NBER WORKING PAPER SERIES

OPTIMAL MIGRATION:
A WORLD PERSPECTIVE

Jess Benhabib
Boyan Jovanovic

Working Paper 12871
http://www.nber.org/papers/w12871

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2007

We thank the NSF for support and Matthias Kredler for doing all the calibrations and producing the associated plots. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Optimal Migration: A World Perspective
Jess Benhabib and Boyan Jovanovic
NBER Working Paper No. 12871
January 2007
JEL No. O0

ABSTRACT

We ask what level of migration would maximize world welfare. We find that skill-neutral policies are never optimal. An egalitarian welfare function induces a policy that entails moving mainly unskilled immigrants into the rich countries, whereas a welfare function skewed highly towards the rich countries induces an optimal policy that entails a brain-drain from the poor countries. For intermediate welfare functions that moderately favor the rich however, it is optimal to have no migration at all.

Jess Benhabib
Department of Economics
New York University
269 Mercer Street, 7th Floor
New York, NY 10003
jess.benhabib@nyu.edu

Boyan Jovanovic
New York University
Department of Economics
269 Mercer Street
New York, NY 10003
and NBER
Boyan.Jovanovic@nyu.edu

Optimal Migration: A World Perspective*

Jess Benhabib and Boyan Jovanovic†

January 17, 2007

Abstract

ABSTRACT: We ask what level of migration would maximize world welfare. We find that skill-neutral policies are never optimal. An egalitarian welfare function induces a policy that entails moving mainly unskilled immigrants into the rich countries, whereas a welfare function skewed highly towards the rich countries induces an optimal policy that entails a brain-drain from the poor countries. For intermediate welfare functions that moderately favor the rich however, it is optimal to have no migration at all.

1 Introduction

All rich industrialized countries severely restrict immigration.[1] While the extent of the restrictions varies by country and by period, they nevertheless are at odds with the basic tenets of free trade, and in deep contradiction with some of the most cherished values of liberal democracy: that there should be no job discrimination based on nationality, ethnicity, race or gender. While we deplore job discrimination directed at citizens, we also design immigration laws that exclude foreign nationals out of our countries and our job market.[2] It follows that there must be costs associated with immigrants that are borne by the citizens of a country, or otherwise the borders would be open.

Several reasons may be given to explain restrictive immigration policies in terms of the costs that immigrants impose on the citizens of a country. The most obvious is

*We thank the NSF for support and Matthias Kredler for doing all the calibrations and producing the associated plots.

†Both authors are at New York University

[1]Freedberg and Hunt (1995) report that all but 100 million of the world's 6 billion people live in the country of their birth.

[2]Protectionist arguments have recently been made against the mobility of capital, on the grounds that some multinational corporations do not pay a "living wage" in third world countries. Yet none of the pundits against outsourcing has advocated opening up of the borders to immigrants in order to improve their lot.

a distributional argument cast in terms of political economy. The median voter whose income derives mostly from wages will wish to keep out the unskilled immigrants who will depress his wage.[3] Others, more controversially, stress the cost of social services that low-skilled immigrants impose on the citizens, or adopt the communitarian view that shared values, customs and culture constitute a social good that would be diluted by immigration, an argument that has often been used to keep the undesirables out.[4] Finally, one can argue that positive output externalities emanating from the average level of human capital will be depressed by immigrants with low human capital stocks who cause congestion, disutility, and have a negative impact on output per capita. Since the goal of this paper is not to identify the nature and scope of the costs of immigration borne by citizens, we will model the latter explanation by externalities emanating from the average level of human capital, a framework which is quite simple, and which may be modified or reinterpreted to capture direct labor-market effects on wages, or negative cultural externalities from low skilled immigrants.

We study the welfare implications of restrictive immigration policies from the world perspective, while allowing for costs of migration to the host and source countries. We ask what the optimal immigration policy would be, given a social welfare function that parametrically weighs the citizens of the industrialized, human-capital-rich countries and those of the third world. One might simply expect that as the welfare weight is continuously shifted from the citizens of the first to the third world, optimal immigration policy, in terms of the proportion of third world citizens allowed to emigrate, would increase continuously. Our results indicate that this is not so: if populations are homogeneous in the skills within a country but differ across countries, there is a threshold relative welfare weight assigned to the third world citizens at which optimal immigration policy shifts from zero immigration to maximal immigration. If populations are heterogenous in skill to labor ratios, then under egalitarian social welfare weights, or, a fortiori, with weights that favor natives of the low average skill country, the optimal policy is to let the least skilled emigrate, up to a threshold skill level, from the low to the high average skill country. A simple and quick calibration shows that this implies that optimally up to 3.2 billion low-skilled people should emigrate from the third world to the OECD.[5] If on the other hand, social welfare weights favor the natives of the high-skill country, optimal immigration policy may be no immigration at all, or an immigration policy that allows only the highly skilled to emigrate from the low- to the high-average-skill country.

[3]See for example Borjas(2003), and Borjas and Katz (2005). For the opposing view see Card (2005). For a more recent reconsideration of this debate, see Peri and Ottaviano (2006).

[4]See the edited volume by Warren F. Schwartz (1995), and in particular the essays in the volume by Jules Coleman and Sarah Harding, and by Michael Trebilcock. For studies suggesting that immigrants do not impose large negative social externalities see National Research Council (1997) and Butcher and Piehl (1998).

[5]The International Organization for Migration estimates that currently there are 191 million transnational migrants worldwide comprising 3% of the global population. See http://www.iom.int/jahia/page254.html

With globalization, pressures to design redistributive and immigration policies that increasingly take a world rather than national perspective are likely to mount. Thus, if the political perspective shifts from a national to an international one, more consistent with values of liberal democracy applied globally, the optimal immigration policy will require a drastic change. Of course other factors, including political costs of policy transition, or political resistance in host countries, may imply a more gradual shift over time. This paper, while allowing for the costs of immigration, shows a basic thrust or tendency calling for a shift in immigration policy as we move towards a world democracy.

2 The Model

Immigrants in our model affect the well-being of the residents of the host country through a group effect. This group effect operates through the effect that immigrants have on "social capital," originally discussed by Coleman (1988)[6]. We define "social capital" as human capital per person, $\bar{h}$, which raises the marginal productivity of human capital h, and we formalize it as in Lucas (1988):

$$\text{Marginal product of human capital} = G\left(\bar{h}\right)$$

where $G' > 0$ and $G'' < 0$. We drop the second factor (physical capital) and assume that the output of a country is

$$Y = G\left(\bar{h}\right) H \tag{1}$$

where H is the total human capital in the country. Evidence supporting this formulation is given by Clark (1987), who attributes $G\left(\bar{h}\right)$ to culture in a multinational setting, and by Rauch (1993) who attributes it to human-capital spillovers at the level of "Standard Metropolitan Statistical Areas". Thus if immigrants decrease the average level of human capital, they depress marginal products, wages and average productivity.[7]

A second, non-market interpretation of G, is one of a cultural externality operating not through production, but through preferences of natives. That is, since agents' utility is a monotone transform of their output or consumption, we can interpret $G\left(\bar{h}\right)$ as an externality acting directly on utility, reflecting a cultural distaste for unskilled immigrants, so that the enjoyment of consumption is diluted in a society

[6] Coleman (1988) describes social capital as falling into three categories: (a) mutual obligations and expectations, (b) social norms (c) information channels and their role in the creation of human capital. The first two may be part of culture.

[7] Recent evidence on externalities of schooling in the U.S. is mixed. Moretti (2004) finds a positive external effect from an increase of college graduates in U.S. cities for 1980-1990, while Acemoglu and Angrist (2001) and Ciccone and Peri (2006) do not find significant externalities from changes in average schooling for U.S. states over the period 1960-1990. More recently Peri and Iranzo (2006) find positive externalities from the share of college graduates.

where the arrival of less skilled immigrants lowers $\bar{h}$. In this sense, culture is a public good that is diminished by the arrival of unskilled immigrants, but enriched by highly skilled immigrants, so that it is not simply xenophobia.[8]

Constant returns and decentralizability.—The production function (1) obeys constant returns to scale in the sense that doubling the number of residents while leaving the distribution of individual human capital h unchanged leaves $\bar{h}$ unaffected, but doubles H and, hence, Y. This allows for a competitive situation in which zero-profit firms (of indeterminate size) hire labor and pay a wage of $G\left(\bar{h}\right)$ per efficiency unit.

Efficiency vs. distribution.—The model has a tension between considerations of efficiency and distribution. If taxes were the distributive tool, the tension would work through incentives. In our model, however, the only distributive tool is migration, and the tension works through the spillover mechanism that induces increasing returns to scale through G:

1. Efficiency requires that production be segregated geographically. This is the content of Proposition 1.

2. The only way to redistribute income in this model is through migration, which requires that we *mix* people of different human capital levels.

Efficiency and redistribution are always in conflict, and this may lead migration to sometimes be zero in spite of the planner's desire to redistribute.

Let $M(h)$ be the world's distribution of human capital, and assume that

$$G(h) = h^{\alpha}. \tag{2}$$

Proposition 1 *World output is maximized when there is complete segregation by h, i.e.,*

$$Y \leq \int h^{1+\alpha} dM(h).$$

Proof. Suppose that there is a location in the world where people are heterogeneous in h. Let the distribution at that location have measure $\mu(h)$, with mean $\bar{h}$. Let the total output at that location be

$$y = G\left(\bar{h}\right)\left(\int h d\mu\right)$$

[8] Another argument for the negative welfare effects of immigration from the host country perspective is given by Lundborg and Segestrom (2002) in the context of a quality ladder model of growth. They show that while under some calibrations immigration may increase R&D and the growth rate in the host country by depressing wages, profits may nonetheless decline because of the aggregate demand effects of the lower wages, making the owners of capital as well as workers worse off.

Then

$$\frac{1}{\int d\mu}y = \bar{h}G\left(\bar{h}\right) = \bar{h}^{1+a} = \left(\frac{\int hd\mu}{\int d\mu}\right)^{1+\alpha} \leq \frac{1}{\int d\mu}\int h^{1+\alpha}d\mu$$

where the inequality follows because $d\mu/\int d\mu$ is a measure adding up to unity, and $h^{1+\alpha}$ is a convex function. Cancelling the multiplicative constant leaves us with

$$y \leq \int h^{1+\alpha}d\mu$$

and the inequality is strict if the support of μ has more than one point. Therefore no location can have heterogeneity of h. ■

This proposition suggests that there should be no mixing of skill levels through migration if the sole objective is to maximize world output. Obviously, the world is fairly segregated by skill. "Social justice" however could be attained without moving people around and, instead, by world-wide redistribution, i.e., foreign aid. Unfortunately extensive foreign aid programs, even though substantial and well-intentioned (2.3 trillion over the last five decades), have failed to alleviate poverty or to raise the standards of living in many of the poor nations. Easterly (2006) documents how the misdirection and mismanagement of foreign aid, due to perverse incentives and insufficient knowledge of local conditions, have resulted in waste rather than the relief of poverty. Therefore we focus on immigration as a means to achieve redistribution and social justice.

Proposition 1 generalizes to a world in which output equals, say, $G\left(\bar{h}\right)k^{\beta}h^{1-\beta}$. On its own, the free mobility of capital will not solve the problem faced by the social planner if, as Lucas (1990) claims, inequality originates in skill differences. The group effect makes factor prices proportional to G, so that they differ across geographic locations. Unlike the Hecksher-Ohlin model, factor prices can no longer be equalized via a flow of capital or goods alone. People must move so as to equalize $\bar{h}$.[9]

[9] In a standard two country, two factor (say human capital and labor) Hecksher-Ohlin model with constant returns in production, no group effects, and where people move with their human capital, Benhabib (1996) shows that if policies are skill-blind, or if populations are homogeneous with respect to their human-capital-to-labor ratio within the countries but differ across countries, the agent with average human capital to labor ratio in the high-average-skill country prefers to let everyone in. Since the argument is symmetric, the agent with the average labor to human capital in the low average skill country also prefers to let everyone in. Thus both of the average agents prefer full immigration. Putting the two points together, and noting that countries initially differ only in average human capital to labor ratios, for any pair of social welfare weights applied to the average agents in each country, the optimal policy is to move all agents to one of the countries. This also demonstrates that in such a model there is no tradeoff between efficiency and social welfare, and optimal immigration is full immigration.

3 Case 1: Two homogeneous countries

So far we have talked of an arbitrary number of locations, but now we specialize to two locations, or "countries", $i = A, B$, with skill levels h_A and h_B. The population of A is normalized to 1, and the population of B to n. In each country there is just one productive input, human capital. In this situation, proposition 1 says that before migration we have efficiency. Migration, if any, occurs *before* production takes place. Each agent would wish to be where G is higher, since then his total wages would be higher there.

Let x denote the probability that a B-native will be allowed to move to A. We shall refer to x as the migration rate, and we denote the average *post*-migration h levels in A and B by $\bar{h}_A$ and $\bar{h}_B$ respectively. Then

$$\bar{h}_A = \frac{h_A + xnh_B}{1 + xn}, \quad \text{and} \quad \bar{h}_B = h_B. \tag{3}$$

Social welfare function and the planner's problem.—The planner is a Stackelberg leader. He announces a policy at the outset, and agents then choose their migration decisions and production takes place. Let θ and $(1 - \theta)$ denote the welfare weights that the planner assigns to utilities of the residents of A and B, respectively. He then chooses x to solve the problem

$$\max_x \left\{\theta U \left(G\left[\bar{h}_A\right] h_A\right) + (1-\theta)\, n \left[xU\left(G\left[\bar{h}_A\right] h_B\right) + (1-x)\, U\left(G\left[\bar{h}_B\right] h_B\right)\right]\right\}.$$

Now assume

$$U(c) = \ln c \tag{4}$$

Letting $g_A(x) = \ln G\left(\frac{h_A + xnh_B}{1+xn}\right)$ and $g_B = \ln G(h_B)$, the problem boils down to choosing x to maximize

$$W(x) = \theta g\left(\bar{h}_A\right) + (1-\theta)\, n\left[x g_A(x) + (1-x)\, g_B\right] \tag{5}$$

subject to (3). The first-order condition is

$$W'(x) = (\theta + [1-\theta]\, nx)\, g'_A + (1-\theta)\, n\,(g_A - g_B) = 0, \tag{6}$$

Then

$$W''(x) = (\theta + (1-\theta)\, nx)\, g''_A + 2\,(1-\theta)\, n g'_A.$$

In that case,

$$W''(x) > 0 \quad \text{iff} \quad \frac{g''_A}{g'_A} < -\frac{2\,(1-\theta)\, n}{\theta + (1-\theta)\, nx}. \tag{7}$$

The Appendix proves the following result:

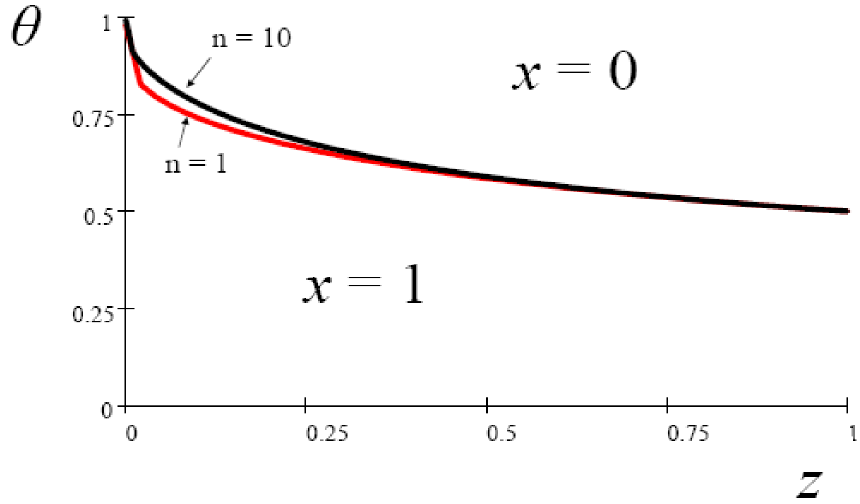

Figure 1: The bang-bang policy in (z, θ) space

Lemma 1 *For any $\alpha > 0$ in (2), $n > 0$ and $\theta \in (0,1)$, if (6) holds, then (7) also does.*

This result implies that the planner's problem cannot have an interior maximum. Rather, the planner's maximum is at a corner: Either $x = 0$ or $x = 1$.

Characterizing the bang-bang solution for x.—Define the initial, date-zero productivity of a B-native relative to that of an A-native by

$$\text{Relative backwardness} \equiv z = \frac{h_B}{h_A}.$$

The optimal policy depends on how backward B is relative to A in terms of skills, and it also depends on n – the population of B relative to A. The following result, proved in the Appendix, is that the form of the optimal policy does not depend on α:

Proposition 2 *The optimal policy is*

$$x = \begin{cases} 0 & \text{if } \theta > \frac{n\left(\ln z - \ln\left(\frac{1+nz}{1+n}\right)\right)}{(1-n)\ln\left(\frac{1+nz}{1+n}\right) + n \ln z}, \\ 1 & \text{otherwise} \end{cases} \tag{8}$$

We plot the indifference locus for $n = 1$ and $n = 10$ in Figure 1.

The planner is more inclined to a policy of immigration if B is poor, and if B is large in terms of population, though the latter is not a quantitatively important consideration. In the plot, the action $x = 0$ is preferred in the North-East quadrant and $x = 1$ is preferred in the South-East quadrant. So, the planner chooses maximal immigration if he cares enough for B (low value of θ), and if B is poor enough (low value of z) and if B is large (high values of n). Empirically, $z = 0.1$ is a good approximation to the average non-OECD income, which means, for $n = 1$, (since in fact x is close to zero) that $\theta \geq 3/4$.

The lesson of this figure is that the "world's planner," if she exists, does not care much for B. Freedberg and Hunt's (1995) numbers tell us that we are, effectively, in the $x = 0$ region. But, since z must be rather small – say 1/10 – the action $x = 0$ is optimal only if θ is at least 0.8. In other words, this outcome we now have is incompatible with even approximately equal weights in the social welfare function.

4 Case 2: Two heterogeneous countries

We now assume that skills are *heterogeneous* in both countries. In general, the world's planner may wish to make immigration policy biased toward some groups in B, but within those groups she may impose neutrality – everyone within a group may then face the same probability of moving from B to A. This subsection poses the problem at this full level of generality.[10]

Let μ_A be the pre-migration mean skills in country A and let the human capital of A's residents be distributed $h \sim F_A(h)$. Let μ_B be the mean skills in country B and let the human capital of B's residents be distributed $h \sim F_B(h)$, with density function $f_B(h)$. Let

$$x = \phi(h)$$

be the probability that a type-h resident of B will be allowed to emigrate to A. That is, $\phi : R \to [0, 1]$.

A *skill-neutral* policy is one in which ϕ is a constant, independent of h. Policies that are not skill neutral are skill biased. Generalizing their definitions in (3) second

[10] The U.S. today follows a mixture of skill-biased policies and skill-neutral policies based on four principles: The reunification of families, the admission of immigrants with needed skills, the protection of refugees, and the diversity of admissions by country of origin. While special legislation now allows for special consideration for medical professionals for example, the majority of legal immigrants enter the US through the family-reunification program. While Canadian policy also allows immigration based on family reunification, preferences stress skills and youth: During 1990–2002, 65 per cent of permanent immigrants to the United States were admitted under family preferences. In Canada, the equivalent proportion was 34 per cent (International Migration and Development: Regional Factsheet, The Americas, http://www.un.org/migration/presskit/factsheet_america.pdf). Similarly Australia heavily emphasizes skills and youth in its preference system for immigrants. See for example http://www.workpermit.com/australia/australia.htm. Recently France has also moved towards a skill biased immigration policy: see http://www.migrationpolicy.org/pubs/Backgrounder2_France.php

period human capital per head in A is

$$\bar{h}_A = \frac{\mu_A + n \int h\phi(h)\, dF_B(h)}{1 + n \int \phi(h)\, dF_B(h)}, \tag{9}$$

and in B it is

$$\bar{h}_B = \frac{\int h [1 - \phi(h)]\, dF_B(h)}{\int [1 - \phi(h)]\, dF_B(h)}. \tag{10}$$

The planner's problem is to choose a function $\phi(h)$ to maximize

$$\theta \int U\left[G\left(\bar{h}_A\right) h\right] dF_A(h) + (1-\theta)\, n \int \left\{\phi(h)\, U\left(G\left[\bar{h}_A\right] h\right) + [1 - \phi(h)]\, U\left(G\left[\bar{h}_B\right] h\right)\right\} dF_B(h). \tag{11}$$

subject to (9) and (10).

4.1 The optimal policy

The rest of the paper will assume that $\mu_A > \mu_B$, that h has no upper bound in the supports of F_A and F_B, and that (4) holds. In this case, the optimal policy generally is skill-biased, and of the "bang-bang" type in the sense that within a group indexed by h, either everyone should migrate or no one should do so. Moreover, the set of types is connected in the sense that if type h_0 is allowed to migrate, then either everyone with h below h_0 is also allowed to migrate, or everyone *above* h_0 is allowed to migrate. The first policy we call "skimming from the bottom" of the F_B distribution; under that policy there exists a cutoff, $\tilde{h}$, such that

$$\phi(h) = \begin{cases} 1 & \text{for } h < \tilde{h} \\ 0 & \text{for } h > \tilde{h} \end{cases} \tag{12}$$

The second "skimming from the top," or simply a "brain-drain" policy:

$$\phi(h) = \begin{cases} 0 & \text{for } h < \tilde{h} \\ 1 & \text{for } h > \tilde{h} \end{cases} \tag{13}$$

The point $\tilde{h}$ is of measure zero and in each case we know only that $0 \leq \phi\left(\tilde{h}\right) \leq 1$, the planner being, in both cases, indifferent about whether $\tilde{h}$ should migrate or not. The rest of this section will prove the following properties of the optimal policy

1. Whenever immigration is positive, it is always skill biased,
2. For θ sufficiently close to unity, the policy is of the form (13),
3. For $\theta < \frac{1}{2}$, the policy is of the form (12), and
4. For some θ's satisfying $\frac{1}{2} < \theta < 1$, the optimal policy may involve no migration.

The rest of this subsection is devoted to proving these claims. Before reading the proof, it is instructive to consider the special case when F_A and F_B are both log-normal as in Figure 3. The region of h-values for which $\phi = 1$ is then the purple-shaded area in Figure 5. We now turn to the proof of proposition 2 and begin with the following lemma.

Lemma 2 *When $U(c) = \ln c$, (11) reduces to*

$$W \equiv \theta_A^* g\left(\bar{h}_A\right) + \theta_B^* g\left(\bar{h}_B\right) \tag{14}$$

subject to (9) and (10), where

$$\theta_A^* = \theta + (1-\theta)\omega n, \quad \theta_B^* = (1-\theta)n(1-\omega), \ \textit{ and } \ \omega \equiv \int \phi(h)\, dF_B.$$

Proof. Substituting for U and leaving out terms that do not depend on ϕ, (11) reads

$$\begin{aligned} & \theta \ln G_A + (1-\theta)\, n \int \{\phi(h)(g_A + h) + (1 - \phi[h])(g_B + h)\}\, dF_B \\ = \ & \theta \ln G_A + (1-\theta)\, n \int \{\phi(h)\, g_A + (1-\phi[h])\, g_B\}\, dF_B + (1-\theta)\, n \int h dF_B. \end{aligned}$$

But the last terms does not depend on ϕ and we are left with (14). ■

Assume that the density f_B exists for all h, and define

$$z(h) = n f_B(h)\phi(h)$$

to be the new control variable that satisfies $z(h): \ R \to [0, \ n f_B(h)]$ for all h. In terms of this control variable in (14) we have

$$\bar{h}_A = \frac{\mu_A + \int h z(h)\, dh}{1 + \int z(h)\, dh}, \ \bar{h}_B = \frac{n\mu_B - \int h z(h)\, dh}{n - \int z(h)\, dh}, \ \text{ and } \ n\omega \equiv Z = \int z(h)\, dh.$$

The constraint set for z is convex. We attach the multiplier λ_0 to the non-negativity constraint, and the multiplier λ_1 to the upper-bound constraint. The planner faces the Lagrangean

$$\mathcal{L} = W + \int \lambda_0(h)\, z(h)\, dh - \int \lambda_1(h)\, z(h).$$

The FOC is

$$\frac{\partial W}{\partial z(h)} = \lambda_1(h) - \lambda_0(h). \tag{15}$$

where $\frac{\partial W}{\partial z(h)}$ is evaluated at the optimal policy, the latter consisting of an entire function $z(.)$. Note that at most one multiplier can be non-zero and that

$$\frac{\partial W}{\partial z(h)} = \left\{ \begin{array}{l} < 0 \Longrightarrow \lambda_0(h) > 0 \text{ and } \phi(h) = 0 \\ > 0 \Longrightarrow \lambda_1(h) > 0 \text{ and } \phi(h) = 1 \end{array} \right. . \tag{16}$$

Now let $\bar{n}_A = 1 + n\omega$ be the post-immigration population of A and $\bar{n}_B = n(1-\omega)$ the post-immigration population of B.

$$\begin{aligned} \tfrac{\partial W}{\partial z(h)} &= \theta_A^* g'(\bar{h}_A)\tfrac{h-\bar{h}_A}{\bar{n}_A} - \theta_B^* g'(\bar{h}_B)\tfrac{h-\bar{h}_B}{\bar{n}_B} + (1-\theta)\left[g(\bar{h}_A) - g(\bar{h}_B)\right] \\ &= \alpha(1-\theta)\left[\ln \tfrac{\bar{h}_A}{\bar{h}_B} + 1 - m + \left(\tfrac{m}{\bar{h}_A} - \tfrac{1}{\bar{h}_B}\right) h\right]. \end{aligned} \quad , \tag{17}$$

where the second equality follows because $g(h) = \alpha \ln h$ and $g'(h) = \alpha / h$, and where

$$m \equiv \frac{\theta + (1-\theta) Z}{(1-\theta)(1+Z)} \qquad \text{and} \qquad Z \equiv \int z(h) dh. \tag{18}$$

By (17), $\frac{\partial W}{\partial z(h)}$ is linear in h,[11] which immediately shows that either (12) or (13) must hold, though possibly with $\tilde{h} = 0$ or $\tilde{h} = +\infty$. That is, the policy is always bang-bang. We now turn to the cases that arise for different values of θ. Before proceeding we assume that that there are enough unskilled people in B so that migration can equalize the average skills, that is $\bar{h}_A = \bar{h}_B$:
Assumption: There exists an $\hat{h} < \infty$ such that

$$\left(\bar{h}_A =\right) \frac{\mu_A + n\int_0^{\hat{h}} h dF_B(h)}{1 + nF_B\left(\hat{h}\right)} = \frac{\int_{\hat{h}}^{\infty} h dF_B(h)}{1 - F_B\left(\hat{h}\right)} \left(= \bar{h}_B\right). \tag{19}$$

Since $\bar{h} = E\left(h_B \mid h_B \geq \hat{h}\right)$, $\hat{h} < \bar{h}_B$. But then $\hat{h} < \bar{h}_A$. The derivative of the LHS of (19) is $\frac{nf_B}{1+nF_B}\left(\hat{h} - h_A\right) < 0$, whereas the derivative of the RHS is $-\frac{f_B}{1-F_B}\left(\hat{h} - h_A\right) > 0$. The two sides of (19) are continuous, and at any solution the LHS must cut the RHS from above. Therefore if $\hat{h}$ exists, it is unique.

The case $\theta = 1$.—In this case the planner cares only about country A. The following policy is then optimal:

Proposition 3 *For $\theta = 1$, the optimal policy is characterized by (13), with $\tilde{h} = \bar{h}_A$.*

Proof. As $\theta \to 1$ in (17)

$$\frac{\partial W}{\partial z(h)} \to \alpha\left(\frac{1}{1+Z}\frac{1}{\bar{h}_A}(h - \bar{h}_A)\right) \gtreqless 0 \quad \text{as} \quad h \gtreqless \bar{h}_A$$

and the claim follows. ■
Concavity of G implies that skill-biased policies help country A less than they harm country B. In spite of this, the planner would allow some immigration from B to A even if $\theta < 1$ as long as θ is sufficiently close to 1. When $\theta = 1$, we found that as long as there are some natives of B with skills exceeding those of the average A-native,

[11] We plot (17) as a function of h in Figure 4.

skill-biased immigration policies will always lead to positive immigration flows. By continuity, we expect this to be true even for some $\theta < 1$, as long as θ is close enough to 1.

The egalitarian case $\theta = 0.5$.—Immigration flows from the low-skill low-wage region B to the high-skill high-wage region A. Now consider the policy that attains $\bar{h}_A = \bar{h}_B$, namely, (12), with $\tilde{h}$ replaced by $\hat{h}$. Evaluated at $\bar{h}_A = \bar{h}_B$, (17) reads

$$\begin{aligned} \frac{\partial W}{\partial z(h)} &= \alpha(1-\theta)\left[1 - m + \left(m\frac{\bar{h}_B}{\bar{h}_A} - 1\right)\frac{h}{\bar{h}_B}\right] \\ &= \alpha(1-\theta)(1-m)\left(1 - \frac{h}{\bar{h}_B}\right) = 0 \end{aligned}$$

because when $\theta = 0.5$, we have $m = 1$. This proves the following result:[12]

Proposition 4 *For* $\theta = 0.5$, *the policy (12), with* $\tilde{h} = \hat{h}$ *as defined in (19) is optimal. Furthermore the skill level of the marginal immigrant,* $h_m = \bar{h}_A = \bar{h}_B$. *This is a* ***skim-the-bottom*** *policy: Only the less-skilled are allowed emigrate.*

The case $\theta < 0.5$.—In this case the planner cares more for B. The following result is proved in the Appendix:

Proposition 5 *For* $\theta < 0.5$ *and* $\tilde{h} \in \left[0, \hat{h}\right]$, *there exists a unique optimal policy of the form (12) with* $\tilde{h} = \tilde{h}^* < \hat{h}$. *Moreover,* $\theta < \frac{1}{2} \Longrightarrow \bar{h}_A\left(\tilde{h}^*\right) < \bar{h}_B\left(\tilde{h}^*\right)$.

We illustrate Propositions 4 and 5 with a calibrated example below. In that example it will be seen that the restriction made in Proposition 5 that $\tilde{h}^* \leq \hat{h}$ is in fact not binding. This is seen by comparing the first panel of Figure 5 with the third panel. In the third panel we see that $\bar{h}_A$ and $\bar{h}_B$ (intersect at $\hat{h} \approx \$12,000$, whereas in the top panel we see that $\tilde{h}^*$ never reaches \$7000.

The interval $\theta \in \left(\frac{1}{2}, 1\right)$ is not covered by the above analysis. Our simulated example shows that there may be θs in that region such that there is no immigration whatsoever: On the one hand, allowing a skilled marginal immigrant would, for those θs, not benefit the natives of A sufficiently to offset the negative brain-drain effect. On the other hand, allowing an unskilled immigrant to emigrate from B does not generate enough benefit to offset the negative externality in A through the lower $\bar{h}_A$.

[12] We assumed in our model that migration flows from B to A. However when $\theta = 0.5$ it is possible to equalize skill levels between A and B with a policy where the most skilled emigrate from A to B. This is because with $\theta = 0.5$ and with regions A and B identical except in their initial average skills, emigration policies that equalize skill levels are optimal, irrespective of who moves. When $\theta \neq 1$, we have symmetry breaking, as illustrated by the case $\theta = 1$ where only zero immigration flows are optimal, because we care differentially about the residents of A and B. Symmetry breaking will also hold for the case $\theta < 0.5$ studied below.

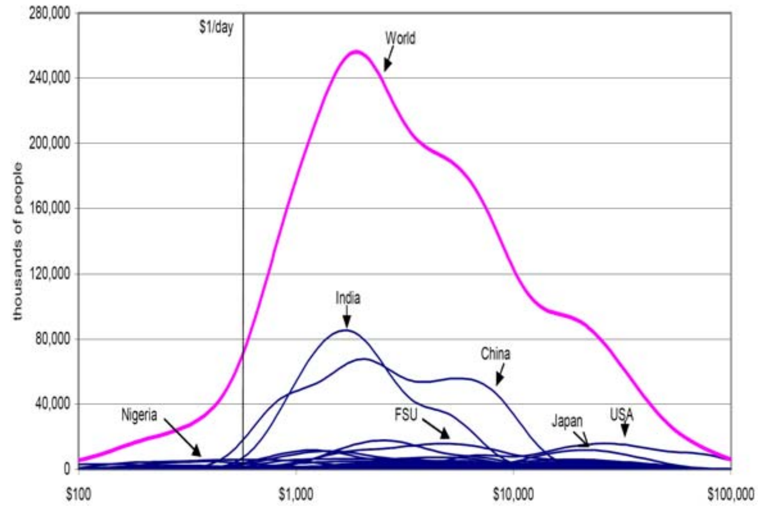

Figure 2: The world income distribution in 2000

5 Calibration and simulation

We now wish to illustrate the optimal policy for various hypothetical values of θ, and for realistic F_A and F_B. We choose "country A" to be the OECD which we shall think of as the developed world. "Country B" will then be the rest of the world. Sala-i-Martin (2006) reports the world distribution in the year 2000, and how it comprises the distributions of income in individual countries. We reproduce these distributions in Figure 2, which shows them to be roughly log-normal in form.

We observe the distribution of income y for each citizen, which we approximate as follows by a log-normal distribution:

$$\mu_{OECD}(\log y) = \ln 20,000, \qquad \sigma_{OECD}(\log y) = \ln 2$$
$$\mu_{Rest}(\log y) = \ln 2,000, \qquad \sigma_{Rest}(\log y) = \ln 2.5$$

These are portrayed in Figure 3. The following equation identifies $\bar{h}$:

$$E(y) = \exp(\mu + \sigma^2/2) = G(\bar{h})E(h) = \bar{h}^{\alpha+1} \Longrightarrow \bar{h}_A = \exp\left(\frac{\mu + \sigma^2/2}{1+\alpha}\right)$$

To infer the human capital, h, of a citizen with income y, we invert the equation $y = G(\bar{h})h = \bar{h}^{\alpha}h$ to get $h = y\bar{h}^{-\alpha}$, i.e., $\ln h = \ln y - \alpha \ln \bar{h}$

Figure 4 plots the RHS of the FOC (17) at the status-quo point at which $Z = 0$, i.e. the point at which there is no migration. The vertical axis measures the marginal benefit of allowing a migrant in; the benefit depends on the migrant's level of h. The figure shows that for some values of θ – say around $\theta = 0.8$, the marginal benefit of migration is negative at all levels of migration. Because the first-order condition is

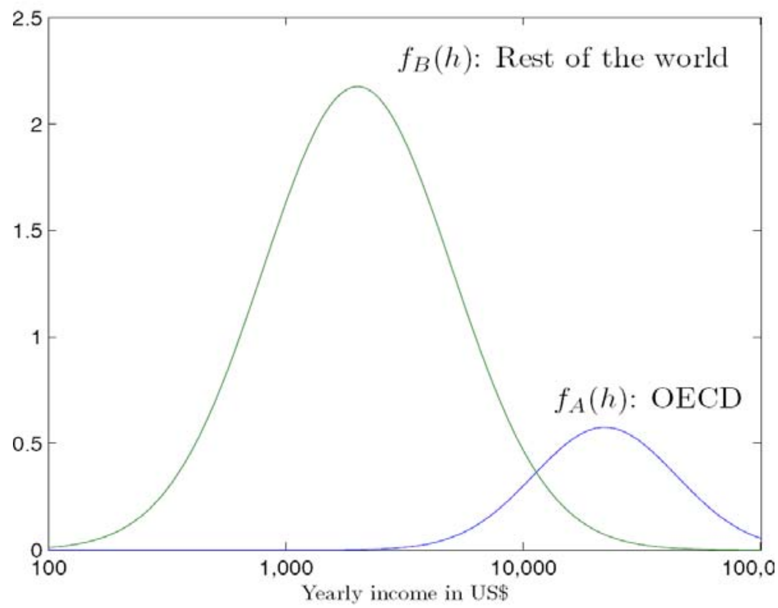

Figure 3: CALIBRATED DISTRIBUTIONS OF A AND B

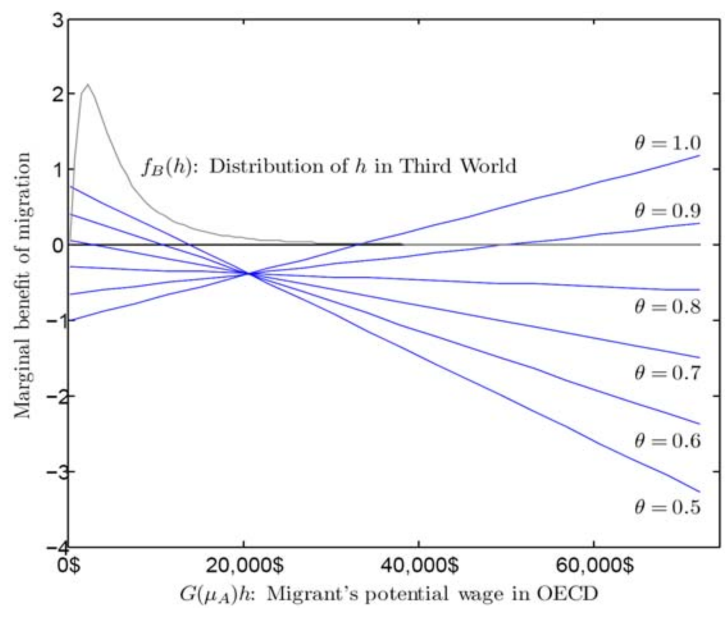

Figure 4: THE FIRST-ORDER CONDITION AT THE STATUS QUO

linear in h, the gain to migrating a worker of type h is either decreasing or increasing in h depending on the sign of $\frac{\theta}{\bar{h}_A} - \frac{1-\theta}{\bar{h}_B}$.

The brain-drain region $\theta \in [\theta_{\text{BD}}, 1]$ — The slope of the FOC changes sign at

$$\theta_{\text{BD}} = \frac{1}{1 + \bar{h}_B / \bar{h}_A} \approx 0.88 \text{ in the calibrated example,}$$

where BD is for brain-drain: If (as is the case in this calibration) F_B has unbounded support, then for any $\theta > \theta_{BD}$, some very smart B-people should go to A, and there will be a brain drain.

The skim-the-bottom region of $\theta \in [0, \theta_{\text{SB}}]$.—There is another threshold, call it θ_{SB}, below which country A will only receive low-h types. Suppose that the lowest level of h in the support of F_B is zero (as, again, is the case in the calibration). Then as shown in Figure 5,

$$\theta_{\text{SB}} = \frac{1+C}{2+C} \approx 0.71$$

where $C = \ln(\bar{h}_A / \bar{h}_B)$. This is also apparent in Figure 4 in which the FOC for $\theta = 0.7$ barely crosses the zero axis in the neighborhood of zero.

The inaction region $\theta \in (\theta_{\text{SB}}, \theta_{\text{BD}})$.—In this region, efficiency losses stemming from the mixing (see Proposition 1) overwhelm the redistributive gains. It is not worth moving the high-skilled B-natives to A because, while this would raise $G\left(\bar{h}_A\right)$, it would reduce $G\left(\bar{h}_B\right)$ by too much. At these intermediate θ's, it is not that the planner does not value the A-natives; he simply values the B-natives too much to allow a brain drain from B to occur.

The optimal policy.—The optimal policy is described in Figure 5. The purple area is the set of people who can move under the optimal policy. The vertical axis in Panel 1 measures $G\left(\mu_A\right) h$, the wage that a migrant of type h would earn in country A assuming that no one else was allowed to move so that average skills in A were at their pre-migration (i.e., current) level of μ_A. For $\theta \leq 0.71$, the unskilled B-natives migrate to A, and for $\theta \geq 0.89$, the skilled B-natives migrate. In between, migration is zero.

The numbers moving are huge. At $\theta = 0.5$, more than half of the B-natives would optimally be moved to A. By comparison, the numbers migrating at high levels of θ are tiny – not much more than the top percentile of h in country B would be allowed to migrate when $\theta = 1$.

6 The effect of immigration on the skill premium

To check the robustness of our conclusions with respect to the introduction of a second skill and physical capital, we now assume that there are two skills and two tasks, task 1 and task 2. We now assume that aggregate output is

$$Y = G\left(\bar{h}\right) K^{1-\beta_1-\beta_2} H_1^{\beta_1} H_2^{\beta_2}$$

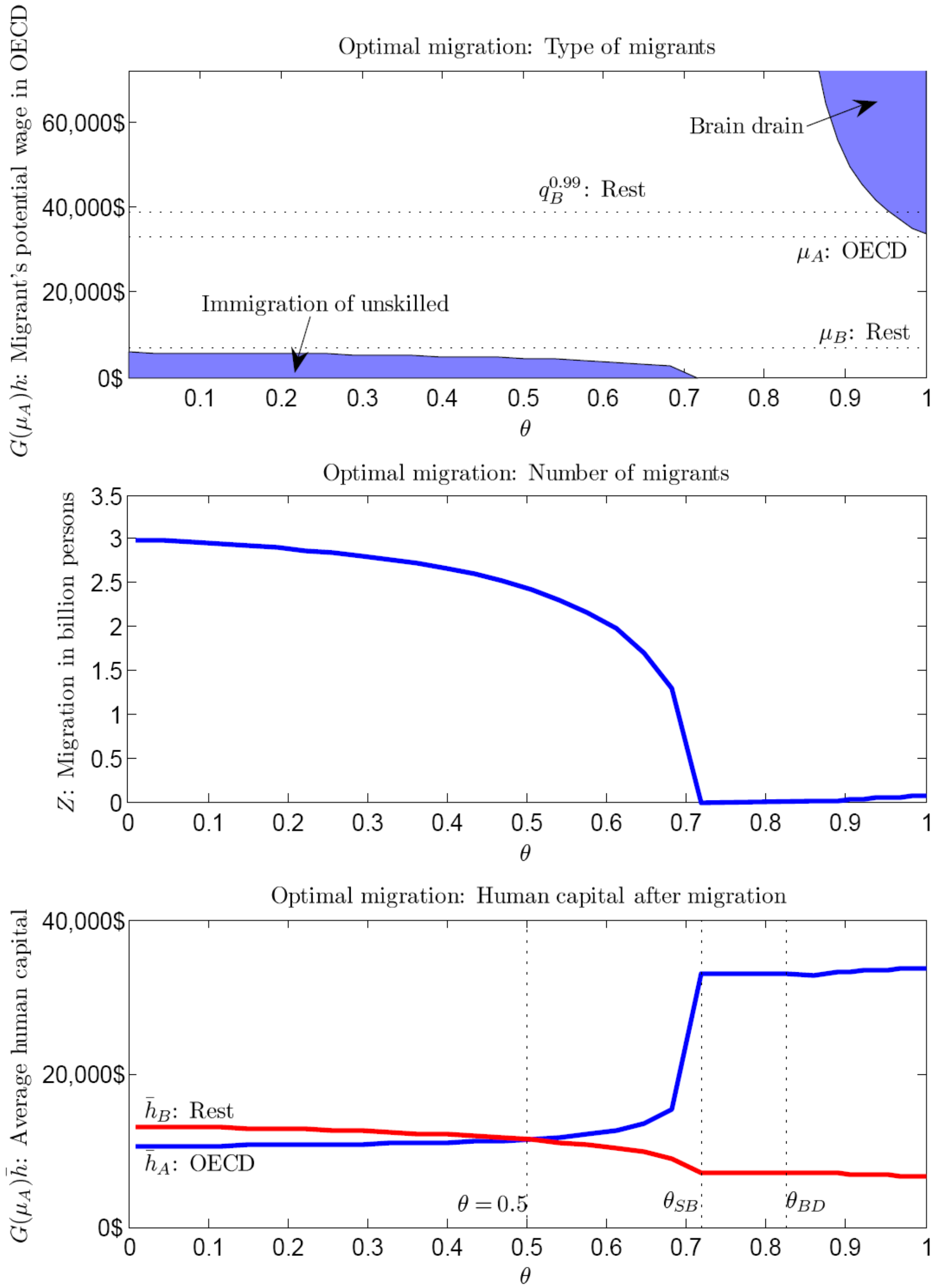

Figure 5: The optimal policy

where G is the same as before, K is capital, and where H_i is the total amount employed of skill i. Given the Cobb-Douglas nature of production, K will affect the level of wages, but not the skill premium.

We suppose that each person has two skills: A person of type h has an amount

$$h_1 = h$$

of type-1 skill, and an amount

$$h_2 = h - \varepsilon.$$

of type-2 skill.

Comparative advantage.—The ratio of type-2 skill to type-a skill, $\frac{h_2}{h_1} = 1 - \frac{\varepsilon}{h}$, is increasing in h, which means that in equilibrium the high-h people will opt to work in task 2, and that low-h people will opt to work in task 2. Let h_m denote the marginal worker. In that case, the aggregate supplies of the two skills are

$$H_1 (h_m, F_B) = \int_0^{h_m} h dF_B (h) \quad \text{and} \quad H_2 (h_m, F_B) = \int_{h_m}^{\infty} (h - \varepsilon) \, dF_B (h)$$

respectively. Let the two skill prices, i.e., the two wages per efficiency unit, be w_1 and w_2. Let us suppose that there is an agent with skill level h_m who is indifferent between the two occupations. Such indifference implies that the agent's income in the two occupations is the same:

$$w_1 h_m = w_2 (h_m - \varepsilon) .$$

Rearranging, we can solve for w_1 in terms of w_2 and h_m:

$$w_1 = \left(1 - \frac{\varepsilon}{h_m} \right) w_2$$

For a person with an arbitrary skill-level h, then, we can express

$$\text{Earnings in occupation } 1 = w_1 h = \left(1 - \frac{\varepsilon}{h_m} \right) w_2 h$$

and

$$\text{Earnings in occupation } 2 = w_2 (h - \varepsilon)$$

and therefore the ratio of earnings in the two occupations is

$$\frac{w_2 (h - \varepsilon)}{w_1 h} = \frac{h - \varepsilon}{h - \varepsilon \frac{h}{h_m}} \quad \begin{cases} > 1 & \text{if} \quad h > h_m \\ < 1 & \text{if} \quad h < h_m \end{cases}$$

Therefore for every ε, the high-skilled prefer occupation 2 and the low-skilled prefer occupation 1.

Letting $\varepsilon \to 0$.—As $\varepsilon \to 0$, we have

$$H_1(h_m, F_B) \to \int_0^{h_m} h dF_B(h) \quad \text{and} \quad H_2(h_m, F_B) \to \int_{h_m}^{\infty} h dF_B(h).$$

Now w_1 and w_2 will converge to w, given by

$$w = \text{MPH}_1 = G(\bar{h})\beta_1 \left(\frac{K}{H_1}\right)^{1-\beta_1-\beta_2} \left(\frac{H_2}{H_1}\right)^{\beta_2} = G(\bar{h})\beta_2 \left(\frac{K}{H_1}\right)^{1-\beta_1-\beta_2} \left(\frac{H_2}{H_1}\right)^{\beta_2-1} = \text{MPH}_2.$$

That is, as $\varepsilon \to 0$, the two wages per unit of skill must become the same, or else people would not be happy in one of the two occupations. This allows us to solve for the factor ratios: $\frac{H_1}{H_2} = \frac{\beta_1}{\beta_2}$.

Equivalence to the one-task case without physical capital.—We shall now show that our results carry over fully to this case. Let $H_1 + H_2 = H$ be the total stock of human capital in a given economy, be it A or B. In that economy,

$$\frac{H_1}{H} = \frac{\beta_1}{\beta_1 + \beta_2}, \text{ and } \frac{H_2}{H} = \frac{\beta_2}{\beta_1 + \beta_2}.$$

That economy's output will therefore be

$$Y = \tilde{G}(\bar{h}) K^{1-(\beta_1+\beta_2)} H^{\beta_1+\beta_2}, \quad \text{where}$$

where

$$\tilde{G}(\bar{h}) = G(\bar{h}) \left(\frac{\beta_1}{\beta_1 + \beta_2}\right)^{\beta_1} \left(\frac{\beta_2}{\beta_1 + \beta_2}\right)^{\beta_2}.$$

The wage per unit of skill, is just $w = \frac{\partial Y}{\partial H_1} = \frac{\partial Y}{\partial H_2}$. Firms still have constant-returns-to-scale production functions, and they make zero profits. In the long run, if K adjusts to a world interest rate or to the steady-state rate determined by discounting, then K is proportional to H and therefore Y is linear in H. Under this assumption from here on, we can apply our above analysis where $Y = \tilde{G}(\bar{h}) H$ with no further changes, so that all our results go through with two tasks and an elastic capital stock.

Skill premia: Pre-migration, short-run, and long-run.—The skill premium in an economy depends on the distribution of h among its residents – call that F – and on the total amount of skill employed in tasks 1 and 2, i.e., $\int_0^{h_m} hdF$ and $\int_{h_m}^{\infty} hdF$. The skill premium for the marginal worker then is

$$p(h_m, F) \equiv \frac{\text{MPH}_2(h_m, F)}{\text{MPH}_1(h_m, F)},$$

which is the skill-price ratio. The ratio of the average wages of workers in the high-h group relative to the low-h group is

$$\text{sp}(h_m, F) \equiv p(h_m, F) \frac{F(h_m) \int_{h_m}^{\infty} hdF}{(1 - F[h_m]) \int_0^{h_m} hdF}$$

and this is what one would ordinarily call "the" skill premium. Now we shall analyze the skill premium in A in three situations: (A) before the migration, (B) right after the migration, and (C) in the long run. The model has no adjustment cost for moving between countries or for changing jobs, and so the following is meant to be only suggestive of what may happen to the skill premium if such costs did exist. The presence of these costs would, however, affect the planner's policy, complicating the analysis, and so we leave them out. But we shall imagine that following the migration, the skill-level of the marginal worker does not change "for a while."

The pre- and post-migration distribution of h in country A.—The pre-migration distribution of h is F_A. Its post-migration distribution depends on θ, and it is

$$\tilde{F}_\theta(h) = \frac{F_A(h) + n\int_0^h \phi(h', \theta)\, dF_B(h')}{1 + n\int_0^\infty \phi(h', \theta)\, dF_B(h')}.$$

The pre- and post-migration h_m in country A.—Denote country A's pre-migration marginal worker's skill by $h_{m,0}$, and its post-migration marginal worker's skill by $h_{m,\theta}$. After the migration, economy A will have distribution $\tilde{F}_\theta$ which, in turn, will dictate a marginal worker $h_{m,\theta}$. But for a while, the marginal worker remains at $h_{m,0}$ so that the total skills in task 1 and 2 would in the short run be $\int_0^{h_{m,0}} h d\tilde{F}_\theta$ and $\int_{h_{m,0}}^\infty h d\tilde{F}_\theta$, respectively. This will generally mean that $p\left(h_{m,0}, \tilde{F}_\theta\right) \neq 1$, and that workers would wish to change tasks. In the long run, however, workers would move, the marginal worker would become $h_{m,\theta}$, and $p\left(h_{m,\theta}, \tilde{F}_\theta\right) \equiv p(\theta)$ would again equal unity, its "long-run" level portrayed in Figure 6.

We now make the following calculations, and plot the results in Figure 6:

The pre-migration premium.—We now imagine economy A starts in a state of equilibrium where all workers are happy in the task they are performing. Then the pre-migration skill premium is $\mathrm{sp}(h_{m,0}, F_A)$. We do not plot this quantity because it can be inferred from the positions of the curves at those values of $\theta \in (0.72, 0.85)$ for which no migration takes place. The middle panel shows that the pre-migration skill premium, $\mathrm{sp}(h_{m,0}, F_A)$, is about 2.5, or 150%.

The post-migration short-run "disequilibrium" skill premium.—This is $\mathrm{sp}\left(h_{m,0}, \tilde{F}_\theta\right)$, the steeper of the two curves in the middle panel of Figure 6. It is negatively related to the skill-level of the immigrant pool. The bulk (but not all) of the premium is caused by the difference in the composition of the two skill groups, and not in the difference in the skill prices, i.e., not by a departure of the skill-price ratio $p\left(h_{m,0}, \tilde{F}_\theta\right)$ from unity; the latter remains below 1.8 for low θs and above 0.75 for high θs.

The post-migration long-run equilibrium skill premium.—Here $p\left(h_{m,\theta}, \tilde{F}_\theta\right)$ is once again unity, and the entire skill premium now originates in skill differences between

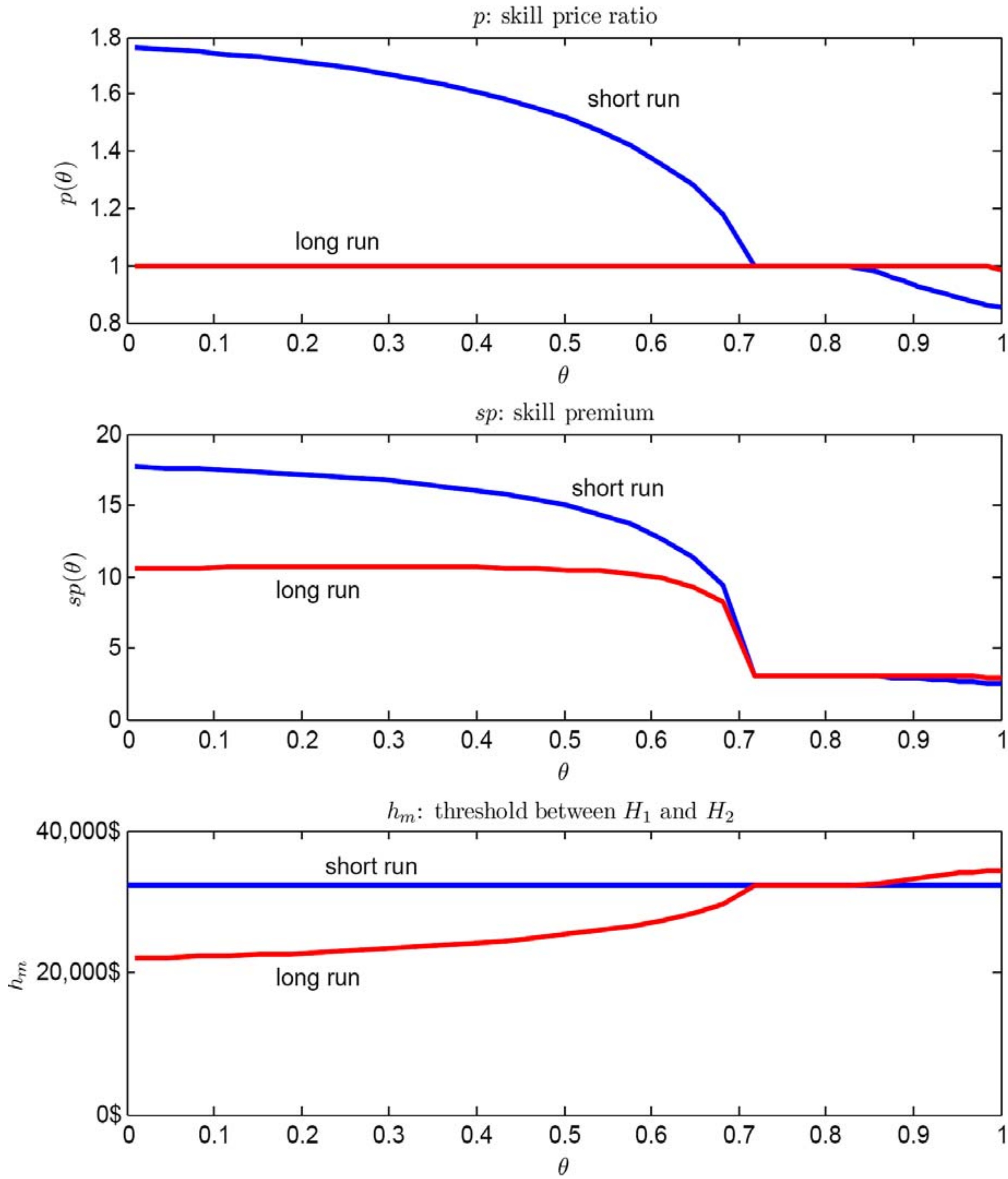

Figure 6: The short-run and long-run effect on the skill premium

the two groups. The skills are so vastly different, however, that $\mathrm{sp}\left(h_{m,\theta}, \tilde{F}_\theta\right)$ gets as high as 10.5 for low θs, and never gets below 3, even when θ is high.[13]

7 The price of high morals in A

If people's skills differ because of poor endowments and not on conscious investment choices, then a low-skill person is not at fault for not being high skilled. Arguably, then, a skill-neutral immigration policy is more "just" because it gives every agents in the poor countries an equal chance to migrate, regardless of his or her skill. The host country, however, generally prefers to admit high-skilled immigrants. We end the analysis by asking how much the host country loses by making its immigration policy skill neutral so that every B-native has the same chance of migrating to A.

In a brain-drain policy, Country A chooses the lowest acceptable level of immigrant skills. Let $\gamma(h_m) = n[1 - F_B(h_m)] =$ the number of immigrants when the marginal immigrant has ability h_m. Then Proposition 3 says that the optimal skill-biased policy yields country A average skill of

$$\bar{h}(h_m) = \frac{\mu_A + \gamma(h_m) E(h_m \mid h \geq h_m)}{1 + \gamma(h_m)}$$

Under the skill-neutral policy, by contrast, the mean ability of the immigrants is not $E(h_m \mid h \geq h_m)$ but μ_B. If that policy admits the same number of people (i.e., $\gamma(h_m)$ people), the post-immigration average skills in A would be

$$H(h_m) = \frac{\mu_A + \gamma(h_m)\mu_B}{1 + \gamma(h_m)}.$$

Then the price of high morals is then the difference in the domestic efficiency wage (i.e., the difference in G) under the two policies:

$$P(h_m) \equiv \bar{h}(h_m) - H(h_m) = \frac{\gamma(h_m)}{1 + \gamma(h_m)}[E(h \mid h \geq h_m) - \mu_B] \geq 0.$$

If the support of F_B (the skill distribution in B) is $[h_{\min}, h_{\max}]$, i.e., a bounded interval, then $P(h)$ is an inverted-U-shaped curve, starting at zero when $h = h_{\min}$ and ending at zero when $h = h_{\max}$. Even some unbounded distributions have this property as long as the tail is not too thick.

The price of morals in the Pareto case.—This distribution allows us to calculate the price analytically. Let the distribution in country B be

$$F_B(h) = 1 - \left(\frac{h}{h_{\min}}\right)^{-\rho}, \tag{20}$$

[13] In estimating the elasticity of wages with respect to the ratio of immigrants to natives, Borjas and Katz (2005) distinguish between the larger short-run effect where the capital stock is assumed to be fixed, and the smaller long-term effect on wages where capital adjusts.

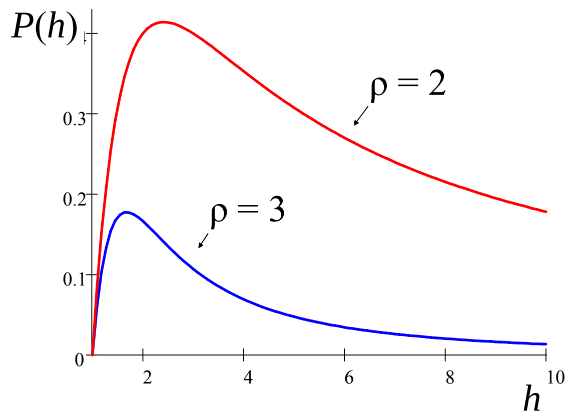

Figure 7: The price of morals for the Pareto case (20) with $h_{\min}$=1

in which case $E\left(\tilde{h} \mid \tilde{h} \geq h\right) = \frac{\rho}{\rho-1}h$ and $\gamma(h) = \left(\frac{h}{h_{\min}}\right)^{-\rho}$. Then $E\left(\tilde{h} \mid \tilde{h} \geq h\right) - \mu_B = \frac{\rho}{\rho-1}(h - h_{\min})$, and therefore

$$P(h) = \frac{1}{1 + \left(\frac{h}{h_{\min}}\right)^{\rho}} \frac{\rho}{\rho - 1}(h - h_{\min}) \geq 0.$$

Figure 7 take the case $h_{\min} = 1$ and plot the result for two values of ρ.The bigger is ρ, the smaller is the advantage of the skill-biased policy. This makes sense because the variance of the Pareto distribution decreases with ρ, and in the limit, as $\rho \to \infty$ all the B-natives become alike. Also, the right tail of the Pareto is thicker for *lower* values of ρ.

While skill-neutral policies sound good "in principle," they do not maximize welfare. Even if migration is the only redistributive tool – and that is what we have been assuming in this paper – skill-biased policies do better for the natives of the host country.

8 Conclusion

Egalitarian optimal immigration policy from the world perspective, taking into account the economic costs of immigration to the host and source countries, may still require a significant and abrupt relaxation of the restrictive immigration policies currently imposed by the rich countries. With increasing globalization, the third world countries are likely to acquire a greater voice and request greater access to world labor markets. It will probably become harder for richer countries to justify their non-discriminatory and redistributive welfare policies at home, while denying access to their labor markets to citizens of poorer countries, basing the exclusion simply on ethnicity and nationality[14]. While the deep contradictions between the democratic values of the West and the limitations on free access to world labor markets based on nationality have only recently began to surface, they are likely to become increasingly apparent in the future, and enter political discourse through international organizations like the UN or the World Bank. Political negotiations and compromises, however, may at best yield a gradual relaxation of restrictions on labor mobility, as in the case of a slowly expanding EU or the phased legalization of illegal immigrants in the US, rather than an abrupt switch to free immigration that an egalitarian parametrization of our model suggests.

References

[1] Acemoglu D. and Angrist J. (2001) "How Large are the Social Returns to Education: Evidence from Compulsory Schooling Laws," in Ben Bernanke and Kenneth Rogoff (Editors), *NBER Macroeconomic Annual* (2000): 9-59.

[2] Benhabib, Jess. "On The Political Economy of Immigration," *European Economic Review* 40 (1996): 1737-44.

[3] Borjas, George J., "The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market," Quarterly Journal of Economics 118 (November 2003); 1335-74.

[4] Borjas, George J., and Lawrence F. Katz, "The Evolution of the Mexican-Born Workforce in the United States," Working Paper 11281, NBER, Cambridge, (April 2005), http://www.nber.org/papers/w11281.

[5] Card, David, "Is the New Immigration Really So Bad?" *Economic Journal* 115, no. 507 (November 2005): F300-23.

[6] Butcher, Kristin F., and Anne Morrison Piehl. "Recent Immigrants: Unexpected Implications for Crime and Incarceration," *Industrial and Labor Relations Review* 51, no. 4 (1998): 654-79.

[7] Clark, Gregory. "Why Isn't the Whole World Developed? Lessons from the Cotton Mills." *Journal of Economic History* 47, no. 1 (March 1987): 141-173.

[8] Ciccone A. and G. Peri "Identifying Human Capital Externalities: Theory with Applications", *Review of Economic Studies* 73 (2006): 381-412.

[14] By contrast, however, the French Interior Minister Sarkozy recently declared "It is the right of our country, like all the great democracies of the world, to choose which foreigners it allows to reside on our territory." See http://www.migrationpolicy.org/pubs/Backgrounder2_France.php

[9] Coleman, James. "Social Capital in the Creation of Human Capital" *American Journal of Sociology* 94, Supplement: *Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure* (1988): S95-S120.

[10] Coleman, Jules, and Sarah Harding. 1995. "Citizenship, the Demands of Justice, and the Moral Relevance of Political Borders." In Schwartz (1995).

[11] Easterly, William. *The White Man's Burden: Why The West's Efforts To Aid the Rest Have Done So Much Ill and So Little Good*. Penguin Press, New York, 2006.

[12] Friedberg, Rachel, and Jennifer Hunt. "The Impact of Immigrants on Host Country Wages, Employment and Growth." *Journal of Economic Perspectives* 9, no. 2 (Spring, 1995): 23 - 44.

[13] Lucas, Robert E., Jr. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22, no. 1 (1988): 3-42.

[14] Lundborg, Per, and Paul S. Segerstrom "The Growth and Welfare Effects of International Mass Migration." *Journal of International Economics* 56 (2002): 177–204.

[15] Moretti, Enrico. "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data." *Journal of Econometrics* 121, no. 1 (2004): 175-212.

[16] National Research Council. *The New Americans: Economic, Demographic, and Fiscal Effects of Immigration.* National Academy Press. Washington, D.C., 1997.

[17] Peri, Giovanni, and Gianmarco Ottaviano, "Rethinking the Effects of Immigration on Wages," NBER WP #12497 (2006) Cambridge.

[18] Peri, Giovanni, and Susana Iranzo, "Schooling Externalities, Technology and Productivity: Theory and Evidence from U.S. States," NBER WP #12440 (2006), Cambridge.

[19] Rauch James E. "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities" *Journal of Urban Economics* 34, no. 3 (November 1993): 380-400.

[20] Sala-i-Martin, Xavier. "The World Distribution of Income." *Quarterly Journal of Economics* 121, no 2 (May 2006): 351-397.

[21] Schwartz, Warren (ed.). *Justice in Immigration.* Cambridge: Cambridge University Press, 1995.

[22] Trebilcock, Michael. "The Case for a Liberal Immigration Policy." in Schwartz (1995).

9 Appendix

9.0.1 Proof of Lemma 1

The proof consists of showing that if (6) holds, then (7) also does. Since $g_A = \alpha\left(\ln\left(h_A + nxh_B\right) - \ln\left(1+nx\right)\right)$, Therefore,

$$\frac{1}{\alpha}g_A'\left(x\right) = \frac{n}{z^{-1}+nx} - \frac{n}{1+nx} < 0,$$

and

$$\frac{1}{\alpha}g_A''\left(x\right) = \frac{d}{dx}\left(\frac{n}{z^{-1}+nx} - \frac{n}{1+nx}\right) = \left(-\left(\frac{n}{z^{-1}+nx}\right)^2 + \left(\frac{n}{1+nx}\right)^2\right) > 0.$$

Therefore

$$\begin{aligned} \frac{g_A''}{g_A'} &= \frac{\left(-\left(\frac{n}{z^{-1}+nx}\right)^2 + \left(\frac{n}{1+nx}\right)^2\right)}{\frac{n}{z^{-1}+nx} - \frac{n}{1+nx}} \\ &= \frac{\left[\left(\frac{n}{1+nx}\right) - \left(\frac{n}{z^{-1}+nx}\right)\right]\left[\left(\frac{n}{1+nx}\right) + \left(\frac{n}{z^{-1}+nx}\right)\right]}{\frac{n}{z^{-1}+nx} - \frac{n}{1+nx}} \\ &= -\left[\left(\frac{n}{1+nx}\right) + \left(\frac{n}{z^{-1}+nx}\right)\right]. \end{aligned}$$

Since

$$\begin{aligned} g_A - g_B &= \alpha\ln\left(\frac{h_A + nxh_B}{1+nx}\right) - \alpha\ln h_B \\ &= \alpha\ln\left(\frac{h_A + nxh_B}{\left(1+nx\right)h_B}\right) \end{aligned} \tag{21}$$

so that

$$\frac{g_A - g_B}{g_A'} = \frac{\alpha\ln\left(\frac{\left(\frac{h_A}{h_B}+nx\right)}{(1+nx)}\right)}{\alpha\left(\frac{n}{z^{-1}+nx} - \frac{n}{1+nx}\right)} = \frac{\ln\left(\frac{\left(z^{-1}+nx\right)}{(1+nx)}\right)}{\left(\frac{n}{z^{-1}+nx} - \frac{n}{1+nx}\right)}$$

Now from the definitions of h_A, h_B, and g_A,

$$
\begin{aligned}
\frac{g_A''}{g_A'} &= \frac{\left(-\left(\frac{n}{z^{-1}+nx}\right)^2+\left(\frac{n}{1+nx}\right)^2\right)}{\frac{n}{z^{-1}+nx}-\frac{n}{1+nx}} \\
&= \frac{\left[\left(\frac{n}{1+nx}\right)-\left(\frac{n}{z^{-1}+nx}\right)\right]\left[\left(\frac{n}{1+nx}\right)+\left(\frac{n}{z^{-1}+nx}\right)\right]}{\frac{n}{z^{-1}+nx}-\frac{n}{1+nx}} \\
&= -\left[\left(\frac{n}{1+nx}\right)+\left(\frac{n}{z^{-1}+nx}\right)\right].
\end{aligned}
$$

So, for (7) to hold, one needs that

$$
-\left[\left(\frac{n}{1+nx}\right)+\left(\frac{n}{z^{-1}+nx}\right)\right] < -\frac{2\,(1-\theta)\,n}{\theta+(1-\theta)\,nx}
$$

or that

$$
\frac{1}{2}\left[\left(\frac{1}{1+nx}\right)+\left(\frac{1}{z^{-1}+nx}\right)\right] > \frac{(1-\theta)}{\theta+(1-\theta)\,nx} = \frac{1}{\frac{\theta}{1-\theta}+nx}
$$

Now (6) implies $\frac{\theta}{(1-\theta)} = -\left(n\left(\frac{g_A-g_B}{g_A'}\right)+nx\right)$, which is equivalent to

$$
\frac{1}{2}\left[\left(\frac{1}{1+nx}\right)+\left(\frac{1}{z^{-1}+nx}\right)\right] > \frac{1}{-\left(n\left(\frac{g_A-g_B}{g_A'}\right)+nx\right)+nx}
$$

or, from (21), to

$$
\frac{1}{2}\left[\left(\frac{1}{1+nx}\right)+\left(\frac{1}{z^{-1}+nx}\right)\right] > \frac{1}{-n\left(\frac{g_A-g_B}{g_A'}\right)} = \frac{-1}{\frac{\ln\left(\frac{(z^{-1}+nx)}{(1+nx)}\right)}{\left(\frac{1}{z^{-1}+nx}-\frac{1}{1+nx}\right)}} = \frac{-\left(\frac{1}{z^{-1}+nx}-\frac{1}{1+nx}\right)}{\ln\left(\frac{(z^{-1}+nx)}{(1+nx)}\right)}
$$

Now, this condition can be re-written as

$$
\begin{aligned}
\ln\left(\frac{(z^{-1}+nx)}{(1+nx)}\right) &> \frac{2\left(\frac{1}{1+nx}-\frac{1}{z^{-1}+nx}\right)}{\left[\left(\frac{1}{1+nx}\right)+\left(\frac{1}{z^{-1}+nx}\right)\right]} = 2\frac{\left(\frac{1}{1+nx}\right)^2-\left(\frac{1}{z^{-1}+nx}\right)^2}{\left(\left(\frac{1}{1+nx}\right)+\left(\frac{1}{z^{-1}+nx}\right)\right)^2} \\
\ln\left(\frac{1}{(1+nx)}\right)-\ln\left(\frac{1}{(z^{-1}+nx)}\right) &> \frac{2\left(\frac{1}{1+nx}-\frac{1}{z^{-1}+nx}\right)}{\left[\left(\frac{1}{1+nx}\right)+\left(\frac{1}{z^{-1}+nx}\right)\right]},
\end{aligned}
$$

or, if we write $A=\frac{1}{(1+nx)}$ and $B=\frac{1}{(z^{-1}+nx)}$, to

$$
\ln(A)-\ln(B) = \ln\left(\frac{A}{B}\right) > 2\left(\frac{A-B}{A+B}\right) = 2\frac{\frac{A}{B}-1}{\frac{A}{B}+1}
$$

That is, we need, for all $y > 1$

$$\ln y > 2\frac{y-1}{y+1}$$

The LHS and RHS are both zero at $y = 1$. Therefore it's enough to show that the first derivative of the LHS is more positive than the first-derivative of the RHS for all y. That is, we are done if we can show that

$$\begin{aligned} \frac{1}{y} &> 2\left(\frac{1}{y+1} - \frac{y-1}{(y+1)^2}\right) = \frac{2}{y+1}\left(1 - \frac{y-1}{y+1}\right) \\ &= \frac{2}{y+1}\left(\frac{y+1-y+1}{y+1}\right) = \frac{4}{(y+1)^2} \end{aligned}$$

So, we need to show that $(y+1)^2 > 4y$, or that $y^2 + 2y + 1 > 4y$, or that $y^2 + 1 > 2y$. Now, the LHS and the RHS both equal 2 when $y = 1$, but the derivative of the LHS, $2y$, always exceeds the derivative of the RHS. Therefore (7) holds. ■

9.0.2 Proof of Proposition 2

If the claim is true, the point of indifference is where $\theta = \frac{n\left(\ln z - \ln\left(\frac{1+nz}{1+n}\right)\right)}{(1-n)\ln\left(\frac{1+nz}{1+n}\right) + n\ln z}$. We shall now show this is so. This is the curve in (θ, z)-space at which the expression (5) is the same at $x = 1$ and at $x = 0$. It is the set of (θ, z) pairs at which

$$\theta g\left(\bar{h}[0]\right) + (1-\theta)\, n\left[g\left(\bar{h}[0]\right) + g\left(\bar{h}_B\right)\right] = \theta g\left(\bar{h}[1]\right) + (1-\theta)\, ng\left(\bar{h}\right)$$

if

$$G^{\theta} G^{*(1-\theta)n} = G^{\theta+(1-\theta)n}$$

Now take $G\left(\bar{h}\right) = \bar{h}^{\alpha}$. Then for $x = 0$, $G = h^{\alpha}$, $G^* = \left(\bar{h}_B\right)^{\alpha}$ and for $x = 1$ $G = \left(\frac{h+nh^*}{1+n}\right)^{\alpha}$. Therefore along the boundary

$$\begin{aligned} h^{\alpha\theta}\left(h^*\right)^{\alpha(1-\theta)n} &= \left(\frac{h+nh^*}{1+n}\right)^{\alpha(\theta+(1-\theta)n)} \\ \left(\frac{h^*}{h}\right)^{-\alpha\theta}\left(h^*\right)^{\alpha(\theta+(1-\theta)n)} &= \left(\frac{1+n\frac{h^*}{h}}{1+n}\right)^{\alpha(\theta+(1-\theta)n)} \qquad (h)^{\alpha(\theta+(1-\theta)n)} \\ z^{(1-\theta)n} &= \left(\frac{1+nz}{1+n}\right)^{\alpha(\theta+(1-\theta)n)} \end{aligned}$$

or

$$\begin{aligned} (1-\theta)\, n\ln z &= (\theta + (1-\theta)\, n)\ln\left(\frac{1+nz}{1+n}\right) \\ n\ln z - n\ln\left(\frac{1+nz}{1+n}\right) &= \theta\left((1-n)\ln\left(\frac{1+nz}{1+n}\right) + n\ln z\right) \end{aligned}$$

from which the claim follows.

9.0.3 Proof of Proposition 5

Proof. For a policy that allows mobility if and only if $h \leq \tilde{h}$, we have

$$\bar{h}_A\left(\tilde{h}\right) \equiv \frac{\mu_A + n\int_0^{\tilde{h}} hdF_B(h)}{1+nF_B(h^*)}, \quad \text{and} \quad \bar{h}_B\left(\tilde{h}\right) \equiv \frac{\int_{\tilde{h}}^{\infty} hdF_B(h)}{1-F_B(h^*)}.$$

Let

$$Q\left(\tilde{h}\right) \equiv \frac{\bar{h}_A\left(\tilde{h}\right)}{\bar{h}_B\left(\tilde{h}\right)}.$$

Since the policy (12) is indexed by a single number, the critical bound, $\tilde{h}$, we can write the criterion as $W\left(\tilde{h}\right)$, i.e., as a function of the real number $\tilde{h}$ alone. Now, as we did in (17), we perform a variation in the **entire** function $z(h)$ around the hypothesized optimum

$$z(h) = \begin{cases} nf_B(h) & \text{for } h < \tilde{h} \\ 0 & \text{for } h > \tilde{h} \end{cases}$$

For this class of bang-bang policies we can write (17) as a function of $\tilde{h}$ and h :

$$\frac{\partial W}{\partial z(h)} = \Psi\left(\tilde{h}, h\right) = \alpha(1-\theta)\left[\ln Q\left(\tilde{h}\right) + 1 - \frac{h}{\bar{h}_B\left(\tilde{h}\right)} - m\left(1 - \frac{h}{\bar{h}_A\left(\tilde{h}\right)}\right)\right]. \quad (22)$$

We shall show that (i) There exists an $\tilde{h} = \tilde{h}^* \in \left[0, \hat{h}\right]$ at which

$$\left.\frac{\partial W}{\partial z(h)}\right|_{h=\tilde{h}=\tilde{h}^*} = \Psi\left(\tilde{h}^*, \tilde{h}^*\right) = 0,$$

and that (ii) The solution is unique for $\tilde{h} \in \left[0, \hat{h}\right]$. Taken together, (i) and (ii) will imply that the optimal policy is (12).

(i) Note that $\bar{h}_A(0) = \mu_A$, $\bar{h}_B(0) = \mu_B$, and $Q(0) = \frac{\mu_A}{\mu_B} > 1$. From (18), at $\tilde{h} = 0$, $m = \frac{\theta}{1-\theta} \leq 1$. Therefore

$$\left.\frac{\partial W}{\partial z(h)}\right|_{h=\tilde{h}=0} = \Psi(0,0) = \alpha(1-\theta)\left[\ln\frac{\mu_A}{\mu_B} + 1 - m\right] > 0.$$

On the other hand, since $\hat{h} < \bar{h}_A\left(\hat{h}\right) = \bar{h}_B\left(\hat{h}\right)$, since $\ln Q\left(\hat{h}\right) = 0$ and since

$$m = \frac{\theta + (1-\theta)n}{(1-\theta)(1+n)} \begin{cases} = 1 & \text{if } \theta = 0.5 \\ < 1 & \text{if } \theta < 0.5 \end{cases},$$

$$\left.\frac{\partial W}{\partial z(h)}\right|_{h=\tilde{h}=\hat{h}}=\Psi\left(\hat{h},\hat{h}\right)=\alpha(1-\theta)\left(1-\frac{\hat{h}}{\bar{h}_A}\right)(m-1)=\begin{cases}0 & \text{if} \quad \theta=\frac{1}{2}\\ <0 & \text{if} \quad \theta<\frac{1}{2}\end{cases}$$

Therefore the continuous function $\left.\frac{\partial W}{\partial z(h)}\right|_{h=\tilde{h}}=\Psi\left(\tilde{h},\tilde{h}\right)$ crosses zero at least once.

(*ii*) *Uniqueness.*—We have shown that the solution to $\left.\frac{\partial W}{\partial z(h)}\right|_{h=\tilde{h}=\hat{h}}=\Psi\left(\tilde{h},\tilde{h}\right)=0$ exists. Now we show that the solution is unique on the interval $\left[0,\hat{h}\right]$. We can write

$$\left.\frac{\partial W}{\partial z(h)}\right|_{h=\tilde{h}}=\Psi\left(\hat{h},\hat{h}\right)=\alpha(1-\theta)\left[\ln Q\left(\tilde{h}\right)+1-m+\left(m-\frac{\bar{h}_A}{\bar{h}_B}\right)\frac{\tilde{h}}{\bar{h}_A}\right]$$

$$\left.\frac{\partial}{\partial\tilde{h}}\frac{\partial W\left(\tilde{h}\right)}{\partial z(h)}\right|_{h=\tilde{h}}=\frac{\partial\Psi\left(\hat{h},\hat{h}\right)}{\partial\tilde{h}}=\alpha(1-\theta)\left(\begin{array}{r}\frac{Q'\left(\tilde{h}\right)}{Q\left(\tilde{h}\right)}-m'\left(\tilde{h}\right)+\left(m'\left(\tilde{h}\right)-Q'\left(\tilde{h}\right)\right)\frac{\tilde{h}}{\bar{h}_A}+\\ +\left(m-\frac{\bar{h}_A}{\bar{h}_B}\right)\left(\frac{\bar{h}_A-\tilde{h}\frac{d\bar{h}_A}{d\tilde{h}}}{\left(\bar{h}_A\right)^2}\right)\end{array}\right)$$

We have $\theta<0.5$. Recall that $Z=\int^{\tilde{h}}z\left(h\right)dh$ and therefore $dZ/d\tilde{h}=z\left(\tilde{h}\right)\in\left[0,nf_B\left(\tilde{h}\right)\right]$. Therefore

$$\begin{aligned}m'\left(\tilde{h}\right) &= \frac{\left[(1-\theta)\left(1+Z\right)\right]\left(1-\theta\right)-\left(\theta+\left(1-\theta\right)Z\right)\left(1-\theta\right)}{\left(\left(1-\theta\right)\left(1+Z\right)\right)^2}z\left(\tilde{h}\right)\\ &= \frac{\left\{\left[(1-\theta)\left(1+Z\right)\right]-\left(\theta+\left(1-\theta\right)Z\right)\right\}\left(1-\theta\right)}{\left(\left(1-\theta\right)\left(1+Z\right)\right)^2}z\left(\tilde{h}\right)\\ &= \frac{\left\{\left[(1-\theta)\right]-\theta\right\}\left(1-\theta\right)}{\left(\left(1-\theta\right)\left(1+Z\right)\right)^2}z\left(\tilde{h}\right)=\frac{\left\{(1-2\theta\right\}\left(1-\theta\right)}{\left(\left(1-\theta\right)\left(1+Z\right)\right)^2}z\left(\tilde{h}\right)\geq 0\end{aligned}$$

So $m'\left(\tilde{h}\right)\geq 0$ if $\theta<0.5$. Remember, by definition, $Q\left(\tilde{h}\right)=\frac{\bar{h}_A}{\bar{h}_B}$. Then

$$Q'\left(\tilde{h}\right)=\frac{\bar{h}_A}{\bar{h}_B}\text{ so }Q'\left(\tilde{h}\right)<0\text{ since by construction }\frac{d\bar{h}_A}{d\tilde{h}}<0\text{ and }\frac{d\bar{h}_B}{d\tilde{h}}>0$$

and so $\left.\frac{\partial}{\partial\tilde{h}}\frac{\partial W\left(\tilde{h}\right)}{\partial z(h)}\right|_{h=\tilde{h}}=\frac{\partial\Psi\left(\hat{h},\hat{h}\right)}{\partial\tilde{h}}$ can be written as

$$\begin{aligned}&\alpha(1-\theta)\left(\frac{Q'\left(\tilde{h}\right)}{Q\left(\tilde{h}\right)}-m'\left(\tilde{h}\right)\left(1-\frac{\tilde{h}}{\bar{h}_A}\right)-Q'\left(\tilde{h}\right)\frac{\tilde{h}}{\bar{h}_A}+\left(m-\frac{\bar{h}_A}{\bar{h}_B}\right)\left(\frac{1-\frac{\tilde{h}}{\bar{h}_A}\frac{d\bar{h}_A}{d\tilde{h}}}{\bar{h}_A}\right)\right)\\ =\;&\alpha(1-\theta)\left(Q'\left(\tilde{h}\right)\left[\frac{1}{Q\left(\tilde{h}\right)}-\frac{\tilde{h}}{\bar{h}_A}\right]-m'\left(\tilde{h}\right)\left(1-\frac{\tilde{h}}{\bar{h}_A}\right)+\left(m-\frac{\bar{h}_A}{\bar{h}_B}\right)\left(\frac{1-\frac{\tilde{h}}{\bar{h}_A}\frac{d\bar{h}_A}{d\tilde{h}}}{\bar{h}_A}\right)\right)\\ =\;&\alpha(1-\theta)\left(Q'\left(\tilde{h}\right)\left[\frac{\bar{h}_B-\tilde{h}}{\bar{h}_A}\right]-m'\left(\tilde{h}\right)\left(1-\frac{\tilde{h}}{\bar{h}_A}\right)+\left(m-\frac{\bar{h}_A}{\bar{h}_B}\right)\left(\frac{1-\frac{\tilde{h}}{\bar{h}_A}\frac{d\bar{h}_A}{d\tilde{h}}}{\bar{h}_A}\right)\right)\end{aligned}$$

Now, at the optimum, $\bar{h}_B - \tilde{h} > 0$, $1 - \frac{\tilde{h}}{\bar{h}_A} > 0$, and $m - \frac{\bar{h}_B}{\bar{h}_A} < 0$. Since $\frac{d\bar{h}_A}{d\tilde{h}} < 0$, $\frac{\partial}{\partial \tilde{h}} \left. \frac{\partial W\left(\tilde{h}\right)}{\partial z(h)} \right|_{h=\tilde{h}} = \frac{\partial \Psi\left(\hat{h},\hat{h}\right)}{\partial \tilde{h}} < 0$, which means that there can be at most one crossing of zero on the interval $\left[0, \hat{h}\right]$, which we denote as $\tilde{h}^*$. ■