Heaven’s Swing Door:
Endogenous skills, migration networks and the effectiveness of quality-selective immigration policies

Simone Bertoli\textsuperscript{a} and Hillel Rapoport\textsuperscript{b}

\textsuperscript{a}CERDI\textsuperscript{†}, University of Auvergne and CNRS
\textsuperscript{b}Paris School of Economics and Bar-Ilan University

Abstract

A growing number of OECD countries are leaning toward the adoption of selective immigration policies, which are expected to raise immigrants’ quality (or education level). This view neglects two important dynamic effects: the role of migration networks, which could reduce immigrants’ quality, and the responsiveness of education decisions to the prospect of migration. We propose a model of self-selection into migration with endogenous education choices that predicts that migration networks and immigrants’ quality can be positively associated when destination countries adopt sufficiently selective immigration policies. Empirical evidence presented as background motivation for this paper suggests that this is indeed the case.

Keywords: migration, self-selection, brain drain, immigration policies, discrete choice models.

JEL classification codes: F22, O15, J61.

\textsuperscript{†}Bd. F. Mitterrand, 65, F63000, Clermont-Fd; email: simone.bertoli@udamail.fr (corresponding author).
1 Introduction

A growing number of OECD countries are leaning toward adopting more restrictive and increasingly “quality selective” immigration policies, which confer better chances of admission at destination to the applicants with a higher level of education. This tendency is apparent from the gradual introduction of points-based immigration systems, first in Canada in 1967, followed by Australia in 1989, New Zealand in 1991 and more recently the United Kingdom in 2008. Elsewhere, immigration policies have also evolved towards becoming more restrictive quantitatively and more selective qualitatively, be it through the introduction of specific visa categories for highly-skilled professionals (e.g., the H1-B visa category in the United States, or the European “Blue Card” project currently in its infancy) or through introducing biased selection criteria making low-skill immigration more difficult while at the same time encouraging permanent high-skill immigration (e.g., France’s short-lived “chosen immigration” reform of 2007).

Observed immigration flows are the result of the combined effect of self-selection (i.e., size and skill composition of a given pool of candidate immigrants) and out-selection (i.e., external selection among existing candidates) mechanisms. The underlying assumption behind quality-selective immigration policies is that more selection will raise immigrants’ average education level. This makes perfect sense from a static standpoint, when we regard the pool of applicants as given. Still, this pool evolves over time as the migration process unfolds: migration networks tend to reduce the moving costs (Massey et al., 1994; Carrington et al., 1996; Munshi, 2003; Kanbur and Rapoport, 2005), thus influencing the self-selection into migration (McKenzie and Rapoport, 2010; Beine et al., 2011a; Bertoli, 2011).

The migration cost-reducing effects of networks is best illustrated by the “swing door”

---

1 Education alone can provide an applicant with up to 25 out of the 67 points that are currently necessary for admission into Canada (Bertoli et al., 2012), and its pivotal role in shaping the chances of admission is magnified by its positive correlation with labor market experience (21 points) and language proficiency (24 points).

2 The decline in the level of education of the immigrants (Borjas, 1999), with its possible contribution to rising inequality and increased pressure on underprivileged segments of the native population (Borjas et al., 2010), has prompted proposals to increase the degree of selectivity of immigration policies (see, for instance, the specific proposals advanced by Borjas, 1999); the immigration reform bill that was introduced in the US Senate in April 2013 contains a provision for the admission of 120,000 immigrants per fiscal year through a merit-based system.
metaphor in our title: while the first migrants to push the door will encounter the most resistance, their followers will be able to enjoy the lower resistance of a swing door in movement. The literature has established that this “swing door” effect is not neutral with respect to the immigrants’ quality: McKenzie and Rapoport (2010) find that the probability of (first time) migration from Mexico to the United States increases with education up to relatively high education levels in communities with small networks (high migration costs), which is consistent with positive self-selection, and, conversely, that migration propensities decrease with education in communities with large networks (low migration costs), which is consistent with negative self-selection. A similar result is obtained by Bertoli (2011) from the analysis of Ecuadorian migration, while Beine et al. (2011a) provide evidence that larger networks translate into more negative self-selection patterns using bilateral data from 195 sending to 31 destination countries.3,4

The “swing door” effect prima facie suggests that the positive effect of an increase in selectivity on the average level of education of the migrants is, at best, short-lived, as the expansion of migration networks progressively erodes the influence of the change in immigration policies on migrants’ quality. Still, the expansion of migration networks does not influence migrants’ quality only through its uneven impact on the decision to self-select into migration of individuals with different levels of education, as it also influences education decisions at origin, which respond to changes in the prospect of migrating (see, inter alia, Mountford, 1997; Stark et al., 1997; Beine et al., 2001). This occurs as the expansion in the size of migration networks increases the chances of opting for migration, thus making the expected return to the investment education more sensitive both to the labor market conditions that prevail at destination and to the chances of admission for potential migrants with different education credentials. This, in turn, suggests that the response of education decisions at origin to an expansion in the size of networks depends on the immigration policies that are adopted at destination, as selective immigration policies can create further

3See also Beine et al. (2011b), Bertoli and Fernández-Huertas Moraga (2012) and Beine and Salomone (2013).

4The structure of fixed effects in Beine et al. (2011a), Bertoli and Fernández-Huertas Moraga (2012) and Beine and Salomone (2013) controls for the dependency of migrants’ quality on immigration policies, but both papers maintain the assumption that the effect of migration networks on quality is independent from immigration policies; see also Antecol et al. (2003) and Jasso and Rosenzweig (2009) for analyses of the relationship between immigration policies and the level of education of the immigrants where the latter is assumed to be independent from the size of migration networks at destination.
incentives to invest in education in order “to be eligible for emigration” (Mountford, 1997, p. 301).

The objective of this paper is to understand whether the adoption of selective immigration policies can produce a change in the sign of the relationship between migration networks and immigrants’ quality. The relevance of this research question is connected to the observation that the legal framework that regulates immigrants’ admission at destination often remains unchanged for long spells of time, as the introduction of reforms is hindered by the contentious and divisive nature of the political debate on immigration in most recipient countries. This, in turn, implies that it is crucial to understand not only how selective immigration policies influence in the short-run the quality of the immigrants, but also how which are the effects they produce in the long-run, when the factors that shape self-selection into migration evolve.

We draw on two recent strands of migration research, namely the literature on networks and self-selection (Borjas, 1987; Chiquiar and Hanson, 2005; McKenzie and Rapoport, 2010; Belot and Hatton, 2012; Fernández-Huertas Moraga, 2013) and the new brain drain literature (Mountford, 1997; Stark et al., 1997; Beine et al., 2001, 2008; Stark and Wang, 2002; Chand and Clemens, 2008; Docquier and Rapoport, 2012) to propose a unified theoretical framework where immigration policies, network effects and endogenous education decisions jointly determine the eventual pattern of migrants’ selection. The model consists of a discrete-choice, random utility-maximization model where heterogeneous individuals in terms of ability make their education decisions while considering the costs, which are influenced by the size of networks, and expected benefits, which depend on foreign wages and on the probability of admission at destination, of migration.

A central result of this paper is to show that quality-selective immigration can enhance immigrants’ quality in the long-run. More precisely, we demonstrate that migration networks and immigrants’ quality can be positively associated under a condition regarding (i) the degree of selectivity of immigration policies, (ii) the initial pattern of migrants’ self-selection on education and (iii) the way education-specific time-equivalent migration costs relate to networks. Interestingly, the possibility of a positive relationship between network size and immigrants’ quality is fully driven by the endogenous response of education decisions at

\footnote{For instance, the “Immigration Reform and Control Act”, which represents the last major reform adopted in the United States, dates back to 1986.}
origin and it is independent of the static gains from increased selectivity.

In any event, these results imply that the relationship between network size and immigrants’ quality should differ according to the type of immigration policy (selective versus non-selective) adopted at destination. This is consistent with the empirical evidence presented as background motivation in Section 2 below. Section 3 and 4, on the other hand, are purely theoretical: Section 3 presents the general theoretical framework and Section 4, which derives the main predictions, focuses on the interplay between networks, immigrants’ quality and the degree of selectivity of immigration policies. Section 5 concludes.

2 Empirical background

In this section we present a number of stylized facts on the dynamic relationship between migration networks and the skill composition of immigration. We draw on data on immigrant stocks disaggregated by country of origin and level of education recently collected by Brücker et al. (2013) for 20 OECD receiving countries.\footnote{The dataset by Brücker et al. (2013) provides seven observations between 1980 and 2010 on the size of bilateral migrant stocks, broken down by gender and level of education, for up to 195 origin countries, and it builds upon the methodology proposed by Docquier et al. (2009) and Defoort (2008).} We use these data as background empirical motivation for our model rather than for testing it empirically as such an ambitious objective would require having data on gross migration flows (rather than stocks) and would also require obtaining comparative data on immigration policies for all main destinations.\footnote{Variations in stocks represent a very noisy measure of gross flows, as they also reflect attrition due to return migration, migration to third countries, and mortality (Docquier and Rapoport, 2012).} However, there is at present no comparative database on immigration policies; Mayda (2010) and Ortega and Peri (2013) represent only partial exceptions in this respect, as they focus only on the openness of immigration policies, while our theoretical model suggests that it is their degree of selectivity that can shape the relationship between networks and migrants’ quality.

We specify a pseudo-gravity model of international migration from an underlying random utility model that describes the location-decision problem faced by prospective migrants (Grogger and Hanson, 2011; Beine et al., 2011a; Ortega and Peri, 2013; Bertoli and Fernández-Huertas Moraga, 2013) to estimate the following selection equation:\footnote{This equation can be derived from an underlying RUM model where the stochastic component follows}
\[ g_{jkt} = \ln \left( \frac{m_{jk}^h}{m_{jk}^l} \right) = \alpha_k \ln(n_{jkt-5}) + d_{jk} + d_{jt} + d_{kt} + \epsilon_{jkt} \] (1)

where \(m_{jk}^h\) and \(m_{jk}^l\) represent the stock of high- and low-educated migrants\(^9\) originating from country \(j\) and residing in destination \(k\) at time \(t\),\(^10\) \(n_{jkt-5}\) represents the size of migration networks at time \(t - 5\),\(^11\) and \(d_{jk}\), \(d_{jt}\) and \(d_{kt}\) represent origin-destination, origin-time and destination-time dummies respectively. The inclusion of origin-destination dummies \(d_{jk}\) allows to purge the relationship between migrants’ quality and the size of migration networks from the confounding influence of dyadic time-invariant factors such as distance, historical relationships, or linguistic and cultural proximity, that can influence migrants’ quality. Similarly, the dummies \(d_{jt}\) and \(d_{kt}\) allow to control for any time-varying factors that are specific either to the origin country \(j\) or to the destination country \(k\). We allow for the coefficient of the size of migration networks, \(\alpha_k\), to vary across destinations \(k\). In the absence of systematic data on immigration policies adopted at destination, this approach represents a parsimonious (albeit crude) way to capture differences in the relationship of interest across countries that have different legal frameworks that regulate incoming migration flows, as in Docquier \textit{et al.} (2012).

Table 1 reports the estimates of the selection equation on an unbalanced panel of 16,521 non-missing observations.\(^{12}\) Specification (1) retains the assumption that the effect of networks does not vary across destinations: the estimates suggest that a 10 percent increase in the scale of bilateral migration networks reduces the ratio between high- and low-educated migrants by 0.97 percent, an effect that is significant at the 1 percent confidence level and is in line with the evidence provided by Beine \textit{et al.} (2011a) using cross-sectional data.\(^{13}\)

---

\(^{9}\)Highly-educated migrants are defined as having college education.

\(^{10}\)We use migration stocks rather than flows as in Grogger and Hanson (2011) and Beine \textit{et al.} (2011a).

\(^{11}\)Notice that we follow Beine \textit{et al.} (2011a) and Bertoli and Fernández-Huertas Moraga (2012) by adding one to the size of migration networks before taking the natural logarithm, to avoid generating missing values.

\(^{12}\)The dataset by Brückner \textit{et al.} (2013) has a total size of 23,280 observations.

\(^{13}\)Notice that the corresponding effect estimated in the baseline specification of the selection equation by Beine \textit{et al.} (2011a) stands at -1.71 percent; the longitudinal dimension of the data by Brückner \textit{et al.} (2013) allows us to include dyadic fixed effects \(d_{jk}\) in (1), and this might contribute to explain the smaller size of our estimated elasticity of migrants’ quality with respect to networks.
Table 1: Migrants’ quality and migration networks

<table>
<thead>
<tr>
<th>Networks interacted with destination dummies</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>0.014</td>
<td>(0.031)</td>
</tr>
<tr>
<td>Austria</td>
<td>-0.116***</td>
<td>(0.031)</td>
</tr>
<tr>
<td>Canada</td>
<td>0.003</td>
<td>(0.023)</td>
</tr>
<tr>
<td>Chile</td>
<td>-0.195***</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Denmark</td>
<td>-0.106***</td>
<td>(0.028)</td>
</tr>
<tr>
<td>Finland</td>
<td>-0.026</td>
<td>(0.032)</td>
</tr>
<tr>
<td>France</td>
<td>-0.066**</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Germany</td>
<td>-0.105***</td>
<td>(0.034)</td>
</tr>
<tr>
<td>Greece</td>
<td>-0.098***</td>
<td>(0.025)</td>
</tr>
<tr>
<td>Ireland</td>
<td>-0.097***</td>
<td>(0.030)</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>-0.065*</td>
<td>(0.037)</td>
</tr>
<tr>
<td>Netherlands</td>
<td>-0.085**</td>
<td>(0.036)</td>
</tr>
<tr>
<td>New Zealand</td>
<td>-0.180***</td>
<td>(0.026)</td>
</tr>
<tr>
<td>Norway</td>
<td>0.033</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Portugal</td>
<td>-0.200***</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Spain</td>
<td>0.025*</td>
<td>(0.015)</td>
</tr>
<tr>
<td>Sweden</td>
<td>-0.178***</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Switzerland</td>
<td>-0.128***</td>
<td>(0.032)</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>-0.072**</td>
<td>(0.032)</td>
</tr>
<tr>
<td>United States</td>
<td>-0.180***</td>
<td>(0.023)</td>
</tr>
</tbody>
</table>

Observations 16,521 16,521
Adjusted $R^2$ 0.888 0.890
$F$-test - 11.96***
Origin-destination dummies Yes Yes
Origin-time dummies Yes Yes
Destination-time dummies Yes Yes

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; standard errors in parentheses; $F$-test on the null hypothesis that the estimated coefficients do not vary across destinations.

Source: Authors’ elaboration on Brücker et al. (2013).
Specification (2) relaxes this assumption, whose validity is rejected by the data: a $F$-test on the null hypothesis that $\alpha_k = \alpha$ for all $k$ rejects it, suggesting that the relationship between migrants’ quality and migration networks varies across destinations.\footnote{If the null hypothesis had not been rejected, then the destination-time dummies $d_{kt}$ would have fully absorbed the influence of selective immigration policies on migrants’ quality $g_{jkt}$.}

Figure 1: Migrants’ quality and migration networks, United States (1985-2010)

![Graph showing the relationship between migrants' quality and migration networks.](source: Authors' elaboration on Brücker et al. (2013).)

Notes: partial regression plot of the residuals of a regression of each of the two variables on origin×year, destination×year and origin×destination dummies; skilled migrants are defined as the migrants with tertiary education; migration networks are defined as the total stock of migrants in the year $t - 5$; the figure is based on 1,095 observations with positive stocks for both skilled and unskilled migrants.

The second data column of Table 1 reveals that the negative elasticity of migrants’ quality with respect to networks identified by Beine et al. (2011a) is notably robust when we allow for heterogeneity across destinations, but it also suggests that the adoption of selective immigration policies could actually be influencing the relationship between networks and quality. Specification (2) reveals that the estimated coefficient $\hat{\alpha}_k$ is negative for 16 out
Figure 2: Migrants’ quality and migration networks, Canada (1985-2010)

Notes: see the notes to Figure 1; the figure is based on 1,018 observations with positive stocks for both skilled and unskilled migrants.

Source: Authors’ elaboration on Brücker et al. (2013).

of 18 non-selective destinations in our sample,\textsuperscript{15,16} while the estimated coefficient is not significantly different from zero for Australia and Canada, two destinations that adopted selective immigration policies since 1989 and 1967, respectively. We also allowed $\alpha_k$ for Australia to vary before and after the introduction of the point-based system: the estimated coefficient is negative before 1989, and positive and significant at the 5 percent confidence level after the adoption of selective immigration policies.\textsuperscript{17}

\textsuperscript{15} The only exceptions are represented by Norway and Spain.

\textsuperscript{16} Notice that we regard New Zealand, whose estimated coefficient of networks is negative and significant, as a non-selective destination, as “the New Zealand system has evolved into a model where entry is granted on the basis of very short-term labour market considerations” (Bertoli et al., 2012, p. 28), and as the points associated to individual characteristics, such as education, is less pronounced than either in Australia or in Canada.

\textsuperscript{17} Results are available upon request.
Figure 1, which plots the variation in the data that identifies $\hat{\alpha}_k$ for the United States, reveals that a 10 percent increase in the size of the bilateral migration networks is associated with a significant 1.8 percent decline in the quality of immigrants, as measured by the ratio of skilled to unskilled immigrants, in the United States. For Canada, which has been adopting selective immigration policies since 1967, the partial-regression plot in Figure 2 does not reveal a negative relationship between the two variables, as the slope of the fitted line stands at 0.003 and it is not significantly different from zero. Similar differences emerge if we compare another destination with selective immigration policies such as Australia with other non-selective destinations, such as France or Germany, as evidenced by our estimates in Table 1. This suggests that an increase in the size of migration networks leads to a worsening in migrants’ quality in nearly all non-selective destinations, while quality does not systematically vary with the scale of migration networks in the few clearly selective countries in our sample. The theoretical model proposed below tries to make sense of these stylized facts and identifies the conditions under which selective immigration policies preserve or improve the quality of immigration even if networks expand over time.

3 The model

We propose a theoretical two-country model that relies on a random utility maximization model à la McFadden (1974) to describe the process of self-selection into migration of individuals who optimally chose their level of education, which can also influence their chances of admission at destination. As in the literature on the beneficial brain drain, we assume that there is an excess supply of potential migrants, so that the destination country randomly selects the immigrants to be admitted through a system of lotteries, which can be specific to applicants with a certain level of education. We assume that the uncertainty about migration, which is related to the outcome of the migration and to the individual preference for migration, captured by the realization of the stochastic component of utility, is not resolved when the individuals have to decide whether or not to invest in education. This implies that individuals take their education decisions on the basis of the known probability.

---

18We regard the United States as a non-selective destination, although they have also introduced some selective channels of entry, such as the H1-B visa program, as the numerical relevance of these entry channels is limited.

19We assume that education is a discrete decision.
of admission at destination and of a rational expectation on the probability that they will find optimal to self-select into migration. This probability depends on the wage differential between the two countries, and on migration costs\(^{20}\) which are negatively related to the size of migration networks (Carrington et al., 1996). An expansion in the size of networks influences both self-selection into migration and education decisions, and it thus changes migrants’ quality, which we define as an increasing function of the ratio between educated and uneducated migrants. The direction of the influence of the networks-induced reduction in migration costs on education decision at origin is \textit{a priori} ambiguous, and it depends on the extent to which educated applicants enjoy a higher probability of admission at destination and on the initial pattern of self-selection on education, which is, in turn, determined by the education-specific wage differential between the two countries and by migration costs.

Our model builds upon Beine et al. (2001), introducing a number of major innovations: (i) we allow moving cost to depend on the size of migration networks as McKenzie and Rapoport (2010) and Beine et al. (2011a), (ii) we include a stochastic term in location-specific utility\(^{21}\) and (iii) we do not normalize the probability of admission at destination of uneducated individuals to zero. We also introduce a key modification in the underlying hypotheses of the model as we assume that the heterogeneity across individuals in innate learning ability results from different time-equivalent costs of education, as in Beine et al. (2008), rather than from a different level of effective units of labor acquired with an education that has an identical cost across agents\(^{22}\).

\section*{3.1 Setup}

We model the choices taken by a mass of two-period lived individuals; in the first period, agents can either devote all of their time to domestic employment, or devote a share \(a \in [0, 1]\) of their time to education. The agents who do not invest in education keep their initial level of human capital unchanged also in the second period of their lives, i.e., \(h_1^i = h_2^i = 1\), while

\(^{20}\)Migration costs include all monetary and non-monetary, e.g., psychological, costs associated to a move to the destination country.

\(^{21}\)See also Borjas (1987), Grogger and Hanson (2011), Beine et al. (2011a), Bertoli and Fernández-Huertas Moraga (2013) and Ortega and Peri (2013) for random utility models applied to migration decisions.

\(^{22}\)Both assumptions give rise to differences across individuals in the net returns to education, but the one that we retain here is more convenient from an analytical perspective, as it implies that educated migrants are not heterogeneous with respect to their productivity.
the individuals who invest in education have a human capital \( h_i^2 \) equal to \((1 + \phi_h)\), with \( \phi_h > 0 \) representing the education premium at origin, in the second period of their lives.\(^{23}\)

Let \( I(i) \) be an indicator function that takes the value of one if the individual \( i \) pursued an education in the first period of his life, and zero otherwise.

We focus on a small open economy. We assume that wages are an exponential function of \( h \); utility \( u^i \) is additively separable and logarithmic in the income received in the two periods; furthermore, utility in the second period also contains an individual- and location-specific stochastic component \( \epsilon^i_j \), with \( j = h, d \). The stochastic component of utility can reflect a preference shock or a shock to the cost of moving, as in Kennan and Walker (2011), or a random term in wages. Notice that, as we retain from the literature (Mountford, 1997; Beine et al., 2001, 2008; Docquier and Rapoport, 2012) the assumption that migration occurs only in the second period,\(^{24}\) the inclusion of a stochastic component of utility in the first period would have no influence on the analysis.\(^{25}\)

The utility associated to a domestic employment in both periods is equal to:\(^{26}\)

\[
   u^h_i[I(i)] \equiv 1 + \ln[1 - I(i)a^i] + V^h_i[I(i)] + \epsilon^i_h
\]

where \( V^h_i[I(i)] = 1 + I(i)\phi_h \) represents the deterministic component of utility at origin in the second period.

In the second period of their lives, individuals self-select into migration; if an individual \( i \) self-selects into migration, he faces a probability \( p_{I(i)} \), which is set by the country of destination, to be admitted at destination. The probability can vary depending on whether the individual \( i \) is educated, and we assume that \( p_1 = p_0 + \pi \), with \( p_0 \in (0, 1] \) and \( \pi \in [0, 1 - p_0] \). The probability \( p_0 \) can be interpreted as reflecting the options to migrate through non-

\(^{23}\)We do not explicitly consider either inter- or intra-generational externalities connected to education that generally motivate the concerns about the effects of international migration in the brain drain literature.

\(^{24}\)Allowing for migration in the first period, at least for uneducated individuals, would require to adopt a dynamic discrete choice models of migration, where agents could revise their location decisions in the second period; this represents an interesting theoretical extension, that goes beyond the scope of this paper.

\(^{25}\)The introduction of heterogeneity in the utility at origin in the first period would not influence education decisions provided that the realizations of the stochastic component of utility are not serially correlated, an assumption that is introduced in dynamic discrete choice models of migration (Kennan and Walker, 2011; Bertoli et al., 2013), as in this case the realization of the first period realization does not convey any information on the future preference for migration.

\(^{26}\)We are normalizing the first-period utility of an individual who does not invest in education to one.
selective channels, e.g., family reunification provisions or undocumented migration. Notice that we have assumed that \( p_0 \) is bounded away from zero, as the country of destination is unable to completely close its borders, so that the baseline probability to migrate \( p_0 \) cannot be indefinitely compressed. The destination country retains control over \( \pi \), which represents the additional chances to be admitted at destination that immigration policies can give to educated applicants.

The utility associated to a domestic employment in the first period and to a foreign employment in the second period is equal to:

\[
 u_d^i[I(i)] \equiv 1 + \ln[1 - I(i)a_i] + V_d[I(i), n] + \epsilon_d^i \tag{3}
\]

The deterministic component of utility at destination \( V_d[I(i), n] \) depends on the wages that a migrant earns at destination and on migration costs, i.e., higher migration costs reduce \( V_d[I(i), n] \). Migration costs are assumed to be a function of the size of migration networks \( n \). We assume that migrants enjoy a non-negative return to education at destination, i.e., \( V_d[1, n] \geq V_d[0, n] \), but we do not introduce assumptions on the sign of the difference between the return to education at home and at destination, which might depend on \( n \). Even if the gross education premium for natives is lower at destination than at origin, this does not pin down the sign of the difference in the net return to education for immigrants, as (i) immigrants might enjoy a different gross education premium than natives (Borjas, 1987), and (ii) time-equivalent migration costs can decrease with education (Chiswick, 1999; Chiquiar and Hanson, 2005; McKenzie and Rapoport, 2010). Furthermore, we also assume that an expansion in the size of migration networks can produce a larger impact on the utility of uneducated than of educated migrants:

\footnote{The size of migration networks \( n \) can be thought of as a function of the scale of past bilateral migration flows; more broadly, \( n \) could also reflect the factors, such as the thickness of information flows between the destination and the origin country, that influence the size of migration costs.}

\footnote{An alternative modeling strategy would have been to assume two separate education-specific networks of size \( n_0 \) and \( n_1 \), with the deterministic component of utility at destination for uneducated (educated) individuals influenced only by the size of \( n_0 \) (\( n_1 \)); this alternative assumption would imply that networks contribute to preserve immigrants’ quality unchanged over time, as the network of educated migrants reduces migration costs only for educated individuals.}

\footnote{Our model encompasses the limiting case where migrants’ utility at destination does not increase with education, i.e., \( V_d[1, n] = V_d[0, n] \), so that educated migrants are exposed to a brain waste (Mattoo et al., 2008).}
\[
\frac{\partial V_d[0,n]}{\partial n} \geq \frac{\partial V_d[1,n]}{\partial n} \quad (4)
\]

The model is solved by backward induction, analyzing first the decision to self-select into migration, conditional upon the education choice that has been made. We then derive the expression for expected utility, which depends on the probabilities of self-selection into migration and of admission at destination, and solve for the utility-maximizing education choice.

### 3.2 Self-selection into migration

An individual \(i\) will self-select into migration if and only if:

\[
V_d[I(i),n] + \epsilon_i^d > V_h[I(i)] + \epsilon_i^h \quad (5)
\]

The probability of self-selection into migration \(q_j\), with \(j = 0, 1\), can be derived by specifying the distributional assumptions on the stochastic component of utility and the information upon which the self-selection decision is based; if we assume that \(\epsilon\) follows an identically and independently distributed Extreme Value Type-1 distribution and that the realizations of both \(\epsilon_i^h\) are \(\epsilon_i^d\) are observed as in Borjas (1987),\(^{30}\) then following McFadden (1974) we can express the probability of self-selection into migration as a function of the deterministic component of utility in the two countries:\(^{31}\)

\[
q_{I(i)} = \frac{1}{1 + e^{V_h[I(i)]-V_d[I(i),n]}} \quad (6)
\]

We assume that education decisions are taken before observing the realizations of the stochastic component of utility, so that each individual anticipates that he will self-select into migration in period 2 with a different probability \(q_{I(i)} \in (0,1)\) depending on his own

---

\(^{30}\)Alternatively, we could assume that the stochastic components of utility are only locally observable, as in Bertoli (2010), and hence the individual \(i\) decides whether to migrate before having observed the realization of \(\epsilon_i^d\); this different information structure would generate the same self-selection probability \(q_{I(i)}\) if we assumed that \(\epsilon\) follows an identically and independently distributed logistic distribution.

\(^{31}\)Notice that the self-selection probabilities in (6) are not influenced by the anticipation of the probability of not being admitted at destination; this would no longer be the case, as in Bianchi (2013), if we introduced a fee of applying for migration.
education decisions in the first period of his life.\textsuperscript{32} We say that a pattern of positive migrants’ self-selection in education occurs if \( q_1 > q_0 \), which occurs if and only if \( V_h(1) - V_d(1, n) < V_h(0) - V_d(0, n) \).

### 3.3 Expected utility

The expected utility corresponding to the two possible education choices \( I(i) = 0, 1 \) can be written as:\textsuperscript{33}

\[
E[u^i[I(i)]] = 1 + \ln [1 - I(i)a^i] + p_{I(i)} \ln \left( e^{V_h[I(i)]} + e^{V_d[I(i), n]} \right) + (1 - p_{I(i)}) V_h[I(i)] + \gamma
\]  

(7)

The expected utility of the individual \( i \) is a linear combination of the expected utility from domestic employment plus the expected utility \( \ln \left( e^{V_h[I(i)]} + e^{V_d[I(i), n]} \right) \) from an unconstrained choice situation (Small and Rosen, 1981), with the weight of the latter term given by the probability to be admitted at destination conditional upon self-selecting into migration.

If we take the partial derivative of (7) with respect to the size of migration networks \( n \), we obtain:

\[
\frac{\partial E[u^i[I(i)]]}{\partial n} = p_{I(i)} q_{I(i)} \frac{\partial V_d[I(i), n]}{\partial n} > 0
\]

The size of the effect depends on (i) the unconditional probability of migration \( p_{I(i)} q_{I(i)} \), and on (ii) the derivative of the deterministic component of utility at destination with respect to networks. Notice that the assumption we introduce on the size of this derivative for educated and for uneducated migrants in (4) does \textit{not} suffice to say whether a marginal increase in networks increases expected utility more for uneducated or for educated individuals, as:

\[
\frac{\partial E[u^i(1)]}{\partial n} = (p_0 + \pi) q_1 \frac{\partial V_d(1, n)}{\partial n} \leq p_0 q_0 \frac{\partial V_d(0, n)}{\partial n} = \frac{\partial E[u^i(0)]}{\partial n}
\]

(8)

with the vector of policy parameters \( \mathbf{p} = (p_0, \pi)' \) determining the direction of the inequality in (8).

\textsuperscript{32}Notice that \( q_{I(i)} \) never attains the two extremes of the interval when the differential between the deterministic component of utility in the two countries is finite.

\textsuperscript{33}See the Appendix A.1 for the derivation.
3.4 Optimal education decisions

An individual $i$ will invest in education if and only if:

$$E[u^i(1)] > E[u^i(0)]$$

which, using (7), can be rewritten as:

$$\ln(1 - a^i) + p_1 \ln \left( e^{V_h(1)} + e^{V_d(1,n)} \right) + (1 - p_1)V_h(1) > p_0 \ln \left( e + e^{V_d(0,n)} \right) + (1 - p_0)$$

This inequality implicitly defines a threshold level of innate learning ability $a(n, p|\theta)$, where $\theta$ is a vector that contains all the other parameters of our model, which separates the individuals with $a^i \in [0, a(n, p|\theta))$ who find optimal to invest in education from the individuals with $a^i \in [a(n, p|\theta), 1]$ whose utility-maximizing choice is not to invest in education.\textsuperscript{35}

With some simple algebraic manipulations on (9), we get that:

$$a(n, p|\theta) = 1 - e^{-\phi_h(1 - q_1)^{p_0 + \pi}} \frac{(1 - q_0)^{p_0}}{p_0}$$

This expression evidences the direct relationship between the share of the population at origin that invests in education and (i) the self-selection probabilities $q_1$ and $q_0$,\textsuperscript{36} and (ii) the immigration policies $p$ adopted by the country of destination.

4 The scale of migration and migrants’ quality

The derivation of the optimal education decision and of the probability of self-selection into migration allows us to study the influence exerted by immigration policies $p$ and the size $n$ of migration networks on migrants’ quality. If we normalize the size of population at origin to one, we can define the scale of migration $f(n, p|\theta)$ simply as:

\textsuperscript{34}Recall that $V_h(0)$ is normalized to unity.

\textsuperscript{35}We are abstracting here from the possibility that private education costs might be endogenous with respect to the prospect to migrate (Docquier et al., 2008; Bertoli and Brücker, 2011).

\textsuperscript{36}The key role played in the analysis by the prevailing pattern of self-selection on education, which contributes to determine the sign of the partial derivative of $a(n, p|\theta)$ with respect to $n$, motivates our choice to write the threshold value of ability in (10) as a function of $q_1$ and $q_0$ rather than in an equivalent form that depends on the deterministic components of utility.
\[ f(n, \mathbf{p}|\theta) = (p_0 + \pi)q_1F[a(n, \mathbf{p}|\theta)] + p_0q_0(1 - F[a(n, \mathbf{p}|\theta)]) \]

where \( F(a) \) represents the cumulative density function of innate learning ability. We can define migrants’ quality as an increasing function of the ratio between the number of educated and of uneducated migrants; following the relevant empirical literature and the suggestive evidence that we provided in Section 2, we define migrants’ quality \( g(n, \mathbf{p}|\theta) \) as the logarithm of this ratio:

\[ g(n, \mathbf{p}|\theta) = \ln \left( \frac{(p_0 + \pi)q_1F[a(n, \mathbf{p}|\theta)]}{p_0q_0(1 - F[a(n, \mathbf{p}|\theta)])} \right) \tag{11} \]

It is straightforward to observe that \( g(n, \mathbf{p}|\theta) \) is an increasing function of \( \pi \), as:

\[ \frac{\partial a(n, \mathbf{p}|\theta)}{\partial \pi} = [a(n, \mathbf{p}|\theta) - 1] \ln(1 - q_1) > 0 \tag{12} \]

When the probabilities of self-selection into migration are unchanged, an increase in \( \pi \) unambiguously increases migrants’ quality.\(^{37}\) Our key interest is to understand whether this positive static effect of selective immigration policies can be preserved over time, when a variation in the size of migration networks \( n \) induces a change in decision to self-select into migration.

### 4.1 Networks and migrants’ quality

A variation in the size of migration networks \( n \) influences migrants’ quality, as it changes both the incentives to self-select into migration, i.e., \( q_1 \) and \( q_0 \), and optimal education decisions, i.e., \( a(n, \mathbf{p}|\theta) \), at origin. If we derive (11) with respect to \( n \), we have that:

\[ \frac{\partial g(n, \mathbf{p}|\theta)}{n} = \frac{\partial[\ln(q_1) - \ln(q_0)]}{\partial n} + \frac{f[a(n, \mathbf{p}|\theta)]}{F[a(n, \mathbf{p}|\theta)](1 - F[a(n, \mathbf{p}|\theta)])} \frac{\partial a(n, \mathbf{p}|\theta)}{\partial n} \geq 0 \tag{13} \]

where \( f(a) \) represents the density function of learning ability \( a \). The first term on the right hand side of (13) describes the impact of a marginal variation in \( n \) on migrants’ quality that goes through migrants’ self-selection, while the second term captures the effect that

\(^{37}\)It is also straightforward to observe that (11) monotonically increases with \( a(n, \mathbf{p}|\theta) \), so that an increase in the share of educated individuals at origin improves migrants’ quality.

\(^{38}\)See the Appendix A.2 for a derivation of the partial derivative of \( a(n, \mathbf{p}|\theta) \) with respect to \( p_0 \).
goes through the induced change in education decisions at origin. We have the following Proposition:

**Proposition 1** If migrants’ are positively self-selected on education, then an increase in the size of migration networks reduces migrants’ quality if education decisions at origin are exogenous.

**Proof.** From (6), we have that:

\[
\frac{\partial [\ln(q_1) - \ln(q_0)]}{\partial n} = (1 - q_1) \frac{\partial V_d(1, n)}{\partial n} - (1 - q_0) \frac{\partial V_d(0, n)}{\partial n}
\]

Given (4), then:

\[
\frac{\partial [\ln(q_1) - \ln(q_0)]}{\partial n} \leq (q_0 - q_1) \frac{\partial V_d(0, n)}{\partial n}
\]

with the right hand side of the inequality above being negative when \(q_1 > q_0\).

Proposition 1 gives us a sufficient condition to sign the first term on the right hand side of (13) if we assume—following the entire literature on immigrants’ self-selection—that the distribution of education is exogenous.\(^{39}\) In such a case, if migrants are positively self-selected on education to begin with,\(^{40}\) then an expansion of migration networks \(n\) reduces the extent of positive self-selection on education.\(^{41}\)

An immediate corollary of Proposition 1 is that, when migrants are positively self-selected on education, an expansion in migration networks can lead to an improvement in migrants’

\(^{39}\)Proposition 1 also applies when an expansion in the size of migration networks produces an identical impact on the deterministic component of utility for educated and uneducated individuals, i.e., (4) holds as an equality.

\(^{40}\)Notice that the aggregate migration data by Docquier et al. (2009) reveal that international migrants are (almost) invariably positively selected in education, i.e. \(p_1 q_1 > p_0 q_0\); as most destination countries covered in their dataset do not adopt selective immigration policies, i.e. \(p_1 \approx p_0\), then this suggests that migrants are positively self-selected on education; this pattern might also be the result of the presence of liquidity constraints that prevent low-educated individuals from migrating (Belot and Hatton, 2012).

\(^{41}\)The same result holds in McKenzie and Rapoport (2010), but there is a key difference between the two models: McKenzie and Rapoport (2010) do not consider a stochastic component of utility, so that the probability of self-selection into migration \(q\) is either 0 or 1 and an expansion of migration networks \(n\) as no impact on the choices of all individuals for whom \(q\) was already equal to 1; in our model, \(q \in (0, 1)\), so that a marginal increase in \(n\) influences the location decisions of all individuals.
quality only if the increase in \( n \) induces an increase in \( a(n, p|\theta) \). As shown by (8), the influence of an increase in \( n \) on the expected return to the investment in education crucially depends on the vector \( p \) that describes the immigration policies adopted by the country of destination. Specifically, we have that the expected return to education is an increasing function of \( n \) if and only if:

\[
\pi > p_0 \left[ \frac{q_0}{q_1} k(n|\theta) - 1 \right]
\]

where \( k(n|\theta) \), which is no lower than 1 because of (4), is defined as:

\[
k(n|\theta) = \frac{\partial V_a(0, n)}{\partial n} / \frac{\partial V_a(1, n)}{\partial n}
\]

If (14) is not satisfied, then an expansion in the size of migration networks \( n \) unambiguously reduces migrants’ quality when migrants are positively self-selected on education, as the negative effect on quality due to the change in the pattern of self-selection described by Proposition 1 is also matched by a reduction in the share of educated individuals at origin. This, in turn, implies that endogenizing education decisions at origin is, per se, not sufficient to alter the prediction of a negative relationship between networks size and migrants’ quality contained in Proposition 1. Why does the initial pattern of self-selection on education contribute to determine whether an expansion in the size of networks improves or worsens the incentives to invest in education in the first period? This occurs because the increase in \( n \) raises utility at destination, and the extent to which this translates into an increase in expected utility depends on the education-specific probabilities \( q_1 \) and \( q_0 \) that migration represents the utility-maximizing alternative. When \( q_1 > q_0 \), this counteracts, and possibly overturns, the stronger effect exerted by the expansion in the size of networks on the utility at destination of uneducated individuals.

We thus have that (14) represents a necessary but not sufficient condition for having a positive relationship between the size of migration networks and migrants’ quality. When (14) holds, the strength of the positive effect on migrants’ quality due to induced increase in the share of the population at origin that invests in education depends on the density \( f(a) \) and on the cumulative density \( F(a) \) of innate learning ability in correspondence to \( a(n, p|\theta) \), as shown by (13). In what follows, we will introduce for analytical convenience the assumption that ability is uniformly distributed over the unit interval, i.e., \( f(a) = 1 \) and \( F(a) = a \). The Appendix A.3 shows that (13) is actually higher under more general distributional
assumptions, that allow for the density $f(a)$ to be either monotonically decreasing with $a$ or non-monotonic and concave in $a$.

### 4.2 Can selective immigration policies preserve migrants’ quality?

Let us introduce the assumption that innate learning ability $a$ is uniformly distributed over the unit interval; in this case, we can simplify (13), and we have that a marginal increase in the size of migration networks $n$ improves migrants’ quality if and only if:

$$\frac{\partial[\ln(q_1) - \ln(q_0)]}{\partial n} a(n, p|\theta)[1 - a(n, p|\theta)] + \frac{\partial a(n, p|\theta)}{\partial n} \geq 0$$

Using the fact that:

$$\frac{\partial a(n, p|\theta)}{\partial n} = [1 - a(n, p|\theta)] \left[(p_0 + \pi)q_1 \frac{\partial V_d(1, n)}{\partial n} - p_0q_0 \frac{\partial V_d(0, n)}{\partial n}\right]$$

and:

$$\frac{\partial[\ln(q_1) - \ln(q_0)]}{\partial n} = (1 - q_1) \frac{\partial V_d(1, n)}{\partial n} - (1 - q_0) \frac{\partial V_d(0, n)}{\partial n}$$

the inequality above can be further simplified as follows:

$$k(n|\theta) \leq \frac{a(n, p|\theta) + [p_0 + \pi - a(n, p|\theta)]q_1}{a(n, p|\theta) + [p_0 - a(n, p|\theta)]q_0} \equiv t(\pi, n|p_0, \theta)$$

(15)

An increase in the size of migration networks improves migrants’ quality if and only if (15) holds. The left hand side of (15) is given by the ratio $k(n|\theta)$ between the partial derivatives of the deterministic components of utility $V_d[1, n]$ and $V_d[0, n]$ with respect to $n$, with $k(n|\theta) \geq 1$. The right hand side of (15) is a function of $\pi$, the differential in the probability to migrate between educated and uneducated agents.\(^{42}\)

We can now derive the main prediction of our theoretical model, which is contained in the following Proposition:

**Proposition 2** Migrants’ quality increases with the size of migration networks when $\pi = 1 - p_0$ and $q_1 \geq q_0$, provided that $k(n|\theta)$ is sufficiently low.

\(^{42}\)A marginal increase in $\pi$ influences the right hand side of (15) both directly, and through the increase it induces in the share $a(n, p|\theta)$ of the population at origin that invests in education, with the sign of the partial derivative of the right hand side of (15) being ambiguous.
Proof. When $q_1 \geq q_0$, then $t(1 - p_0, n|p_0, \theta) > 1$. This entails that there are values of $k(n|\theta) \geq 1$ such that (15) holds.

Figure 3 provides a graphical representation of Proposition 2: the area in dark gray identifies all admissible combinations of $k(n|\theta)$ and $\pi$ that give rise to a positive relationship between the size of migration networks and migrants’ quality, as $k(n|\theta) < t(\pi, n|p_0, \theta)$. Proposition 2 ensures that, when $q_1 \geq q_0$, this area is always non-empty, as $t(\pi, n|p_0, \theta) > 1$ when $\pi = 1 - p_0$, i.e., there are no restrictions on the migration of educated individuals. We also know that the graph of the function $t(\pi)$ lies above the graph of (14), which represents a necessary but not sufficient condition to have a positive relationship between migrants’ quality and $n$. As demonstrated in the Appendix A.3, this result does not hinge on the assumption that ability $a$ is uniformly distributed, as assuming that $f(a)$ is a (i) linear, (ii) triangular or (iii) a quadratic and concave function of $a$, so that $f(a)$ can also be non-monotonic in $a$, would actually produce a larger set of combinations of $k(n|\theta)$ and $\pi$ for which migrants’ quality increases with $n$.\textsuperscript{43}

The evidence presented in Section 2 on the absence of a negative relationship between migrants’ quality and the size of migration networks in countries that adopt selective immigration policies suggests that Proposition 2 is based on hypotheses that are plausible from an empirical perspective. Specifically, if we were to use Proposition 2 to interpret the estimates in Section 2, then we could say that non-selective destinations, such as the US or France, have immigration policies that are characterized by a differential in the probability of admission for educated and uneducated applicants such that $k(n|\theta) \geq t(\pi, n|p_0, \theta)$, so that an expansion of networks invariably reduces the quality of the immigrants that they receive. Australia and Canada, on the other hand, provide a reward to education $\pi$ in terms of higher chances of admission such that $k(n|\theta) < t(\pi, n|p_0, \theta)$ for a large number of origin countries, so that there is no systematic relationship between the size of migration networks and the quality of the migrants. An increase in $\pi$ is not neutral with respect to the scale of incoming migration flows, and hence the evolution of the size of migration networks $n$; this, in turn, entails that destination countries could be implicitly facing a trade-off between

\textsuperscript{43}Our definition of migrants’ quality is independent from their level of innate learning ability $a$, which exerts no influence on productivity and wages, as in Beine et al. (2008); the average level of ability of the immigrants needs not to be a monotonically decreasing function of the size of migration networks, even when a destination country adopts non-selective immigration policies, i.e., $\pi = 0$. 

21
the quantity $f(n, p|\theta)$ and the evolution of quality $g(n, p|\theta)$ over time when choosing how much to reward the education of the applicants in terms of better chances of admission at destination. This observation is consistent with the fact that Australia and Canada have among the highest shares of foreign-born in the total population among OECD countries.

5 Concluding remarks

The model proposes a possible rationale to explain why, as suggested by the empirical evidence provided that motivates our paper, quality-selective immigration policies can be dynamically effective and neutralize the otherwise adverse effect of migration networks on migrants’ self-selection. The central prediction of our model is that migration networks and immigrants’ quality can be positively associated under a set of conditions regarding the degree of selectivity of immigration policies, the prevailing pattern of migrants’ self-selection on education, and the way time-equivalent migration costs by education level relate to networks. Our main testable implication is that the relationship between network size and immigrants’ quality should vary with the type of immigration policy (selective versus non-selective) at
destination. Bringing the model to the data is currently out of reach due to binding data constraints; in particular, the stringency and selectivity dimensions of immigration policies are very imperfectly captured in existing datasets, and data on gross bilateral migration flows by education are hardly available.
References


BERTOLI, S., H. BRÜCKER, G. FACCHINI, A. MAYDA, AND G. PERI (2012): “Understanding Highly Skilled Migration in Developed Countries: The Upcoming Battle for Brains,” in Brain Drain and Brain Gain: The Global Competition to Attract High-Skilled Migrants,


A Derivation of some theoretical results

A.1 Expected utility

Expected utility depends both on the probability of self-selection \( q_{I(i)} \) and on the probability \( p_{I(i)} \) to be admitted at destination. Formally, we have that:

\[
E \left[ u^i[I(i)] \right] = p_{I(i)} E \left[ V_d[I(i), n] + \epsilon_d^i \mid \epsilon_d^i - \epsilon_h^i > V_h[I(i)] - V_d[I(i), n] \right] + \nonumber \]

\[
(1 - q_{I(i)}) E \left[ V_h[I(i)] + \epsilon_h^i \mid \epsilon_d^i - \epsilon_h^i \leq V_h[I(i)] - V_d[I(i), n] \right] + 
\]

\[
q_{I(i)}(1 - p_{I(i)}) E \left[ V_h[I(i)] + \epsilon_h^i \mid \epsilon_d^i - \epsilon_h^i > V_h[I(i)] - V_d[I(i), n] \right] \quad (A.1)
\]

Here, we can exploit a key result from the literature on discrete choice models to simplify (A.1): when the stochastic component of utility is i.i.d. EVT-1, then the expected utility from choosing any of the possible alternatives, conditional upon the fact that the chosen alternative is a utility-maximizing one, does not vary across alternatives (de Palma and Kilani, 2007). Specifically, this implies that the first two expected conditional utilities on the right hand side of (A.1) coincide. The expected utility from an unconstrained choice situation is equal to the Euler’s constant \( \gamma \) plus the logarithm of the sum of the exponentiated values of the deterministic component of utility in the various alternatives (Small and Rosen, 1981). In our case, this implies that:

\[
E \left[ V_d[I(i), n] + \epsilon_d^i \mid \epsilon_d^i - \epsilon_h^i > V_h[I(i)] - V_d[I(i), n] \right] = 
E \left[ V_h[I(i)] + \epsilon_h^i \mid \epsilon_d^i - \epsilon_h^i \leq V_h[I(i)] - V_d[I(i), n] \right] = 
1 + \ln[1 - I(i)\alpha] + \ln (e^{V_h[I(i)]} + e^{V_d[I(i), n]}) + \gamma
\]

These two results entail that we know the value of the first two terms on the right hand side of (A.1); if the decision to migrate is not subject to the restrictions imposed by immigration policies, then the expected utility of migrants coincides with the expected utility of stayers, even if the two countries are characterized by different wages. The size of the wage differential influences the probability of self-selecting into migration, but the utility-maximizing location decisions imply that stayers and migrants enjoy the same expected level of utility.

The third term corresponds to the expected utility of the individuals who could not opt for their utility-maximizing location, as they self-selected into migration but did not get

\[44\text{Theorem 2.4 in Cardell (1997) already established the invariance of the distribution of conditional utility when the choice set includes two alternatives, as in our model.}\]
admitted at destination. We can notice that the (unconditional) utility from staying at home in the second period can be expressed as:

\[ V_h[I(i)] = q_{I(i)}E \left[ u^i_{h2}[I(i)]|u^i_{a2}[I(i)] > u^i_{h2}[I(i)] \right] + \\
(1 - q_{I(i)})E \left[ u^i_{h2}[I(i)]|u^i_{a2}[I(i)] \leq u^i_{h2}[I(i)] \right] \]

This implies that:

\[ E \left[ V_h[I(i)] + \epsilon^i_h \mid \epsilon^i_d - \epsilon^i_h > V_h[I(i)] - V_d[I(i), n] \right] = \\
\frac{V_h[I(i)]}{q_{I(i)}} - \frac{1 - q_{I(i)}}{q_{I(i)}} \left[ \ln \left( e^{V_h[I(i)]} + e^{V_d[I(i), n]} \right) + \gamma \right] \]

This eventually allows us to rewrite expected utility as:

\[ E \left[ u^i[I(i)] \right] = 1 \ln(1 - I(i)a^i) + \\
p_{I(i)} \ln \left( e^{V_h[I(i)]} + e^{V_d[I(i), n]} \right) + (1 - p_{I(i)})V_h[I(i)] + \gamma \]

**A.2 The relationship between \( a(n, p|\theta) \) and \( p_0 \)**

An increase in \( p_0 \) exerts an ambiguous influence on the threshold value of ability. Taking the partial derivative of (10) with respect to \( p_0 \), we get:

\[ \frac{\partial a(n, p|\theta)}{\partial p_0} = [1 - a(n, p|\theta)] \ln \left( \frac{1 - q_0}{1 - q_1} \right) \leq 0 \quad (A.2) \]

The sign of this inequality depends on the pattern of migrants’ self-selection on education: specifically, we have that (A.2) is positive if and only if \( q_1 > q_0 \). This ambiguity in the sign of (A.2) is due to the fact that a marginal increase in \( p_0 \) also increases \( p_1 = p_0 + \pi \); if we consider an increase in \( p_0 \) that is matched by a simultaneous identical reduction in \( \pi \), so that \( p_1 \) remains unchanged, then the share of the population that invests in education unambiguously declines, as from (A.2) and (12) we have that:

\[ \frac{\partial a(n, p|\theta)}{\partial p_0} - \frac{\partial a(n, p|\theta)}{\partial \pi} = [1 - a(n, p|\theta)] \ln(1 - q_0) < 0 \]

**A.3 Alternative distributional assumptions on \( a \)**

Let us rewrite here the partial derivative of migrants’ quality with respect to \( n \) in (13):
$$\frac{\partial g(n, p|\theta)}{n} = \frac{\partial[\ln(q_1) - \ln(q_0)]}{\partial n} + \frac{f[a(n, p|\theta)]}{F[a(n, p|\theta)](1 - F[a(n, p|\theta)])} \frac{\partial a(n, p|\theta)}{\partial n} \geq 0$$

Any density function \( f(a) \) that satisfies:

$$\forall a \in [0, 1]: f(a) \geq \frac{F(a)[1 - F(a)]}{a - a^2} \quad (A.3)$$

gives rise to a higher value of (13), when the necessary condition (14) holds, than under a uniform distribution of ability. It can be easily proved analytically that any density function that is linear in \( a \) satisfies (A.3). Linear density functions take the form \( f(a) = \beta a + \eta \), with \( \beta \in [-2, 2] \) and \( \eta = 1 - \beta/2 \), with the two conditions on \( \beta \) and \( \eta \) ensuring that \( f(a) \) does represent a density function, i.e., \( f(a) \geq 0 \) and \( F(1) = 1 \). When \( f(a) = \beta a + \eta \), (A.3) can be rewritten as:

$$\beta a + \eta \geq \frac{\beta^2 a^2 + \eta a}{2} \left( 1 - \frac{\beta^2 a^2 - \eta a}{2} \right)$$

$$\beta^2 a^3 - \beta(1 - \eta)a^2 + \left( \frac{\beta}{2} - \eta + \eta^2 \right) a \geq 0 \quad (A.4)$$

Multiplying both sides of (A.4) by \( a(1 - a) \), developing the products and moving terms around we get:

$$\frac{\beta^2}{4} a^3 - \beta(1 - \eta)a^2 + \left( \frac{\beta}{2} - \eta + \eta^2 \right) a \geq 0$$

Using the restriction that \( \eta = 1 - \beta/2 \), with some simple algebraic manipulations we get:

$$\frac{\beta^2}{4} a(a - 1)^2 \geq 0$$

a condition that is always satisfied for any \( a \in [0, 1] \).

We can also consider triangular density functions, where \( f(a) \) first increases and then decreases linearly with \( a \), which take the form:

$$f(a) = \begin{cases} \beta a + \eta, & a \in [0, 1/2] \\ \beta(1 - a) + \eta, & a \in (1/2, 1] \end{cases}$$

with \( \beta \in [0, 4] \) and \( \eta = 1 - \beta/4 \), with the two conditions ensuring also in this case that \( f(a) \geq 0 \) and \( F(1) = 1 \). When \( a \in [0, 1/2] \), we can substitute \( \eta = (1 - \beta/4) \) in (A.5), and with some simple algebraic manipulations we get:\footnote{We are assuming that \( \beta > 0 \); if \( \beta = 0 \), then \( f(a) \) is simply the uniform distribution.}

\[45\]
\[ \frac{\beta^2}{4} a \left[ \left( a - \frac{1}{2} \right)^2 + \frac{1}{\beta} \right] \geq 0 \]

which clearly holds as \( \beta > 0 \). As both sides of (A.3) are symmetric around \( a = 1/2 \) when \( f(a) \) is a triangular function of \( a \), then we can conclude that (A.3) holds for \( a \in [0,1] \). Similarly, numerical simulations reveal that any quadratic and concave density function \( f(a) \) also satisfies (A.3) for any threshold value of ability.\(^{46}\) Figure A.1 plots the two sides of (A.3) for some examples of quadratic density functions.

Figure A.1: Graphical representation of (A.3) for quadratic density functions

Notes: the three solid and the three dashed lines represent respectively the left and the right hand side of (A.3), which have been drawn for \( \alpha = -3, \beta = 2, 3, 4 \) and \( \eta = 1 - \alpha/3 - \beta/2 \).

Hence, any density function that is linear, triangular or quadratic and concave in \( a \) leads to a value of (13), when the necessary condition (14) holds, that is larger than under a uniform distribution for \( a \), independently of the threshold value of ability.

\(^{46}\)Quadratic and concave density function take the form \( f(a) = \alpha a^2 + \beta a + \eta \), with \( \alpha \in [0,-6], \beta \in [-(4/3)\alpha - 2, -(2/3)\alpha + 2] \) and \( \eta = 1 - \alpha/3 - \beta/2 \), with the conditions on the three parameters \( \beta \) and \( \eta \) ensuring that \( f(a) \geq 0 \) and \( F(1) = 1 \).