\underline{\theta}\right) - c\left(\underline{\theta}\right)\right), \tag{49}$$

$$\pi'(\theta) \geq 0,\ V_A(\theta) - V_B(\theta) + c(\theta) \geq 0,\ \pi'(\theta)\left[V_A(\theta) - V_B(\theta) + c(\theta)\right] = 0. \tag{50}$$

Since $v'/\theta = 1 - T'$ and $e^H(\theta) = v'/\left(l_A v''\right)$, rearranging (46) yields

$$\frac{T'}{1-T'} = -\frac{\iota}{\theta f}\left(1 + \frac{1}{e^H(\theta)}\right). \tag{51}$$

Integration of (47) between θ and $\overline{\theta}$ gives

$$\iota(\theta) = \iota\left(\overline{\theta}\right) - \int_\theta^{\overline{\theta}} \iota'(\theta)\, d\theta = \iota\left(\overline{\theta}\right) - \int_\theta^{\overline{\theta}} \left[f - \pi'\right] d\theta = \iota\left(\overline{\theta}\right) - 1 + F(\theta) + \int_\theta^{\overline{\theta}} \pi' d\theta, \tag{52}$$

which is plugged into (51). The proof is completed in two steps.

Step 1: $\int_\theta^{\overline{\theta}} \pi' d\theta = \int_{\theta \in \left[\theta, \overline{\theta}\right] \cap \Theta^{PC}} f(\theta)\, d\theta = \mu\left(\left[\theta, \overline{\theta}\right] \cap \Theta^{PC}\right)$.

Consider any non-empty open interval in $\left[\underline{\theta}, \overline{\theta}\right]$ where $\pi' > 0$. By Property 4, $T = \overline{c}$ and $T' = 0$. Hence, $\iota(\theta) = 0$ by (51) so that $\iota'(\theta) = 0$. Therefore, (47) reads $\pi'(\theta) = f(\theta)$. The equality above holds because (PC) cannot be active at isolated points.

Step 2: $\iota\left(\overline{\theta}\right) = 0$.

If (PC) is inactive at $\overline{\theta}$, $\iota\left(\overline{\theta}\right) = 0$ because of (48). If (PC) is active at $\overline{\theta}$, there is an $\varepsilon > 0$ such that it is active on $I_{\varepsilon>0} = \left(\overline{\theta} - \varepsilon, \overline{\theta}\right]$. By Property 4, $T' = 0$ on I_ε. Hence, by (51), $\iota\left(\overline{\theta}\right) = 0$. $\square$

Proof of Proposition 4. We proceed in two steps.

Step 1: Necessary conditions. z_A is control variable; V_A and $R(\theta) := \int_{\underline{\theta}}^{\theta} T\left(z_A(\tau)\right) dF(\tau)$ are state variables. Since $T := z_A - x_A$, Leibnitz's rule yields $R'(\theta) = \left(z_A\left(\overline{\theta}\right) - x_A(\theta)\right) f(\theta)$. The isoperimetric contraint (TR) is taken into account through R' and the boundary conditions $R\left(\underline{\theta}\right) = 0$ and $R\left(\overline{\theta}\right) = \overline{G}$. It is not necessary to take x_A explicitly into account since it is uniquely determined by V_A and $z_A : x_A = h(V_A, z_A; \theta)$, with $\partial x_A/\partial V_A = 1/u'_x$ and $\partial x_A/\partial z_A = s$. The Hamiltonian and Lagrangian are respectively

$$\begin{aligned} H^M &= \phi_\rho(V_A) f + \iota u'_\theta + \gamma(z_A - x_A) f, \\ L^M &= H^M + \pi'(V_A - V_B + c). \end{aligned}$$

[10]The necessary conditions are often stated for state variables which are fixed at the initial point, which is not the case presently. We have used Seierstad and Sydsaeter (1987, Theorem 5, pp. 185, Eq. 30b) (referred to as (S-S)) to obtain (49). This remark applies to the other proofs in the paper.

We apply Theorem 2 in Seierstad and Sydsaeter (1987, p. 332-333 and p. 335)[11]. We call junction points θ_j $(j = 1, ..., N)$ points where (PC) becomes or ceases to be active as well as $\underline{\theta}$ and $\widehat{\theta}$. As $\partial u'_\theta/\partial z_A = u''_{\theta z} + s u''_{\theta x} = -u'_x s'_\theta$, and $\partial u'_\theta/\partial V_A = u''_{\theta x}/u'_x$, necessary conditions include:

$$\begin{aligned}
\partial H^M/\partial z_A &= 0 \Leftrightarrow \iota u'_x s'_\theta - \gamma(1-s) f = 0, && (53)\\
\partial L^M/\partial V_A &= -\iota' \Leftrightarrow \iota'(\theta) = -\phi'_\rho(V_A) f - \iota u''_{\theta x}/u'_x - \pi' + \gamma f/u'_x, && (54)\\
\partial L^M/\partial R &= -\gamma' \Leftrightarrow \gamma' = 0, && (55)\\
& \quad (48),\ (49),\ (50), && (56)\\
\iota\left(\theta_j^-\right) - \iota\left(\theta_j^+\right) &= \pi\left(\theta_j^+\right) - \pi\left(\theta_j^-\right) \geq \ (= 0 \text{ if } V_A(\theta_j) - V_B(\theta_j) + c(\theta_j) > 0). && (57)
\end{aligned}$$

(53)-(56) are standard. (57) is a jump condition.

As $s = 1 - T'$, $T' = \iota u'_x s'_\theta / (\gamma f)$ by (53). In addition, using basic calculus, $\left[1 + e^M(\theta)\right]/e^H(\theta) = -\theta s'_\theta / s$. Hence,

$$\frac{T'}{1-T'} = -\frac{\iota u'_x}{\gamma \theta f}\frac{1+e^M(\theta)}{e^H(\theta)} = -\frac{\iota u'_x}{\gamma(1-F(\theta))}\frac{1+e^M(\theta)}{e^H(\theta)}\frac{1-F(\theta)}{\theta f(\theta)}. \tag{58}$$

If $(.;\tau)$ means evaluation at $(x_A(\tau), z_A(\tau);\tau)$, integrating (54) between θ and $\overline{\theta}$ yields

$$\iota(\theta) = \iota\left(\overline{\theta}\right) + \int_\theta^{\overline{\theta}} \left(\phi'_\rho(V_A(\tau)) f(\tau) + \pi'(\tau) - \frac{\gamma f(\tau)}{u'_x(.;\tau)}\right) \widetilde{\Psi}_{\theta\tau} d\tau, \tag{59}$$

with $\widetilde{\Psi}_{\theta\tau} := \exp \int_\theta^\tau u''_{\theta x}(.;\tau')/u'_x(.;\tau')\, d\tau'$. By (48), (49), and (59),

$$\iota\left(\underline{\theta}\right) - \iota\left(\overline{\theta}\right) = \int_{\underline{\theta}}^{\overline{\theta}} \left[\phi'_\rho(V_A(\tau)) f(\tau) + \pi'(\tau) - \frac{\gamma f(\tau)}{u'_x(.;\tau)}\right] \widetilde{\Psi}_{\theta\tau} d\tau \leq 0. \tag{60}$$

Rearranging the latter inequality, we check that $\gamma > 0$.

The following relation has been proved by Saez (2001, p. 227):

$$\Psi_{\theta\tau} := \frac{u'_x(.;\theta)}{u'_x(.;\tau)}\widetilde{\Psi}_{\theta\tau} = \exp \int_\theta^\tau \left(1 - \frac{e^M(\tau')}{e^H(\tau')}\right)\frac{z'_A(\tau')}{z_A(\tau')} d\tau'. \tag{61}$$

We use it and (59) to get

$$-\frac{\iota(\theta) u'_x(.;\theta)}{\gamma} = \int_\theta^{\overline{\theta}} \left[1 - \left(\phi'_\rho(V_A(\tau)) + \frac{\pi'(\tau)}{f(\theta)}\right)\frac{u'_x(.;\tau)}{\gamma}\right] \Psi_{\theta\tau} dF(\tau) - \frac{\iota\left(\overline{\theta}\right) u'_x(.;\theta)}{\gamma}, \tag{62}$$

which is plugged into (58) to complete the proof. □

Proof of Corollary 1. We have $u_x = 1$, $e^H = e^M$, and thus $\Psi_{\theta\tau} = 1$. Hence, $A = 1 + 1/e^H$ and

$$B(\theta) = 1 - \frac{1}{1-F(\theta)} \int_\theta^{\overline{\theta}} \frac{\phi'_\rho(V_A(\tau)) f(\theta) + \pi'(\tau)}{\gamma} d\tau - \frac{\iota\left(\overline{\theta}\right)}{\gamma(1-F(\theta))}. \qquad \square$$

[11] We provide further details about how the Theorem we referred to is applied. Since the adjoint variables are assumed to have a finite number of jump discontinuities and be $\mathscr{C}^1$ elsewhere, the "almost necessary conditions" (p. 335) are in fact necessary. In the theorem, $q' = \lambda$, which is piecewise continuous, is our π'. Hence, q is our π and λ our π'. Consequently, their $\beta^k = \pi\left(\tau_k^+\right) - \pi\left(\tau_k^-\right)$ (note that $\gamma(\theta)$ is constant and $\partial\left(V_A - V_B + c\right)/\partial R = 0$). We then employ their Eq. (5.37) to get (57).

B.2 National and Resident Points of view

We prove Properties 7–9 in three steps. Both first steps derive restrictions on the solution in $\widehat{\theta}$ to Problem 3. The third one establishes the properties.

Step 1: Resident point of view. We state necessary conditions for Problem 4. $\zeta_A := z'_A$ is control variable; z_A, V_A and R are state variables; η, ι and γ are adjoint variables. (SOIC) is transformed into $g(\zeta_A) \geq 0$ to avoid dealing with singular solutions, where g is a $\mathscr{C}^2$-function s.t. $g' > 0$ and $g(0) = 0$. The Hamiltonian and Lagrangian are respectively

$$\begin{aligned} H^R &= \phi_\rho (V_A) f + \eta \zeta_A + \iota u'_\theta + \gamma (z_A - x_A) f, \\ L^R &= H^R + \pi' [V_A - V_B + c] + \kappa g(\zeta_A). \end{aligned}$$

Under our regularity conditions, a solution to Problem 4 must satisfy[12]:

$$\partial L^R / \partial \zeta_A = 0 \Leftrightarrow \eta + \kappa g'(\zeta_A) = 0, \tag{63}$$
$$\eta' = -\partial L^R / \partial z_A \Leftrightarrow \eta' = \iota u'_x s'_\theta - \gamma (1-s) f, \tag{64}$$
$$\iota' = -\partial L^R / \partial V_A \Leftrightarrow \iota' = -\phi'_\rho f - \iota u''_{\theta x} / u'_x + \gamma f / u'_x - \pi', \tag{65}$$
$$\gamma' = -\partial L^R / \partial R \Leftrightarrow \gamma' = 0, \tag{66}$$
$$\pi' \geq 0,\ V_A - V_B + c \geq 0,\ \pi' (V_A - V_B + c) = 0, \tag{67}$$
$$\kappa \geq 0,\ g(\zeta_A) \geq 0,\ \kappa g(\zeta_A) = 0, \tag{68}$$
$$\eta(\underline{\theta}) = \eta\left(\widehat{\theta}\right) = 0, \tag{69}$$
$$\iota(\underline{\theta}) \leq 0 \ (= 0 \text{ if } V_A(\underline{\theta}) - V_B(\underline{\theta}) + c(\underline{\theta}) > 0), \tag{70}$$
$$\iota\left(\widehat{\theta}\right) \geq 0 \ \left(= 0 \text{ if } V_A\left(\widehat{\theta}\right) - V_B\left(\widehat{\theta}\right) + c\left(\widehat{\theta}\right) > 0\right), \tag{71}$$
$$\iota\left(\theta_j^-\right) - \iota\left(\theta_j^+\right) = \pi\left(\theta_j^+\right) - \pi\left(\theta_j^-\right) \geq (= 0 \text{ if } V_A(\theta_j) > V_B(\theta_j) - c(\theta_j)). \tag{72}$$

It is easy to check that $\gamma > 0$. We note that

$$\eta\left(\widehat{\theta}^-\right) \zeta_A\left(\widehat{\theta}\right) = 0. \tag{73}$$

Indeed, either there is bunching at $\widehat{\theta}$ and then $\zeta_A\left(\widehat{\theta}\right) = 0$ or there is no bunching. In the later case, $\zeta_A\left(\widehat{\theta}\right) > 0$ and ζ_A continuous at $\widehat{\theta} \Rightarrow \exists \widetilde{\theta} < \widehat{\theta}$ s.t. $\forall \theta \in]\widetilde{\theta}, \widehat{\theta}]$, $\zeta_A(\theta) > 0$ and thus $\forall \theta \in]\widetilde{\theta}, \widehat{\theta}]$, $g(\zeta_A(\theta)) > 0$. Therefore, $\forall \theta \in]\widetilde{\theta}, \widehat{\theta}]$, $\kappa(\theta) = 0$ by (68) so that $\forall \theta \in]\widetilde{\theta}, \widehat{\theta}]$, $\eta(\theta) = 0$ by (63). We define

$$\vartheta^R\left(\widehat{\theta}\right) := H^R\left(\widehat{\theta}^-\right) - f\left(\widehat{\theta}\right) \mathscr{W}^R_{A,\rho}\left(\widehat{\theta}\right) + \left[\pi\left(\widehat{\theta}\right) - \pi\left(\widehat{\theta}^-\right)\right] \left[V'_A\left(\widehat{\theta}\right) - V'_B\left(\widehat{\theta}\right) + c'\left(\widehat{\theta}\right)\right]. \tag{74}$$

[12](S-S), Note 6, p. 375. We note that η and γ are continuous because $\partial (V_A - V_B + c) / \partial z_A = 0$ and $\partial (V_A - V_B + c) / \partial R = 0$.

If $\widehat{\theta}=\widehat{\theta}^{N}$, it must be :

$$\vartheta^{i}\left(\widehat{\theta}\right)\leq 0 \text{ if } \underline{\theta}\leq\widehat{\theta}^{i}<\overline{\theta},\ \vartheta^{i}\left(\widehat{\theta}\right)\geq 0 \text{ if } \widehat{\theta}^{i}=\overline{\theta}, \tag{75}$$

where $i=N$. Since $\widehat{\theta}^{i}$ cannot be $\underline{\theta}$[13],

$$\begin{aligned}\vartheta^{i}\left(\widehat{\theta}\right) &> 0,\ \forall\widehat{\theta}>\underline{\theta}\Rightarrow\widehat{\theta}^{i}=\overline{\theta}, &(76)\\ \vartheta^{i}\left(\overline{\theta}\right) &< 0\Rightarrow\widehat{\theta}^{i}<\overline{\theta}, &(77)\end{aligned}$$

where $i=N$.

It remains to examine the sign of ϑ^{R}. We use (73), the continuity of x_A, z_A, f, η, γ, V_A, and the definition of T to get

$$\vartheta^{R}\left(\widehat{\theta}\right)=\left[\phi_{\rho}\left(V_A\right)-\mathscr{W}^{R}_{A,\rho}\right]f+\iota\left(\widehat{\theta}^{-}\right)V_A'+\gamma T\left(z_A\right)f+\left[\pi-\pi\left(\widehat{\theta}^{-}\right)\right]\left[V_A'-V_B'+c'\right], \tag{78}$$

where all functions, here and in the following, are evaluated at $\widehat{\theta}$ except otherwise stated.

Step 2: National Point of View. As previously, we state necessary conditions for Problem 4. The control and state variables are the same as from the Resident point of view. The Hamiltonian and Lagrangian are respectively

$$\begin{aligned}H^{N} &= \phi_{\rho}\left(V_A\right)f+\eta\zeta_A+\iota u_{\theta}'+\gamma\left(z_A-x_A\right)f,\\ L^{N} &= H^{R}+\pi'\left[V_A-V_B+c\right]+\kappa g\left(\zeta_A\right).\end{aligned}$$

A solution to Problem 4 must also satisfy (63)–(71)[14]. We define[15]

$$\vartheta^{N}\left(\widehat{\theta}\right)=H^{N}\left(\widehat{\theta}^{-}\right)+\left[\pi-\pi\left(\widehat{\theta}^{-}\right)\right]\left[V_A'-V_B'+c'\right]-\phi_{\rho}\left(V_B-c\right)f. \tag{79}$$

If $\widehat{\theta}=\widehat{\theta}^{N}$, it must be (75), in which $i=N$. Rearranging as in Step 1,

$$\vartheta^{N}\left(\widehat{\theta}\right)=\left[\phi_{\rho}\left(V_A\right)-\phi_{\rho}\left(V_B-c\right)\right]f+\gamma T\left(z_A\right)f+\iota\left(\widehat{\theta}^{-}\right)V_A'+\left[\pi-\pi\left(\widehat{\theta}^{-}\right)\right]\left[V_A'-V_B'+c'\right]. \tag{80}$$

Step 3: Properties.

Proof of Property 7. We want to establish that: $\vartheta^{i}\left(\widehat{\theta}\right)>0,\ \forall\widehat{\theta}>\underline{\theta}$. Consider the allocation solution to Problem 4 for $\widehat{\theta}\in(\underline{\theta},\overline{\theta}]$ and the corresponding (γ,ι,π,T,V_A). (a) Since ι is continuous at $\widehat{\theta}$, $\iota\left(\widehat{\theta}^{-}\right)=\iota\left(\widehat{\theta}\right)\geq 0$ by (71) and $\pi\left(\widehat{\theta}\right)=\pi\left(\widehat{\theta}^{-}\right)$. Hence, the last terms of (78)

[13] Assume $\widehat{\theta}^{i}=\underline{\theta}$ at the optimum. In this case, A's tax policy is the laissez-faire. All individuals with productivity greater than $\underline{\theta}$ are in B where they receive their reservation utility. Since $c>0$, they would be better in A. Since making them relocated to A is feasible and does not alter the utility level of the $\underline{\theta}$ individuals, $\widehat{\theta}^{i}=\underline{\theta}$ does not yield a Pareto optimum. A contradiction.

[14] Indeed, since $\widehat{\theta}$ is given, the social welfare of the nationals staying in B is constant.

[15] (S-S), Theorem 16, p. 397-398. The last term in (79) is due to the scrap value function.

and (80) are nil. (b) By (FOIC), $\iota\left(\widehat{\theta}^{-}\right)V_A'\left(\widehat{\theta}\right)\geq 0$. (c) Using Property 6 and the assumption made, $\gamma T\left(z_A\left(\widehat{\theta}\right)\right)>0$. (d) As regards the resident point of view, it remains to prove that $\phi_\rho\left(V_A\right)-\mathscr{W}^R_{A,\rho}\geq 0$. As $\phi'>0$, (FOIC) yields:

$$\phi_\rho\left(V_A\left(\widehat{\theta}\right)\right)\geq\frac{1}{F\left(\widehat{\theta}\right)}\int_{\underline{\theta}}^{\widehat{\theta}}\phi_\rho\left(V_A\right)dF\left(\theta\right)\equiv\mathscr{W}^R_{A,\rho}\left(\widehat{\theta}\right). \tag{81}$$

(e) From the national point of view, $V_A\left(\theta\right)\geq V_B\left(\theta\right)-c\left(\theta\right),\ \forall\theta\leq\widehat{\theta}$, and thus $\phi_\rho\left(V_A\left(\theta\right)\right)\geq\phi_\rho\left(V_B\left(\theta\right)-c\left(\theta\right)\right),\ \forall\theta\leq\widehat{\theta}$, because $\phi_\rho'>0$. (f) Finally, by (78) and (80), $\vartheta^i\left(\widehat{\theta}\right)>0$ for all $\widehat{\theta}>\underline{\theta}$, and thus $\widehat{\theta}^i=\overline{\theta}$ by (76), $i=\{N,R\}$. □

Proof of Property 8. Consider the allocation solution to Problem 4 for $\widehat{\theta}\in(\underline{\theta},\overline{\theta}]$ and the corresponding (γ,ι,π,T,V_A). There is a non-degenerate interval containing $\widehat{\theta}$ where (PC) is active. Hence $V_A'=V_B'-c'$ at $\widehat{\theta}$. Moreover, by Property 1, $T\left(z_A\left(\widehat{\theta}\right)\right)>0$. We then use (d)–(f) in the proof of Property 7. □

Proof of Property 9. We employ (78) and (80) to write $\vartheta^i\left(\overline{\theta}\right),\ i=\{R,N\}$. We then use (77), in which $V_A=V_B-c$ and $V_A'<V_B'-c'$ at $\overline{\theta}$. We replace $\mathscr{W}^i_{A,\rho}\left(\overline{\theta}\right)$ by $W^M_{A,\rho}$. □

C NUMERICAL SIMULATIONS

Preferences are given by (15) and (19). We consider the Mercantilist point of view and use sufficient conditions to provide numerical results. These conditions[16] allow jump discontinuities in ι. Our strategy is to look for a solution such that:

Assumption 5. *In $[\underline{\theta},\overline{\theta}]$, (i) ι is continuous and piecewise continuously differentiable; (ii) $\pi'\left(\theta\right)$ is piecewise continuous; (iii) z_A is strictly increasing.*

If we find such a solution, we will not have to take complex jump conditions and (SOIC) explicitly into account. Under Assumption 5, Problem 3 is an optimal control problem with one control variable l_A and two state variables, V_A and R, with adjoint variables ι and γ respectively. The Hamiltonian is $H=\phi_\rho\left(V_A\right)+\iota l_A^{1+1/e}/\theta+\gamma(\theta l_A-x_A)f$ and the Lagrangian is $L=H+\pi'\left(V_A-V_B+c\right)$.

Lemma 2. *Under Assumption 5, sufficient conditions for Problem 3 are:*

$$\partial H/\partial l_A=0\Leftrightarrow\iota\left(1+1/e\right)l_A^{1/e}/\theta^2+\gamma\left(1-l_A^{1/e}/\theta\right)f=0, \tag{82}$$

$$\partial L/\partial V_A=-\iota'\Leftrightarrow\iota'\left(\theta\right)=-\phi_\rho'\left(V_A\right)f+\gamma f-\pi', \tag{83}$$

$$H \text{ concave in } l_A \text{ and in } V_A,\ V_A-V_B+c \text{ quasi-concave in } V_A, \tag{84}$$

and (48)-(50).

[16](S-S), Theorem 1, pp. 317-318.

The conditions in (84) are satisfied. The quasi-concavity is obvious. In addition, since $s = 1 - T' = l_A^{1/e}/\theta$, (82) yields (82) yields (51) in which $e^H = e$. Hence, by (82),

$$\frac{\partial^2 H}{\partial l_A^2} = \frac{l_A^{1/e-1}}{\theta^2 e}\left[\iota\left(1+1/e\right) - \gamma\theta f\right] \leq 0 \Leftrightarrow -\frac{\iota\left(1+1/e\right)}{\gamma\theta f} \geq -1 \Leftrightarrow \frac{T'}{1-T'} \geq -1 \text{ for } \theta \neq 0. \quad (85)$$

As $T' < 1$ under Assumption 1, (85) is satisfied, which establishes the concavity of H in l_A. Finally, H is concave in V_A because $\partial^2 H/\partial V_A^2 = \phi''\left(V_A\right) f < 0$.

When $\rho \to \infty$, the reader can check that $\gamma = 1$ and $\phi'_\rho\left(V_A\right) = 0$. Therefore:

$$\frac{T'}{1-T'} = \frac{1+1/e}{\theta f}\left(1 - F\left(\theta\right) - \int_\theta^{\overline{\theta}} \pi'\left(\tau\right) d\tau - \iota\left(\overline{\theta}\right)\right) \text{ and } \iota' = f - \pi'. \quad (86)$$

In addition, we have $l_A\left(\theta\right) = \theta^e\left(1-T'\right)^e$, $V_B\left(\theta\right) - c\left(\theta\right) = \theta^{e+1}/\left(1+e\right) - c\left(\theta\right)$ and $V_B'\left(\theta\right) - c'\left(\theta\right) = \theta^e - c'\left(\theta\right)$. We choose $\underline{\theta} = 0$. In computations, one can proceed as follows.

$(1a.)$ Assume that Assumption 1 holds. That allows us to apply Lemma 2. $(1b.)$ Assume that, if $\Theta^{PC} \neq \emptyset$, θ^* is such that $\pi' > 0$ for all $\theta > \theta^*$. (2.) Choose an arbitrary value of θ^*. (3.) Use (17), (FOIC) and (PC) to derive z_A and $T\left(z_A\right)$ for $\theta > \theta^*$ and compute $T'\left(z_A\right) = \left(dT'/d\theta\right)/z_A'\left(\theta\right)$. The values of $\iota\left(\overline{\theta}\right)$ and $\int_\theta^{\overline{\theta}} \pi'\left(\tau\right) d\tau$ are then obtained from (86),

$$\iota\left(\overline{\theta}\right) = -\frac{T'\left(z_A\left(\overline{\theta}\right)\right)}{1-T'\left(z_A\left(\overline{\theta}\right)\right)}\frac{e\overline{\theta} f\left(\overline{\theta}\right)}{1+e}, \quad (87)$$

$$\int_\theta^{\overline{\theta}} \pi'\left(\tau\right) d\tau = 1 - F\left(\theta\right) - \frac{T'\left(z_A\left(\theta\right)\right)}{1-T'\left(z_A\left(\theta\right)\right)}\frac{e\theta f}{1+e} - \iota\left(\overline{\theta}\right). \quad (88)$$

Compute $\iota\left(\theta\right)$ for $\theta^* < \theta < \overline{\theta}$, $\iota\left(\theta\right) = -\frac{T'}{1-T'}\frac{\theta f}{1+1/e}$, derive $\iota'\left(\theta\right)$ and use (83) to get $\pi' = f - \iota'\left(\theta\right)$ for all $\theta \geq \theta^*$. (4.) The marginal tax rates are obtained from (86) for $\theta < \theta^*$. (5.) Compute l_A and derive $V_A\left(\underline{\theta}\right)$ from (TR). Indeed, since

$$V_A\left(\theta\right) = V_A\left(\underline{\theta}\right) + \int_{\underline{\theta}}^{\theta} l_A^{1+1/e}\left(\tau\right)/\tau d\tau, \quad (89)$$

by integration of (FOIC) and $T\left(\theta l_A\right) = \theta l_A - v\left(l_A\right) - V_A$,

$$\int_{\underline{\theta}}^{\overline{\theta}} T\left(z_A\right) f d\theta = \int_{\underline{\theta}}^{\overline{\theta}}\left[\theta l_A - v\left(l_A\right) - V_A\left(\underline{\theta}\right) - \int_{\underline{\theta}}^{\theta}\frac{l_A^{1+1/e}\left(\tau\right)}{\tau} d\tau\right] f d\theta.$$

Note that the LHS is $\overline{G}$ and apply Fubini's theorem to get

$$V_A\left(\underline{\theta}\right) = \int_{\underline{\theta}}^{\overline{\theta}}\left[\theta l_A f - \left(1 - F\left(\theta\right)\right)\frac{l_A^{1+1/e}}{\theta} - \frac{l_A^{1+1/e}}{1+1/e} f\right] d\theta - \overline{G}. \quad (90)$$

(6.) Employ (90) and (89) to compute $V_A\left(\theta\right)$. (7.) Check that $(1b.)$ and (6) are satisfied. If it is not the case, change the value of θ^* by an increment and start again. Otherwise, check that $(1a.)$ is satisfied. If it is the case, the candidate solution is an optimal solution.

References

BODIN, J. (1578): *Réponse Au Paradoxe de Monsieur de Malestroit*. Paris.

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.

CHANDER, P., AND L. WILDE (1998): "A General Characterization of Optimal Income Tax Enforcement," *Review of Economic Studies*, 65, 165–183.

D'AUTUME, A. (2000): "L'imposition optimale du revenu: une application au cas français," *Revue Française d'Economie*, 3, 3–63.

DGI (2005): "Les délocalisations de contribuables (personnes physiques), Bilan des départs recensés pour l'année 2003," Discussion paper, MINEFI.

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

EBERT, U. (1992): "A Reexamination of the Optimal Nonlinear Income Tax," *Journal of Public Economics*, 49(1), 47–73.

GUESNERIE, R., AND J. SEADE (1982): "Nonlinear Pricing in a Finite Economy," *Journal of Public Economics*, 57, 157–179.

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.

HINDRIKS, J. (1999): "The Consequences of Labour Mobility for Redistribution: Tax Vs. Transfert Competition," *Journal of Public Economics*, 74, 215–234.

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

JULLIEN, B. (2000): "Participation Constraints in Adverse Selection Models," *Journal of Economic Theory*, 93, 1–47.

LAFFONT, J.-J., AND D. MARTIMORT (2002): *The Theory of Incentives*. Princeton University Press.

LASLIER, J.-F., A. TRANNOY, AND K. VAN DER STRAETEN (2003): "Voting under Ignorance of Job Skills of Unemployed: The Overtaxation Bias," *Journal of Public Economics*, 87, 595–626.

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

LEWIS, T., AND D. SAPPINGTON (1989): "Countervailing Incentives in Agency Problems," *Journal of Economic Theory*, 49, 294–313.

MAGGI, G., AND A. RODRIGUEZ-CLARE (1995): "On Countervailing Incentives," *Journal of Economic Theory*, 66, 238–263.

MIRRLEES, J. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38(2), 175–208.

——— (1974): "Notes on welfare economics, information and uncertainty," in *Essays in economic behavior under uncertainty*, ed. by M. Balch, M. McFadden, and S. Wu, pp. 243–258. North Holland.

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

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

PAGE, F., AND P. MONTEIRO (2003): "Three Principles of Competitive Nonlinear Pricing," *Journal of Mathematical Economics*, 39, 63–109.

PARFIT, D. (1984): *Reasons and Persons*. Oxford University Press.

PIASER, G. (2003): "Labor Mobility and Income Tax Competition," CORE Discussion Paper 2003/6, UCL.

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

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

SEADE, J. (1982): "On the Sign of the Optimum Marginal Income Tax," *Review of Economic Studies*, 49, 637–643.

SEADE, J. K. (1977): "On the shape of optimal tax schedule," *Journal of Public Economics*, 7, 203–235.

SEIERSTAD, A., AND K. SYDSAETER (1987): *Optimal Control Theory with Economic Applications*. North-Holland, Amsterdam.

SIMULA, L., AND A. TRANNOY (2006a): "Optimal Linear Income Tax When Agents Vote with their Feet," *FinanzArchiv*.

——— (2006b): "Optimal Non-Linear Income Tax When Agents Vote with their Feet: An Illustration on French Data," *mimeo, IDEP-GREQAM*.

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.

TUOMALA, M. (1990): *Optimal Income Tax and Redistribution*. Clarendon Press, Oxford.

WEYMARK, J. (1986): "A Reduced-Form Optimal Nonlinear Income Tax Problem," *Journal of Public Economics*, 30, 199–217.

——— (1987): "Comparative Static Properties of Optimal Nonlinear Income Taxes," *Econometrica*, 55, 1165–1185.

WILSON, J. D. (1980): "The Effect of Potential Emigration on the Optimal Linear Income Tax," *Journal of Public Economics*, 14, 339–353.

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