International migration, imperfect information, and brain drain

Vianney Dequiedt a, Yves Zenou b,∗

a Université d’Auvergne
b Stockholm University, IFN and GAINS

A R T I C L E   I N F O

Article history:
Received 21 September 2011
Received in revised form 8 October 2012
Accepted 10 October 2012
Available online 7 November 2012

JEL classification:
D51
J61
F22
O12

Keywords:
Asymmetric information
Self-selection of migrants
Skill-biased migration

A B S T R A C T

We consider a model of international migration where skills of workers are imperfectly observed by firms in the host country and where information asymmetries are more severe for immigrants than for natives. Because of imperfect information, firms statistically discriminate high-skilled migrants by paying them at their expected productivity. The decision of whether to migrate or not depends on the proportion of high-skilled workers among the migrants. The migration game exhibits strategic complementarities, which, because of standard coordination problems, lead to multiple equilibria. We characterize them and examine how international migration affects the income of individuals in sending and receiving countries, and of migrants themselves. We also analyze under which conditions there is positive or negative self-selection of migrants.

© 2012 Elsevier B.V. All rights reserved.

"How many immigrants does the United States want? And which types of immigrants should the country admit?" George J. Borjas (1999).

1. Introduction

Even though international migration is quite uncommon,1 it generates a lot of controversial debates. The current policy debate about international immigration in the United States focuses in fact almost entirely on high-skilled workers. This is not unique to the U.S. In Germany, for example, since the government approved Chancellor Gerhard Schroeder’s “green card” plan in 2000, the law gives 20,000 high-skilled immigrants 5 year temporary work permits in order to ease the perceived shortage of IT workers. Similar proposals of making temporary work permits more easily available for high-skilled immigrants are taking place in Great Britain, Ireland and even Sri Lanka. This can explain the overall tendency for migration rates to be much higher for the highly-skilled. Between 1990 and 2000, the total number of foreign-born individuals legally residing in the OECD member countries has been multiplied by 1.4, with a larger increase for highly skilled migrants (×1.64) than for low skilled migrants (×1.14) (Docquier and Marfouk, 2006). In high-income receiving countries, the concern is that the wrong individuals are trying to get in (Borjas, 1999), though this position has generated controversy (Card, 2005). It is not, however, clear if high-skilled migration always has a positive or negative effect on the source country. When there are positive spillovers associated with human capital (Lucas, 1988) or education is financed through taxation (Bhagwati and Rodriguez, 1975), the emigration of skilled labor can in fact hinder economic development (Benhabib and Jovanovic, 2012).

There is an interesting literature on international migration that has mainly focused on the difficulty of evaluating the skills and abilities of migrants and how this affects the labor market outcomes of both migrants and natives (see Section 2 below). The main objective and contribution of the present paper are theoretical and are to extend and generalize the theory of migration with imperfect information to account for statistical discrimination and the resulting possibility of multiple equilibria. We are also able to match several

1 See Fig. 1 in Hanson (2009) which, using data compiled by the United Nations, shows that in 2005 individuals residing outside of their country of birth comprised just 3% of the world’s population.

2 For overviews on these issues, see Bhagwati and Hanson (2009) and Hanson (2009).
empirical facts about international migration and to provide testable empirical predictions.

To be more precise, we develop a model of international migration from a "Southern" country to a "Northern" country (for example, from Mexico to the United States), where the North has an absolute advantage in terms of productivity and skill premium. Migrants can be high-skilled or low-skilled. Any worker can be assigned to two different tasks. In the simple task, all workers in a given country have the same low productivity whereas, in the more complex task, a high-skilled worker's productivity is higher than a low-skilled worker's productivity. The skill/ability of each migrant is not observed by Northern firms. The school record is, instead, observed but it is an imperfect signal of the skill/ability of the worker. As a result, Northern firms perform a noisy test. High-skilled migrants always pass the test while only a fraction of low-skilled migrants pass it. Firms statistically discriminate high-skilled migrants by paying workers who passed the test at their average productivity. This productivity is lower than the high-skilled migrants real productivity since some low-skilled workers (those who have passed the test) are wrongly considered as high-skilled. The quality of the test could be a proxy of the cultural distance between the two countries. It is clearly more difficult for the UK to assess the skill of a migrant from Kazakhstan than from India, given the past colonial history between the UK and India.

In this context, we consider a two-stage model, where, in the first stage, workers in the South decide whether to move and pay the migration cost, while, in the second stage, firms offer wages to the immigrant and native workers who are in the country. In the first stage, the decision of whether to migrate or not depends on the proportion of high-skilled workers among the migrants. In equilibrium, anticipations about the percentage of migrants of each type are rational, i.e. the anticipated productivity of migrants is equal to the true productivity of migrants. The migration game exhibits strategic complementarities, which, because of standard coordination problems, lead to multiple equilibria. The existence of multiple equilibria illustrates the coordination problem among migrants. Indeed, if they anticipate that no high-skilled worker migrates, then they anticipate that all migrants will be paid at the low-skilled native wage and those anticipations are self-confirming. If they anticipate that some high-skilled workers migrate, then they anticipate that migrants who successfully pass the screening test will be paid more than the low-skilled native wage. Depending on the value of the parameters of the model, several levels of high-skilled workers migration may be self-confirming. We show that two types of equilibria emerge. There is a high-discrimination equilibrium for which all immigrants are offered low-skilled tasks, irrespective of the outcome of the screening test. In that case, no high-skilled workers migrate; only the low-skilled ones. There is also a low-discrimination equilibrium, where all migrants who pass the screening test are offered high-skilled tasks, whereas all immigrants who fail the screening test are offered low-skilled tasks.

We characterize these equilibria and examine how international migration affects the incomes of individuals in sending and receiving countries, and of migrants themselves. We find that the quality of the screening test (i.e. the "social or cultural" distance between the two countries) affects wages as well as the skill composition of the migrants. The better is this test, the higher is the wage of high-skilled migrants and the "better" is the quality of migrants. Indeed, when the test becomes better, high-skilled migrants are more likely to migrate because they will be paid at their "real" productivity. Similar results are obtained for the ex ante proportion of low-skilled workers in the South and in the North and in the productivity difference between the North and the South.

We then endogenize the skill/education decision. In that case, in the first stage, without knowing their migration costs, workers in the South have to decide whether to be educated (high-skill) or not (low-skill). In the second stage, migration costs are revealed to workers and given their education level, workers in the South decide whether to move to the North and pay the migration costs or to stay in the South. In the third stage, firms offer wages to the immigrant and native workers who are in the North. Compared to the previous model, there are now two types of coordination problems. As in the benchmark model, coordination problems arise in the migration decision since workers do not know how many high-skilled Southern workers will migrate to the North. On top of this, coordination problems arise in the education decision since Southern workers do not know how many workers in the South will be educated. As a result, when deciding to be educated or not, they have to anticipate both how many workers will be educated in the South and how many high-skilled and low-skilled workers will migrate to the North.

As before, we have multiple equilibria, with high- and low-discrimination equilibria. In this model, the prospects of international migration can have positive effects for the South since some workers will decide to be educated or skilled because of the job opportunities in the North but then decide not to migrate because their migration costs will be too high. We completely characterize the high and the low-discrimination equilibrium and show, for example, that an increase in the skill premium between the South and the North decreases the equilibrium fraction of low-skilled workers in the South.

2. Related literature

As stated above, there is a small literature on the effect of asymmetric information on migration (Katz and Stark, 1984, 1986, 1987a,b; Kwok and Leland, 1982). The closest paper to ours is the one by Katz and Stark (1987a). They consider a model in which heterogeneous workers (in terms of skills) from a poor country consider to migrate to a rich country. They assume that foreign employers are less well-informed than the migrants about the workers' skills and statistically discriminate by giving the same average wage (or productivity) to all migrants, whatever their skills. Contrary to us, they focus on the differences between the perfect information and the asymmetric information cases. Their main result shows that the skill composition of the workforce can differ between the two regimes. In the perfect information case, it can be that low-skilled and high-skilled migrants migrate but not workers with intermediate skills whereas this is never possible in the asymmetric case since, if it is beneficial for a migrant of a given skill to migrate, then it is automatically true for all migrants of a lower skill. This (pooling equilibrium) result is driven by the fact that employers statistically discriminate but also because all workers have the same migration cost. They then extend this model to allow for workers to signal their skill, assuming that the signalling cost is the same for all workers. They show that the top-skill individuals are the most likely to signal their quality. Because of signalling, the authors can retrieve a similar result than the one found in the perfect information case. Indeed, the equilibrium migration pattern that emerges is characterized by the fact that the least skilled migrate without signalling, the intermediate group does not migrate and the highly skilled migrate with a signal (this is shown in a numerical example but not proved formally for the general case). Finally, in the last part of their paper, Katz and Stark (1987a) introduce the possibility that the true skill of migrants can be discovered after some time. They show that, in this case, more high-skilled workers will migrate. In more recent papers, Stark (1995b) and Chau and Stark (1999) investigate the latter issue by focusing on return migration. As in the other models, because of unknown information about skills, local employers give to new migrants

3 The number of individuals migrating from the “South” to the “North” increased from 14 million in 1960 to 60 million in 2000 (Özden et al., 2011). They define the “North” as Australia, New Zealand, Japan, Canada, the United States, the EU-15, and other nations of the European Free Trade Association, and the “South” as all other countries. So “North” corresponds to countries that were developed, and “South” to countries that were developing, in 1960.
a wage based on the average product of the group of migrants. However, after some time, skills are discovered and only the low-skilled workers go back to their home country.

Compared to this literature, our contribution is to extend and generalize the theory of migration with imperfect information to account for statistical discrimination and the resulting possibility of multiple equilibria. We are also able to totally characterize all the possible equilibria and to endogenize the education/skill decision of migrants.

The concept of strategic complementarity in high-skilled emigration decisions has also been studied by De la Croix and Docquier (2012) in a different context a la Lucas (1988) with endogenous total factor productivity (TFP). In their model, if high-skilled workers anticipate a large (resp. low) brain drain, they anticipate low (resp. high) TFP levels and migrate more. A novelty in this paper is that they apply quantitative theory to a large set of countries and identify the cases of coordination failure. They also find that there are two stable equilibria but the authors mostly characterize the interior equilibrium (with positive flows of high-skilled and low-skilled migrants). Our model is clearly related to that of De la Croix and Docquier (2012) since in both models anticipation of productivity and wages in the host country is at the heart of the migration decision. Our model is, however, different since imperfect information and statistical discrimination are the key driving forces of the equilibrium determination. Kreckemeier and Wrona (2011) also present a model of strategic complementarities but with two-way migration between similar countries. They mainly focus on welfare issues. Finally, Hendricks (2001) and Giannetti (2003) also have strategic complementarities in the migration decision so that the basic selection mechanism of high-skilled individuals into emigration is similar. They focus, however, on very different issues.

To summarize, the objective and contribution of our paper are mainly theoretical and are to extend the theory of migration with imperfect information. We are also able to match some empirical facts observed in the real-world. To be more precise, our theoretical contribution is twofold:

(1) To have a general theory of migration under imperfect information. It is true that the previous research cited above (such as Katz and Stark, 1984, 1986, 1987a,b; Kwok and Leland, 1982) develops the same initial idea: firms in the North which do not observe the skills of migrants and thus pool them together by statistically discriminate them. However, in all these models: (i) there is only a partial equilibrium analysis; (ii) there is not a complete characterization of all possible equilibria, i.e. it is not said under which conditions on the parameters one or two equilibria will emerge. These are keys for the analysis and for the empirical predictions; (iii) there is no signal and Bayesian updating since firms have no information and just pool migrant workers together. Here, firms observe an imperfect signal, the school record, and update in a Bayesian way their beliefs about migrants so that high-ability workers are never mistaken for low-skilled workers while low-ability ones may be.

(2) To have a theory of education and migration under uncertainty. This creates two types of coordination problems, i.e. about the ex ante (initial) proportion of high-skilled workers in the South and the ex post (after migration) proportion of high-skilled workers who migrate to the North. We believe that this is the first theoretical paper that proposes a general theory of education and migration under uncertainty.

3. The model

We consider an economy consisting of two countries, each populated by a unit mass of risk-neutral workers. One country (N, the “North”) has a technological advantage over the other (S, the “South”), reflected by the fact that the productivity is higher in firms located in N than in firms located in S.

Workers differ in productivity, and there are two types of workers, with high (H) and low (L) skills. Any worker can be assigned to two different tasks. In the simple task, all workers in a given country i have the same productivity, \( \lambda_i \), whereas in the more complex task, a high-skilled worker’s productivity is \( \lambda_{hi} \) while a low-skilled worker’s productivity is \( \lambda_{li} \), where \( \lambda_{hi} > \lambda_{li} \). Productivity is higher in the North in all tasks. In particular, \( \lambda_{hi} = \lambda_{hi}^N \) whereas, \( \lambda_{hi} > \lambda_{hi}^S \) and \( \lambda_{li} = \lambda_{li}^N \) whereas, \( \lambda_{li} > \lambda_{li}^S \). Note that the parameter \( \lambda \) captures absolute cross-country productivity differences, whereas \( \psi \) parameterizes differences in the extent of wage inequality. In particular, \( \psi < 1 \) \((\psi > 1)\) means that the skill premium is lower (higher) in the North than in the South. We will assume that \( \psi > \max(\frac{\lambda_{hi}}{\lambda_{li}}, 1) \). This ensures that (i) the skill premium is positive in the North and (ii) native skilled workers in the North earn a higher wage than skilled workers in the South, implying that it is possible to have migration of skilled workers in equilibrium.

The skill/ability of each migrant is not observed by Northern firms. The school record is, instead, observed but it is an imperfect signal of the skill/ability of the worker. Indeed, the signal (school record) is assumed, for simplicity, to perfectly reveal the type of native workers within each country. In country S (resp. country N), there are \( \beta_i \) (resp. \( \gamma_i \)) workers of type L and \( (1 - \beta_i) \) (resp. \( (1 - \gamma_i) \)) workers of type H. Firms practice statistical discrimination among migrant workers and workers are paid at their marginal product. Northern firms perform a noisy test such that high-skilled migrants always pass this test while only a fraction \( 1 - \sigma \) of low-skilled migrants pass it. Firms statistically discriminate high-skilled migrants by paying workers who passed the test at their average productivity. This productivity is lower than the high-skilled migrants real productivity since some low-skilled workers (those who have passed the test) are wrongly considered as high-skilled.

Let us justify the fact that firms practice statistical discrimination among migrants workers and that this discrimination in based on origin–country (rather than on race or ethnicity). There is an important literature in labor economics trying to test whether ethnic minorities are discriminated against or not. It is, however, very difficult to be sure if you are capturing discrimination or something else, unless you have some kind of experiments, like for example, sending matched CVs that vary in only one variable (for example the name) to employers in response to job advertisements (see for instance, Neumark, 1996; Bertrand and Mullainathan, 2004). It is even more difficult to test whether discrimination is taste based (Becker, 1957) or statistical (Arrow, 1973; Coate and Loury, 1993; Phelps, 1972). To our knowledge, there is no convincing empirical test capable to disentangle between these two aspects of discrimination. This is even truer about migrant workers since experiments are difficult to implement. However, we can justify the fact that firms practice statistical discrimination among migrant workers and that this discrimination is based on the country of origin (rather than race or ethnicity) by looking at outcomes. If skilled migrants are statistically discriminated against and this discrimination is based on the country of origin, then it has to be that their wage is pooled with other (less skilled) migrants coming from the same country. In other words, it has to be that the wages (or the type of jobs) of skilled workers coming from a country has to be on average lower (or worse) than the wages of high-skilled natives. Also, in Proposition 3 below, we show that the difference in wages between high-skilled native and high-skilled migrant workers is higher, the lower is \( \sigma \) the quality of information about migrants’ skills. We can, in fact, provide evidence on this. Indeed, the empirical study by Matttoo et al. (2008), who use the 2000 US Census, show that it is mainly skilled migrants from non-English speaking countries with poor-quality education systems who struggle to find skilled jobs. In other words, most skilled migrants do find a skilled job (see Table 1 in Gibson and McKenzie, 2011) unless they come from countries that are very different culturally and in terms of education system. Similarly, Özden (2006) finds that high-skilled migrants do not always use their

\footnote{For a very complete overview, see Altonji and Blank (1999), and more recently Lang and Lehmann (forthcoming), and Boeri and van Ours (forthcoming).}
skills after they immigrate. He also finds “striking” differences in the earnings of skilled immigrants from difference source countries, confirming the idea that discrimination is based on the country of origin and not race or ethnicity. Özden (2006) finds that cultural similarities (captured by a low \( \sigma \) in our model) between source and destination countries (like speaking the same language or having a similar educational system) improve skilled migrants’ earnings. In other words, if there was no statistical discrimination based on imperfect information about migrants’ skills or productivity, then we should observe that all high-skilled workers, migrants and natives, should be paid the same wage and should end up in the same high-skilled jobs. This is not what we observe in the real-world and this is why it is reasonable to assume that firms statistically discriminate high-skilled migrants based on their country of origin, especially those coming from countries that are culturally very different.

We assume that workers can migrate at a cost \( c \). This cost is individual-specific, and its c.d.f. is assumed to be i.i.d. across types. In particular, the density function is uniformly distributed over the interval \([0, c]\). The functional form is intended to obtain closed-form solutions. Note that a positive migration flow will always be observed as long as there are wage differentials across countries. Clearly, migration flows will only go from \( S \) to \( N \).

The timing of the model is the following. In the first stage, workers in the South decide whether to move and pay the migration costs. These costs are assumed to be sunk. In the second stage, firms offer wages to the immigrant and native workers who are in the country.

We proceed backward and first concentrate on the second stage of the model. We introduce the realistic assumption that information asymmetries are more severe for immigrants than for natives. More precisely, we assume that the school record of an immigrant is imperfectly observed (or imperfectly understood) by firms in the host country. In particular, we denote by \( \sigma \in [0, 1] \) the probability that Northern firms observe a negative school record for a low-skilled immigrant (for a low-skilled native, \( \sigma = 1 \)). The realization of what firms observe is unknown to the worker when he decides to migrate. We denote by \( \Sigma^+ \) the event “to migrate, go through a screening process in the North and not being found with a negative record”. For example, \( P(\Sigma^+ | J) \), for \( J \in \{H, L\} \), is the joint probability that a worker of type \( J \) (i) migrates and (ii) is screened and not found to have a negative school record. Similarly, \( P(\Sigma^- | J) \), is the joint probability that a worker of type \( J \) (i) migrates and (ii) is screened and found to have a negative school record. Clearly, \( P(\Sigma^- | H) = 0 \). Also, \( P(J|\Sigma^-) \) is the probability of having skill \( J \) given that this worker has migrated, went through a screening process in the North and not being found with a negative record. A similar interpretation can be made for \( P(J|\Sigma^+) \).

An immigrant whose negative school record is observed by a Northem firm will earn a wage \( w_L^M \) to be determined in equilibrium. Thus, low-skilled workers in the South contemplating moving to the North have the following expected wage:

\[
w_L = \sigma \lambda w_L + (1 - \sigma) w_L^M
\]  (1)

High-skilled immigrants earn a safe wage \( w_H^M \). This is the same wage that is earned by low-skilled immigrants whose negative school record has passed undetected by the employing firm in the North. Thus, in equilibrium, firms will pay immigrants with a “clean record” a wage equal to

\[
w_H = P(\{H|\Sigma^+\}) \psi \lambda w_H + P(\{L|\Sigma^+\}) \lambda w_L
\]  (2)

In other words, Northern firms facing a group of workers whose individual productivity is unknown (they have all passed the test) offer the same wage to all. This wage is equal to the average productivity of the group. Observe that the implied productivity of a low skilled worker in a complex task in the host country is \( \lambda w_L \), the same as the productivity of a low skilled worker in the simple task. It is because low-skilled workers cannot adapt to new technologies and thus have the same productivity whatever the task they are assigned to. Observe that this assumption has no impact on the results as long as the productivity of low-skilled workers is lower than that of the high-skilled in the complex task.

4. The migration game

We now come back to the first stage of the model where each type of workers decides to migrate or not. For simplicity, we assume \( \sigma \) to be sufficiently large to ensure that (given the other parameters of the model) neither all low-skilled nor all high-skilled workers migrate from the South to the North.

This first stage is a strategic form game played by a continuum of players: the workers in country \( S \). There are two types of players, the low-skilled and the high-skilled workers and each player is characterized by its type and its individual-specific cost of migration \( c \). Each player has two strategies: to stay in country \( S \), the South, or to migrate to country \( N \), the North. A high-skilled worker with migration cost \( c \) obtains \( w_H^S \) if he stays in the South and \( w_H^N - c \) if he migrates. A low-skilled worker with migration cost \( c \) obtains \( w_L^S \) if he stays in the South and \( (1 - \sigma) w_L^N + \alpha \lambda w_L - c \) if he migrates. Strategic interactions arise because the wage \( w_L^N \) (positively) depends on the proportion of high-skilled workers among the migrants, i.e. \( w_L^S \) depends on the strategy profile of all workers in country \( S \).

Suppose we fix the strategies of all low-skilled workers. The reduced game played by high-skilled workers exhibits strategic complementarities because \( w_L^N \) is increasing with the number of high-skilled workers who migrate, as can be seen from Eq. (2), and so is the payoff of a high-skilled worker contemplating migration. Therefore, the more high-skilled workers choose to migrate, the higher the incentives of other high-skilled workers to migrate. This remark suggests that multiple equilibria are likely to emerge in the migration game. We investigate this issue below and determine the pure strategy Nash equilibria of the first stage game.

In this game, all the strategic interactions go through the wage \( w_L^N \). A high-skilled worker prefers to stay in the South if and only if \( w_L^S \geq w_L^N - c \). This relation defines a threshold cost \( c^+ = \max \{w_H^N - w_L^S, 0\} \) such that workers with a cost larger (smaller) than \( c^+ \) stay (migrate). Given the uniform distribution, \( c^+ / \sigma \) is the proportion of high-skilled workers who migrate. Similarly, a low-skilled worker prefers to stay in the South if and only if \( w_L^S \geq w_L^N - c \) where \( w_L^S \) is defined by Eq. (1). This inequality defines a threshold \( c = \max \{(1 - \sigma) (w_L^N - w_L^S) + \alpha \lambda w_L (\lambda - 1), 0\} \), implying that \( c / \sigma \) is the proportion of low-skilled workers who move to the North.

For technical convenience, we assume that \( \psi \lambda w_H - w_L < c \). This implies that even in the limit case where immigrants suffer no discrimination some workers of both types stay in the South.

We can now establish the following Lemma.

**Lemma 1.**

(i) The probability that an immigrant of a high- and low-skilled type for whom Northern firms did not detect a negative test score is respectively given by:

\[
P(H|\Sigma^+) = \frac{c^+ (1 - \beta)}{c^+ (1 - \beta) + c^- (1 - \sigma) \beta}
\]

\[
P(L|\Sigma^-) = \frac{c^- (1 - \sigma) \beta}{c^- (1 - \sigma) \beta + c^+ (1 - \beta)}
\]

where \( c^+ = \max \{w_H^N - w_H^S, 0\} \) and \( c^- = \max \{(1 - \sigma) (w_L^N - w_L^S) + \alpha \lambda w_L (\lambda - 1), 0\} \).  

\(^5\) See Topkis (1998), Milgrom and Roberts (1990), and Vives (1990) for an overview of the literature on games with strategic complementarities.
(ii) In equilibrium, the wage earned by an immigrant who passed successfully the screening test in the North satisfies

$$w_H^* = \frac{c'(1-\beta)\psi\lambda w_H + c'(1-\sigma)\beta w_H}{c'(1-\beta) + c'(1-\sigma)\beta} \equiv \eta^*(w_H^*)$$

where \(\eta^*(w_H^*)\) is defined in the range \(w_H^* \in [w_L, \psi\lambda w_H]\) and has the following properties

\[
\eta^*(w_H^*) = \lambda w_H \quad \text{if} \quad w_H^* \in [w_L, w_M]
\]

\[
\eta^*(w_H^*) > 0, \eta^*(w_H^*) \leq 0 \quad \text{if} \quad w_H^* \in [w_M, \psi\lambda w_H]
\]

**Proof.** See Appendix A.

As stated above, the wage (Eq. (3)) corresponds to the average productivity of workers who have passed the test and thus illustrates the statistical discrimination policy implemented by Northern firms against high-skilled immigrants. Anticipations of workers concerning the value of \(w_H^*\) correspond to anticipations concerning the decisions of other high- and low-skilled workers in the migration game and therefore correspond to anticipations concerning the expected productivity of migrants (i.e. the function \(\eta^*(w_H^*)\)). For instance if a worker anticipates \(\eta^*(w_H^*) \in [w_L, w_M]\), then he knows that only low-skilled workers will migrate since the high-skilled workers are better off staying at home (thus \(c' = 0\) and \(c'' > 0\)). Firms will therefore pay them \(\lambda w_L\) like the local low-skilled workers. If this worker anticipates a higher wage, \(\eta^*(w_H^*) \in [w_M, \psi\lambda w_H]\), then he knows that high-skilled workers start to migrate (\(c'\) and \(c''\) are now both strictly positive). As a result, there is a positive monotonic relationship between the \(\eta^*(w_H^*)\) and the expected productivity of migrants. The concavity stems from the fact that when wages increase, both high- and low-skill workers are induced to migrate so that at the margin the expected productivity increases less and less.

In equilibrium, anticipations are rational and \(\eta^*(w_H^*) = w_H^*\), i.e. the anticipated productivity of migrants is equal to the true productivity of migrants. We define an equilibrium such that all immigrants are offered low-skilled tasks, irrespective of the outcome of the screening test as a *high-discrimination* equilibrium. In this equilibrium, \(w_H^* = \lambda w_L\). We define an equilibrium such that all immigrants who pass the screening test are offered high-skilled tasks, whereas all immigrants who fail the screening test are offered low-skilled tasks as a *low-discrimination* equilibrium. In the latter equilibrium, the wage is determined by the solution of the equation \(w_H^* = \eta(w_H^*)\) (see Lemma 1) in the range \(w_H^* \in [w_L, \psi\lambda w_H]\). Finally, we define an equilibrium such that \(w_H^* = \psi\lambda w_H\) as a *no-discrimination* equilibrium. This latter case will never happen in equilibrium, unless \(\sigma\) is equal to 1.

**Proposition 1.** Let

\[
\phi(\lambda) = \frac{w_H(1 + \beta) - w_L(1 + \lambda) + 2\sqrt{\beta(w_H - w_L)(w_H - \lambda w_L)}}{\lambda(1-\beta)w_H^*},
\]

where \(\psi^*(\lambda) < 0\). Then,

1. If \(\lambda > w_M/w_L\), then, for all \(\sigma \in (0, 1)\), there exists a unique stable low-discrimination equilibrium.
2. If \(\lambda < w_M/w_L\) and \(\psi > \phi^*(\lambda)\), then there exist a stable high-discrimination equilibrium and an unstable low-discrimination equilibrium.
3. If \(\lambda > w_M/w_L\) and \(\psi < \phi^*(\lambda)\), then, for all \(\sigma \in (0, 1)\), there exists a stable high-discrimination equilibrium and two low-discrimination equilibria, one stable and one unstable.

**Proof.** See Appendix A.

The existence of multiple equilibria illustrates the coordination problem among migrants. If they anticipate that no high-skilled worker migrates, then they anticipate that all migrants will be paid \(\lambda w_L\) when \(\lambda w_L \leq w_H\), those anticipations are self-confirming, i.e. they turn out to be correct. If they anticipate that some high-skilled workers migrate, then they anticipate that migrants who successfully pass the screening test will be paid more than \(\lambda w_L\). Depending on the value of the parameters of the model, several levels of high-skilled workers migration may be self-confirming. Figs. 1, 2 and 3 provide an illustration of Proposition 1.

For each graph, we report the limit cases of \(\sigma = 0\), including the limit case of \(\sigma = 0\), where the graph of the function \(\eta^*(w_H^*)\) is stepwise linear, with \(\eta^*(w_H^*) = \lambda w_H\) for \(w_H^* \leq w_L\) and \(\eta^*(w_H^*) = \lambda w_H\) for \(w_H^* \geq w_H\). The function \(\eta^*(w_H^*)\) is strictly decreasing in \(\sigma\), for any \(w_H^* > w_L\). For any \(\sigma > 1\), including the limit case of \(\sigma = 0\), the function \(\eta^*(w_H^*)\) is strictly concave and its graph is smooth. Furthermore, \(\eta^*(w_H^*) < \lambda w_H\) for any \(w_H^* \in [0, \lambda w_H]\).

![Fig. 1. The case of large productivity differences between N and S (large \(\lambda\).](image-url)
Fig. 1 describes case 1 in the Proposition, i.e., $\lambda > w_H / w_L$. In this case, due to the large cross-country productivity difference, some high-skilled workers would be prepared to migrate even if they were offered the low-skilled task in the North. But this implies that high-discrimination is not sustainable in equilibrium. If workers passing the screening test were offered the low-skilled wage $\lambda w_L$, their average productivity would exceed $\lambda w_L$, ruling out the existence of a high-discrimination equilibrium. Therefore, and due to the concavity of the function $\eta$, there exists a unique low-discrimination equilibrium described by point $L$ in Fig. 1.

Fig. 2 describes case 2 in the Proposition, i.e., $\lambda b w_H / w_L$ and $\psi b \phi(\lambda)$. In this case, the nature of the set of equilibria depends on the extent of the informational asymmetry. When $\sigma$ is low, $(\sigma < \sigma(\lambda, \psi))$, i.e. large informational asymmetry, the only equilibrium features high discrimination (point $H'$ in Fig. 2). This is a standard case of “market for lemons” (Akerlof, 1970). The informational asymmetry drives out of the market high-skilled immigrants. No equilibrium in which immigrants are offered high-skilled tasks is sustainable for low values of $\sigma$, and only the high-discrimination equilibrium, described by point $H'$ in Fig. 2, is sustained. If firms in the North were to offer higher salaries to immigrants passing the screening test, they would be swamped by a large proportion of low-skilled immigrants that are undistinguishable from the high-skilled ones. Relatively low productivity differences together with a low skill premium in the North cause the migration flow to be dominated by the incentive for low-skilled workers to migrate in the hope of being pooled with the high-skilled ones. While the average skill of immigrants increases in response to higher salaries, the increase is not steep enough to sustain a low-discrimination equilibrium. This is due to the joint effect of a low $\lambda$, a low $\psi$ and a low $\sigma$. The figure also shows the knife-edge case ($\sigma = \sigma(\lambda)$) where the graph of the function $\eta(w_M)$ is tangent to the 45° line and there are two equilibria (described by points $H'$ and $K'$ in Fig. 2). Whenever $\sigma > \sigma(\lambda, \psi)$, there exist three equilibria (described by points $H'$, $M'$ and $L'$ in Fig. 2), two of them ($M'$ and $L'$) featuring low discrimination. These emerge because informational imperfections are now less severe. In this case, there are multiple self-fulfilling beliefs. In the low-discrimination equilibrium ($L'$), workers in the South expect that, in the second-stage, firms in the North will offer high wages to those passing the test. It is then optimal for high-skilled workers with low mobility costs to migrate. Firms offering high-skilled tasks to immigrant at the equilibrium wage will,
on average, be satisfied with the immigrants’ performance. In the high-discrimination equilibrium (\(H^*\)), instead, high-skilled workers do not move since they expect low wages.

Of the two low-discrimination equilibria (\(M^*\) and \(L^*\) in Fig. 2), only the one with the highest wage (\(L^*\)) is stable to wage perturbations. In particular, consider an equilibrium like \(M^*\) in Fig. 2. If migrant workers anticipate a slightly higher (lower) wage than the one corresponding to \(M^*\), the average productivity of the pool of immigrants who passed the test would increase (decrease) by more than the initial wage increase (decrease). Thus, firms would be induced to offer an even higher (lower) wage, and so on, until the equilibrium \(L^* (H^*)\) is reached.

Fig. 3 describes case 3 in the Proposition, i.e., \(\lambda < w_M/w_L\) and \(\psi > \phi (\lambda)\). In this case, there are multiple equilibria irrespective of \(\alpha\). As before, the low-discrimination equilibrium corresponding to \(M^*\) in Fig. 3 is unstable. Intuitively, the skill premium \(\psi\) in the North is large enough to guarantee that the average quality of the pool of immigrants responds sufficiently to increases in the wage to sustain an equilibrium with low discrimination.

In the migration game, decisions to migrate can be strategic complements or substitutes depending on the types of the migrants. Everything else being equal, the more high-skilled workers migrate, the more it pays for both types of workers to migrate. And the more low-skilled workers migrate the less it pays for both types of workers to migrate. However, the migration game is essentially a game with strategic complementarities as we explain now. Suppose we modify slightly the timing of our game and suppose that workers in the South take their migration decision sequentially: first, high-skilled workers decide to migrate or not and second, after observing the number of high-skilled migrants, low-skilled workers take their migration decision. We shall argue that this game is strategically equivalent to our migration game for high-skilled workers and that, after applying backward induction, the first-stage exhibits strategic complementarities.\(^6\)

**Proposition 2.** Suppose high-skilled workers in the South simultaneously decide to migrate or not before low-skilled workers decide simultaneously to migrate or not and that low-skilled workers observe the high-skilled workers decisions,

1. For any fixed number of high-skilled migrants, the subgame played by low-skilled migrants has only one equilibrium,
2. When they correctly anticipate the equilibrium behavior of low-skilled workers, high-skilled migrants play a first-stage game that exhibits strategic complementarities,
3. The subgame perfect equilibria of this sequential migration game coincide with the equilibria of the simultaneous migration game.

**Proof.** See Appendix A.

**Proposition 2** highlights the fact that strategic complementarities in the migration decisions of high-skilled migrants are crucial for the analysis. These complementarities explain the multiplicity of equilibria and suggest that equilibria can be Pareto ranked. In fact, the higher the equilibrium wage \(w_M^*\), the higher the welfare of all workers in the South. It should be clear that coordination problems are at the heart of the existence of multiple equilibria and the fact that high-skilled migrants cannot coordinate with low-skilled migrants is very costly for them.

Stark (1995a, Chap. 4) has proposed a way to coordinate migrants in a Pareto-optimal way. If high-skilled migrants could form a cohesive group and act jointly, they should be willing to make a transfer to the low-skilled migrants to induce them not to migrate. He argues that it is a win–win situation since high-skilled migrants would earn as much as the natives, i.e., \(\lambda w_L\) (in our model) minus the transfer, and pay a transfer that exactly compensates the difference between the net-wage abroad, \(w_M - c\), and the wage at home, \(w_L\). If such a transfer exists, then this could be implemented and would be clearly Pareto optimal. In our model, it would be difficult to implement it for two reasons. First, how can all high-skilled migrants (i.e., all high-skilled migrants who have a migration cost lower than \(c^+\)) in a given country coordinate themselves and form a big cohesive group in order to pay transfers to all possible low-skilled migrants? Also, other problems, such as free riding, will arise. Second, because we have a distribution of migration costs, to pay an individual transfer to each worker, the migration cost of each low-skilled migrant must be perfectly observed so that all workers will a cost higher than \(c^-\) should not be given any transfer will the others should get a transfer exactly equals to \(w_M - c\), which is individual specific.

Docquier and Rapoport (1998) have also criticized the Stark’s proposal by highlighting the importance of different communities within a country. In other words, even though Northern firms may pool all migrants from the same country together and give then the same wage, Docquier and Rapoport argue that high-skilled workers from one community in the Southern country cannot make a strategic transfer to low-skilled workers from another community in the same country. For example, high-skilled Kurds from Turkey migrating to Germany will certainly not give transfers to low-skilled non-Kurds from Turkey to prevent them to move to Germany. However, German employers, who cannot distinguish between Kurd and non-Kurd Turks will pool them together and give them the same wage based on their average productivity. The fact that firms’ pooling may not exactly overlap with real migrants’ groups is an important issue but cannot be addressed here since we do not differentiate between different communities in the South.

To complete the study, we establish some comparative statics results. There are basically two types of equilibria. The high-discrimination equilibrium where all migrants are paid \(\lambda w_L\) is not very interesting and the comparative statics results are straightforward. As a result, we focus on the stable low-discrimination equilibrium and restrict attention to parameter values for which it exists.

**Proposition 3.** When it exists, the stable low-discrimination equilibrium wage \(w_M^*\) offered to migrants who successfully pass the test satisfies

\[
\frac{\partial w_M^*}{\partial \sigma} \geq 0, \quad \frac{\partial w_M^*}{\partial \psi} \geq 0, \quad \frac{\partial w_M^*}{\partial \beta} \leq 0.
\]

When \((1 - \sigma) \lambda > 1\), it also satisfies

\[
\frac{\partial w_M^*}{\partial \lambda} \geq 0.
\]

**Proof.** See Appendix A.

These results are quite intuitive. When the test becomes better, i.e. \(\sigma\) increases, the pool of migrants is of better quality and therefore their expected productivity increases. As a result, \(w_M^*\) also increases. The intuition is similar for the skill premium \(\psi\) and \(\beta\) the percentage of low-skilled workers in the South. Indeed, when \(\psi\) increases, more high-skilled workers migrate and thus \(w_M^*\) increases. When \(\beta\) increases, the quality of the migrants decreases and so does \(w_M^*\). Interestingly, the effect of \(\lambda\), the absolute productivity difference between the two countries, on \(w_M^*\) is ambiguous. Indeed, when \(\lambda\) increases, both skilled and unskilled workers are attracted to the North and thus the average quality of migrants can increase or decrease, depending which force dominates the other. If, however, \((1 - \sigma) \lambda > 1\), meaning that \(\lambda\) has to be quite high (since \(1 - \sigma < 1\)), then the net effect is positive.

All the comparative statics results are clearly specific to the imperfect information setting since we look at the effect of the parameters on the wage \(w_M^*\) of high-skilled and seemingly high-skilled migrants. In a world of perfect information and migration, since low- and high-skilled
workers will not be pooled together, we would not obtain the same result. In particular, the ambiguous impact of \( \lambda \) on \( w_{rms}^m \) cannot be obtained in a model with perfect information (and not statistical discrimination) since the wages of low and high-skilled migrants will just be different.

It is interesting to observe that, in our simple framework where high- and low-skilled workers are perfect substitutes, Northern firms statistically discriminate the high-skilled immigrants by offering them a lower wage than the local high-skilled workers \( \{w_{rms}^m \leq \phi \lambda w_H\} \). Proposition 3 shows that the difference in wages between high-skilled native and high-skilled migrant workers is higher, the lower is \( \sigma \), the quality of information about migrants’ skills. This result could be interpreted as the fact that high-skilled migrants have difficulties to find a good job that fits their skills because of the imperfect information and signal about their skills. This is in line with the empirical study of Mattoo et al. (2008) who use the 2000 US Census to show that it is mainly skilled migrants from non-English speaking countries with poor-quality education systems who struggle to find skilled jobs. In other words, most skilled migrants do find a skilled job (see Table 1 in Gibson and McKenzie, 2011) unless they come from countries that are very different culturally and in terms of education system (which is captured by \( \sigma \) in our model).

An another consequence of the model is that, in any low-discrimination equilibrium, low-skilled migrants earn on average a higher wage than their local counterparts since \( w_{rms} \geq \lambda w_H \). This is because they pool, with some probability, with the high-skilled workers from the same country. This is not what is usually observed in the real-world. We can, however, reinterpret our model so that this feature will not be present anymore. Consider a continuum of workers from the South who have two characteristics: education (or school record) and ability. There are two levels of education, high and low, and two levels of ability, high and low. Education (or school record) is a perfect signal of education in the home country (both in the South and in the North). In the North (but also in the South), because education is a perfect signal of ability, low-educated native workers always have low-ability whereas high-educated native workers have a high ability. However, this is not true for migrants. Low-educated workers always have low-ability whereas high-educated workers can have either a high or a low ability. In other words, a high-education diploma from the South has some value in the North but it is a noisy signal in terms of ability. Note that this does not mean that there are absolute different levels of abilities for migrants whether they are in the South or in the North. This means that when a high-educated Southern individual lives in the South, her education level corresponds exactly to her ability because her diploma is adapted to the South. When this high-educated Southern individual migrates to the North, her education level is not always well adapted to the job requirement in the North. Consider, for example, an engineer from Iraq. In her country, her diploma reveals her ability because her skills (education) are adapted to local jobs. This is not always true in the North, say the US. It may be that this Iraqi engineer has some unobserved ability that helps her perform well on the American job (high ability) or, on the contrary, she has difficulty to adapt to the job requirement in the US and will perform poorly (low ability). This is clearly unobserved by Northern firms. So, for highly-educated migrants, what matters in the North is their ability not their school record. As a result, when a migrant migrates to the North with a low educational level, she is always assigned to the low-skilled job since the firm knows with certainty that she has a low ability and this worker obtains a wage of \( \lambda w_H \). When a migrant has a high education level, she goes through a test. If the migrant has a high ability, then she always passed this test. If the migrant has a low ability, she will be detected with probability \( \sigma \). Observe that, when a high-educated individual from the South wants to migrate to the North, she knows her own ability (high or low) in the North. There are still strategic complementarities but now it is between high-educated low and high ability workers. Observe also that, when considering whether to migrate or not, low-educated workers know with certainty that their wage in the North will be \( \lambda w_H \), high-educated low-ability workers know that their expected wage is \( w_{rms}^m \) while high-educated high-ability workers consider \( w_{rms}^m \) in their migration decision. Assume that the fraction of low-ability educated workers from the South is \( \beta \) and that the fraction of low-educated workers in the South is \( \theta \). The equilibrium is calculated as before but there are now three categories of migrant workers: low-educated, high-educated with low ability and high-educated with high ability migrants. The results are clearly the same since there is no interaction between the low and the high-educated workers. In terms of characterization of equilibria, there will be, however, one major difference: low-educated migrants will now be paid exactly as the low-educated natives, that is \( \lambda w_H \). On the other hand, as before, high-educated migrants (whether they have a low or a high ability) will always earn less than the native high-educated workers since \( w_{rms}^m < \lambda \psi w_H \).

We are now able to discuss the issue of brain drain in our model. Before migration, there were \( 1 - \beta \) high-skilled workers in the South. After migration, \( (1 - \beta)(1 - \psi) \) high-skilled workers stay in the South while \( (1 - \beta)(\psi) \) high-skilled workers migrate to the North. As a result, there is positive selection if, after migration, the number (or fraction) of high-skilled workers from the South is higher in the North than in the South, i.e. \( (1 - \beta)(\psi) > (1 - \beta)(1 - \psi) \). This is equivalent to \( c^+ > \psi/c \). This also implies that (i) high-skilled workers are more migratory than the low-skilled and (ii) the after-migration proportion of high-skilled in the South is lower than the before-migration proportion. There is brain drain as soon as some high-skilled workers migrate, and there is positive selection if the brain drain exceeds low-skilled emigration. Since \( c^+ = w_{H,i} - w_{L,i} \), using Proposition 3, this means that the higher is \( \sigma \) (meaning that the more similar are the two countries), or the higher is the skill premium \( \psi \), or the higher is \( \lambda \), the absolute productivity difference between the two countries (assuming \( (1 - \sigma)\lambda > 1 \)), or the lower is the initial percentage of low-skilled workers in the South, \( \beta \), the more likely that there will be a brain drain. This is because all these parameters positively influence the wage \( w_{rms}^m \) of high-skilled migrants and thus their migration decision. If, for example, the North and the South are very similar countries (in terms of language or educational system) so that \( \sigma \) is close to 1, then Northern firms will more easily evaluate the skills of migrants. Thus, high-skilled migrants will be less likely to be pooled with low-skilled migrants and more likely receive a wage similar to high-skilled native workers. This will trigger more high-skilled migration to the North.

Let us focus on the determination of the skill composition of the immigrants. In high-discrimination equilibria, only low-skilled workers migrate. We focus therefore on cases where a low-discrimination equilibrium exists and is stable, and study the selection effects. Given the uniform distribution, the migration flow consists of \( (1 - \beta)(w_{rms}^m - w_H)/\psi \) high-skilled workers and \( \beta(1 - \sigma)(w_{rms}^m - w_L) + \sigma(\lambda - 1)/\psi \) low-skilled workers.

**Proposition 4.** The proportion of high-to-low skill immigrants in the stable low-discrimination equilibrium is equal to:

\[
\frac{(1 - \beta) c^+}{\beta} = \frac{\Psi}{\psi} \left( \frac{w_{rms}^m - w_H}{C_0/C_1} \right) \left( \frac{w_{H,i} - w_{L,i}}{w_{H,i} - w_{L,i} \lambda(\psi - 1)/\psi} \right)
\]

This proportion satisfies:

\[
\frac{\partial H}{\partial \psi} \geq 0, \quad \frac{\partial H}{\partial \psi} \geq 0, \quad \frac{\partial \psi}{\partial \beta} \leq 0.
\]

When \( (1 - \sigma)\lambda > 1 \), it also satisfies

\[
\frac{\partial H}{\partial \lambda} \geq 0.
\]

\footnote{For very complete overviews on brain drain and development, see Gibson and McKenzie (2011) and Docquier and Rapoport (2012).}
Given the comparative statics of the equilibrium wage \( w_M \) (see Proposition 3) in a stable low-discrimination equilibrium and since \( \left( \frac{H}{T} \right) \) is increasing in \( w_M \), it is straightforward to show that the equilibrium proportion of high-to-low skill immigrants is increasing in \( \sigma, \psi \) and decreasing in \( \beta \). Thus, the model predicts that the pool of immigrants from countries that have a relatively low skill premium compared to recipient countries (high \( \psi \)) or are characterized by lower informational barrier (high \( \sigma \)) or a large proportion of skilled workers (low \( \beta \)) will be more skilled. The condition \( (1 - \alpha) \lambda > 1 \) is likely to be verified when either the informational asymmetries are important \( (\sigma \text{ small}) \) and/or the productivity gap between the two countries is big \( (\lambda > 1) \). When it is verified, the equilibrium proportion of high-to-low skill immigrants is increasing in \( \lambda \); it is larger when the country of origin is poorer.

More generally, in our model, we find that there is a positive association between skilled and unskilled migrant workers. In their survey article, Gibson and McKenzie (2011) identify eight questions about brain drain. One the key question is (see page 114): “Is there a positive association between skilled and unskilled migration?”. They show (see their Fig. 1 and the empirical evidence cited in their paper) that levels of skilled and unskilled migration actually have a strong positive association. They find that countries that sent relatively many high-skill migrants to one country also sent relatively many low-skill migrants to the same country. Similarly, when a country increases the number of high-skill migrants it sends to a recipient country, the number of low-skill migrants to the same country also increases. The underlined economic mechanism is, however, not clear. They argue that “both skilled and unskilled migrants are likely to determine their migration decisions in part on the basis of the institutional characteristics of their home country and the presence of common links with potential destination countries”. Our theoretical framework can provide some intuition of this empirical result. In our model, there is a positive association between skilled and unskilled migrants because the decision whether to migrate or not depends on the proportion of high-skilled workers among the migrants, meaning that there are strategic complementarities in the migration decision game. In other words, the higher is the number of high-skilled migrants, the higher is the number of low-skilled migrants because the latter benefits from the former by increasing their wages in the host country since the average productivity of (seemingly) high-skilled migrants increases with the proportion of high-skilled migrants.

In the specific model of Borjas (1987), based on the Roy model, a fall in the income in the United States or an increase in migration costs (here a fall in \( \lambda \)), implies that fewer workers migrate. However, it does not change the skill composition of the workers that migrate. It is clear that, in our model, a change in \( \lambda \) affects both the skill composition of the migrants and the number of migrants. In fact, when \( \lambda \) is very high (large income differences between \( N \) and \( S \)), most of the migrants will be highly skilled (Fig. 1). Indeed, in this case, there is only a low-discrimination equilibrium in which Northern firms are prepared to employ immigrants in high-skilled jobs, although paying them less than natives for an identical job. In particular, all workers passing the screening process are assigned to high-skilled jobs. Their wage is determined according to statistical discrimination. High-skilled workers, in turn, anticipate that good job opportunities exist in the North, and a share of them decides to migrate. The pool of immigrants is, in this equilibrium, superior in terms of average skill and there is more migration. When \( \lambda \) decreases and reaches intermediate values (Fig. 2), more people migrate to \( N \), the rich country, since both high- and low-skilled workers migrate. Finally, when \( \lambda \) is further reduced and reach small values (Fig. 3), depending of the value of \( \sigma \) the skill composition and the number of migrants will be affected.

Let us investigate the issue of self-selection of migrants.

**Proposition 5.** There is positive (negative) self-selection, if and only if \( c^+ > c^- (c^+ < c^-) \), i.e.,

\[
w_M > (\frac{(1-\beta)w_H - w_M (1 - \sigma \lambda)}{1-\beta-\beta(1-\sigma)})
\]

(4)

The proof of this proposition is straightforward. Indeed, in equilibrium, there are \( (1 - \beta)c^+ \) high-skilled migrants out of the \( (1 - \beta)c^- + \beta c^- \) total number of migrants. As a result, the proportion of high-skilled migrants is the ratio of these two quantities. Similarly, the proportion of high-skilled non migrants is the ratio between the number of non-migrants that stay in the home country, i.e., \( (1 - \beta)(c^- - c^+) \), and the total individuals that do not migrate, which is: \( (1 - \beta)(c^- - c^+) + \beta(c^- - c^-) \). Thus, there is positive self-selection if and only if:

\[
(1-\beta)c^- + \beta c^- > (1-\beta)(c^- - c^+) + \beta(c^- - c^-)
\]

(5)

Solving this equation leads to \( c^+ > c^- \), which, using the values of \( c^+ \) and \( c^- \) gives condition (4).

Observe that condition (4) compares an endogenous variable \( w_M \) on the left hand side (LHS) with exogenous variables on the right hand side (RHS). As such, the intuition of this condition is difficult to see through at first sight. We know, however, how \( w_M \) varies with the different parameters (see Proposition 4) and we can calculate how the RHS of Eq. (4) varies with the parameters. So, for example, if \( (1 - \sigma) \lambda > 1 \), then both the LHS and the RHS of Eq. (4) increase and the result is indeterminate. On the other hand, it can be checked that the RHS of Eq. (4) is decreasing with \( \psi \) while the LHS (\( w_M \)) is increasing with \( \sigma \) (Proposition 4). As a result, the higher is \( \sigma \), the more likely condition (4) holds and the more likely there is positive self-selection, i.e. \( c^+ > c^- \). For \( \psi \), the skill premium, it is easier since the RHS of Eq. (4) does not depend on \( \psi \). As a result, the higher is \( \psi \), the more likely there is positive self-selection.

5. **Endogenous skills**

So far, we have assumed that the fraction of low-skilled and high-skilled workers in the South, \( \beta \) and \( 1 - \beta \), were exogenously given. In this section, we endogenize this decision. The timing of the model is now as follows. In the first stage, without knowing their migration costs, workers in the South decide whether to be educated (high-skill) or not (low-skill). In the second stage, migration costs are revealed to workers and given their education level, workers in the South decide whether to move to the North and pay the migration costs or to stay in the South. In the third stage, firms offer wages to the immigrant and native workers who are in the North. Compared to the previous model, there are now two types of coordination problems. As before, coordination problems arise in the migration decision since workers do not know how many high-skilled Southern workers will migrate to the North. On top of this, coordination problems arise in the education decision since Southern workers do not know how many workers in the South will be educated. As a result, when deciding to be educated or not, they have to anticipate both how many workers will be educated in the South and how many high-skilled and low-skilled workers will migrate to the North.

As before, let us solve the model backward. Stages 3 and 2 have already been solved in the previous sections for fixed \( \beta \) and \( 1 - \beta \). Let us now solve the first stage, that is workers’ education decision. Assume that the cost of education is denoted by \( s \) and is uniformly distributed on the support \( [0, S] \). Assume also that workers do not know their migration cost \( c \) when deciding upon education. This is a reasonable assumption since individuals make education decisions when they are young and migration decisions when they are usually older. However, the migration cost, which reflects age, family status, etc., is usually not known when people decide to be educated. In
this context, the expected utility of being educated (high-skill) for a person with an education cost \( s \) is given by:

\[
EU_{H}(s) = \int_{0}^{c} \frac{1}{c} (W_{M} - c) \, dc + \int_{c}^{1} \frac{1}{c} W_{H} \, dc - s
\]  

(6)

Indeed, if this person has a migration cost lower than \( c^{+} \), she will migrate to the North, obtain a wage of \( W_{M} \) and occurs a sunk cost of \( c \) while, if it is above \( c^{+} \), she will not migrate and obtain the Southern wage equals to \( W_{L} \). In both case, the worker pays the education cost \( s \). Similarly, the expected utility of not being educated (low-skill) is equal to:

\[
EU_{L}(s) = \int_{0}^{c} \frac{1}{c} (W_{M} - c) \, dc + \int_{c}^{1} \frac{1}{c} W_{L} \, dc
\]  

(7)

The interpretation is similar to Eq. (6) by noting that the migration threshold value for low-skilled workers is \( c^{-} \) and not \( c^{+} \) and that they don’t pay the education cost \( s \).

We can therefore determine a threshold value of \( s \), denoted by \( \overline{s} \), such that \( EU_{H}(\overline{s}) = EU_{L} \). We obtain:

\[
\overline{s} = c^{+} W_{H}^{-} - c^{+} W_{M}^{-} + \frac{c^{-2} - c^{+2}}{2c} + \frac{c^{+} - c^{-}}{c} + W_{H} - W_{L}
\]  

(8)

where \( c^{\pm} = \max(W_{M} - W_{L}, 0) \), \( c^{-} = \max(\alpha W_{L} (1 - \alpha) + (1 - \alpha)(W_{M} - W_{L}), 0) \), \( W_{M}^{-} = \alpha W_{L} + (1 - \alpha) W_{M} \), and \( W_{L}^{-} \) is defined by Eq. (3). In this case, all individuals with a \( s \) lower than \( \overline{s} \) will be educated while, the others, will not. Observe that the fraction of low-skilled workers in the South is now given by \( \beta = 1 - \overline{s} / \overline{s} \) whereas the fraction of high-skilled workers in the South is given by \( I = \delta = \overline{s} / \overline{s} \).

We can analyze here an interesting idea developed in the literature about brain gain: the decision of individuals to invest in education reacts to the prospect of future migration, and that not all those who choose to increase their education because of the chance they may actually end up migrating. Gibson and McKenzie (2011) surveyed different empirical papers that, indeed, find that it is the case in several developing countries. These empirical studies find some plausible source of exogenous variation in migration opportunities or returns to education abroad to identify a “causal effect”. See, for example, Shrestha (2010) who uses a change in the educational requirements of Gurkha British Army recruits in 1993, which increased the returns to education abroad for Gurkha men in Nepal, or Docquier et al. (2010) who show that after the 1995 Bosman ruling, a ruling by the European Court of Justice that increased the opportunities for African soccer players to play in Europe, the skill level of African soccer leagues most likely to supply players to Europe increased.

In our model, all individuals from the South who have an education cost \( s < \overline{s} \) and a migration cost \( c > c^{+} \) will end up investing in education without migrating to the North. It is because these individuals do not know their migration costs when deciding to invest in education and anticipate that they will eventually migrate to the North and obtain a very high return from education, i.e. a wage \( W_{M} \), which is much higher than the low-skilled wage in the home country, \( W_{L} \), and in the North, \( W_{M} \). Our mechanism is quite different to that of Katz and Rapoport (2005) where migration is constrained by policy barriers, so that not all workers who would like to migrate are able to do so. In that case, a brain gain effect can occur even with open borders if people acquire education for the option value of migrating and then do not all exercise this option.

As we have seen in the previous sections, two types of equilibria emerge: a low-discrimination and a high-discrimination equilibrium. Let us first focus on the high-discrimination equilibrium.

5.1. High-discrimination equilibrium

In this case, all workers in the South, when deciding whether to educate themselves or not, rationally anticipate that the firms in the North will treat them equally (high and low-skilled workers) and give them a wage \( w_{H} = w_{M} = \lambda W_{H} \) independent of their skills. In this equilibrium, no high-skilled worker from the South will migrate to the North so that \( c^{+} = 0 \) and thus \( c^{-} = W_{L} (\lambda - 1) \). The condition that guarantees that no skilled worker from the South will never migrate to the North is therefore:

\[
\lambda W_{L} < w_{H}
\]  

(9)

We can determine the expected utility of being educated for a person with an education cost \( s \) in the high-discrimination equilibrium as:

\[
EU_{H}(s) = \int_{0}^{\lambda W_{L} (\lambda - 1)} \frac{1}{c} W_{M} \, dc - s
\]  

Similarly, the expected utility of being non-educated in the high-discrimination equilibrium is:

\[
EU_{L}(s) = \int_{0}^{\lambda W_{L} (\lambda - 1)} \frac{1}{c} W_{L} \, dc
\]  

We have the following proposition:

**Proposition 6.** If

\[
w_{H} > \max \left\{ w_{L} + \frac{(1-\lambda)^{2}}{2\sigma} \lambda W_{L} \right\}
\]  

(10)

holds, there exists a unique stable high-discrimination equilibrium for all \( \sigma \in (0, 1) \) for which \( 1 - \beta = 1 - \overline{s} / \overline{s} \) workers from the South get educated where

\[
\overline{s} = W_{H} - W_{L} - \frac{(1-\lambda)^{2}}{2\sigma} \lambda W_{L}
\]  

(11)

The proof of this proposition is straightforward. Indeed, condition (10) guarantees that, in the high-discrimination equilibrium, both Eq. (9) holds and \( \overline{s} > 0 \). The threshold \( \overline{s} \) is calculated by solving \( EU_{H}(\overline{s}) = EU_{L} \). Condition (10) is intuitive since it says that, for a high-discrimination equilibrium to exist, i.e. (rational) expectations are fulfilled in equilibrium, the high-skilled wage in the South, \( w_{H} \), has to be high enough because it will ensure that some workers will find it worth to educate themselves and that they do not want to migrate to the North. In the high-discrimination equilibrium, the higher is \( w_{H} \), the high-skilled wage in the home country, the more spread is the migration cost (i.e. the higher is \( \sigma \)), the lower is \( w_{L} \), the low-skilled wage in the home country, and the smaller is the productivity difference \( \lambda \) between the two countries, the higher is the number of educated and skilled workers in the home country (i.e. the higher is \( \overline{s} \)). What is interesting here is that \( \lambda \), the productivity difference between the North and the South, still has an impact on the fraction of educated workers in the South, even though nobody migrates in equilibrium. Here the impact is negative since nobody migrates so larger productivity deters rather than encourages education.

5.2. Low-discrimination equilibrium

This case is clearly more complicated to solve. The threshold value \( \overline{s} \) is now defined by Eq. (8). Plugging the value of \( c^{+} = W_{M} - W_{L} \), \( c^{-} = \alpha W_{L} (\lambda - 1) + (1 - \alpha)(W_{M} - W_{L}) \) and \( W_{M} = \alpha W_{L} + (1 - \alpha) W_{M} \) into Eq. (8) allows us to define \( \overline{s} \) as a function only of \( W_{M}^{-} \), i.e.,

\[
2\sigma \overline{s} = \sigma(2-\sigma)(W_{M}^{-})^{2} - 2[w_{H} - (1-\sigma)(1-\lambda W_{L})]W_{M}^{+} + W_{L} - (1-\sigma)^{2}W_{L}^{+} + 2\sigma(W_{H} - W_{L})
\]  

(12)

Since \( \overline{s} = \overline{s}(1-\beta) \), we can write this equation as:

\[
\beta = \frac{\sigma(2-\sigma)(W_{M}^{-})^{2} - 2[w_{H} - (1-\sigma)(1-\lambda W_{L})]W_{M}^{+} + W_{L} - (1-\sigma)^{2}W_{L}^{+} + 2\sigma(W_{H} - W_{L})}{2\sigma \overline{s}}
\]  

(13)
This is our first key equation, which is referred to as the education determination equation.

In the low-discrimination equilibrium, the wage in the North is still determined by \( w_{N} \), which is given by Eq. (3). This equation can be written in terms of \( \beta \) as a function of \( w_{N} \). We obtain:

\[
\beta = \frac{c'(w_{N}' - \psi \lambda w_{H})}{c'(w_{M}' - \psi \lambda w_{H}) - c'(1 - \sigma)(w_{M}' - \lambda w_{L})} \quad (14)
\]

By plugging the value of \( c' = w_{N}' - w_{N} = w_{M}' - w_{M} \) and \( \sigma = \omega_{L}(\lambda - 1) + (1 - \sigma)(w_{M}' - w_{L}) \) into this equation, we obtain:

\[
\beta = \frac{(w_{N}' - w_{N}(1 + \phi_{N}) - \lambda \psi \lambda w_{H})}{(w_{M}' - w_{M}(1 + \lambda_{1}(1 - 2\sigma)) - w_{M}(1 + \phi_{M})w_{H}(1 - \lambda \sigma)(1 - \sigma)\lambda \psi \lambda w_{H})} \quad (15)
\]

This is our second key equation, which is referred to as the wage determination equation.

By combining these two Eqs. (13) and (15), we obtain the equilibrium value of \( w_{N}' \) for the low-discrimination equilibrium for which \( w_{N}' \in [w_{M}', \psi \lambda w_{H}] \). It is easily seen that solving this equation leads to a fourth-degree polynomial in \( w_{N}' \). However, by using the properties of each function (see, in Appendix A, the proof of Proposition 7), only three cases may appear: either there is no equilibrium, or one equilibrium or two equilibria. Denote by \( \beta^{*} > 0 \) and \( (w_{N}')^{*} \) the equilibrium values. In the following proposition, we show under which conditions there is a unique low-discrimination equilibrium.

**Proposition 7.** Assume that \( \forall \) and \( \exists \) are not too large. Then, if

\[
\psi > \max \left\{ \frac{w_{N}}{w_{M}} \frac{2}{X(2 - \sigma)} \right\},
\]

there exists a unique stable low-discrimination equilibrium for which \( \beta^{*} > 0 \) and \( w_{N}' < (w_{N}')^{*} < \psi \lambda w_{H} \).

**Proof.** See Appendix A.

[Fig. 4 displays this equilibrium where the non-monotonic hump-shape concave curve is the education determination Eq. (13) while the continuous decreasing curve is the wage determination Eq. (15). The conditions given in Proposition 7 guarantees that we have this configuration with a unique interior equilibrium.]

We can here perform some simple comparative statics. First, we can analyze how \( \psi \), which parameterizes differences in wage inequality between the North and the South, affects the equilibrium variables. Observe that we have assumed that \( \psi > \max \{w_{N}, w_{M}\} \) (this is encompassed in Eq. (16)), which means that the skill premium is higher in the North than in the South. In the low-discrimination equilibrium, if \( \psi \) increases so that the skill premium is even higher in the North than in the South, then only the wage determination Eq. (15) is affected. **Proposition 3** has shown that there were a positive relationship between \( w_{N}' \) and \( \psi \), which implies that there is also a positive relationship between \( \beta \) and \( \psi \) for the wage determination equation. As a result, in Fig. 4, the continuous decreasing curve of the wage determination equation is shifted to the right, so that, in equilibrium, \((w_{N}')^{*}\) increases but \( \beta^{*} \), the fraction of skilled (educated) workers in the South decreases. This is an interesting result since it says that, if \((w_{N}')^{*}\) increases due to a higher skill premium in the North, then there is more brain drain so that there will be less educated workers in the South. This is because when \( \psi \) increases, all workers in the North are better paid but the decision to be educated is not directly affected by this increase. As a result, in equilibrium, the wage for the high-skilled and the seemingly high-skilled workers, \( w_{N}' \), increases, which triggers more skilled migration so that \( \beta^{*} \), the equilibrium fraction of low-skilled workers in the South, decreases.

Consider now \( \lambda \), the absolute productivity difference between the North and the South. From Proposition 3, we know that there is a positive relationship between \( w_{N}' \) and \( \lambda \) (assuming that \( (1 - \sigma)\lambda > 1 \)), so that the continuous decreasing curve of the wage determination equation is shifted to the right. From Eq. (13), we can see that \( \lambda \) negatively affects \( \beta \) (i.e. better productivity difference between the North and the South encourages workers to educate themselves more so that the fraction of low-educated workers \( \beta \) decreases). This implies that the hump-shape education determination curve is shifted downward. As a result, the equilibrium wage in the North \((w_{N}')^{*}\) increases but the effect on the equilibrium fraction of low-educated workers, \( \beta^{*} \), is ambiguous. This is because there are two opposite forces at work. On the one hand, because people are better paid in the North, there is more skilled migration and thus less educated workers at home. On the other hand, the ex ante (before migration) incentive to be educated increases so that more people become educated. This is an interesting effect because it shows that large difference in productivities between two countries where migration is possible has always a positive effect on the decision to be educated but may result in less educated workers in the South because of the incentives to migrate to the North. This gives a clear mechanism of our discussion about the brain gain presented above (Gibson and McKenzie, 2011) where the decision of individuals to invest in education reacts to the prospect of future migration.

### 6. Stylized facts and empirical predictions

The essence of our information-theoretic model is statistical discrimination. Because what often distinguishes international from internal migration is in part how visible the foreign-born are, the issue of statistical discrimination and its implication for who migrates should be explored empirically. Most of the US studies show that, initially (i.e. when they arrive), immigrant earnings are below the ones of native (see e.g. Borjas, 1999; Carliner, 1980; Chiswick, 1978). The reasons put forward by these authors are the following. When immigrants arrive in the United States, they lack many of the skills that are valued by American employers. These US-specific skills include language, educational credentials, and information on what the best-paying jobs are and where they are located.

Let us now expose the salient features and stylized facts of international migration that our model predicts:

1. There is always a wage discrepancy between native and immigrant high-skilled workers. In other words, high-skilled immigrants are always less paid than their native counterparts. This wage difference between high-skilled native and migrant workers decreases with \( \sigma \). This result could be interpreted as the fact that high-skilled migrants have more difficulties to find a good job that fits their skills the less similar are the home and the host country. This is in line with the empirical study of Matttoo et al. (2008) who use the 2000 US Census to show that it is mainly skilled migrants from non-English speaking countries with poor-quality education systems who struggle to find skilled jobs. In other words, most skilled migrants do find a skilled job (see Table 1 in Gibson and McKenzie, 2011) unless they come from countries that are very different culturally and in terms of education system (which is captured by \( \sigma \) in our model).

---

6 The empirical studies listed below use the concept of schooling/education while our model uses the concept of skills. However, in Section 4, we show that our model can also be interpreted in terms of schooling. As a result, when confronting our model with empirical evidence, we will not differentiate between skills and schooling.

7 A good example of this is the immigration of Russian people to Israel. Most of them were highly qualified (PhDs, medical doctors...) but lack of educational credentials. When they arrive, they were allocated to low-skilled jobs. Weiss et al. (2003) show that on average, Russian immigrants can expect lifetime earnings to fall short of the lifetime earnings of comparable natives by 57%. For a recent survey on educational mismatch and labor-market outcomes of migrants, see Piracha and Vadean (2012).
There is a positive association between skilled and unskilled migrant workers. Gibson and McKenzie (2011) show (see their Fig. 1 and the empirical evidence cited in their paper) that levels of skilled and unskilled migration have a strong positive association. They find that countries that sent relatively many high-skill migrants to one country also sent relatively many low-skill migrants to the same country. Similarly, when a country increases the number of high-skill migrants it sends to a recipient country, the number of low-skilled migrants to the same country also increases.

The more culturally similar is the home and the host country (in terms of language and educational system), and/or the higher is the skill premium, and/or the higher is the absolute productivity difference between the two countries and/or the lower is the initial percentage of low-skilled workers in the home country, the more likely that there will be a brain drain.

The lower is the skill premium between the home and the receiving country, the higher is the fraction of educated workers in the home country, i.e., the more likely there is brain gain.

There is more likely to be a positive self-selection of migrants in the host country, the more culturally similar is the home and the host country (in terms of language and educational system) and/or the higher is the skill premium in the host country. Belot and Hatton (forthcoming) find that factors such as linguistic, cultural and geographical proximity are stronger determinants of positive selection of migrants than factors such as the relative returns to skills, poverty in source countries or immigration policies in receiving countries.

7. Policy issues

The immigration policy strongly influences the skill composition of immigrants. Let us discuss three possible policies in the context of our model.

7.1. Costly signalling

Diminishing the informational asymmetries faced by northern firms is one way to increase the skill composition of migrants. This can be done by letting the private sector develop a signalling activity that helps employers identify the skills of migrants. Suppose that, in addition to the school record, high-skilled workers can perfectly signal their skill at a cost $s$. The setting we consider here is the one developed in Section 4. We shall assume that the cost of signalling is the same for all workers in the South. By incurring this cost $s$ in addition to the migration cost $c$, high-skilled workers can guarantee that they will be paid $\psi\lambda w_H$ by northern firms because they will not be pooled with low-skilled workers. Provided $s < \psi\lambda w_H - w_H$, this possibility rules out the high-discrimination equilibrium. This is so because some high-skilled workers prefer to pay the signalling cost and migrate rather than stay in their home country. Thus the high-discrimination equilibrium is replaced by an equilibrium with signalling in which the same number of low-skill workers migrate and $(\psi\lambda w_H - s - w_H)/c$ high-skilled workers migrate, signal themselves and are paid the same wage as their native counterparts. This equilibrium with signalling entails a better skill composition of migrants than the high-discrimination equilibrium. When the equilibrium wages $w_M$ identified when there is no signalling opportunity are such that $\psi\lambda w_H - s > w_M$, low-discrimination equilibria are also ruled out and replaced by the equilibrium with signalling described above which is the only equilibrium of the game. In the equilibrium with signalling, high-skilled workers prospects increase while those of low-skilled workers decrease. As a result, the skill composition of migrants is improved. When $\psi\lambda w_H - s = w_M$, the signalling opportunity does not, however, influence the behavior of high-skilled workers that anticipate the equilibrium wage $w_M$. The low discrimination equilibrium stays an equilibrium issue of the game with signalling. In this case, the signalling opportunity does not change the skill composition of migrants.
7.2. Tax on low-skilled jobs

There is no reason to assume that the regulator in the North is better informed than the firms about the skills of the migrants. Therefore, in order to discourage low-skilled migration and increase the skill composition of migrants, directly taxing low-skilled migrants is impossible. One indirect possibility consists in taxing migrants with a negative school record or equivalently in our model, taxing firms that employ migrants with a negative school record. Suppose the regulator in the North imposes a tax \( t \) on migrants with a negative school record. We can replicate the analysis conducted in Section 4 as follows. The expected wage of a low-skilled worker in the South contemplating migration is now given by

\[
W^+_M = \sigma(\lambda w_l - t) + (1 - \sigma)W^+_M.
\]

The threshold cost \( c^- \) that is relevant for low-skilled workers becomes

\[
c^- = (1 - \sigma)W^+_M + \sigma(\lambda w_l - t) - w_l.
\]

For a fixed \( W^+_M \), this threshold is decreasing in \( t \), which implies that the function \( \eta(W^+_M) \) decreases as \( t \) increases because the convex combination that defines \( \eta(W^+_M) \) puts more weight on the highest productivity as \( t \) increases. By an argument similar to the one used to prove Lemma 1, we deduce that the equilibrium wage of migrants with a positive school record in the stable low-discrimination equilibrium is increasing with \( t \) and so is the equilibrium proportion of high-to-low skill immigrants.

7.3. No discrimination policy

Suppose that the regulator in the North tries to increase the prospective wage of high-skilled workers in the South contemplating migration by enforcing a no-discrimination policy. According to this policy, firms in the North shall not pay high-skilled migrants a wage different from \( \psi w_M \), the wage of their native counterparts. When \( \sigma < 1 \), the striking effect of this policy is to make high-skilled workers migration less attractive. The only equilibrium situation that remains is the high-discrimination outcome where migrants are only offered low-skilled jobs. Competitive firms cannot offer high-skilled jobs to migrants that have a positive school record because their productivity is strictly less than \( \psi w_M \) (when \( \sigma < 1 \), the pool of migrants with a positive school record contains low-skilled workers) and the firms cannot afford to pay them \( \psi w_M \). Therefore competitive firms offer low-skilled jobs to all migrants, which decreases the expected wage of high-skilled migrants.

Appendix A

Proof of Lemma 1.

(i) Define \( P(M|J; t) \in (H, I) \), as the probability that a randomly chosen worker type \( J \) decides to migrate. From Bayes’ rule

\[
P(H|\Sigma^+) = \frac{P(\Sigma^+|H)P(H)}{P(\Sigma^+|H)P(H) + P(\Sigma^+|L)P(L)}
\]

\[
P(L|\Sigma^+) = \frac{P(\Sigma^+|L)P(L)}{P(\Sigma^+|L)P(L) + P(\Sigma^+|H)P(H)}
\]

Next, observe that \( P(\Sigma^+|L) = c^- (1 - \sigma) \). Also, \( P(\Sigma^+|H) = c^+ \) as all high-skilled workers pass the foreign screening test. Finally, recall that \( P(H) = 1 - \beta \) and \( P(L) = \beta \). Combining these expressions, it is immediate to obtain the probabilities \( P(H|\Sigma^+) \) and \( P(L|\Sigma^+) \) as in the first part of the Lemma.

(ii) Since workers are paid the expected wage conditional on observables, then

\[
w^+_M = P(H|\Sigma^+)\psi w_M + P(L|\Sigma^+)\lambda w_L
\]

Using the expressions above, one gets immediately \( W^+_M \) as in the Lemma. As for the properties of \( \eta(w^+_M) \), recall that, if \( w^+_M \in [w_l, w_M] \), then \( c^- = 0 \) and \( c^+ > 0 \). Hence \( \eta(w^+_M) = \lambda w_l \) in this range. If \( w^+_M \in [\psi w_M, \psi w_M] \), then

\[
\eta(w^+_M) = \frac{(w^+_M - \psi w_M)(1 - \gamma)\psi w_M + \gamma w_M (1 + \psi \lambda)}{(w^+_M - w_l)(1 - \beta) + (1 - \gamma)(w^+_M - w_M) + \gamma \psi w_M (1 - \sigma) \lambda w_l}
\]

Standard differentiate shows that \( \eta(w^+_M) \) is increasing and concave in this range.

Proof of Proposition 1. We start by proving that, if \( \sigma \) is sufficiently large, then a low-discrimination equilibrium exists. Suppose \( \sigma = 1 \). Then, Eq. (3) in Lemma 1 implies \( \eta(w^+_M) = \psi w_M \) for \( w^+_M \geq w_M \) (and, hence, \( \eta(w^+_M) = 0 \)). In particular, \( \eta(\psi w_M) = \psi \lambda w_M \), implying that a no-discrimination equilibrium exists in this case. Next, note that \( \eta(w^+_M) \) is continuously increasing in \( \sigma \). Furthermore, \( \eta(w^+_M) \) is also a continuous function of \( \sigma \). Hence, as \( \sigma \rightarrow 1 \), \( \eta(\psi w_M) = \psi \lambda w_M \) and \( \eta(\psi \lambda w_M) \rightarrow 0 \) (implying, in particular, that \( \eta(\psi \lambda w_M) < 1 \) in a neighborhood of \( \sigma = 1 \)). Therefore, there exists \( \varepsilon > 0 \) such that \( \psi \lambda w_M - \varepsilon \in (0, w^+_M) \). This proves that a low-discrimination equilibrium exists if and only if \( \varepsilon \leq w^+_M / \psi w_M \).

We continue the proof according to the numbering in the Proposition.

1. First, note that, for any \( \sigma > 1 \), \( \eta(\psi w_M) = \psi w_M \). Next, observe that \( \lambda > w_M / \psi \lambda w_M \) implies that \( \eta(\psi w_M) \geq w_M > w_M \). Then, the continuity of \( \eta(w^+_M) \) establishes the existence of a low-discrimination equilibrium. Furthermore, since \( \eta(w^+_M) \) is increasing and concave for all \( w^+_M > w_M \), then the equilibrium is unique. Part 1 of the Proposition is, therefore, established.

2. Consider, next, the range where \( \lambda > w_m / \psi \lambda w_M \). Assume \( \sigma = 0 \), and recall that \( \eta(\sigma) \) is strictly increasing in \( \sigma \) (in the range \( \omega^+_M \in [w_L, \psi w_M] \)). Then, if a low-discrimination equilibrium exists for \( \sigma = 0 \), such equilibrium also exists for all positive \( \sigma's \). More formally, a low-discrimination equilibrium, given \( \sigma = 0 \), exists if there exists value(s) of \( w^+_M \) such that, for some \( w^+_M > w^+_M > w^+_M \),

\[
w^+_M = \lambda \left( \frac{(w^+_M - w_M)(1 - \beta)\psi w_M + \gamma w_M (1 + \psi \lambda)}{(w^+_M - w_l)(1 - \beta) + (1 - \gamma)(w^+_M - w_M) + \gamma \psi w_M (1 - \sigma) \lambda w_l} \right)
\]

Multiplying both sides of this expression by the denominator of the right-hand side yields the following quadratic equation

\[
w^+_M^2 - (\psi w_M (1 - \beta) (1 + \psi \lambda) + \beta w_M (1 + \lambda)) w^+_M + \lambda \left( \psi w_M^2 (1 - \beta) + \beta w_M^2 \right) = 0.
\]
The roots of Eq. (17) are real if and only if
\[ \Delta = (w_H(1-\beta) + \psi \lambda + \beta w_M(1 + \lambda))^2 - 4 \lambda (\psi w_H(1 - \beta) + \beta w_M^2) \geq 0. \]

This inequality holds if either
\[ \psi < \frac{w_H(1 + \beta) - w_I(1 + \lambda) - 2 \sqrt{\beta (w_H - w_I)(w_H - \lambda w_M)}}{\lambda(1 - \beta)w_H} = \phi_1, \]
or
\[ \psi > \frac{w_H(1 + \beta) - w_I(1 + \lambda) + 2 \sqrt{\beta (w_H - w_I)(w_H - \lambda w_M)}}{\lambda(1 - \beta)w_H} = \phi(\lambda). \]

The range \( \psi < \phi_1 \) can, however, be ruled out, since it implies that both roots (in \( w_M^* \)) of Eq. (17) are smaller than \( w_M \). Thus, no low-discrimination equilibrium exists for \( \sigma = 0 \) if \( \psi < \phi_1 \). A low-discrimination equilibrium (under the assumption that \( \lambda > w_M/w_I \)) therefore exists for \( \sigma = 0 \) (and, a fortiori, for any \( \sigma > 0 \)) if \( \psi > \phi(\lambda) \).

3. Finally, consider the range where \( \lambda > w_M/w_I \) and \( \psi < \phi(\lambda) \). Then, part 2 of the proposition established that no low-discrimination equilibrium exists for \( \sigma = 0 \). However, we know that a low-discrimination equilibrium exists for \( \sigma > 1 \). Moreover, \( \eta(\cdot) \) is continuously increasing and concave in \( \sigma \) in the range of \( w_M^* \in [w_M, \lambda w_M] \). Thus, a unique low-discrimination equilibrium exists.

Stability in our setting refers to the properties of the tâtonnement process where migration decisions adjust to wages and wages in turn adjust to migration decisions. An equilibrium is stable in our game when \( \eta(\frac{w_M^*}{w_L}) < 1 \) or when the function \( \eta(\cdot) \) crosses the diagonal from above. An equilibrium is unstable when \( \eta(\frac{w_M^*}{w_L}) > 1 \) or when the function \( \eta(\cdot) \) crosses the diagonal from below. Stability properties of the different equilibria as mentioned in the proposition are straightforward to check.

Proof of Proposition 2. Suppose the migration of high-skilled workers is such that a number \( c^* \in [c, 2] \) decided to migrate. The simultaneous migration decision of low-skilled workers will result in a wage \( w_M^*(c^*) \) for those who pass the screening test. The equilibrium condition for the migration of low-skilled workers is
\[ w_M^* = \frac{c^*(1 - \beta) \psi \lambda w_H + (1 - \sigma)(w_M^* - w_I) + \sigma w_L(1 - \lambda) \beta \lambda w_M}{c^*(1 - \beta) + ((1 - \sigma)(w_M^* - w_I) + \sigma w_L(1 - \lambda) \beta) \lambda}, \]

where the left-hand side is the expected productivity of a migrant who passed the test, when low-skilled workers in the South expect such a wage to be \( w_M^* \). It is easy to see that the right-hand side is decreasing in \( w_M^* \), so that there exists a unique equilibrium value for \( w_M^*(c^*) \). This shows the first point in the Proposition.

Next we show that the equilibrium value \( w_M^*(c^*) \) is an increasing function of \( c^* \). Suppose not. This means that increasing the number of high-skilled workers decreases the expected productivity of a migrant who passed the test. This, in turn, means that an increase in the number of high-skilled migrants unambiguously increases the number of low-skilled migrants. However, a lower \( w_M^* \) must induce a lower number of low-skilled migrants since it decreases the prospects of low-skilled workers: a contradiction. Therefore we know that \( w_M^*(c^*) \) is an increasing function and the game played by the high-skilled workers in the first stage is a game with strategic complementarities.

Finally, because our setting is one with a continuum of players, no worker has an influence on aggregate variables such as the expected productivity of migrants. Therefore, playing first does not give any advantage to the high-skilled workers.

Proof of Proposition 3. The comparative statics results for the \( w_M^* \) corresponding to the stable low-discrimination equilibrium are obtained by studying the function \( \eta(w_M^*) \) given by Eq. (3). At this equilibrium, the function \( \eta(\cdot) \) verifies locally
\[ \eta(x) \equiv x \iff x \leq w_M^*. \]

Therefore it is sufficient to show that, for a fixed \( w \), \( \eta(w) \) is increasing (resp. decreasing) in a parameter to prove that \( w_M^* \) is increasing (resp. decreasing) in that parameter.

At a low-discrimination equilibrium, both types of workers migrate and we know that
\[ \eta(w_M^*) = \frac{(w_M^* - w_I)(1 - \beta) \psi \lambda w_I + ((1 - \sigma)(w_M^* - w_L) + \sigma(\lambda - 1)w_I)(1 - \sigma) \lambda w_I}{(w_M^* - w_I)(1 - \beta) + ((1 - \sigma)(w_M^* - w_L) + \sigma(\lambda - 1)w_I)(1 - \sigma) \beta}. \]

It is straightforward to show that the function \( \eta(\cdot) \) is increasing with \( \psi \). It is decreasing with \( \beta \) because \( \psi w_M \geq w_I \) and increasing \( \beta \) gives more weight to \( \psi w_M \) in the convex combination that defines \( \eta(\cdot) \). By the same token, and because \( w_M^* \geq w_L \), increasing \( \sigma \) lowers the weight attributed to \( w_I \) in the convex combination and therefore increases \( \eta \).

When \( \lambda \) increases, both the weight attributed to \( \lambda w_I \) and the value of \( \lambda w_I \) increase. Moreover, the term
\[ \frac{(w_M^* - w_I)(1 - \beta) \psi \lambda w_I}{(w_M^* - w_I)(1 - \beta) + ((1 - \sigma)(w_M^* - w_L) + \sigma(\lambda - 1)w_I)(1 - \sigma) \beta} \]

is increasing in \( \lambda \) when \( (1 - \sigma)(w_M^* - w_L) - \sigma w_I > 0 \). Because we know that \( w_M^* \geq \lambda w_I \), we can deduce that \( \eta(w_M^*) \) is increasing with \( \lambda \) when \( (1 - \sigma) > 1.10 \]

Proof of Proposition 7. As stated in the text, by combining Eqs. (13) and (15), we obtain the equilibrium value of \( w_M^* \) for the low-discrimination equilibrium. Let us study these two equations and show that they intersect only once.

Let us first study the education determination equation, which is given by Eq. (13). This is a second-order polynomial in \( w_M \). Let us denote this function as \( \beta = f(w_M^*) \). The discriminant of this second-order polynomial is given by:
\[ \Delta = \frac{(w_H - (1 - \sigma)(1 - \lambda) w_I)_{1}^{2}}{c^*(2 - \sigma)} - \frac{\alpha(2 - \sigma)}{c^*(2 - \sigma)} \times \left[w_H - (1 - \lambda) w_L^2 + 2 \Sigma (w_H - w_I) + \tau \Sigma \right] \]

The discriminant \( \Delta \) is strictly positive if and only if:
\[ [1 - \sigma(2 - \sigma)]w_H^2 + (1 - \lambda) w_L^2 > 2 \sigma c^*(2 - \sigma)(w_H - w_I) + 2(1 - \sigma)(1 - \lambda) w_H w_I + \alpha c^*(2 - \sigma) \Sigma \]

This inequality is true if \( c^* \) and \( \Sigma \) are small enough, which we assume. As a result, the second-order polynomial \( f(w_M^*) \) is a concave function with two real positive roots, which are given by:
\[ \frac{w_M^*}{\alpha(2 - \sigma)} = \frac{2(w_H - (1 - \sigma)(1 - \lambda) w_I) - 2 \Sigma \sqrt{\Delta}}{\alpha(2 - \sigma)} > 0 \]

\[ \frac{w_M^*}{\alpha(2 - \sigma)} = \frac{2(w_H - (1 - \sigma)(1 - \lambda) w_I) + 2 \Sigma \sqrt{\Delta}}{\alpha(2 - \sigma)} > 0 \]

10 Of course, the equilibrium value of \( w_M^* \) solves a quadratic equation and can be computed analytically. It is therefore possible to find the necessary and sufficient condition on the parameters for \( w_M^* \) to be increasing with \( \lambda \) in the stable low-discrimination equilibrium. However, the interpretation of this condition would be difficult and would involve too many parameters. For simplicity we prefer the simpler sufficient condition found above.
We have
\[ \frac{\partial \beta}{\partial w_M} = f'(w_M^*) = -\frac{\sigma(2-\sigma)}{\sigma} w_M^* + \frac{1}{\sigma} [w_M^* (1-\sigma)(1-\lambda\sigma)w_L] \]
so that
\[ f'(w_M^*) > (\sigma) 0 \iff w_M^* > (\sigma) 0 \]

Furthermore,
\[ f(0) = -\frac{1}{\sigma} [w_M^* (1-\sigma)^2 + 2\sigma(w_M^* - w_L) + \sigma \beta] < 0 \]

In Fig. 4, we plot this function \( \beta = f(w_M^*) \).

Let us now study the wage determination equation, which is given by Eq. (15) and that we denote by \( \beta = h(w_M^*) \). In the benchmark model (without education choice), we have studied the relationship between \( w_M^* \) and \( \beta \) which is given by Eq. (3). Let us denote this relationship by \( w_M^* = g(\beta) \). In Proposition 3, we have shown that this relationship was negative, i.e. \( g'(\beta) \leq 0 \). Since Eq. (15) given by \( \beta = h(w_M^*) \) is just the inverse function, i.e. \( \beta = h(w_M^*) = g^{-1}(w_M^*) \), we have that \( h'(w_M^*) \leq 0 \) since
\[ h'(w_M^*) = (g^{-1})'(w_M^*) = \frac{1}{g'(\beta)} \leq 0 \]

It also easily verified that:
\[ h(0) = \frac{\psi \lambda \sigma}{(1-\lambda\sigma)(1-\sigma)w_M^* + \psi \lambda w_H} > 0 \]
\[ \lim_{w_M^* \to -\infty} h(w_M^*) = \frac{1}{(1-\sigma)^2 + 1} > 0 \]
and
\[ g(0) = w_M^*(\beta = 0) = \psi \lambda w_H \]

In Fig. 4, we plot this function \( \beta = h(w_M^*) \). It can cross the other curve \( \beta = f(w_M^*) \) in only three ways: only once (one equilibrium), or twice (two equilibria) or not at all (no equilibrium). We here focus on the case described in Fig. 4, i.e. a unique low-discrimination equilibrium, which we denote by \((w_M^*)^*\). We need now to check that this equilibrium \((w_M^*)^*\) is such that \( w_M^* < (w_M^*)^* < \psi \lambda w_H \). We know that \( (w_M^*)^* < \psi \lambda w_H \) (see Fig. 4). We need therefore to show that: \((w_M^*)^* > w_H\), which is equivalent to that \((w_M^*)^* > w_H\).

Using the value of \( \Delta \) in Eq. (18), this inequality can be written as:
\[ \omega \psi \lambda \sigma + 4(1-\sigma)(1-\lambda\sigma)w_M^* + \psi \lambda w_H > 4(1-\sigma)(1-\lambda\sigma)^2 w_L^2 \]

which is equivalent to:
\[ (1-\sigma)w_M^* > (1-\lambda\sigma)w_L \]
which is always true since \( w_M^* > w_L \) and \( 1-\sigma > 1-\lambda\sigma \) since \( \lambda > 1 \).

Let us summarize the conditions we need for this equilibrium to exist and to be unique:

(i) \( \varepsilon \) and \( \delta \) have to be not too large for the discriminant \( \Delta \) to be strictly positive;
(ii) \( \psi > \max \left\{ \frac{\lambda \sigma}{\psi \lambda \sigma}, \frac{\lambda \sigma}{\psi \lambda \sigma} \right\} \), which guarantees that the skill premium is positive in the North and that \((w_M^*)^* < \psi \lambda w_H\) holds.
(iii) \( \frac{\lambda \sigma}{\psi \lambda \sigma} > \frac{1}{1-\psi \lambda \sigma} \), which guarantees that \( \psi \lambda w_H < \frac{\lambda \sigma}{\psi \lambda \sigma} (1-\lambda\sigma)w_M^* \) holds.

Conditions (i) and (ii) can be put together since \( \psi > \frac{\lambda \sigma}{\psi \lambda \sigma} \) is equivalent to \( -\psi \lambda \sigma < -1 \). This implies that \( \frac{\lambda \sigma}{\psi \lambda \sigma} > \frac{1}{1-\psi \lambda \sigma} \) since \( \frac{1}{1-\psi \lambda \sigma} > 1 \). In other words, if \( \psi > \frac{\lambda \sigma}{\psi \lambda \sigma} \), holds, then \( \frac{\lambda \sigma}{\psi \lambda \sigma} > \frac{1}{1-\psi \lambda \sigma} \) is always satisfied. So the only conditions left for the low-discrimination equilibrium to be unique such that \( w_M^* < (w_M^*)^* < \psi \lambda w_H \) and \( \beta > 0 \) are:

(i) \( \varepsilon \) and \( \delta \) have to be not too large;
(ii) \( \psi > \max \left\{ \frac{\lambda \sigma}{\psi \lambda \sigma}, \frac{\lambda \sigma}{\psi \lambda \sigma} \right\} \).

This is what is stated in the proposition.\( \blacksquare \)

References