Credible Redistribution Policy and Skilled Migration *

Roc Armenter, *Federal Reserve Bank of Philadelphia*
Francesc Ortega, *Universitat Pompeu Fabra*

December 2009

Abstract

We analyze the joint determination of income redistribution and migration flows across fiscally independent regions. In our model, regional governments lack commitment so their policy announcements must be credible, and redistribution between skilled and unskilled workers is bounded by informational constraints. In any given region, the welfare of all workers is increasing in the share of skilled workers, as after-tax incomes increase for both skilled and unskilled workers. When skilled workers are more geographically mobile than unskilled ones, the endogenous response of redistribution policy can induce regional agglomeration of skilled workers. We also find that the equilibrium features symmetry-breaking if migration costs are relatively low; and that worker mobility tends to amplify pre-existing welfare differences in income and welfare across regions.

Keywords: Agglomeration, Migration, Redistribution

JEL Codes: E61, J61

1 Introduction

The process of geographical concentration of skilled labor may be at the heart of sustained economic development. Lucas (2004) formalizes this idea, building on Eaton and Eckstein (1997), in a model where existing skills are a key input in the production of new skills. While empirically it is well established that skilled labor gravitates toward areas where skills are already abundant (Moretti (2004) and references therein), it is less clear which are the particular mechanisms driving skill agglomeration in the data. As Duranton and Puga (2004) point out, the implications for efficiency and policy often depend crucially on the specific micro-foundation generating skill agglomeration.

*The views expressed in the paper are those of the authors and do not necessarily reflect the views at the Federal Reserve Bank of Philadelphia or the Federal Reserve System. Any errors or omissions are the responsibility of the authors.

This paper introduces a new agglomeration mechanism: we show how endogenous income redistribution among fiscally independent regional authorities can lead to skill agglomeration. Consider two regions that differ in the skill composition of their workforce , and operate an identical constant-returns-to-scale technology. In our model *after-tax* income for skilled workers is higher in the region with a higher share of skilled workers. The lower skill wage premium translates into lower need for redistribution: the lower taxes paid by skilled workers more than compensate for the lower *pre-tax* skilled income. As a result, footloose skilled workers will migrate from the skill-scarce to the skill-abundant region.

We propose a model where bilateral migration flows and regional redistributive policies are jointly determined. We crucially depart from previous literature by focusing on *credible policies.* That is, we dismiss the ability of governments to commit to policies that are not in the best interest of the after-migration workforce in the region. We model the policy decision as the decision of a benevolent government subject to some informational constraints. The worker skill level is private information, and thus the tax schedule can only be a function of workers' earnings. We solve the government's decision problem using a modified version of Mirrlees (1971). Equilibrium policies are sub-game perfect: each regional government's policy must be optimal given the final workforce in the region. Turning to migration flows, skilled workers decide whether and where to migrate subject to a set frictions. Each skilled worker receives one idiosyncratic opportunity to migrate, which specifies a mobility cost and a potential destination region. Skilled workers then compare the welfare gain of migration with their idiosyncratic mobility cost, correctly forecasting the redistribution policy at destination. Importantly, in our model unskilled workers are geographically less mobile than skilled workers. For simplicity, we assume that migration costs for unskilled workers are prohibitively high.[1]

We show that the informational constraint limits the amount of redistribution. Specifically, a government that can tax individuals based on their (unobserved) type would implement larger redistribution that is feasible in our environment. In our equilibrium allocations skilled workers consume (and work) more than unskilled workers. Another important feature of our model is that the welfare of skilled workers in a region is an increasing function of the share of skilled workers in that region. As noted earlier, an inflow of skilled workers reduces pre-tax income inequality, which leads to lower redistributive taxation. We show that the resulting allocation delivers higher utility to both skilled and unskilled residents. We emphasize that if redistributive taxation were *fixed*, both pre-tax and after-tax skilled income would be reduced by an inflow of skilled workers into the region. Regional skill agglomeration is a direct implication of the previous result. As skilled workers move into a region, redistributive taxes fall, after-tax skilled income rises, and the region becomes even more attractive to skilled workers. The agglomeration process stops only when the mobility cost

[1] Armenter and Ortega (2009) analyze a richer environment where both types of workers are mobile.

faced by the marginal migrant becomes too high.

We highlight two important implications of the theory. First, for low mobility costs, the equilibrium features symmetry-breaking, that is, regional differences in income and welfare arise among ex-ante identical regions.[2] Inequality across regions is thus a necessary part of (stable) equilibria when mobility costs are low. Second, skilled worker mobility amplifies any pre-existing regional differences in income and utility. Regions with a technological or human capital disadvantage will suffer a brain drain, as their skilled workers migrate to more developed regions. The nature of regional differences (TFP differences versus skill endowments) is not innocuous. When TFP differences are the main source of heterogeneity across regions, migration flows tend to improve productive efficiency and social welfare. We also note that if regional differences in TFP are very large, initial differences in skill endowments may be irrelevant to migration flows.

Our paper is closely related to two strands of literature. First, it makes a contribution to the literature on agglomeration economies. The review by Duranton and Puga (2004) lays out a taxonomy for the numerous micro-foundations proposed in the literature. The first category consists of agglomeration based on *sharing* indivisible inputs, productivity gains derived from a large variety of intermediate inputs, specialization and cost reduction through learning-by-doing, or sharing risks. Second, agglomeration economies can also be based on *matching* producers and consumers more efficiently. Finally, the third category is based on the creation and diffusion of *knowledge*. Our skill agglomeration mechanism is driven by endogenous redistribution and, in a sense, belongs to the category of agglomeration based on *sharing*. Taking as given the demand for income redistribution in a region, an inflow of skilled (rich) workers increases the tax base. This allows for improved sharing of the redistribution burden, allowing for lower taxes. As noted earlier, in addition to this tax-base effect skilled in-migration will also have an effect on the desired size of redistributive transfers.

Our paper is also related to the vast public-economics literature on taxation of mobile factors and, in particular, labor. Since Tiebout (1956), economists and political scientists have been interested in the consequences of policy competition on efficiency in environments with mobile factors. Tiebout (1956) suggested that voter mobility will induce local governments to efficiently supply public goods. In a different vein, the highly influential work of Oates (1972) forcefully argued that policy competition would lead to a *race to the bottom* in taxation. Zodrow and Mieszkowski (1986) formalize the argument in a model with distortionary capital taxes. Their work has contributed to the widely spread belief that factor mobility will induce underprovision of public goods. However, other authors have qualified this prediction. Epple and Romer (1991) argue that significant local redistribution can be sustained once voters internalize the impact of migration flows on the price of a fixed factor. Cai and Treisman (2005) note

[2] Symmetry-breaking arises in several economic environments. See Matsuyama (2002) and Mookherjee and Ray (2003), and references therein.

that regional heterogeneity may lead to divergent tax policies, rather than convergence to a lower bound. Our paper is also related to the work of Fernandez and Rogerson (2003) and others on education finance systems. Typically, these studies exogenously restrict the government's redistribution technology, which helps in making quantitative predictions. Finally, our paper is also related to Armenter and Ortega (2009). They extend the model presented here by allowing both skilled and unskilled workers to be geographically mobile. Their model needs to be solved numerically and the results depend crucially on the pattern of regional heterogeneity fed into the model. They calibrate the model using data for US states.

The paper is organized as follows. In the next Section we introduce the model and then analyze optimal redistribution policy in a closed economy in Section 3. Section 4 takes on labor mobility and our definition of a equilibrium. Our analysis of a equilibrium with credible policies and its most important properties is in Section 5. Section 6 contains a two-region economy illustrating the results. Section 7 concludes.

2 The Economy

We consider a world economy consisting of $R = \{1, 2, ..., N\}$ regions. In each region $r \in R$, there are two types of workers: unskilled and skilled, denoted by subscripts $i = 1$ and $i = 2$, respectively. Each region r starts with a measure $e_i^r > 0$ of workers of each type. After all migration decisions have been made, the measure of workers of type i in region r is denoted n_i^r.

Definition 1 *A workers' distribution* $n = \{n_1^r, n_2^r\}_{r \in R}$ *is feasible if*

$$\sum_{r \in R} n_i^r = \sum_{r \in R} e_i^r \tag{1}$$

for $i = 1, 2$ *and* $n_i^r \geq 0$ *for all* $r \in R$, $i = 1, 2$.

Let non-negative vector $x^r = (c_1^r, c_2^r, l_1^r, l_2^r)$ denote an allocation for region r where c_i^r and l_i^r denote consumption and hours worked by an agent of type i in region r. Each worker owns one unit of time to be used for work or leisure. We let $x = \{x^r\}_{r \in R}$ be a world allocation. In the tradition of the political economy of taxation pioneered by Metzler and Richard (1981) we allow workers to make labor supply choices. Specifically, we assume that the preferences of both types of workers are represented by a common utility function defined over consumption and hours of work, $U(c_i, l_i)$. To save on notation we shall often write $U(x_i^r)$ with the understanding that $x_i^r = (c_i^r, l_i^r)$. Utility function $U(c_i, l_i)$ is assumed to be differentiable and strictly concave, with $U_c > 0$, $U_l < 0$, $U_{cc} < 0$, and $U_{ll} < 0$. We also assume a non-positive cross-derivative: $U_{cl} \leq 0$.[3].

[3]That is, we assume that consumption and leisure are complements. Note that the case of separable utility is included.

Under these assumptions, indifference curves in the labor-consumption, (l, c), space are increasing and strictly convex.

Production is carried out by combining two factors of production, skilled and unskilled labor, as summarized by the region-specific production function $F^r(n_1^r l_1^r, n_2^r l_2^r)$. We assume that function F^r is differentiable, constant returns to scale, strictly concave, and satisfies $F_{12}^r > 0$ as well as the appropriate Inada conditions. Since all regions produce the same undifferentiated good there is no reason to trade.[4] We can now define feasible allocations.

Definition 2 *An allocation x^r is feasible given (n_1^r, n_2^r) if*

$$n_1^r c_1^r + n_2^r c_2^r \leq F^r(n_1^r l_1^r, n_2^r l_2^r) \tag{RC}$$

and non-negativity constraints on hours and consumption.

We assume that unskilled workers can only supply unskilled labor as they are not qualified to perform certain tasks. Skilled workers, though, can work both in skilled and unskilled tasks. As we explain in detail below, the government cannot discriminate by worker type, which is unobservable. We show below that the incomplete information constrains the degree of income redistribution.

Throughout the paper, we restrict our attention to economies where the marginal product of skilled labor is higher than the marginal product of unskilled labor. In other words, skilled workers always earn a higher pre-tax wage rate. More specifically, define the skilled-to-unskilled worker ratio by $\eta^r = \frac{n_2^r}{n_1^r}$. We will concentrate on situations where skilled workers are relatively scarce. Formally, we make restrictions on primitives so that in equilibrium $\eta^r \leq \bar{\eta}^r$, for all regions, where $\bar{\eta}^r$ is given by

$$F_1^r(1, \bar{\eta}^r) = F_2^r(1, \bar{\eta}^r) - \varepsilon \tag{2}$$

for some small $\varepsilon > 0$. We show below that, for any skill ratio $\eta^r \leq \bar{\eta}^r$, the marginal product of skilled labor will be above the marginal product of unskilled labor. It follows that allocations inducing skilled workers to take on unskilled jobs are inefficient.

3 Optimal Redistribution in a Closed Economy

Redistribution policy in our model is endogenously determined as the decision of a regional fiscal authority which looks after the welfare of its residents. We start by studying the problem of optimal redistribution policy for a given workforce (n_1, n_2). To ease notation we drop the superscripts indexing each region.

We do not exogenously restrict the tax instruments available to the fiscal authority. In particular, we allow for non-linear tax schedules and hence progressive income taxation. We assume, though, that worker's types are unobservable so the tax schedule can only be a function of the task performed by

[4] The production function can also be viewed as the reduced form of a more general production function with additional factors (capital) that are perfectly mobile across regions. The key is to assume that each region faces a perfectly elastic supply of each of those factors.

each worker and the associated earnings. This constrains redistribution policy. Since skilled workers can perform unskilled tasks, a very aggressive redistribution policy would lead skilled workers to take up unskilled tasks.

We state the optimal redistribution policy problem as a classic Mirrlees (1971) direct taxation problem. This reduces the problem to choosing feasible allocations subject to a set of incentive compatibility constraints. These constraints ensure that all workers truthfully reveal their type. In our case, only skilled workers can mislead the government. Hence, the only incentive-compatibility constraint is that a skilled worker cannot be worse off than an unskilled worker.

Definition 3 *A feasible allocation $x = (c_1, l_1, c_2, l_2)$ is incentive compatible if*

$$U(c_1, l_1) \leq U(c_2, l_2). \tag{IC}$$

Our incentive compatibility constraint is different from the usual formulation in Mirrlees' problems because we assume that the task choice (working in a skilled or unskilled task) is observable. This approach retains the efficiency-redistribution trade-off while being very convenient when we extend the analysis to multiple regions. The optimal redistribution policy problem requires picking the incentive-compatible allocation providing the highest social welfare given the current workforce (n_1, n_2). Constant returns to scale in production allows us to formulate the problem in terms of the skilled-unskilled ratio η. We label the resulting allocation as *second-best*.

Definition 4 *An allocation $x = (c_1, l_1, c_2, l_2)$ is second-best given $\eta = \frac{n_2}{n_1}$ if it solves*

$$\max\{U(c_1, l_1) + \eta U(c_2, l_2)\} \tag{SBP}$$

subject to

$$c_1 + \eta c_2 \leq F(l_1, \eta l_2),$$
$$U(c_1, l_1) \leq U(c_2, l_2).$$

and non-negativity constraints for x.

3.1 Redistribution with full information

Before documenting the properties of second-best allocations, examining the problem of optimal redistribution under complete information (first-best allocation) provides a useful benchmark. In our setup this implies simply dropping the incentive-compatibility constraint (IC), which clearly improves the government's redistribution technology.

Definition 5 *We say that an allocation x is first-best given $\eta = \frac{n_2}{n_1}$ if it solves*

$$\max_x U(c_1, l_1) + \eta U(c_2, l_2) \tag{FBP}$$

subject to (RC) and non-negativity constraints for x.

Problem (FBP) is a standard concave program over a convex set. It follows that the first-order conditions are necessary and sufficient:

$$U_c(c_1, l_1) = U_c(c_2, l_2)$$
$$MRS_i(c_i, l_i) = F_i(l_1, \eta l_2)$$

for $i = 1, 2$, where $MRS_i = -U_l(c_i, l_i)/U_c(c_i, l_i)$. Not surprisingly, in the first-best allocation the marginal utility of consumption is equalized across worker types. In addition, the marginal rate of substitution between labor and consumption for each type of worker is equalized to the corresponding marginal product of labor.

The first-best allocation will not equate the welfare of both types of workers. As long as skilled labor is more productive, skilled workers will be called to supply more work hours. Generally speaking, equating the marginal utility of consumption will not "compensate" skilled workers for having less leisure. The next proposition states that in first-best allocations skilled workers are strictly worse off than unskilled workers.

Proposition 1 *Let $\eta \leq \bar{\eta}$ and x be a first best allocation given η. Then unskilled workers enjoy higher welfare than skilled workers:*

$$U(c_1, l_1) > U(c_2, l_2). \tag{3}$$

Proof. In the Appendix ■

It is instructive to illustrate this Proposition with an example. Suppose that utility is separable in consumption and labor supply. Now recall that the first-best allocation calls for complete equalization of the marginal utility of consumption across the two skill types. Of course under separable utility this implies equalizing consumption levels. The first-best allocation also requires equating the marginal rate of substitution to the marginal product for each type of labor. Under our assumptions on utility and technology, the government chooses an allocation requiring skilled workers (who have a higher marginal product) to supply more work than unskilled workers. As a result, the level of utility enjoyed by unskilled workers surpasses the utility from the consumption-leisure bundle targeted to skilled workers.

Such an extreme degree of redistribution may not be feasible. In particular, there are natural constraints on taxation arising from incomplete information. It is important to show that our skill agglomeration mechanism operates even in the presence of such constraints on redistribution.

3.2 Redistribution under incomplete information

As we have just seen, first-best allocations are not incentive-compatible. This implies that the informational friction (types are not observable) effectively constrains the government's ability to redistribute income from rich to poor. We now characterize second-best allocations, which turns out to be remarkably simple. Policy models with linear tax rates are often hindered by implementability constraints shaping non-convex choice sets. In contrast, we can assert the necessity and sufficiency of the first-order conditions associated to the second-best problem (SBP).

Proposition 2 *The first-order conditions associated with problem (SBP) are necessary and sufficient to characterize second-best allocations.*

Proof. In the Appendix. ■

Proposition 1 already made clear that first-best allocations are not incentive compatible. We next show that the incentive constraint (IC) is binding in second-best allocations.[5]

Proposition 3 *Let $\eta \leq \bar{\eta}$. Then for any second-best allocation x, the incentive compatibility constraint (IC) is binding:*

$$U(c_1, l_1) = U(c_2, l_2). \tag{4}$$

Proof. Assume otherwise. Then the necessary first-order conditions for the first and second-best allocations coincide. By the sufficiency of both sets of conditions (see Proposition 2), it implies second-best allocations are also solutions to the FBP problem. But Proposition 1 implies $U(c_1, l_1) > U(c_2, l_2)$, violating the incentive compatibility constraint ■

Equipped with Propositions 2 and 3, we can characterize the second-best allocation with just four equations. The first-order conditions associated with problem (SBP) yield that the labor supply is not being distorted, that is,

$$MRS_i(c_i, l_i) = F_i(l_1, \eta l_2) \tag{5}$$

for $i = 1, 2$. These two conditions in addition to the binding resource (RC) and incentive-compatibility (IC) constraints fully characterize the set of second-best allocations as long as $\eta \leq \bar{\eta}$.

Let us now highlight some properties of second-best allocations. First, we confirm that there is a positive skill premium, that is, the marginal product of a skilled worker is higher than that of an unskilled worker. Second, skilled workers work more than unskilled ones, and are "compensated" with higher consumption. Their consumption, though, is strictly less than their pre-tax labor income. In other words, the regional government redistributes income from rich to poor residents. The degree of redistribution implicit in second-best allocations is the maximum feasible in the presence of the informational constraint.

Proposition 4 *Let $\eta \leq \bar{\eta}$. In any second-best allocation x given η,*

[5] Under inelastic labor supply first-best and second-best allocations coincide. In addition, consumption and utility levels are equated between skilled and unskilled individuals in the same region. Intuitively, taxation is not distortionary hence the regional government redistributes income aggressively to the point of equating after-tax income and consumption levels.

1. *There is a strictly positive skill premium:*

$$F_1(l_1, \eta l_2) < F_2(l_1, \eta l_2). \tag{6}$$

2. *Skilled workers consume more ($c_2 > c_1$) and supply more labor ($l_2 > l_1$) than unskilled workers.*

3. *Skilled worker consumption c_2, which is equal to after-tax income, is strictly less than pre-tax income:*

$$c_2 < F_2(l_1, \eta l_2) l_2. \tag{7}$$

We conclude this section by providing a result that will be key in the multi-region analysis. In a no-redistribution (laissez-faire) economy an inflow of skilled workers increases the income (and thus welfare) of unskilled workers because of complementarity in production. At the same time the income and welfare of skilled workers decreases. However, this is not the case in second-best allocations. *Both* types of workers benefit from an inflow of the scarce skilled labor.[6] This result is formalized in the next proposition.

Proposition 5 *Let $\eta < \eta' \leq \bar{\eta}$ and let x and x' be second-best allocations given η and η', respectively. Then*

$$U(c_2, l_2) < U(c_2', l_2') \tag{8}$$

and

$$U(c_1, l_1) < U(c_1', l_1'). \tag{9}$$

Proof. In the Appendix. ■

The intuition for the result is the following. In the second-best allocation the government is taxing skilled workers to the limit, that is, the incentive-compatibility constraint is binding. An inflow of skilled workers into the region produces a surplus that allows the government to increase the level of consumption of the relatively poor unskilled individuals. To avoid skilled workers from switching to unskilled tasks the government needs to raise their welfare as well.[7] In the next section we present a particular decentralization that shows explicitly how the government adjusts taxation in response to skilled in-migration.

[6] The distinction between redistribution allocations (first-best or second-best) and the laissez-faire allocation plays a fundamental role in the public debate over immigration policy. It is widely recognized that immigration flows with an arbitrary skill composition deliver an aggregate *immigration surplus* (Berry and Soligo 1969, Borjas 1994). However, in the absence of appropriate income redistribution mechanisms (e.g. laissez-faire) any immigration flow will generate winners and losers. As a result, the political economy of immigration policy is highly contentious (Benhabib 1996, Ortega 2005, Ortega 2009).

[7] We note that the statement in the Proposition is also true if we replace second-best allocations by first-best allocations. Namely, an inflow of skilled workers leads to a change in the first-best allocations that entails higher utility for both types of workers. But, as already argued, the degree of income redistribution implicit in first-best allocations is unrealistically high: it makes skilled workers worse off than unskilled ones. The actual Proposition makes a stronger since the statement holds even when the government's redistribution ability is constrained by unobserved worker types.

3.3 Decentralization

Here we show that second-best allocations can be decentralized into a competitive equilibrium with lump-sum taxes. Hence, there is no need for distortionary taxation in our economy despite the binding informational friction.

Proposition 6 *Let x be a second-best allocation given $\eta \leq \bar{\eta}$. Then there exists a lump sum tax τ and wages rates (w_1, w_2) such that allocation x can be decentralized as a competitive equilibrium defined by*

1. *Pair (c_1, l_1) solves the unskilled household problem:*

$$\max U(c_1, l_1) \;\; s.t. \;\; c_1 \leq w_1 l_1 + \eta\tau, \tag{10}$$

with $c_1 \geq 0$, $l_1 \geq 0$.

2. *Pair (c_2, l_2) solves the skilled household problem:*

$$\max U(c_2, l_2) \;\; s.t. \;\; c_2 \leq w_2 l_2 - \tau \tag{11}$$

with $c_2 \geq 0$, $l_2 \geq 0$.

3. *Wages equal marginal products:*

$$w_1 = F_1(l_1, \eta l_2),$$
$$w_2 = F_2(l_1, \eta l_2).$$

Proof. In the Appendix. ■

We will use later the definition of competitive equilibrium in Proposition 6. We conclude with several observations. The decentralization of second-best allocations in the previous proposition is not unique, as it depends on the tax instruments available to the government. That is the main reason for stating our main results in terms of allocations. The advantage from assuming a particular tax system is that it helps provide intuition for our results. For instance, we can rephrase the statements in Proposition 6 as follows: an inflow of skilled workers ($\eta < \eta'$) leads to higher unskilled wages and lower skilled wages ($w_2/w_1 > w_2'/w_1'$). With a *fixed* tax τ on skilled tasks, this change in wages would lead to a situation where unskilled workers enjoy higher utility than skilled ones. In order to avoid violating the incentive-compatibility constraint the government needs to reduce the tax on workers performing skilled tasks ($\tau > \tau'$). The tax cut need not imply a fall in the transfer received by each individual performing unskilled tasks ($\eta\tau$). This is because the lower tax on skilled tasks may be offset by the increase in the tax base, that is, the number of individuals working in skilled tasks.

4 Labor Mobility and Equilibrium

We consider an environment where skilled workers are more geographically mobile than unskilled ones. This asymmetry has substantial empirical support.[8] For simplicity we make the extreme assumption that unskilled workers have prohibitively high migration costs.

Specifically, we assume that each skilled worker in region r receives an *opportunity to migrate*, (r', m), specifying a destination region $r' \neq r$ and a migration cost m in terms of utility. Each region generates migration opportunities equally, that is, a fraction $1/(N-1)$ of skilled workers born in state r receive opportunities to migrate to each other region r'. Hence, the number of potential migrants from r to $r' \neq r$ is given by

$$\frac{e_2^r}{N-1}. \tag{12}$$

Migration cost m is idiosyncratic, drawn from a distribution with cumulative distribution function $D(m)$ with $D'(m) > 0$ for all $m \geq 0$ and $D(0) = 0$. In equilibrium if a skilled worker born in region r with migration opportunity $(r', \bar{m})$ chooses to migrate, all other workers from the same region with opportunities (r', m) entailing lower migration costs, $m \leq \bar{m}$, will migrate as well.[9]

Let $\delta(r, r')$ be the fraction of potential migrants that actually move from r to r'. The whole matrix of (gross) migration flows from one region to the others can be summarized by function $\delta : R^2 \to [0, 1]$ where $\delta(r, r) = 0$. For each pair of regions (r, r'), we can define the migration cost paid by the marginal migrant as follows,

$$\mu(\delta(r, r')) \equiv D^{-1}(\delta(r, r')). \tag{13}$$

We note that $\mu(x)$ is unbounded as $x \to 1$, is differentiable for $x > 0$, and $\mu'(x) > 0$. An environment with higher mobility costs implies a higher $\mu(x)$ for all $x > 0$.

Given migration flows δ, the native skilled workforce that remains in region r is given by

$$e_2^r \left(1 - \frac{1}{N-1} \sum_{r' \in R} \delta(r, r')\right). \tag{14}$$

Total inflows into the region are given by

$$\sum_{r' \in R} \frac{\delta(r', r)}{N-1} e_2^{r'}. \tag{15}$$

[8] Bound and Holzer (2000) and Chiquiar and Hanson (2005) provide evidence of lower migration costs for skilled workers within US states and internationally, respectively.

[9] It is possible to entertain richer environments but only at a significant cost in terms of tractability. Armenter and Ortega (2009) analyze a richer environment where both skilled and unskilled workers receive opportunities to migrate, solving for the equilibrium numerically. Provided that average migration costs are higher for unskilled workers their economy behaves in a similar manner to the one analyzed here.

Hence, the final skilled workforce in region r is

$$n_2^r = e_2^r \left(1 - \frac{1}{N-1} \sum_{r' \in R} \delta\left(r, r'\right)\right) + \sum_{r' \in R} \frac{\delta\left(r', r\right)}{N-1} e_2^{r'}. \qquad (16)$$

Before proceeding to our equilibrium with endogenous redistribution, it is useful to define an equilibrium for any given set of feasible policies $\{\tau_r\}_{r \in R}$.

Definition 6 *An equilibrium given policies $\{\tau_r\}_{r \in R}$ is a set of allocations $\{x_r\}_{r \in R}$, a pattern of migration flows $\delta : R^2 \to [0,1]$ and a worker distribution $\{n_1^r, n_2^r\}_{r \in R}$ with $\eta^r \leq \bar{\eta}^r$ for all $r \in R$, such that*

1. *For every $r \in R$, x_r is a competitive equilibrium given τ_r and $\{n_1^r, n_2^r\}$,*
2. *The worker distribution is feasible and is given by equation (16) and $n_1^r = e_1^r$ for all $r \in R$.*
3. *For each $r \in R$, all individually profitable moves from r to r' take place, that is,*
$$U(x_2^{r'}) - U(x_2^r) \leq \mu(\delta(r, r')), \qquad (17)$$
with equality if $\delta(r, r') > 0$, for all $r' \neq r$.

Condition 3 states the rationality of migration decisions. Migration takes place from region r to r' until the marginal migrant is indifferent. Obviously, there are no migration flows (from r to r') if migration is not profitable for a potential migrant with zero cost: $U(x_2^{r'}) - U(x_2^r) < 0$. Note that each individual migrant takes policies (allocations) as given.

In an equilibrium given policies, the bilateral gross flow between any two given regions are equal to the net flow.[10] That is, workers only flow in one direction between any pair of regions; in our notation, the product $\delta\left(r, r'\right) \delta\left(r', r\right) = 0$. It is possible, though, that a region may be attracting workers from one region while simultaneously shedding workers to other regions. Hence, overall gross flows in a region are not necessarily equal to net flows. Clearly, since migration costs are non-negative, workers only move to higher welfare regions (though not necessarily to the highest). Thus it is possible to order all regions according to their levels of utility. Bilateral migration flows only take place in one direction along this ranking.

Let us illustrate the implications of these observations using a simple example. Consider a three-region environment (regions 1, 2 and 3) and suppose that in equilibrium there are migration flows from region 2 to region 1 ($\delta(2,1) > 0$) and from region 3 to region 2 ($\delta(3,2) > 0$). Then it must be the case that there is migration from region 3 to region 1 as well ($\delta(3,1) > 0$). The reason is that equilibrium utilities in this example satisfy

$$U(x_2^1) > U(x_2^2) > U(x_2^3). \qquad (18)$$

[10] The models in Coen-Pirani (2006) and Lkhagvasuren (2006) generate gross flows that are larger than net flows.

Furthermore, it will be the case that some workers native to region 3 migrate to location 1 and others to region 2. Clearly, the argument can be extended to any number of regions. In general, we can always construct a ranking of regions (possibly with ties) in terms of (skilled) worker welfare. It will always be the case that a region with a given equilibrium welfare level will suffer outflows toward all regions ranking higher in terms of welfare.

We conclude the section by stating the uniqueness of the equilibrium for exogenously given policies $\{\tau_r\}_{r \in R}$. As will be clear later, when redistribution policies are endogenous there will often be multiple equilibria.

Proposition 7 *There is a unique equilibrium given feasible policies* $\{\tau_r\}_{r \in R}$.

Proof. In the Appendix. ■

5 Credible Redistribution Policies

We start this section by defining our concept of equilibrium under worker mobility when redistribution policies are endogenous. Essentially, a policy equilibrium requires workers to make individually optimal migration decisions and the resulting allocations to be second-best given the final workers' distribution. It is useful to visualize our equilibrium concept as a sequential game. First, workers decide whether to migrate and then each region decides its redistribution policy. The requirement that the allocations be second-best is akin to subgame perfection. It rules out situations where a region makes non-credible redistribution promises in order to attract skilled workers.

Definition 7 *A credible policy equilibrium is an equilibrium given policies* $\{\tau_r\}_{r \in R}$ *such that for every* $r \in R$ *allocation* x^r *is second best given* $\{n_1^r, n_2^r\}$.

The restriction to credible redistribution policies prevents a "race to the bottom" where regions keep undercutting each others' taxes to attract skilled workers. Since allocations in a credible policy equilibrium are second-best, Proposition 4 implies a positive tax and, indeed, as much redistribution as is allowed by the informational constraint.

There will typically be multiple credible policy equilibria. This follows from Proposition 5: the welfare of mobile (skilled) workers implied by second-best allocations is increasing in the skilled-unskilled ratio. To see this consider a symmetric two-region world with relatively low mobility costs. Clearly, there is a credible policy equilibrium where both regions feature the same second-best allocation. In this case, there is no migration since skilled workers get exactly the same utility in both regions. Yet other credible policy equilibria are possible. Say some skilled workers move from one region to the other. The region receiving the inflows experiences an increase in its skilled-unskilled ratio. As a result, (skilled) utility increases in that region, which validates individual migration choices. The process continues until, eventually, migration costs become high enough so as to eliminate the incentive to migrate. At equilibrium the two regions differ in their workforce composition and redistribution levels. The region receiving skilled inflows ends up with higher welfare for all its residents, a lower skilled-unskilled wage ratio, and a lower lump-sum tax on skilled workers.

We argue that the no-flows equilibrium is unstable when migration costs are low. In order to refine our equilibrium concept, we introduce a definition of local stability. Loosely speaking, imagine a small measure of skilled workers accidentally moving to another region. A candidate equilibrium worker allocation will be stable if the displaced individuals end up regretting the move, that is, if their increase in utility is lower than the mobility cost they incur.

Definition 8 *A credible policy equilibrium* $\{x, n, \delta, \tau\}$ *is locally stable if, for any pair of regions* $r, r' \in R$*, there exists* $\bar{\varepsilon} > 0$ *such that for all* $\varepsilon \in (0, \bar{\varepsilon})$*,*

$$U\left(\tilde{x}_2^{r'}\right) - U(\tilde{x}_2^r) \leq \mu\left(\delta\left(r, r'\right) + \varepsilon\right) \tag{19}$$

where bundles $\tilde{x}^r$ *and* $\tilde{x}^{r'}$ *are second-best given* $(n_1^r, n_2^r - \varepsilon e_2^r)$ *and* $\left(n_1^{r'}, n_2^{r'} + \varepsilon e_2^r\right)$*, respectively.*

5.1 Symmetry-breaking

Under low mobility costs, stable credible policy equilibria are symmetry-breaking, that is, regions with identical fundamentals necessarily end up with different allocations.[11]

Proposition 8 *Consider* R *identical regions, that is,* $e_1^r = e_1^{r'}$*,* $e_2^r = e_2^{r'}$*,* $F^r = F^{r'}$ *for all* $r, r' \in R$*. For mobility costs small enough, the symmetric distribution of workers,* $n^r = n^{r'}$ *for all* $r, r' \in R$*, is not a locally stable credible policy equilibrium.*

Proof. In the Appendix. ■

Going back to our symmetric two-region example, the migration of a small measure of workers would raise welfare in their destination region. If the welfare gain is larger than the mobility cost, then the symmetric equilibrium is not stable and one of the two regions will necessarily feature a higher skill ratio and a higher level of utility. The next section provides an example illustrating this result.

5.2 Amplification of Differences

In a stable credible policy equilibrium allocations will differ across (ex ante identical) regions. However, we do not know which regions will gain and which ones will lose from worker mobility. This is obvious in the case of identical regions but even when the environment is asymmetric our definition of local stability leaves the door open to equilibrium indeterminacy.

[11] Obviously, with high enough mobility costs, the unique equilibrium is symmetric and there is no migration.

We present an equilibrium selection based on a simple tâtonnement argument. Imagine there are just two regions, A and B, with the same initial labor endowments. Region A has a technological advantage such that, in autarky, skilled workers are strictly better off than in region B. In this situation there are two locally stable equilibria. In one equilibrium region A gains some skilled workers; in the other, region B gains enough skilled workers to overcome its technological disadvantage. One can think of the tâtonnement refinement as follows: the equilibrium must be achieved by specifying a sequence of arbitrarily small sets of workers; for each set in the sequence, workers always move to the region with higher welfare at that point. The equilibrium with flows from A to B can be reached no matter how small the groups are. In contrast, the second equilibrium requires that a large group moves to the region with lower ex-ante welfare, region B, in order to turn around the welfare ranking of regions.

Let us now formalize this idea. In essence, an equilibrium will be admissible if one can construct a sequence of worker distributions and second-best allocations converging to it where no skilled worker ever moves toward a region with lower welfare.

Definition 9 *A credible policy equilibrium $\{x, n, \delta, \tau\}$ is admissible if there exists a sequence of feasible worker distributions, allocations and migration flows $\{n_j, x_j, \delta_{j+1}\}_{j=0}^{\infty}$ such that*

1. *Allocations x_j are second best given n_j for all $j \geq 0$.*

2. *Migration flows satisfy*
$$U\left(x_j^r\right) \leq U\left(x_j^{r'}\right) \tag{20}$$
if $\delta_{j+1}\left(r, r'\right) > 0$ for all pairs $r, r' \in R$ and all $j \geq 0$.

3. *The worker distribution satisfies*
$$n_{2j}^r = e_2^r \left(1 - \frac{1}{N-1} \sum_{r' \in R} \delta_j\left(r, r'\right)\right) + \sum_{r' \in R} \frac{\delta_j\left(r', r\right)}{N-1} e_2^{r'} \tag{21}$$
for $j \geq 0$ with $n_0^r = e_0^r$ for all $r \in R$.

4. *Allocations converge to x.*

Equipped with the admissibility refinement, we can now state the second key property of the model: admissible equilibrium outcomes always amplify initial differences in primitives. Following up on our previous example, in the only admissible equilibrium high-TFP region A increases its skill ratio at the expense of region B. The initial welfare gap between the two regions is exacerbated by the inflows of skilled workers from the low to the high-TFP region. The next Proposition formalizes this idea in a two-region setup.

Proposition 9 *Consider two regions $R = \{A, B\}$ with*

$$U\left(\tilde{x}_2^A\right) > U\left(\tilde{x}_2^B\right) \tag{22}$$

where $\tilde{x}_2^r$ is the bundle for skilled workers in the second-best allocation given labor endowments (e_1^r, e_2^r). Any admissible equilibrium $\{x, n, \delta, \tau\}$ features

$$U\left(x_2^A\right) - U\left(x_2^B\right) \geq U\left(\tilde{x}_2^A\right) - U\left(\tilde{x}_2^B\right) \tag{23}$$

with strict inequality if $n \neq \left(e_1^A, e_2^A, e_1^B, e_2^B\right)$.

Proof. In the Appendix. ■

In a generic economy with R economies, not every pair of regions will necessarily diverge in welfare terms. It can be shown, though, that the welfare differences between the top and bottom regions will increase.

Observe that when regions only differ in their technologies, worker mobility improves overall production efficiency; workers flow to the region with superior technology, amplifying cross-region differences. When regions only differ in their labor endowments, (skilled) workers flow away from the region where they are scarce. In this case, worker mobility widens the gap between the marginal products of skilled labor in the two regions. We return to this point in the Conclusions, where we discuss welfare implications. We also note that our finding of amplification of differences is reminiscent of Cai and Treisman (2005), in the context of capital mobility. However, in their economy there is divergence only when initial asymmetries are large enough. In contrast, endogenous policy induces symmetry-breaking in our environment.

Throughout we have assumed that only skilled workers are geographically mobile. We note that the results on symmetry-breaking and welfare divergence can be generalized to a setup where unskilled workers are also mobile, as long as labor flows are skill-biased, that is to say, as long as skilled workers are overrepresented in migration flows (Armenter and Ortega 2009).[12]

6 A Two-region Example

We present here a two-region economy $R = \{1, 2\}$ that illustrates the main results of the paper. First, we consider how allocations and taxes vary with the skill ratio. Then we solve numerically for credible redistribution policy equilibria.

[12] If we assume instead that only unskilled workers are mobile several features of our equilibrium are overturned. In the case of identical regions case, the symmetric equilibrium is now locally stable: any small migration by unskilled workers would lead to a reduction in the destination region skilled-unskilled ratio. As a result, these individuals would regret the move, even if migration costs were zero. In a world with asymmetric regions, workers still flow to the region with higher utility at autarky. But now migration flows tend to reduce interregional differences in income and utility.

6.1 Specification

Let us assume separable log utility:

$$U(c,l) = \log(c) + \psi \log(1-l) \tag{24}$$

for $\psi > 0$. Preferences are assumed to be identical across regions. The production function is CES:

$$F^r(l_1, l_2) = \theta^r \left[(1-\alpha^r) l_1^\rho + \alpha^r l_2^\rho\right]^{1/\rho} \tag{25}$$

with $\rho \in (0,1)$, $\alpha^r \in (0,1)$ and $\theta^r > 0$. We consider two dimensions of heterogeneity across regions: differences in overall labor productivity, θ^r, and in the relative demand for skilled labor, α^r. Our baseline economy features symmetric regions, with $\theta^r = 1$, $\alpha^r = .44$, and $\rho = .4$.[13]

Throughout the exercise, we assume that both regions are endowed with one unit of unskilled workers: $e_1^1 = e_1^2 = 1$. We assume that mobility costs are drawn from a Pareto distribution. In that case the mobility cost function is then given by

$$\mu(\delta) = (1-\delta)^{-k} - 1 \tag{26}$$

with $k > 0$.

6.2 Allocations in a Closed Economy

We first compute second-best allocations as a function of the skill ratio $\eta = n_2/n_1$. We also take a look at the corresponding decentralization as a competitive equilibrium. These comparative statics are at the core of the mechanics of our endogenous policy equilibrium.

Figure 1 plots the second-best allocation as a function of the skilled-unskilled ratio. Respectively, in the first two graphs the solid line corresponds to unskilled worker's consumption and labor supply, the dashed line are the analogous variables for skilled workers. As stated in Proposition 4, skilled workers consume and work more than unskilled workers. As the skilled-unskilled ratio increases, skilled labor becomes less scarce and the gap between the bundles for the two types of workers narrows. The third graph plots the wage rates. We note that there is a positive skill wage premium and the premium is decreasing in the skilled-unskilled ratio.

We point out that, following an inflow of skilled workers and keeping constant the tax on the income from skilled tasks, the combination of the fall in skilled labor supply and the fall in the wage rate would lead to a steep reduction in the consumption and utility of skilled workers. However, Proposition 5 makes clear that skilled worker's welfare is *increasing* in the skilled-unskilled ratio. This can be seen in the first graph in Figure 1, which reveals that skilled consumption

[13] These parameter values are based on the calibration in Armenter and Ortega (2009) using data for US states.

in second-best allocations is an increasing function of the skilled-unskilled ratio despite the decreasing pre-tax labor income for skilled workers.[14]

The second row of graphs in Figure 1 displays output per worker, the tax rate on workers performing skilled tasks, and transfers per recipient (that is, individuals in unskilled jobs). Output per worker increases in the skilled-unskilled ratio. Both taxes and transfers decrease with the skill ratio: as the wage gap disappears, so does the demand for tax-based redistribution. We also note that the schedule for the transfer per recipient is flatter than the one for the tax rate, reflecting the increase in the tax base as the skilled-unskilled ratio increases.

6.3 Credible policy equilibrium

We start by computing the equilibrium in the symmetric case, $F^1 = F^2$ and $e_2^1 = e_2^2 = 0.5$. This will illustrate the symmetry-breaking property of credible policy equilibria with low mobility costs. Figure 2 introduces a useful diagram in the case of two regions. In the horizontal axis we have the skilled-unskilled ratio of region 1, $\eta^1 = n_2^1/n_1^1$. If no workers move, the skilled-unskilled ratio in both regions is equal to 0.5 and it is indicated with a vertical dotted line. If the skilled-unskilled ratio in region 1 increases, the skilled-unskilled ratio for region 2 falls. The equilibrium mobility condition implies that $\eta^2 = 1 - \eta^1$. The solid line computes the difference in welfare between the two regions at each given skilled-unskilled ratio

$$DU\left(\eta^1\right) = U^1\left(c\left(\eta^1\right), l\left(\eta^1\right)\right) - U^2\left(c\left(1-\eta^1\right), l\left(1-\eta^1\right)\right). \tag{27}$$

This is an increasing function of η^1, as welfare in region 1 increases with the skilled-unskilled ratio in the region. The function takes value 0 at $\eta^1 = 0.5$ because of the symmetry of the two regions.

The dashed line is the cost of migration of the marginal migrant at each given skilled-unskilled ratio η^1. We express the mobility cost in terms of the welfare in region 2, so we use negative numbers when the workers in region A are incurring in the mobility costs. In other words, the mobility equilibrium condition is satisfied when the solid and dashed lines cross, $DU\left(\eta^1\right) = \mu\left(\eta^1 - .5\right)$.

Figure 2 depicts the case of high mobility costs. Note that the only equilibrium involves zero flows as welfare differences are not enough to compensate for mobility costs at any skill ratio. Figure 3 displays the economy with low mobility costs. Now the solid and dashed lines intersect at three points. First, there is an equilibrium, indicated by letter A, with zero flows: the two countries are ex-ante identical so there are no gains from migration even for individuals with zero mobility costs. However, there are two additional equilibria, B and B'.[15] In these equilibria skilled workers migrate up to the point where the utility gain equates the mobility cost of the marginal migrant. In these two equilibria the

[14] Whether skilled worker consumption actually increases depends on parameters. It is always the case that skilled workers increase their leisure as the skilled-unskilled ratio increases. If leisure is heavily weighted in the preferences, it is possible that consumption actually falls.

[15] Equilibrium B and B' are symmetric: $\eta_B^1 = \eta_{B'}^2$.

two regions display different utility levels despite being ex-ante identical. It is easy to see in Figure 3 why the symmetric equilibrium is not locally stable. Any small deviation in skilled-unskilled ratio η^1 would make it profitable to move into one region or the other, eventually converging to one of the asymmetric equilibria, B or B'.

Let us now illustrate the divergence result. Suppose region 1 has superior technology, that is, let $\theta^1 > \theta^2$. Clearly, this implies higher labor productivity in region 1 for both types at each skilled-unskilled ratio. Figure 4 illustrates this example, using $\theta^1 = 1.01\theta^2$. First, we note that there is no equilibrium with zero migration flows since region 1 delivers strictly higher welfare than region 2: $DU(0.5) > 0$. Hence, skilled workers wish to move and some of them have low enough migration costs to make migration profitable. Now the only credible policy equilibrium features skilled migration from region 2 to region 1, which widens the utility gap between the two regions. There is no guarantee, though, that the asymmetry leads to a unique equilibrium. Figure 5 shows an economy with technology differences featuring three equilibria.[16] Equilibrium B is locally unstable, but both equilibria A and C are locally stable. However, equilibrium A is intuitively more appealing: region 1, which had a technological advantage at equal skill ratios. attracts workers. Equilibrium B requires that a large mass of workers decide to migrate to region 2 despite its lower ex-ante welfare. Our equilibrium stability refinement selects equilibrium A.

The agglomeration mechanism is highlighted when regions differ only in the skill composition of the initial population. We assume region 1 is endowed with more skilled workers than region 2, $e_1^2 = 0.55 > e_2^2 = 0.45$, and assume that technologies are identical in the two regions ($F^1 = F^2$). This economy is depicted in Figure 6. We have shifted the axis to cross at $(0.55, 0)$, the point of zero migration flows, which is no longer an equilibrium. Because of the ex-ante advantage of a higher skilled-unskilled ratio, the only credible policy equilibrium leads to higher welfare in region 1 at the expense of region 2. In other words, skilled workers flow to region 1 despite i) skilled labor being initially more abundant in this region, ii) the lack of regional differences in technology, and iii) constant returns to scale in production. Clearly, it is the endogeneity of income redistribution policies that drives skill agglomeration here.[17]

7 Conclusions

We conclude with two final remarks. First, we briefly discuss the welfare consequences of tax competition among regional governments. Second, we discuss the implications of our analysis for the ongoing European integration process.

[16]To be more precise, we let $\theta^1 = 1.01\theta^2$ and $\alpha^1 = .43 < \alpha^2 = .44$.

[17]We also note that there is partial sorting of individuals by income/skills across regions. When migration costs are low, poor/unskilled workers (who benefit from redistribution) agglomerate in a region and rich/skilled workers (who bear the burden or redistribution) agglomerate in the other. Note also that redistributive taxation on skilled jobs is higher in the region where unskilled workers agglomerate. This is reminiscent of the sorting in Tiebout (1956), as pointed out by a referee.

The welfare implications of our analysis are ambiguous, in the sense that they are highly dependent on parameter values. Let us sketch though the main trade-off implicit in policy decentralization. As a benchmark we compare our equilibrium, which involves regionally decentralized redistribution policies, to an alternative scenario where a federal government chooses a common redistribution policy that applies throughout the federation. The federal government faces a federation-wide budget constraint. The degree of worker mobility, and all other primitives are identical in both scenarios.[18] For simplicity we assume that there is no room for regional authorities to top off or undo the federal policy. We first sketch a situation where a centralized tax authority can improve upon the decentralized outcome. Consider the case of two regions with identical labor endowments and identical technologies. As argued earlier, the decentralized equilibrium with credible policies leads to symmetry-breaking. The resulting regional inequality in utility levels is a source of inefficiency. Obviously, the centralized outcome delivers higher welfare in this situation. It is not hard though to come up with examples where regional decentralization delivers an allocation that is socially more desirable than the centralized outcome. Consider the case where the two regions differ only in their labor endowments. In particular, region A has a balanced skill distribution while in region B skilled workers are severely scarce. In this context, a centralized government will not be able to implement much redistribution from rich to poor because of the binding incentive constraint in region A. For appropriate parameter values, the centralized outcome entails a degree of within-region income inequality delivering lower social welfare than the decentralized allocation.

We now turn to a prediction of our theory that may be relevant in the context of the European integration process. Europe has taken bold steps toward liberalizing labor mobility in the last few decades. However, coordination of fiscal policies remains largely at the discretion of national governments and their electorate. Indeed, as of today, the size of the national welfare states differs widely across the European Union. An important question is whether integration will have an effect on the degree of income redistribution undertaken by each member state. It is often argued that increased labor mobility will eventually lead to a *race to the bottom* in redistribution levels, in the sense of dramatic downward convergence. As the argument goes, skill-scarce regions will sharply reduce income taxation in order to attract qualified workers. The remaining regions will then be forced to respond by reducing taxation themselves to avoid a damaging brain drain. Eventually, income redistribution will be reduced to its minimum everywhere, and no country will succeed in attracting skilled workers.

An implication of our analysis is that these aggressive cuts in redistribution policies are not credible: once workers have incurred the cost of relocation, governments would have a strong incentive to renege on their promises. Recognizing the constraints imposed by credibility, we think that a race to the

[18] We note that if the federal government is allowed to set a different policy for each region, a centralized tax authority can always improve upon the decentralized outcome. However, that is not a realistic scenario since it requires discriminating among individuals performing the same task because of their region of residence.

bottom in national-level redistribution policies is unlikely. Indeed it is possible that, over time, we observe a process of skill agglomeration and a widening of the dispersion in the size of national welfare states across the European Union. Of course, policy divergence and skill agglomeration are not granted, as they depend on parameters. We hope the framework laid out here provides the basis for careful quantitative studies addressing this important question.

References

[1] **Armenter, Roc and Ortega, Francesc.** "Credible Redistributive Policies and Migration across US States." *Working Paper, Universitat Pompeu Fabra,* 2009.

[2] **Benhabib, J.**, 1996. "On the Political Economy of Immigration." European Economic Review 40 (9), pp. 1737-1743.

[3] **Berry, R.A. and R. Soligo**, 1969, Some welfare aspects of international migration, Journal of Political Economy 71, 778-794.

[4] **Bewley, Truman F.** "A Critique of Tiebout's Theory of Local Public Expenditures." *Econometrica,* 1981, *49*(3), pp.713-740.

[5] **Borjas, G.**, 1994. "The Economics of Immigration," Journal of Economic Literature, December 1994, pp. 1667-1717.

[6] **Brown, C., Oates, W.**. "Assistance to the poor in a federal system." *Journal of Public Economics*, 32(3), pp. 307—330, 1987.

[7] **Bound, J., Holzer, H.J.** "Demand Shifts, Population Adjustments, and Labor Market Outcomes during the 1980s." *Journal of Labor Economics*, vol. 18, no. 1.

[8] **Buchanan, J.M. and Goetz, J.C.** "Efficiency Limits of Fiscal Mobility: An Assessment of the Tiebout Model." *Journal of Public Economics,* 1972, *1,* pp.25–44.

[9] **Cai, Hongbin and Treisman, Daniel.** "Does Competition for Capital Discipline Government? Decentralization, Globalization and Public Policy." *American Economic Review,* 2005, *95*(3), pp.817-30.

[10] **Chiquiar, Daniel and Hanson, Gordon.** "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States." textitJournal of Political Economy, vol. 113(2), pages 239-281, April 2005.

[11] **Coen-Pirani, Daniele.** "Understanding Worker Gross Flows Across U.S. States." Working Paper, Carnegie Mellon University, 2007.

[12] **Cremer, Helmut and Pestieau, Pierre.** "Social Insurance, majority voting, and labor mobility." Journal of Public Economics, 68. 1998.

[13] **Cremer, Helmut and Pestieau, Pierre.** "Factor Mobility and Redistribution: A Survey." Working Paper, University of Toulouse, 2003.

[14] **Duranton, Gilles and Puga, Diego.** "Micro-foundations of urban agglomeration economies," in J. V. Henderson and J.-F. Thisse (eds.) Handbook of Regional and Urban Economics, Vol. 4, 2004. Amsterdam: North-Holland, 2063-2117.

[15] **Eaton, Jonathan, and Eckstein, Zvi.** "Cities and Growth: Theory and Evidence from France and Japan," Regional Sci. and Urban Econ. 27 (August): 443474, 1997.

[16] **Epple, Dennis and Romer, Thomas.** "Mobility and Redistribution." Journal of Political Economy, 99 (4), 1991, pp.828-58.

[17] **Fernandez, Raquel and Rogerson, Richard.** "Equity and Resources: An Analysis of Education Finance Systems." Journal of Political Economy, 20003, 111: 858-897.

[18] **Hindricks, Jean.** "Mobility and Redistributive Politics." Journal of Public Economic Theory, 3 (1), 2001, pp.95-120.

[19] **Lkhagvasuren, Damba.** "Big Locational Differences in Unemployment Despite High Labor Mobility." Working Paper, Northwestern University, 2007.

[20] **Lucas, Robert E.** "Life Earnings and Rural-Urban Migration," Journal of Political Economy, 2004, vol. 112, no. 1.

[21] **Matsuyama, Kiminori.** "Explaining Diversity: Symmetry-Breaking in Complementarity Games." *American Economic Review*, 2002, *92*, pp.241-246.

[22] **Mirrlees, J.** "An Exploration in the Theory of Optimal Income Taxation." *Review of Economic Studies,* 1971*, 38,* pp.175-208.

[23] **Mookherjee, D. and Ray, D.** "Persistent Inequality," Review of Economic Studies 70, 369-393, 2003.

[24] **Moretti, Enrico.** "Human Capital Externalities in Cities," Handbook of Regional and Urban Economics, North Holland-Elsevier, 2004.

[25] **Oates, W.E.** Fiscal Federalism. Harcourt Brace Jovanovich, New York, 1972.

[26] **Ortega, F.**, 2005. "Immigration Quotas and Skill Upgrading in the Labor Force." Journal of Public Economics 89 (9-10), pp. 1841-1863.

[27] **Ortega, F.**, 2009. "Immigration, Citizenship, and the Size of Government." IZA DP 4528.

[28] **Peri, G.**. "International Migrations: Some comparisons and Lessons for the EU" in *The European Economy in an American Mirror*, B. Eichengreen, M. Landesmann and D. Stiefel Editors, Routledge, Taylor and Francis, Oxford UK, 2007.

[29] **Stiglitz, Joseph E.** "The Theory of Local Public Goods." in *The Economics of Public Services*, M.S. Feldstein and R.F.Inman, eds., London, 1977.

[30] **Tanaka, R.** "School Choice and Long-Run Inequality," *mimeo*, 2001.

[31] **Tiebout, Charles M.** "A Pure Theory of Local Expenditures." *Journal of Political Economy,* 1956, *64*(5), pp.416-424.

[32] **Wilson, J. and Wildasin, D.** "Capital tax competition: bare or boom," *Journal of Public Economics*, 88, 2004.

[33] **Zodrow, George R. and Mieszkowski, Peter.** "Pigou, Tiebout, Property Taxation, and the Underprovision of Local Public Goods." *Journal of Urban Economics,* 1986, *19,* p.356-70.

A Proofs

Proof of Proposition 1. Assume x is a first best allocation with $U(c_1, l_1) \leq U(c_2, l_2)$. Consider first the case $l_2 > l_1$. Then $U(c_1, l_1) \leq U(c_2, l_2)$ implies $c_2 > c_1$. Using the properties of U,

$$U_c(c_1, l_1) > U_c(c_2, l_1) \geq U_c(c_2, l_2) \tag{28}$$

so x does not satisfy the necessary first order conditions for first best allocations.

Consider now the case $l_1 = l_2$. Necessary first order conditions $U_c(c_1, l_1) = \lambda$ and $U_c(c_2, l_2) = \lambda$ imply $c_1 = c_2$. But since $F_1(1, \eta) < F_2(1, \eta)$, necessary first order conditions also require $-\frac{U_l(c_1,l_1)}{U_c(c_1,l_1)} < -\frac{U_l(c_2,l_2)}{U_c(c_2,l_2)}$. Hence x is not first best.

Finally, consider the case $l_1 > l_2$. Concavity implies that

$$\frac{1}{1+\eta} U(x_1) + \frac{\eta}{1+\eta} U(x_2) \leq U\left(\frac{1}{1+\eta} x_1 + \frac{\eta}{1+\eta} x_2\right). \tag{29}$$

Next we show that allocation $\tilde{x}$, given by

$$\tilde{x}_1 = \tilde{x}_2 = \frac{1}{1+\eta} x_1 + \frac{\eta}{1+\eta} x_2 \tag{30}$$

satisfies (RC) with a strict inequality sign. By construction, $c_1 + \eta c_2 = \tilde{c}_1 + \eta \tilde{c}_2$. On the production side, we have that $l_1 < l_2$ implies

$$\frac{F\left(1, \eta \frac{l_2}{l_1}\right)}{1 + \eta \frac{l_2}{l_1}} < \frac{F(1,\eta)}{1+\eta} \tag{31}$$

for $\eta < \bar{\eta}$ (to see this, differentiate $F(1, \eta)$ with respect to η). We can then rearrange

$$F\left(1, \eta \frac{l_2}{l_1}\right) < F(1,\eta) \frac{1 + \eta \frac{l_2}{l_1}}{1+\eta} \tag{32}$$

and multiplying both sides by l_1,

$$F(l_1, \eta l_2) < F(1,\eta) \frac{l_1 + \eta l_2}{1+\eta} = F\left(\tilde{l}_1, \eta \tilde{l}_2\right). \tag{33}$$

Therefore,

$$\tilde{c}_1 + \eta \tilde{c}_2 = c_1 + \eta c_2 = F(l_1, \eta l_2) < F\left(\tilde{l}_1, \eta \tilde{l}_2\right). \tag{34}$$

Summarizing, allocation $\tilde{x}$ is at least as good as x and satisfies the (RC) with a strict inequality sign. Hence x cannot be first best allocation ■

Proof of Proposition 2. The first order conditions of problem (SBP) are

$$\begin{aligned}
(1-\mu) U_c(c_1, l_1) &= \lambda \\
(1-\mu) U_l(c_1, l_1) &= -\lambda F_1(l_1, \eta l_2) \\
(\eta+\mu) U_c(c_2, l_2) &= \lambda \eta \\
(\eta+\mu) U_l(c_2, l_2) &= -\lambda \eta F_2(l_1, \eta l_2) \\
\lambda \left[c_1 + \eta c_2 - F(l_1, \eta l_2)\right] &= 0 \\
\mu \left[U(c_2, l_2) - U(c_1, l_1)\right] &= 0
\end{aligned}$$

for $\lambda \geq 0$ and $\mu \geq 0$.

Consider the alternative program

$$\max_{u_1, u_2, x} u_1 + \eta u_2 \tag{35}$$

subject to

$$\begin{aligned} u_1 &\leq u_2, \\ u_1 &\leq U(x_1), \\ u_2 &\leq U(x_2), \\ c_1 + \eta c_2 &\leq F(l_1, \eta l_2). \end{aligned}$$

We show that an allocation x is second best if and only if there exists u_1 and u_2 such that $\{u_1, u_2, x\}$ solve (35). If any solution $\{u_1, u_2, x\}$ to (35) satisfies $u_1 = U(x_1)$ and $u_2 = U(x_2)$, our claim follows trivially. Assume that x solves (35) but $u_1 < U(c_1, l_1)$ and $u_2 = U(c_2, l_2)$ (obviously $u_2 < U(c_2, l_2)$ will never be a solution). Construct now an alternative allocation with the same work hours but $u_1 = U(c_1', l_1)$, with $c_1' = c_1 - \varepsilon$, $c_2' = c_2 + \varepsilon/\eta$, and $u_2' = U(c_2', l_2)$. Allocation $x' = (c_1', c_2', l_1, l_2)$ satisfies (RC) but $u_2 \leq U(c_2, l_2) < u_2'$ and $u_1 \leq u_2 < u_2'$. Clearly, $\{u_1, u_2', x'\}$ contradicts $\{u_1, u_2, x\}$ being a solution to (35).

The program (35) is concave over a convex set, hence the necessary first order conditions

$$\begin{aligned} 1 &= \alpha + \beta_1 \\ \eta &= \beta_2 - \alpha \\ \beta_1 U_c(x_1) &= \phi \\ \beta_2 U_c(x_2) &= \eta\phi \\ -\beta_1 U_l(x_1) &= \phi F_1(l_1, \eta l_2) \\ -\beta_2 U_l(x_2) &= \eta\phi F_2(l_1, \eta l_2) \\ \alpha[u_1 - u_2] &= 0 \\ \beta_1[u_1 - U(x_1)] &= 0 \\ \beta_2[u_2 - U(x_2)] &= 0 \\ \phi[c_1 + \eta c_2 - F(l_1, \eta l_2)] &= 0 \end{aligned}$$

are also sufficient for the solution to program (35).

Let x be an allocation satisfying the first order conditions associated with problem (SBP). It is straightforward to show that there exist α, β_1, β_2, ϕ, u_1 and u_2 such that allocation x also satisfies the necessary and sufficient conditions for (35). Hence x is a solution to (35) and x is a second best allocation ■

Proof of Proposition 4. We first prove part 1. Assume that second best allocation x has

$$F_1\left(1, \eta\frac{l_2}{l_1}\right) \geq F_2\left(1, \eta\frac{l_2}{l_1}\right). \tag{36}$$

The properties of F and $\eta < \bar{\eta}$, imply $l_2 > l_1$. The incentive compatibility constraint implies then that $c_2 > c_1$. Strict concavity of U implies that if

$c_2 > c_1$, $l_2 > l_1$, then $-\frac{U_l(c_2,l_2)}{U_c(c_2,l_2)} > -\frac{U_l(c_1,l_1)}{U_c(c_1,l_1)}$. But then x is incompatible with the necessary first order conditions of problem (SBP) since $MRS_2 > MRS_1$ implies that $F_2 > F_1$, contradicting our initial hypothesis.

Now we prove the second part. By first order conditions for second best allocation, $MRS(c_2, l_2) > MRS(c_1, l_1)$. Since $U(c_1, l_1) = U(c_2, l_2)$ and indifference curves are strictly convex, we have that $(c_2, l_2) >> (c_1, l_1)$

To prove that for any $\eta \leq \bar{\eta}$ second best allocations x satisfy $c_2 < F_2(l_1, \eta l_2)$ consider the set $A = \{(c, l) : c \leq F_2(l_1, \eta l_2)(l - l_2) + c_2\}$. Since $MRS(c_2, l_2) = F_2(l_1, \eta l_2)$ and preferences are strictly concave, for any $(c, l) \in A$, $U(c, l) \leq U(c_2, l_2)$, with equality sign iff $c = c_2$ and $l = l_2$. Therefore $(c_1, l_1) \notin A$ since the incentive compatibility constraint is binding and $l_1 \neq l_2$ as Proposition 8 indicates. This implies

$$c_1 > c_2 + F_2(l_1, \eta l_2)(l_1 - l_2) \tag{37}$$

and since $F_1(l_1, \eta l_2) < F_2(l_1, \eta l_2)$,

$$c_1 - F_1(l_1, \eta l_2) l_1 > c_2 - F_2(l_1, \eta l_2) l_2. \tag{38}$$

Using constant returns to scale, the resource constraint can be written as

$$(c_1 - F_1(l_1, \eta l_2) l_1) + \eta (c_2 - F_2(l_1, \eta l_2) l_2) = 0 \tag{39}$$

therefore $c_2 < F_2(l_1, \eta l_2) l_2$. ∎

Proof of Proposition 5. We show that second best allocation x is feasible at η'. Note that

$$F(l_1, \eta' l_2) - F(l_1, \eta l_2) = F_2(l_1, \hat{\eta} l_2) l_2 (\eta' - \eta) \tag{40}$$

where $\hat{\eta} \in [\eta, \eta']$ by the Taylor theorem. Using the concavity of F,

$$F(l_1, \eta' l_2) - F(l_1, \eta l_2) > F_2(l_1, \eta' l_2) l_2 (\eta' - \eta). \tag{41}$$

Since the resource constraint is binding

$$F(l_1, \eta' l_2) - c_1 - \eta c_2 > F_2(l_1, \eta' l_2) l_2 (\eta' - \eta) \tag{42}$$

or

$$F(l_1, \eta' l_2) - c_1 - \eta' c_2 > (F_2(l_1, \eta' l_2) l_2 - c_2)(\eta' - \eta). \tag{43}$$

Since we proved that $F_2(l_1, \eta l_2) l_2 - c_2 > 0$ in Proposition 4 for all η, we can pick an arbitrarily close η' without loss of generality such that $F_2(l_1, \eta' l_2) l_2 - c_2 > 0$. Then allocation x satisfies the resource constraint with strict inequality sign when the skill ratio is η'.

By continuity, there exists $\hat{c}_2 > c_2$ such that $F(l_1, \eta' l_2) > c_1 + \eta' \hat{c}_2$. It is clear then that $\hat{x} = \{c_1, \hat{c}_2, l_1, l_2\}$ is feasible and incentive compatible with $U(c_1, l_1) + \eta U(c_2, l_2) < U(c_1, l_1) + \eta U(\hat{c}_2, l_2)$. Since allocations x' cannot do worse than $\hat{x}$, and the incentive constraint is binding for η', the result follows ∎

Proof of Proposition 6.

The sufficient conditions for a competitive equilibrium allocation are

$$\begin{aligned} MRS(c_1, l_1) &= F_1(l_1, \eta l_2), \\ MRS(c_2, l_2) &= F_2(l_1, \eta l_2), \\ c_1 + \eta c_2 &= F(l_1, \eta l_2), \\ c_2 &= F_2(l_2, \eta l_2)\, l_2 - \tau. \end{aligned}$$

It is trivial to show that there exists τ such that competitive equilibrium allocations satisfy $U(c_1, l_1) = U(c_2, l_2)$. The proposition follows from the sufficiency of first-order conditions. ■

Proof of Proposition 7. Assume there exists two equilibria given the set of feasible policies $\{\tau_r\}_{r\in R}$, denoted $\{x, n, \delta\}$ and $\left\{\tilde{x}, \tilde{n}, \tilde{\delta}\right\}$, with distinct worker distributions. Let $R_1 = \{r : \tilde{n}_2^r > n_2^r\}$ and $R_2 = \{r : \tilde{n}_2^r < n_2^r\}$. Since n and $\tilde{n}$ are distinct feasible worker distributions, both sets are non-empty. From the definition of competitive equilibrium given τ_r is clear that the skilled worker welfare is strictly decreasing in η_r and therefore in n_2^r. Therefore, for all $r \in R_1$, $U(\tilde{x}_2^r) < U(x_2^r)$ and for $r \in R_2$, $U(\tilde{x}_2^r) > U(x_2^r)$.

Take any pair $\{r, r'\}$ with $r \in R_1$ and $r' \in R_2$. Hence

$$U\left(\tilde{x}_2^{r'}\right) - U\left(x_2^{r'}\right) > U(\tilde{x}_2^r) - U(x_2^r) \tag{44}$$

and

$$U\left(\tilde{x}_2^{r'}\right) - U(\tilde{x}_2^r) > U\left(x_2^{r'}\right) - U(x_2^r). \tag{45}$$

If $\delta(r, r') > 0$, the mobility equilibrium conditions imply $\tilde{\delta}(r, r') > \delta(r, r')$, since

$$\mu\left(\tilde{\delta}(r, r')\right) = U\left(\tilde{x}_2^{r'}\right) - U(\tilde{x}_2^r) > U\left(x_2^{r'}\right) - U(x_2^r) = \mu(\delta(r, r')) \tag{46}$$

and μ is strictly increasing. Conversely, if $\delta(r', r) > 0$,

$$\begin{aligned} U(\tilde{x}_2^r) - U\left(\tilde{x}_2^{r'}\right) &< U(x_2^r) - U\left(x_2^{r'}\right), \\ U(\tilde{x}_2^r) - U\left(\tilde{x}_2^{r'}\right) &< \mu(\delta(r', r)), \\ \mu\left(\tilde{\delta}(r', r)\right) &\leq \mu(\delta(r', r)) \end{aligned}$$

and hence $\tilde{\delta}(r', r) \leq \delta(r', r)$ (the equality comes because $U(\tilde{x}_2^r) - U\left(\tilde{x}_2^{r'}\right)$ can be negative). Finally, if $\delta(r', r) = \delta(r, r') = 0$, then $U\left(\tilde{x}_2^{r'}\right) - U(\tilde{x}_2^r) > 0$ so $\tilde{\delta}(r, r') > 0$.

Add up all skilled workers in set R_1 for equilibrium $\tilde{\delta}$ and $\tilde{n}$,

$$\sum_{r\in R_1} \tilde{n}_2^r = \sum_{r\in R_1} \left\{ e_2^r \left(1 - \frac{1}{N-1} \sum_{r'\in R} \tilde{\delta}(r, r')\right) + \sum_{r'\in R} \frac{\tilde{\delta}(r', r)}{N-1} e_2^{r'} \right\}. \tag{47}$$

Note that flows within set R_1 must sum zero,

$$\sum_{r\in R_1}\left\{\sum_{r'\in R_1}\frac{\tilde{\delta}\left(r',r\right)}{N-1}e_2^{r'}-\frac{1}{N-1}e_2^{r}\left(\sum_{r'\in R_1}\tilde{\delta}\left(r,r'\right)\right)\right\}=0. \tag{48}$$

Hence

$$\sum_{r\in R_1}\tilde{n}_2^r=\sum_{r\in R_1}\left\{e_2^r\left(1-\frac{1}{N-1}\sum_{r'\in R_2}\tilde{\delta}\left(r,r'\right)\right)+\sum_{r'\in R_2}\frac{\tilde{\delta}\left(r',r\right)}{N-1}e_2^{r'}\right\}. \tag{49}$$

Now we use the inequalities derived above. We have $\tilde{\delta}\left(r,r'\right)\geq\delta\left(r,r'\right)$ for $r\in R_1$ and $r'\in R_2$, so

$$\sum_{r\in R_1}\tilde{n}_2^r\leq\sum_{r\in R_1}\left\{e_2^r\left(1-\frac{1}{N-1}\sum_{r'\in R_2}\delta\left(r,r'\right)\right)+\sum_{r'\in R_2}\frac{\tilde{\delta}\left(r',r\right)}{N-1}e_2^{r'}\right\}. \tag{50}$$

We have $\tilde{\delta}\left(r',r\right)\leq\delta\left(r',r\right)$ for $r\in R_1$ and $r'\in R_2$,

$$\begin{aligned}\sum_{r\in R_1}\tilde{n}_2^r &\leq \sum_{r\in R_1}\left\{e_2^r\left(1-\frac{1}{N-1}\sum_{r'\in R_2}\delta\left(r,r'\right)\right)+\sum_{r'\in R_2}\frac{\delta\left(r',r\right)}{N-1}e_2^{r'}\right\}\\ &= \sum_{r\in R_1}n_2^r.\end{aligned}$$

But $\tilde{n}_2^r>n_2^r$ for all $r\in R_1$ ■

Proof of Proposition 8. It is trivial to show that there is a credible policy equilibrium with a symmetric workers' distribution. We have shown that second-best allocations are a differentiable function of the skill ratio, $x^r(\eta^r)$. Proposition 5 makes clear that the skilled worker welfare is increasing in η^r. Therefore,

$$\frac{\partial U(x^r(\eta^r))}{\partial n_2^r} > 0. \tag{51}$$

For an arbitrarily small $\varepsilon > 0$ we can then conclude that

$$U(\tilde{x}_2^{r'}) - U(\tilde{x}_2^r) > 0 \tag{52}$$

for allocations $\tilde{x}^r$ and $\tilde{x}^{r'}$ that are second-best given $(n_1^r, n_2^r - \varepsilon e_2^r)$ and $\left(n_1^{r'}, n_2^{r'} + \varepsilon e_2^r\right)$, respectively. The mobility costs at $\delta = \varepsilon$ are $\mu(\varepsilon)$ which can be made arbitrarily close to zero as $D'(\varepsilon)$ is arbitrarily close to zero as well. Hence there exists D such that

$$U(\tilde{x}_2^{r'}) - U(\tilde{x}_2^r) > \mu(\varepsilon) \tag{53}$$

and therefore the symmetric workers' distribution cannot constitute a locally stable equilibrium ■

Proof of Proposition 9. Since $U(\tilde{x}_2^A) > U(\tilde{x}_2^B)$, any sequence leading to an admissible equilibrium will need to feature region A gaining skilled workers. Hence welfare in region A will be strictly higher than $U(\tilde{x}_2^A)$ and in region B will be strictly lower than $U(\tilde{x}_2^B)$. The result follows ■

B Figures

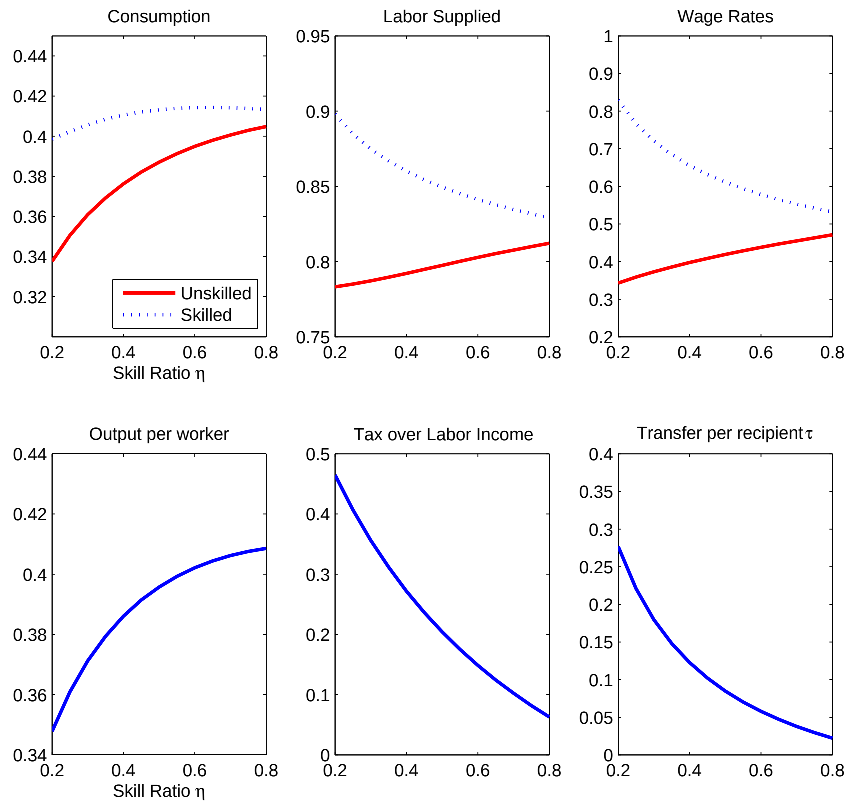

Figure 1: Second best allocation and competitive equilibrium decentralization

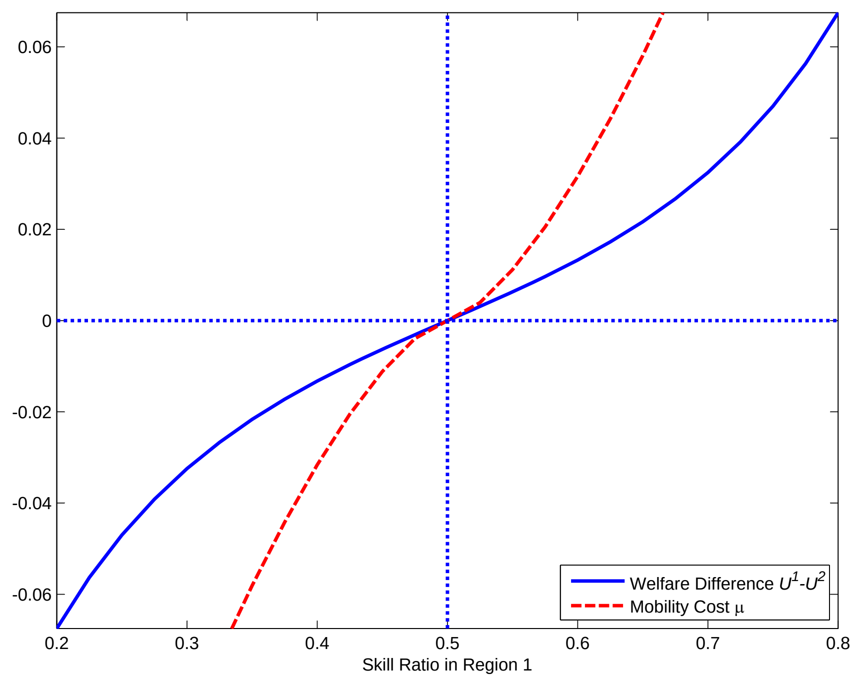

Figure 2: Two region equilibria: Symmetric regions with high mobility costs

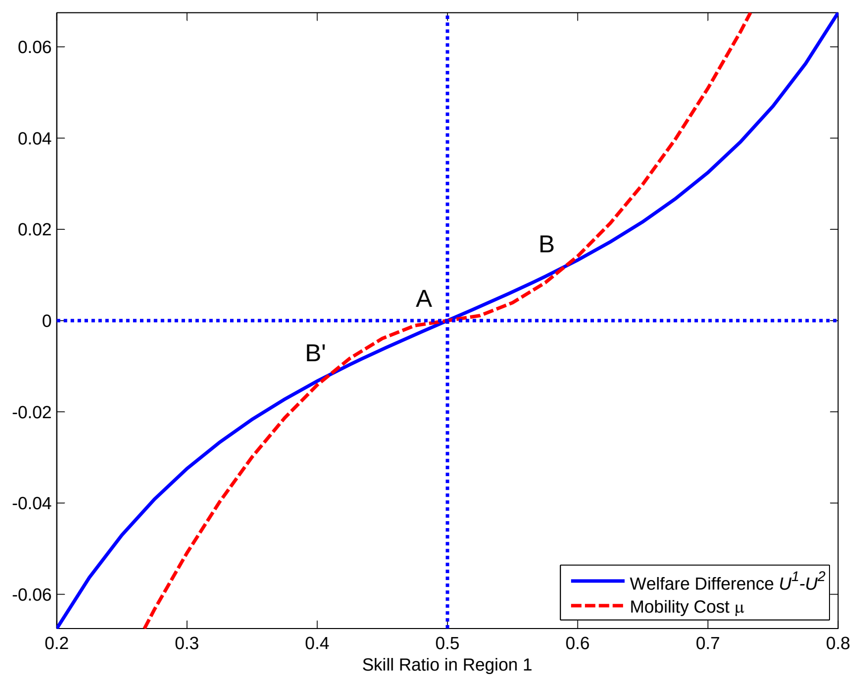

Figure 3: Two region equilibria: Symmetric regions with low mobility costs

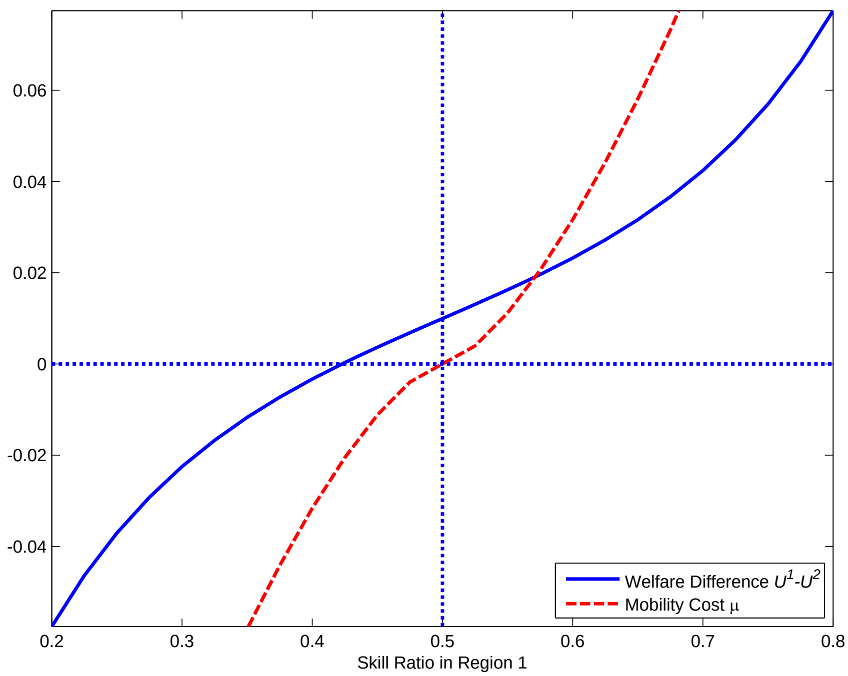

Figure 4: Two region equilibria: Technology differences $\theta^1 > \theta^2$

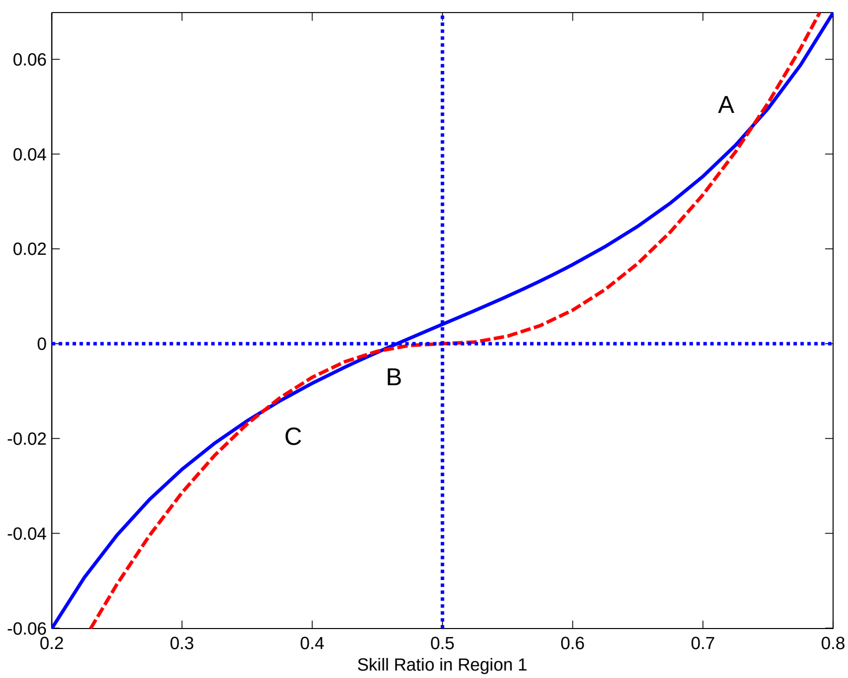

Figure 5: Two region equilibria: Technology differences $\theta^1 > \theta^2, \alpha^1 < \alpha^2$

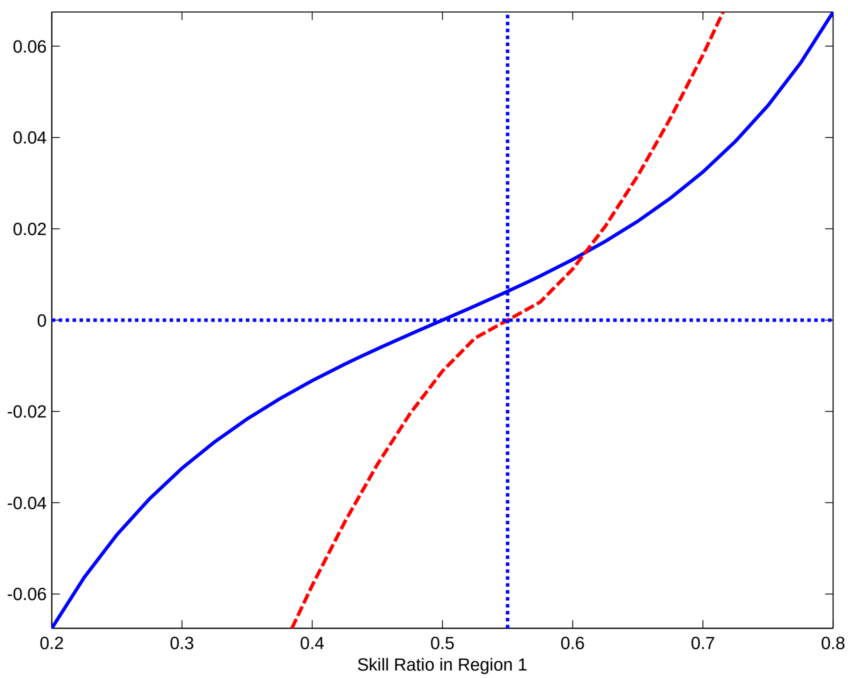

Figure 6: Two region equilibria: Endowment differences $e_2^1 > e_2^2$