# Nonlinear Taxes for Spatially Mobile Workers*

Charles Blackorby, Craig Brett† and Ana Cebreiro

February 2005

Blackorby: Department of Economics, University of Warwick, Coventry United Kingdom and GREQAM, Centre de la Vieille Charité, 2 rue de la Charité, 13002 Marseille, France.

Brett: Canada Research Chair in Canadian Public Policy, Department of Economics, Mount Allison University, Sackville, New Brunswick, Canada.

Cebreiro: Dotecon, 105-106 New Bond Street, London W1S 1DN, United Kingdom.

* Financial assistance from The Canada Research Chairs Programme and The Social Sciences and Humanities Research Council of Canada is gratefully acknowledged. We have benefitted from discussions with Ron Beattie and from the comments of seminar participants at McMaster University and Vanderbilt University.

† Corresponding author: Department of Economics, Mount Allison University, 144 Main Street, Sackville, New Brunswick, E4L 1A7, Canada. Email: cbrett@mta.ca. Phone: +1 (506) 364-2532. Fax: +1 (506) 364-2625.

**Abstract:**

This article examines optimal nonlinear taxes with worker mobility. We devise a model in which two qualitatively different types of workers choose to live in either of two jurisdictions. Residence decisions and labour-leisure choices are influenced by the tax schedules operating in the two regions. Within worker category, individuals may differ in attachment to one or the other of the regions. We show how worker mobility affects the optimal tax schedule, and address the spatial distribution of the population under optimal taxation. We derive a stringent necessary condition for the optimal tax solution to generate the laissez-faire distribution of workers.

Keywords: nonlinear income taxes, residential mobility, bunching, optimal redistributive taxation.

JEL Classification: D82, H21, H73, R23.

## 1. Introduction

Policy makers are conscious of the potential mobility of their workforce when deciding on tax policy. This is particularly so in the case of income taxation. It is sometimes argued that high rates of taxation can act as a catalyst for exit from a region. This exit may be cause for concern to policy makers, especially if exit is more pronounced among highly skilled workers. While it is easy to imagine adjustments to income tax schedules that might help a country retain a larger share of its skilled workforce, it is not clear *a priori* that such initiatives represent any improvement in the overall tax system. Indeed, the adjustments required to keep mobile workers at home might seriously undermine attempts at redistribution. This issue has been highlighted in the work of Bhagwati and Hamada (1989), but Mirrlees (1989) cautions against using this line of argument to justify low rates of taxation.

In this article, we offer a model of a nonlinear tax system in a way that allows a systematic analysis of optimal taxation when a "brain drain" is possible. The model we present is a variant of the Stiglitz (1982) optimal nonlinear income tax model with endogenous wages. There are two types of labour used in production, and two regions in which production may take place. Within each region the relative wages of the two types of labour are determined, in part, by the supply of both types of labour. Labour supply depends on the labour choice of each worker and on the number of workers of each type in a region. Following Mansoorian and Myers (1993), we allow workers to have a level of attachment to one or the other of the regions so that a region does not empty due to tiny differences in wage rates across borders. Instead, marginal changes in tax policy induce small changes in the composition of the workforces in each region. These changes affect both the size of the tax base and the relative wages in the two regions. The exact nature of these changes depends on how attachment to home varies across worker type and on the production technology. Because attachment to home is private information, some workers will enjoy a kind of location rent.

Despite the relative complexity of the economic environment, some clear conclusions about the effect of residential mobility on tax policy are derived. We show that inter-regional mobility *per se* does not provide an additional motivation for marginal distortions.

The choice among a finite set of places of residence is determined by a comparison of the total utility each place provides. Any desire a planner may have to change the composition of the population must be carried out by changing the distribution of total utility. Marginal distortions beyond those necessitated by imperfect information are an inefficient way to redistribute total utility. However, this does not mean that a change in residential mobility has no effect on the magnitude of marginal tax rates. Worker mobility has indirect effects through the labour supply. As in Saez (2002), the elasticity of the supply of each type of labour is affected by both changes in hours worked and changes in where work is carried out.

While the qualitative features of tax schedules are not greatly affected by the mobility of workers, the spatial distribution of population is affected by tax schedules. Unless regions have identical compositions of workers, any attempt to redistribute between workers of differing skills is *de facto* a redistribution across borders. This, in turn, induces a movement of workers across borders. We show that this phenomenon can lead to identical (in terms of productivity) workers in differing regions receiving different allocations of goods at a utilitarian optimum. Moreover, it is sometimes impossible to prevent a redistribution between workers of differing skills within the same region from having spillover effects in another. Redistributions may change the relative incentives workers of differing skills have to move.

The remainder of the article is organized as follows. The next section outlines the most general model considered in this article. Sections 3 and 4 describe, respectively, the general properties of implementable and optimal tax systems. Section 5 focuses attention on an economy with fixed producer wages in which it is easier to present our results on the spatial distribution of workers. Section 6 offers concluding remarks. Proofs are gathered in an Appendix.

## 2. The Model

The idealized economy considered in this article consists of two regions, labelled $A$ and $B$. These regions are populated by workers of two types, 1 and 2. We assume that the labels are assigned so that workers of type 2 earn higher wages. In general, each region is home to some portion of the individuals of each type. A single consumption good is produced in each region. For simplicity, we assume that it is the same good in each region. This good is called $z$. Production in region $j$ is denoted $z^j$. Production requires the use of the labour of workers of both types, and is governed by the aggregate production functions

$$z^j = f^j(L_1^j, L_2^j), \quad j = A, B, \tag{1}$$

where $L_i^j$ is the aggregate amount of labour of type $i$ used in region $j$. An aggregate firm in each region is assumed to choose the mix of labour inputs in order to maximize profit. This maximization process induces a demand for labour of each type. The aggregate supply of labour is influenced by the labour supply decisions of residents and by the (endogenous) number of residents of type $i$ in region $j$. Wages are determined in this competitive environment.[1] The wage rate for workers of type $i$ residing in region $j$ is denoted $w_i^j$. The produced commodity is taken as the numeraire.

Workers are assumed to differ from one another in two ways. First, as is common in nonlinear income tax models, each has a skill type, either 1 or 2. Second, each worker has a residential preference parameter $\theta \in [0, 1]$. The lower the value of $\theta$, the more the individual prefers to live in region $A$, all else equal. Individuals may have differing valuations of region-specific amenities, or they may possess a set of characteristics that allow them to achieve satisfaction more easily in one region than in the other. Language skills are an obvious example of the latter. Personal relationships or identification with a specific group or community may also contribute to residential preference. An individual is identified by the ordered pair $(i, \theta)$ of his or her characteristics; this ordered pair is called his or her type. The total mass of individuals of each skill type is one. Within skill types, attachment to home is distributed according to the cumulative distribution functions $G_1(\theta)$ and $G_2(\theta)$. These distributions have density functions $g_1(\theta)$ and $g_2(\theta)$, respectively. The set of workers residing in region $A$ is denoted $\mathcal{A}$; the set residing in region $B$, $\mathcal{B}$. $\mathcal{A}_i$ and $\mathcal{B}_i$ denote the set of workers of skill type $i$ working in regions $A$ and $B$, respectively.

[1] Firms are assumed to maximize profit, so wage rates are equal to values of marginal products of labour along the labour demand curves.

Given their region of residence, workers choose labour supply. They have preferences over consumption of the produced good, $x$, labour supply, $l$, and region of residence represented by the function

$$U(x,l,\theta) = \begin{cases} v(x,l) + h(\theta), & \theta \in \mathcal{A} \\ v(x,l), & \theta \in \mathcal{B}. \end{cases} \tag{2}$$

The function $v(x,l)$ is assumed to be strictly concave, increasing in $x$, decreasing in $l$ and twice continuously differentiable. The function $h(\theta)$ measures additional utility accruing to an individual who chooses to reside in region $A$ rather than region $B$, which is negative for those who prefer $B$ to $A$. Consistency in the interpretation of the parameter $\theta$ requires that $h$ be decreasing. We also assume that $h$ is differentiable. Individuals are wage takers, so the before-tax income of a worker of skill type $i$ residing in region $j$ who supplies $l$ units of labour is $y_i^j = w_i^j l$. Thus, preferences over $x$, $y$ and $\theta$ are represented by

$$V(x,y,\theta) = \begin{cases} v\big(x, \frac{y}{w}\big) + h(\theta), & \theta \in \mathcal{A} \\ v\big(x, \frac{y}{w}\big), & \theta \in \mathcal{B}. \end{cases} \tag{3}$$

These preferences satisfy the standard single-crossing property in $(x,y)$-space.

A single taxation authority sets tax policy in the two regions simultaneously.[2] It cannot observe individual characteristics, nor can it observe labour supply. It can observe before-tax income and choice of residence, but it cannot mandate hours or location of work. The best it can do is offer a schedule of before-tax and after-tax income combinations in each region. Individuals then choose their region of residence and hours of work to maximize utility. The resulting after-tax and before-tax income of a person of type $(i,\theta)$ who chooses to reside in region $j$ is denoted by $(x_i^j(\theta), y_i^j(\theta))$.[3] There is no *a priori* requirement that all individuals

[2]This rules out strategic behaviour on the part of governments in an attempt to produce a more desirable composition of population. While studying strategic behaviour is undoubtedly interesting, the case of coordinated tax policy is not yet understood. An understanding of the current problem should shed some light on what to expect in strategic settings.

[3]Lemmas 1-3 below make it clear that it is not necessary to specify $(x_i^j(\theta), y_i^j(\theta))$ if individual $(i,\theta)$ does not reside in region $j$. We make the very weak assumption that the functions $x_i^j(\theta)$ and $y_i^j(\theta)$ are integrable with respect to $G_i(\theta)$.

in a given region of the same work type choose the same allocation.

Because individual choices are made from an anonymous budget set, they satisfy the following self-selection conditions:

$$v\Big(x_i^A(\theta), \frac{y_i^A(\theta)}{w_i^A}\Big) + h(\theta) \geq v\Big(x_k^A(\hat{\theta}), \frac{y_k^A(\hat{\theta})}{w_i^A}\Big) + h(\theta), \quad \forall i, k = 1, 2, \ \theta, \hat{\theta} \in \mathcal{A}; \tag{4}$$

$$v\Big(x_i^A(\theta), \frac{y_i^A(\theta)}{w_i^A}\Big) + h(\theta) \geq v\Big(x_k^B(\hat{\theta}), \frac{y_k^B(\hat{\theta})}{w_i^B}\Big), \quad \forall i, k = 1, 2, \ \theta \in \mathcal{A}, \ \hat{\theta} \in \mathcal{B}; \tag{5}$$

$$v\Big(x_i^B(\theta), \frac{y_i^B(\theta)}{w_i^B}\Big) \geq v\Big(x_k^B(\hat{\theta}), \frac{y_k^B(\hat{\theta})}{w_i^B}\Big), \quad \forall i, k = 1, 2, \ \theta, \hat{\theta} \in \mathcal{B}; \tag{6}$$

$$v\Big(x_i^B(\theta), \frac{y_i^B(\theta)}{w_i^B}\Big) \geq v\Big(x_k^A(\hat{\theta}), \frac{y_k^A(\hat{\theta})}{w_i^A}\Big) + h(\theta), \quad \forall i, k = 1, 2, \ \theta \in \mathcal{B}, \hat{\theta} \in \mathcal{A}. \tag{7}$$

Relations (4) and (6) state that any individual weakly prefers the allocation they receive to the allocation chosen by other persons residing in the same region. These are very similar to the incentive compatibility conditions found in one-region optimal taxation models. Relations (5) and (7) are less familiar. These are satisfied if an individual prefers his or her allocation to what could be received in the other region. The best alternative for a potential migrant need not be the allocation designed for members of its work type residing in another region. Incentive compatibility includes the requirement that a person not be tempted to simultaneously change region of residence and misrepresent his or her underlying skill. In the language of nonlinear tax models, individuals must not find it in their interest to move-and-mimic.

The total supply of labour of a given skill type in a particular region comprises the time spent at work of all workers of that skill type who choose to live in that region. In a competitive equilibrium, aggregate labour demand does not exceed labour supply, or

$$L_i^A \leq \int_{\theta \in \mathcal{A}_i} \frac{y_i^A(\theta)}{w_i^A} d\theta, \quad i = 1, 2 \quad \text{and} \quad L_i^B \leq \int_{\theta \in \mathcal{B}_i} \frac{y_i^B(\theta)}{w_i^B} d\theta, \quad i = 1, 2. \tag{8}$$

## 3. Incentive Compatible Taxation

The implications of anonymous taxation on the structure of allocations in a one-region economy are well-documented.[4] As we will soon show, many of the canonical results carry over to economies with mobile residents. However, the existing literature is silent on the nature of residential choices. We now turn our attention to examining the incentive compatible distributions of population.

While residential preferences are continuously distributed, the choice of locations is limited. The taxation authority has no instrument in its arsenal to target location preference directly. For instance, if $\theta$ reflects linguistic ability alone, the government does not give tax credits for language training or capabilities. Given the preferences (3), the marginal rate of substitution between $x$ and $y$ is independent of $\theta$. Thus, labour-leisure choices do not vary systematically with $\theta$. The possibility of differentiating among individuals of the same skill type settled in a particular region according to residential preference is severely limited, as the following Lemma makes explicit.

**Lemma 1** The self-selection constraints imply that all individuals of the same skill type living in the same region receive the same utility from consumption and labour. Specifically,

$$v\Big(x_i^j(\theta), \frac{y_i^j(\theta)}{w_i^j}\Big) = v\Big(x_i^j(\hat{\theta}), \frac{y_i^j(\hat{\theta})}{w_i^j}\Big) \quad \forall i = 1, 2,\ j = A, B,\ \theta, \hat{\theta} \in [0, 1]. \tag{9}$$

Lemma 1 states that all individuals of a given skill type residing in the same region are on the same indifference curve in $(x, y)$-space. If (3) is interpreted as a numerical representation of utility then low-$\theta$ individuals living in the region $A$ are better off than their high-$\theta$ compatriots. Because it is impossible to discriminate on the basis of $\theta$, there are rents to be gained from living in one's *a priori* more preferred location. The government has no means of measuring or taxing these rents.[5] Lemma 1 leaves open the possibility of individuals of the same skill type choosing different allocations on the same $(x, y)$-indifference curve.

[4] See Guesnerie and Seade (1982), Stiglitz (1982) and Weymark (1986) for thorough treatments of this issue.

[5] Leaving location rents to some workers is not unique to the current problem. Osmundsen (1999) and Osmundsen et al. (1995) present models in which location rents are important. In these models, workers

The location rents afforded to individuals residing in region $A$ are decreasing in $\theta$, all else equal. Lemma 1 guarantees that the non-location portion of utility is constant for all residents of region $A$ of a given skill type. Thus, it seems intuitive — and the following Lemma confirms — that if an individual of type $(i, \theta)$ chooses to live in region $A$ then a person of type $(i, \hat{\theta})$ with $\hat{\theta} < \theta$ would also reside in region $A$.

**Lemma 2** The self-selection constraints imply residential sorting according to $\theta$ for each work type. Specifically, there exists some $\theta_i^*$ such that $\mathcal{A}_i = [0, \theta_i^*]$, $i = 1, 2$.

All residents of region $A$ face the same set of consumption-before tax income possibilities should they move to region $B$. Individuals of the same work-type have identical preferences over these allocations. If labour market opportunities in region $B$ do not attract individuals with a relatively weak attachment to region $A$ then those same opportunities cannot attract workers with a relatively strong attachment to region $A$.[6]

Lemma 2 can also be viewed from a screening perspective. The incentive compatibility constraints require that individuals not find it in their interest to move-and-mimic. Because workers may use the move-and-mimic strategy, the planner cannot design harsh penalties for workers who choose the "wrong" region. Such penalties can be avoided by claiming the allocation on offer for a worker of some "correct" residential preference. Moving to region $B$ is, in a sense, an outside option for residents of region $A$. The nature of residential mobility requires this outside option to be implementable. As Jullien (2000) shows, the structure of feasible allocations is often greatly simplified when the outside option is implementable. Lemma 2 describes the exact nature of this simplification for the economy under study.

If both jurisidictions are populated with individuals of each skill type, then $\theta_1^*$ and $\theta_2^*$ are determined by the condition

$$v\Big(x_i^A(\theta_i^*), \frac{y_i^A(\theta_i^*)}{w_i^A}\Big) + h(\theta_i^*) = v\Big(x_i^B(\theta_i^*), \frac{y_i^B(\theta_i^*)}{w_i^B}\Big), \tag{10}$$

can supply labour in more than one jurisdiction, thereby transmitting some information about location preference.

[6]Lemma 2 allows for $\theta_i^*$ to be zero or one, so the question of the existence of a person with a strong enough attachment to region $A$ to reside there has no bearing on the validity of the Lemma.

which states that the marginal individual of each skill type is indifferent between the two locations. Equation (10) is perhaps the most natural indifference condition, as it relates the utility gained by individuals of the same skill type in the two locations. However, in order for (10) to describe the residential sorting in this economy, it must be the case that the move-and-mimic strategy is inferior to simply moving and consuming the bundle intended for individuals of one's own skill type in the other region. The following Lemma states, loosely, that movers have no incentive to mimic as long as non-movers have no incentive to mimic, thereby justifying equation (10).

**Lemma 3** If the within-region and within-skill type self-selection constraints are satisfied then all the self-selection constraints are satisfied. Specifically, if (4) and (6) hold for all $i$ and $j$ and (5) and (7) hold for $i = k$ then (5) and (7) hold for $i \neq k$.

## 4. Optimal Taxation

The taxation authority has a Paretian objective function, defined over the utilities of all individuals residing in both regions. It designs tax policy to maximize the value of this objective subject to the incentive compatibility conditions and economy-wide (inter-regional) market clearing conditions. Technical feasibility of the final allocation of resources requires that aggregate consumption be no greater than aggregate production. The production sector of the economy is described by the relations (1) and (8). Using Lemma 2, the overall materials balance constraint is given by

$$\int_0^{\theta_1^*} x_1^A(\theta)g_1(\theta)d\theta + \int_0^{\theta_2^*} x_2^A(\theta)g_2(\theta)d\theta + \int_{\theta_1^*}^1 x_1^B(\theta)g_1(\theta)d\theta + \int_{\theta_2^*}^1 x_2^B(\theta)g_2(\theta)d\theta \leq$$
$$f^A\Big(\int_0^{\theta_1^*} \frac{y_1^A(\theta)}{w_1^A}g_1(\theta)d\theta, \int_0^{\theta_2^*} \frac{y_2^A(\theta)}{w_2^A}g_2(\theta)d\theta\Big) + f^B\Big(\int_{\theta_1^*}^1 \frac{y_1^B(\theta)}{w_1^B}g_1(\theta)d\theta, \int_{\theta_2^*}^1 \frac{y_2^B(\theta)}{w_2^B}g_2(\theta)d\theta\Big). \tag{11}$$

As is the tradition in optimal tax models, we work with a materials balance constraint rather than with a government budget constraint. This requires the assumption of complete taxation of pure profit. Moreover, use of the aggregate constraint (11) implies that transfers of resources across regional boundaries are permitted and are carried out to an optimal level.

As noted in the discussion of Lemma 1, self-selection alone is not enough to guarantee that all individuals of the same skill type residing in the same region receive the same allocation. However, the planner can do no better than to grant the same allocation to all individuals of the same skill type in the same region. The following Lemma provides a justification for this strategy.

**Lemma 4** For any allocation of resources $\{(x_i^j(\theta), y_i^j(\theta))\}_{j=A,B;\, i=1,2}$ satisfying (11) and (4)-(8) there exists an alternative allocation of resources $\{(\tilde{x}_i^j(\theta), \tilde{y}_i^j(\theta))\}_{j=A,B;\, i=1,2}$ for which the following statements are all true.

i. All individuals are indifferent between the original and alternative allocations.

ii. No individual changes regions between the original and alternative allocations.

iii. The alternative allocation satisfies (11) and (4)-(8).

iv. There exists $\tilde{x}_i^j$ and $\tilde{y}_i^j$ such that for all $i$ and $j$: $\tilde{x}_i^j(\theta) = \tilde{x}_i^j$ and $\tilde{y}_i^j(\theta) = \tilde{y}_i^j$.

Lemma 4 is an extension of Proposition 2 of Brito et al. (1990, p.67) to a continuous population. In their terminology, all workers of the same skill type living in the same region form a self-selection cycle, as each is indifferent between their allocation and that of any other member of this group. Lemma 4 states that the planner can optimally bunch these workers at the same allocation. The proof of Lemma 4 makes it clear that such bunching creates a surplus of consumption goods. At this point, though, we have not shown that the planner can always feasibly dispense such a surplus to make some workers better off. Thus, we have not ruled out optimal allocations of other kinds. Despite this possibility, we choose, for the sake of simplicity, to focus on these bunching solutions.

In view of Lemma 4, the optimal taxation problem can be viewed as a requirement to select four $(x, y)$ pairs — one for each skill type of worker in each region — and four wage rates. Although not choice variables *per se*, wage rates are determined endogenously. Thus, they are under the implicit control of the taxation authority. We shall focus on the case of a central taxation authority with utilitarian objectives. In this circumstance, the planner wishes to maximize

$$\begin{aligned}\mathcal{W} :=& v\Big(x_1^A, \frac{y_1^A}{w_1^A}\Big)G_1(\theta_1^*) + \int_0^{\theta_1^*} h(\theta)g(\theta)d\theta + v\Big(x_1^B, \frac{y_1^B}{w_1^B}\Big)[1 - G_1(\theta_1^*)] \\ &+ v\Big(x_2^A, \frac{y_2^A}{w_2^A}\Big)G_2(\theta_2^*) + \int_0^{\theta_2^*} h(\theta)g(\theta)d\theta + v\Big(x_2^B, \frac{y_2^B}{w_2^B}\Big)[1 - G_2(\theta_2^*)].\end{aligned} \tag{12}$$

The constraints faced by the planner take on relatively simple forms, given Lemma 4 and the results of Section 3. The materials balance constraint is given by

$$\begin{aligned}& x_1^A G_1(\theta_1^*) + x_1^B[1 - G_1(\theta_1^*)] + x_2^A G_2(\theta_2^*) + x_2^B[1 - G_2(\theta_2^*)] \\ \le & f^A\Big(\frac{y_1^A}{w_1^A}G_1(\theta_1^*), \frac{y_2^A}{w_2^A}G_2(\theta_2^*)\Big) + f^B\Big(\frac{y_1^B}{w_1^B}[1 - G_1(\theta_1^*)], \frac{y_2^B}{w_2^B}[1 - G_2(\theta_2^*)]\Big).\end{aligned} \tag{13}$$

The labour market equilibrium conditions are

$$\frac{y_i^A}{w_i^A}G_i(\theta_i^*) \ge L_i^A(w_1^A, w_2^A) \quad \text{and} \quad \frac{y_i^B}{w_i^B}[1 - G_i(\theta_i^*)] \ge L_i^B(w_1^B, w_2^B), \quad i = 1, 2. \tag{14}$$

Given the utilitarian nature of the objective function, the solution will be "redistributive" in the sense of Guesnerie (1995, pp.222-224), so that we consider only the downward self-selection constraints

$$v\Big(x_2^j, \frac{y_2^j}{w_2^j}\Big) \ge v\Big(x_1^j, \frac{y_1^j}{w_2^j}\Big), \quad j = A, B. \tag{15}$$

In view of equation (10) and Lemmas 3 and 4, the allocation of population across jurisdictions is governed by[7]

$$\theta_i^* = \phi(x_i^A, y_i^A, w_i^A, x_i^B, y_i^B, w_i^B) := h_i^{-1}\Big(v\Big(x_i^B, \frac{y_i^B}{w_i^B}\Big) - v\Big(x_i^A, \frac{y_i^A}{w_i^A}\Big)\Big), \quad i = 1, 2. \tag{16}$$

To summarize, the planner's problem may be formulated as

$$\max_{x_1^A, x_1^B, x_2^A, x_2^B, y_1^A, y_1^B, y_2^A, y_2^B, w_1^A, w_1^B, w_2^A, w_2^B} \mathcal{W} \quad \text{s.t.} \quad (13)\text{-}(16). \tag{17}$$

[7]We assume throughout that $\theta_i^* \in (0, 1)$ for $i = 1, 2$. This assumption is satisfied if both types of labour are essential for production in each region and/or residential preferences are sufficiently salient in workers' decision making. If $\theta_i^* \in \{0, 1\}$, then marginal changes in tax policy would typically have no effect on residential decisions for workers of skill type $i$.

The qualitative features of the optimal nonlinear income tax schedule are, perhaps surprisingly, very similar to those that are obtained in the one-region nonlinear tax problem. In order to state these properties, and to compare them with their single region counterparts, it is necessary to introduce a few more pieces of notation. Let $\lambda$ be the shadow value of the constraint (13), also known as the shadow value (in utility terms) of public funds; let $\mu^j$ be the multiplier associated with the self-selection constraint in jurisdiction $j$, and let $\psi_i^j$ be the shadow price of labour of skill type $i$ in region $j$. Finally, let $v_{x_1}^j$ denote the marginal utility of consumption of workers of skill type 1 residing in region $j$, and let $\hat{v}_{x_1}^j$ denote the marginal utility of consumption a worker of type 2 residing in region $j$ would enjoy whilst mimicking a worker of skill type 1.[8] With this notation in mind, we state the main features of the tax schedule in jurisdiction $A$. The tax schedule in jurisdiction $B$ has similar properties.

**Proposition 1** The following statements are true at a solution to the planner's problem (17).

i. Workers of skill-type 2 residing in region $A$ receive a marginal wage subsidy. In particular,
$$-\frac{v_{l_2}^A}{v_{x_2}^A} = w_2^A + \frac{\psi_2^A}{\lambda} > w_2^A. \tag{18}$$

ii. Workers of skill-type 1 residing in region $A$ have marginal rate of substitution between labour and consumption given by
$$-\frac{v_{l_1}^A}{v_{x_1}^A} = \frac{w_1^A \lambda G_1(\theta_1^*) - \mu^A \hat{v}_{l_1}^A \frac{w_1^A}{w_2^A} + \psi_1^A G_1(\theta_1^*)}{\lambda G_1(\theta_1^*) + \mu^A \hat{v}_{x_1}^A}. \tag{19}$$

The results reported in Proposition 1 are exact replicas of those reported by Stiglitz (1982) for a single jurisdiction. As in Stiglitz, more highly skilled workers receive a wage subsidy. The slight differences in form between equations (18) and (19) and their counterparts in Stiglitz is due to the allowance we make for decreasing returns to scale. The motivation for the marginal subsidy at the top is exactly the motivation outlined by Stiglitz: to increase

[8] Analogous expressions for individuals of type 2, or denoting the marginal utility of labour also appear in the sequel. To save space, the reader is asked to supply the analogous interpretations.

the labour supply of highly skilled workers, thereby enhancing redistribution by reducing the wage gap between workers of different skills. Ostensibly, mobility has no effect on the formulae for the optimal tax rates. This echoes the findings of Wilson (1992), who studies the effects of mobility into or out of a single region on the optimal tax schedule in that region. In that situation, the optimal tax scheme also exhibits marginal subsidies at the top.

It stands to reason that residential decisions have no direct bearing on marginal tax rates. Marginal changes in residential choices have no direct utility consequences, as marginal residents are indifferent between the two jurisdictions. Moreover, marginal distortions in policy instruments are incapable of producing welfare-enhancing changes in residential decisions. On the one hand, location is determined by differences in total utility, not by differences in the marginal utilities of various commodities. Residential decisions may be altered by redistributing total utility. Additional marginal distortions are counter-productive, as they inhibit the efficient distribution of utility. Nevertheless, marginal changes in policy instruments have a marginal effect on the spatial distribution of workers (hence, on relative wages and the demand for consumption), because the set of workers indifferent between locations is of measure zero. Saez (2004) has shown the importance of the measure of indifferent workers on the qualitative features of optimal taxation in the context of occupational choice, which is just a different kind of discrete labour market choice. Residential mobility has indirect effects, however, operating through the shadow value of labour and the marginal value of public expenditure.

## 5. Fixed Producer Wages

One way to isolate the effects of residential mobility on the optimal tax structure is to work with the familiar fixed wage model. This model specializes the productions functions to

$$z^j = w_1^j l_1^j + w_2^j l_2^j \quad j = A, B, \tag{20}$$

which exhibit constant returns to scale and linear isoquants in $(l_1, l_2)$-space. With these assumptions, the profit maximizing choice of labour inputs is undefined and the labour market clearing conditions (8) are redundant. The planner no longer has even implicit control over producer wages, and the planner's problem reduces to

$$\max_{x_1^A, x_1^B, x_2^A, x_2^B, y_1^A, y_1^B, y_2^A, y_2^B} \mathcal{W} \quad \text{s.t.} \quad (15),\ (16),\ \text{and}$$

$$[y_1^A - x_1^A]G_1(\theta_1^*) + [y_1^B - x_1^B][1 - G_1(\theta_1^*)] + [y_2^A - x_2^A]G_2(\theta_2^*) + [y_2^B - x_2^B][1 - G_2(\theta_2^*)] \geq 0. \tag{21}$$

The results contained in Proposition 1 carry over with some straightforward modifications. For completeness, we record the analogous results in the following Proposition.

**Proposition 2** The following statements are true at a solution to the planner's problem (21).

i. Workers of skill-type 2 residing in region $A$ face a zero marginal tax rate on labour income. In particular,

$$-\frac{v_{l_2}^A}{v_{x_2}^A} = w_2^A. \tag{22}$$

ii. Workers of skill-type 1 residing in region $A$ pay a positive marginal rate of income tax. In particular,

$$-\frac{v_{l_1}^A}{v_{x_1}^A} = w_1^A \left( \frac{\lambda G_1(\theta_1^*) - \mu^A \hat{v}_{l_1}^A \frac{1}{w_2^A}}{\lambda G_1(\theta_1^*) + \mu^A \hat{v}_{x_1}^A} \right) < w_1^A. \tag{23}$$

Proposition 2 recasts the main findings of Proposition 1. Even if the planner wishes to induce some migration by highly skilled workers as part of the optimal scheme, marginal distortions are not to be recommended. Marginal distortions are an inefficient way to redistribute utility. The familiar positive marginal tax rate at the bottom is easily seen in this version of the model. It arises for exactly the same reason as the distortion in a one-region model: asymmetric information about work type.

While differences in marginal rates of taxation are not sufficient to cause migration, differences in real living standards are. The optimal tax system may induce a distribution of population different from the one that would prevail under laissez-faire. While the model presented here is too general to provide a straightforward comparison of the no-tax population distribution with the second-best optimal distribution, some insight into the forces at play is gained by studying a further specialization of the model that features identical wage distributions in the two regions and a particular normalization of the residential attachment functions.

**Assumption 1** $w_i^A = w_i^B$ and $h_i(\frac{1}{2}) = 0$, for $i = 1, 2$.

The laissez-faire outcome under Assumption 1 is easy to describe. Workers in each region choose identical work-consumption bundles and the mass of workers of types 1 and 2 residing in region $A$ are $G_1(\frac{1}{2})$ and $G_2(\frac{1}{2})$, respectively. The within-region composition of the workforce is determined by the shapes of the distributions $G_1(\theta)$ and $G_2(\theta)$. Moreover, equality of wages implies it is always possible to select $(x_i^A, y_i^A) = (x_i^B, y_i^B)$ for $i = 1, 2$ in the second-best problem, ensuring a post-tax distribution of population identical to the laissez-faire distribution. The two self-selection constraints would then collapse to one. The next Proposition states, however, that it is optimal to have identical allocations in the two regions only under very special circumstances.

**Proposition 3** Let Assumption 1 hold. Then $(x_i^A, y_i^A) = (x_i^B, y_i^B)$ for $i = 1, 2$ holds at a solution to (21) only if

$$\frac{G_1(\frac{1}{2})}{1 - G_1(\frac{1}{2})} = \frac{G_2(\frac{1}{2})}{1 - G_2(\frac{1}{2})}. \tag{24}$$

Proposition 3 states that individuals of the same skill type living in different regions are bunched at the optimum only if the relative shares of workers of each type under laissez-faire is the same in each region. Differentiated allocations across regions induce an allocation of population different than under laissez-faire. The only way for workers of skill type 2 to simultaneously face a zero marginal income tax rate and be on the same indifference curve in $(x, y)$-space is to consume the same bundle. Hence, according to Proposition 3, they must be on different indifference curves in $(x, y)$-space, and $h(\theta_2^*)$ cannot be zero at the optimum.[9] The key to unlocking the intuition behind Proposition 3 is this close relationship between

[9]While the statement of Proposition 3 does not rule out differentiation among workers of skill type 1 alone, it can easily be shown that such an outcome would either violate a self-selection constraint or leave one slack.

identical allocations and the laissez-faire distribution of population. The planner does not desire a redistribution of population for its own sake, nor does it necessarily wish to make a transfer from one region to another. However, when (24) is violated a redistribution of income between skill types causes a redistribution of income between regions. For example, if region A has a larger share of type 1 residents than does region $B$, a redistribution toward type 1 workers transfers purchasing power from region $B$ to region $A$. This increased purchasing power in region $A$ induces migration into region $A$. The planner is indifferent to marginal changes in the spatial distribution of population because the marginal individuals are indifferent between the two locations. Hence, it is willing to accept a small distortion in residential choice in order to enhance transfers to the less skilled.

Despite addressing very similar issues — namely, within-skill-type bunching — Lemma 4 and Proposition 3 generate very different results. Lemma 4 follows from the intuition of Brito et al. (1990) that if two workers are at different points on the same indifference curve in $(x, y)$-space then the planner might just as well offer each of them the cheaper of the two bundles. The self-selection conditions require that all workers of the same skill type residing in the same region be on the same indifference curve. Because region of residence is observable, incentive compatibility does not require workers of the same skill type residing in different regions to be on the same indifference curve. Indeed, Proposition 3 implies that such workers are seldom on the same indifference curve at a second-best optimum.

## 6. Conclusion

This article has presented a model of optimal taxation when workers of all types are potential migrants. Our focus has been on describing the economy-wide optimal tax system. We find that potential migration does not overturn the main findings of the optimal tax literature regarding the qualitative features of the marginal tax rates. The choice of where to work and live is not a marginal decision, so changes in the marginal tax rate *per se* have no effect on location decisions. Redistribution of total income across borders, however, does encourage migration. We have shown how redistribution across income classes often necessitates redistribution across borders. This, in turn, implies that the laissez-faire distribution of population may not be a feature of an economy with the optimal tax mix.

Because this is a work in normative economics, we view our results as providing a possible benchmark against which to assess the outcomes of strategic models of taxation, rather than as a description of any existing policy situation. Yet, our analysis has uncovered some economic forces that can shape policy. In Canada, for example, the Federal government is largely (though not entirely) responsible for personal income taxation. It also controls redistributive levers such as Employment Insurance. Because workforce composition differs among the provinces and territories, redistributions across income classes are implicit inter-regional transfers. There is nothing in our results to suggest that the consequences this may have for the distribution of workers across provinces are necessarily efficiency destroying.

A potentially fruitful extension of this analysis would be the study of designing federal structures in which provincial budget constraints are affected by national redistributive policies. If a federal government were forced to take the consequences of its decisions on provincial coffers into account, and was unable to fully neutralize these decisions with lump sum payments, the features of the tax system may be significantly altered. As Cebreiro (2002) has argued in a non-federal setting, forcing budget balance on a region-by-region basis may conflict with production efficiency and seriously alter the shape of the optimal tax schedule. Introducing strategic behaviour on the part of sub-national governments would complicate matters in further, yet unexplored, ways.

**Appendix**

Proof of Lemma 1:

Take any $i = 1, 2$, $j = A, B$ and $\theta, \hat{\theta} \in [0, 1]$. By (4),

$$v\Big(x_i^j(\theta), \frac{y_i^j(\theta)}{w_i^j}\Big) - v\Big(x_i^j(\hat{\theta}), \frac{y_i^j(\hat{\theta})}{w_i^j}\Big) \geq 0; \tag{25}$$

$$v\Big(x_i^j(\hat{\theta}), \frac{y_i^j(\hat{\theta})}{w_i^j}\Big) - v\Big(x_i^j(\theta), \frac{y_i^j(\theta)}{w_i^j}\Big) \geq 0 \tag{26}$$

The left-hand side of (25) is the additive inverse of the left-hand side of (26), so both of these expressions must be zero. Equation (9) follows for workers in region $A$. A similar argument using (6) establishes the result for workers residing in region $B$. $\square$

Proof of Lemma 2:

The Lemma is false only if there exists a pair $((i, \theta), (i, \hat{\theta}))$ with $\hat{\theta} > \theta$, $(i, \theta) \in \mathcal{B}_i$ and $(i, \hat{\theta}) \in \mathcal{A}_i$ Suppose, by way of contradiction, that such a pair exists. By (5),

$$v\Big(x_i^A(\hat{\theta}), \frac{y_i^A(\hat{\theta})}{w_i^A}\Big) + h(\hat{\theta}) \geq v\Big(x_i^B(\theta), \frac{y_i^B(\theta)}{w_i^B}\Big). \tag{27}$$

By (7),

$$v\Big(x_i^B(\theta), \frac{y_i^B(\theta)}{w_i^B}\Big) \geq v\Big(x_i^A(\hat{\theta}), \frac{y_i^A(\hat{\theta})}{w_i^A}\Big) + h(\theta). \tag{28}$$

Combining (27) and (28) gives

$$v\Big(x_i^A(\hat{\theta}), \frac{y_i^A(\hat{\theta})}{w_i^A}\Big) + h(\hat{\theta}) \geq v\Big(x_i^A(\hat{\theta}), \frac{y_i^A(\hat{\theta})}{w_i^A}\Big) + h(\theta), \tag{29}$$

which implies $h(\hat{\theta}) \geq h(\theta)$. This contradicts the decreasingness of $h$. $\square$

Proof of Lemma 3:

Only the proof that (5) holds for $i \neq k$ is presented. The proof regarding (7) is identical, save for notation.

Select $\theta \in \mathcal{A}$, $\hat{\theta} \in \mathcal{B}$, and $i \neq k$. By hypothesis, (5) holds for any two workers of the same skill type, so

$$v\Big(x_i^A(\theta), \frac{y_i^A(\theta)}{w_i^A}\Big) + h(\theta) \geq v\Big(x_i^B(\hat{\theta}), \frac{y_i^B(\hat{\theta})}{w_i^B}\Big). \tag{30}$$

By (6)

$$v\Big(x_i^B(\hat{\theta}), \frac{y_i^B(\hat{\theta})}{w_i^B}\Big) \geq v\Big(x_k^B(\hat{\theta}), \frac{y_k^B(\hat{\theta})}{w_i^B}\Big). \tag{31}$$

Combining inequalities (30) and (31) establishes the result. □

Proof of Lemma 4:

Consider the case of workers of skill type 1 residing in region $A$. Let $(x_1^A(\theta), y_1^A(\theta))$ be the part of a candidate optimal allocation designed for that group. This candidate optimum must satisfy the feasibility and self-selection constraints.

Hold the allocation for all other skill-residence combinations constant. By Lemma 1, there exits some constant $k$ such that

$$v\Big(x_1^A(\theta), \frac{y_1^A(\theta)}{w_1^A}\Big) = k, \quad \forall \theta \in \mathcal{A}_1. \tag{32}$$

Define the function $x^k(y)$ by

$$x^k(y) = x \quad \leftrightarrow \quad v\Big(x, \frac{y}{w_1^A}\Big) = k \tag{33}$$

The graph of $x^k(\cdot)$ in $(x, y)$-space is the level set of the function $v(\cdot)$ on which all members of $(x_1^A(\theta), y_1^A(\theta))$ are found. The properties of $v(\cdot)$ ensure that $x^k(\cdot)$ is well-defined and strictly convex.

Because the allocation functions are bounded, there exists a greatest lower bound and a least upper bound for $y_1^A(\theta)$. Call these values $y^{\inf}$ and $y^{\sup}$, respectively. Because $\mathcal{A}_1$ is an interval, it is measurable. Its measure with respect to $dG(\theta)$ is $G(\theta_1^*)$. Thus, there exists a $\tilde{y}_1^A \in [y^{\inf}, y^{\sup}]$ such that[10]

$$G(\theta_1^*)\tilde{y}_1^A = \int_0^{\theta_1^*} y_1^A(\theta) dG(\theta). \tag{34}$$

If each worker in $\mathcal{A}_1$ were offered $\tilde{y}_1^A$ the total before-tax income (production) of this group would be the same as in the initial allocation.

Define $\tilde{x}_1^A = x^k(\tilde{y}_1^A)$, which is the $x$-coordinate of the indifference curve on which workers in $\mathcal{A}_1$ are situated corresponding to the $y$-coordinate $\tilde{y}_1^A$. Then

$$\int_0^{\theta_1^*} \tilde{x}_1^A dG(\theta) = G(\theta_1^*) x^k(\tilde{y}_1^A) = G(\theta_1^*) x^k\Big(\int_0^{\theta_1^*} \frac{y_1^A(\theta)}{G(\theta_1^*)} dG(\theta)\Big). \tag{35}$$

[10]This result follows from Exercise 2 on p.86 of Berberian (1965).

Because $x^k(\cdot)$ is strictly convex, Jensen's inequality may be applied to (35) to conclude

$$\int_0^{\theta_1^*} \tilde{x}_1^A dG(\theta) < G(\theta_1^*) \int_0^{\theta_1^*} \frac{x^k(y_1^A(\theta))}{G(\theta_1^*)} dG(\theta) = \int_0^{\theta_1^*} x^k(y_1^A(\theta)) dG(\theta) \tag{36}$$

Thus replacing the candidate allocations for workers of skill type 1 residing in region $A$ by $(\tilde{x}_1^A, \tilde{y}_1^A)$ results in an allocation with the same total before-tax income and lower aggregate consumption than the candidate allocation. Thus, the replacement allocation satisfies the materials balance constraint.

By construction, all members of $\mathcal{A}_1$ are indifferent between $(\tilde{x}_1^A, \tilde{y}_1^A)$ and their allocation in the candidate optimum. Thus, they (weakly) prefer $(\tilde{x}_1^A, \tilde{y}_1^A)$ to any allocation chosen by other individuals in the candidate optimum.

It remains to show that workers outside of $\mathcal{A}_1$ prefer their original allocations to $(\tilde{x}_1^A, \tilde{y}_1^A)$. This is immediate if $(\tilde{x}_1^A, \tilde{y}_1^A)$ coincides with one of the original allocations for members of $\mathcal{A}_1$. Otherwise, select two elements of the original allocation $(x_1^A(\theta), y_1^A(\theta))$ and $(x_1^A(\theta'), y_1^A(\theta'))$ satisfying $(x_1^A(\theta), y_1^A(\theta)) \ll (\tilde{x}_1^A, \tilde{y}_1^A) \ll (x_1^A(\theta'), y_1^A(\theta'))$. Such a pair is guaranteed to exist unless all $(x_1^A(\theta), y_1^A(\theta))$ were equal in the original allocation. But single crossing of indifference curves in $(x, y)$-space ensures that, because all workers outside $\mathcal{A}_1$ prefer their original allocations to both $(x_1^A(\theta), y_1^A(\theta))$ and $(x_1^A(\theta'), y_1^A(\theta'))$, they must also prefer their original allocation to $(\tilde{x}_1^A, \tilde{y}_1^A)$. $\square$

Proof of Proposition 1:

The proof relies heavily on the first order optimality conditions. We begin by presenting an argument that allows some simplification of the optimality conditions. The first order condition associated with $x_1^A$ is

$$\begin{aligned} & v_{x_1}^A G_1(\theta_1^*) + v\Big(x_i^A, \frac{y_i^A}{w_i^A}\Big) g_1(\theta_1^*) \frac{\partial \theta_1^*}{\partial x_1^A} + h(\theta_1^*) g_1(\theta_1^*) \frac{\partial \theta_1^*}{\partial x_1^A} - v\Big(x_i^B, \frac{y_i^B}{w_i^B}\Big) g_1(\theta_1^*) \frac{\partial \theta_1^*}{\partial x_1^A} - \mu^A \hat{V}_{x_1}^A \\ & + \lambda \left[ -G_1(\theta_1^*) - x_1^A g_1(\theta_1^*) \frac{\partial \theta_1^*}{\partial x_1^A} + x_1^B g_1(\theta_1^*) \frac{\partial \theta_1^*}{\partial x_1^A} + \frac{\partial f^A}{\partial l_1^A} \frac{y_1^A}{w_1^A} g_1(\theta_1^*) \frac{\partial \theta_1^*}{\partial x_1^A} - \frac{\partial f^B}{\partial l_1^B} \frac{y_1^B}{w_1^B} g_1(\theta_1^*) \frac{\partial \theta_1^*}{\partial x_1^A} \right] \\ & + \left[ \psi_1^A \frac{y_1^A}{w_1^A} - \psi_1^B \frac{y_1^B}{w_1^B} \right] g_1^*(\theta_1^*) \frac{\partial \theta_1^*}{\partial x_1^A} = 0. \end{aligned} \tag{37}$$

By (10), the second through fourth terms of (37) sum to zero. Equality of the wage rate with marginal product in each region allows one to simplify the expression in square brackets, yielding

$$\begin{aligned} &v^A_{x_1}G_1(\theta^*_1) - \lambda G_1(\theta^*_1) + \lambda\Big[(y^A_1 - x^A_1 - y^B_1 + x^B_1)g_1(\theta^*_1)\frac{\partial\theta^*_1}{\partial x^A_1}\Big] \\ &+\left[\psi^A_1\frac{y^A_1}{w^A_1} - \psi^B_1\frac{y^B_1}{w^B_1}\right]g^*_1(\theta^*_1)\frac{\partial\theta^*_1}{\partial x^A_1} - \mu^A\hat{v}^A_{x_1} = 0. \end{aligned} \tag{38}$$

The same steps may be applied to any of the other optimality conditions. The first order conditions associated with $x^A_2$, $y^A_1$ and $y^A_2$ imply, respectively,

$$\begin{aligned} &v^A_{x_2}G_2(\theta^*_2) - \lambda G_2(\theta^*_2) + \lambda\Big[(y^A_2 - x^A_2 - y^B_2 + x^B_2)g_2(\theta^*_2)\frac{\partial\theta^*_2}{\partial x^A_2}\Big] \\ &+\left[\psi^A_2\frac{y^A_2}{w^A_2} - \psi^B_2\frac{y^B_2}{w^B_2}\right]g^*_2(\theta^*_2)\frac{\partial\theta^*_2}{\partial x^A_2} - \mu^A v^A_{x_2} = 0; \end{aligned} \tag{39}$$

$$\begin{aligned} &\frac{v^A_{l_1}}{w^A_1}G_1(\theta^*_1) + \lambda G_1(\theta^*_1) + \lambda\Big[(y^A_1 - x^A_1 - y^B_1 + x^B_1)g_1(\theta^*_1)\frac{\partial\theta^*_1}{\partial y^A_1}\Big] + \frac{\psi^A_1}{w^A_1}G_1(\theta^*_1) \\ &+\left[\psi^A_1\frac{y^A_1}{w^A_1} - \psi^B_1\frac{y^B_1}{w^B_1}\right]g^*_1(\theta^*_1)\frac{\partial\theta^*_1}{\partial y^A_1} - \frac{\mu^A}{w^A_2}\hat{v}^A_{l_1} = 0; \end{aligned} \tag{40}$$

$$\begin{aligned} &\frac{v^A_{l_2}}{w^A_2}G_2(\theta^*_2) + \lambda G_2(\theta^*_2) + \lambda\Big[(y^A_2 - x^A_2 - y^B_2 + x^B_2)g_2(\theta^*_2)\frac{\partial\theta^*_2}{\partial y^A_2}\Big] + \frac{\psi^A_2}{w^A_2}G_1(\theta^*_1) \\ &+\left[\psi^A_2\frac{y^A_2}{w^A_2} - \psi^B_2\frac{y^B_2}{w^B_2}\right]g^*_2(\theta^*_2)\frac{\partial\theta^*_2}{\partial y^A_2} + \frac{\mu^A}{w^A_2}v^A_{l_2} = 0. \end{aligned} \tag{41}$$

By (16),

$$\frac{\partial\theta^*_i}{\partial x^A_i} = -h^{-1\prime}_i(\cdot)v^A_{x_i} \quad \text{and} \quad \frac{\partial\theta^*_i}{\partial y^A_i} = -h^{-1\prime}_i(\cdot)\frac{v^A_{l_i}}{w^A_i}, \quad i = 1, 2. \tag{42}$$

Thus, (39) and (41) imply

$$\begin{aligned} &v^A_{l_2}\left[G_2(\theta^*_2) + \mu^A - \Big[\lambda(y^A_2 - x^A_2 - y^B_2 + x^B_2) + \psi^A_2\frac{y^A_2}{w^A_2} - \psi^B_2\frac{y^B_2}{w^B_2}\Big]g_2(\theta^*_2)h^{-1\prime}_i(\cdot)\right] \\ &+\psi^A_2 G_2(\theta^*_2) = -\lambda w^A_2 G_2(\theta^*_2); \end{aligned} \tag{43}$$

$$\begin{aligned} &v^A_{x_2}\left[G_2(\theta^*_2) + \mu^A - \Big[\lambda(y^A_2 - x^A_2 - y^B_2 + x^B_2) + \psi^A_2\frac{y^A_2}{w^A_2} - \psi^B_2\frac{y^B_2}{w^B_2}\Big]g_2(\theta^*_2)h^{-1\prime}_i(\cdot)\right] \\ &= \lambda G_2(\theta^*_2). \end{aligned} \tag{44}$$

Dividing (43) by (44) gives equality in (18). Positiveness of $\psi_2^A$ establishes the inequality in (18). A similar calculation using (38) and (40) establishes (19). $\square$

Proof of Proposition 2:

The equalities are easily derived from (18) and (19) by setting $\psi_j^i = 0$ for all $i$ and $j$, which follows from the absence of the constraints (8). The proof of the inequality in (23) starts with recognizing that the mimicker is on the same indifference curve as the agent of skill type 2, but that $(x_1^A, y_1^A) \ll (x_2^A, y_2^A)$, so that $\frac{-\hat{v}_{l_1}^A}{\hat{v}_{x_1}^A} < w_2^A$. Thus,

$$\lambda G_1(\theta_1^*) - \mu^A \hat{v}_{l_1}^A \frac{1}{w_2^A} < \lambda G_1(\theta_1^*) + \mu^A \hat{v}_{x_1}^A. \tag{45}$$

Hence, the term in parentheses in (23) is less than one, establishing the inequality. $\square$

Proof of Proposition 3:

Suppose that $(x_i^A, y_i^A) = (x_i^B, y_i^B)$ for $i = 1, 2$ at a solution to (21). Then, by Assumption 1, $\theta_1^* = \theta_2^* = \frac{1}{2}$. Moreover, the terms in square brackets in (38) and (39) are zero, yielding

$$v_{x_1}^A G_1\big(\frac{1}{2}\big) - \lambda G_1\big(\frac{1}{2}\big) = \mu^A \hat{v}_{x_1}^A, \tag{46}$$

$$v_{x_2}^A G_2\big(\frac{1}{2}\big) - \lambda G_2\big(\frac{1}{2}\big) = -\mu^A v_{x_2}^A, . \tag{47}$$

Likewise, the optimality conditions associated with $x_1^B$ and $x_2^B$ reduce to:

$$v_{x_1}^B \Big[1 - G_1\big(\frac{1}{2}\big)\Big] - \lambda \Big[1 - G_1\big(\frac{1}{2}\big)\Big] = \mu^B \hat{v}_{x_1}^B, \tag{48}$$

$$v_{x_2}^B \Big[1 - G_2\big(\frac{1}{2}\big)\Big] - \lambda \Big[1 - G_2\big(\frac{1}{2}\big)\Big] = -\mu^B v_{x_2}^B, . \tag{49}$$

Because the allocations are assumed to be symmetric across regions, we may divide (46) by (48), and (47) by (49) to give

$$\frac{G_1\big(\frac{1}{2}\big)}{1 - G_1\big(\frac{1}{2}\big)} = \frac{\mu^A}{\mu^B} = \frac{G_2\big(\frac{1}{2}\big)}{1 - G_2\big(\frac{1}{2}\big)}, \tag{50}$$

which establishes the result. $\square$

**References**


Berberian, S. (1965), *Measure and Integration*, New York: Chelsea.

Bhagwati, J., and K. Hamada (1989), "Tax policy in the presence of immigration," J. Bhagwati and J.D. Wilson, eds., *Income Taxes and International Mobility*, 113–140, Cambridge: MIT Press.

Brito, D., J. Hamilton, S. Slutsky, and J. Stiglitz (1990), "Pareto efficient tax structures," *Oxford Economic Papers* 42, 61–77.

Cebreiro, A. (2002), *Optimal Redistributive Taxation with Factor Mobility*, Ph.D. Thesis, University of Essex.

Guesnerie, R. (1995), *A Contribution to the Pure Theory of Taxation*, Cambridge: Cambridge University Press.

Guesnerie, R., and J. Seade (1982), "Non linear pricing in a finite economy," *Journal of Public Economics* 17, 157–179.

Jullien, B. (2000), "Participation constraints in adverse selection models," *Journal of Economic Theory* 93, 1–47.

Mansoorian, A., and G. Myers (1993), "Attachment to home and efficient purchases of population in a fiscal externality economy," *Journal of Public Economics* 52, 117–132.

Mirrlees, J. (1989), "Migration and optimal income taxes," J. Bhagwati and J.D. Wilson, eds., *Income Taxes and International Mobility*, 179–200, Cambridge: MIT Press.

Osmundsen, P. (1999), "Taxing internationally mobile workers: a case of countervailing incentives," *International Tax and Public Finance* 6, 149–164.

Osmundsen, P., G. Schjelderup, and K. Hagen (1995), "Personal income taxation and welfare appropriation in an open economy," *Discussion Paper* 15, Institute of Economics, Norwegian School of Economics and Business Administration.

Saez, E. (2002), "Optimal income transfer programs: intensive versus extensive labour supply responses," *Quarterly Journal of Economics* 117, 1039–1073.

Saez, E. (2004), "Direct or indirect tax instruments for redistribution: short-run versus long-run," *Journal of Public Economics* 88, 503–518.

Stiglitz, J. (1982), "Self-selection and Pareto efficient taxation," *Journal of Public Economics* 17, 213–240.

Weymark, J. (1986), "A reduced-form optimal nonlinear income tax problem," *Journal of Public Economics* 30, 199–217.

Wilson, J.D. (1992), "Optimal income taxation and international personal mobility," *American Economic Review* 82, 191–196.