Dipartimento di Scienze Economiche
Università degli Studi di Firenze

Working Paper Series

# The informational structure of migration decision and migrants' self-selection

Simone Bertoli

Working Paper N. 03/2010
March 2010

Dipartimento di Scienze Economiche, Università degli Studi di Firenze
Via delle Pandette 9, 50127 Firenze, Italia
www.dse.unifi.it

The findings, interpretations, and conclusions expressed in the working paper series are those of the authors alone. They do not represent the view of Dipartimento di Scienze Economiche, Università degli Studi di Firenze

Stampato in proprio in Firenze dal Dipartimento Scienze Economiche
(Via delle Pandette 9, 50127 Firenze) nel mese di March 2010,
Esemplare Fuori Commercio Per il Deposito Legale
agli effetti della Legge 15 Aprile 2004, N.106

# The informational structure of migration decision and migrants' self-selection[1]

**Simone Bertoli**

IAB, Institute for Employment Research
Weddigenstr. 20-22
90478, Nuremberg, Germany
e-mail: sb3151@gmail.com

**Abstract.** This paper derives the implications for migrants' self-selection in unobservables that arise from the introduction of uncertainty in the decision problem that would-be migrants face. We show that if one lifts the assumption introduced in Borjas (1987) that foreign wages are known before the migration decision is taken, then the case for the so-called refugee sorting narrows down considerably, while negative selection becomes a more likely outcome. A greater dispersion of income at destination no longer suffices to predict that immigrants will obtain a higher average income than natives.

**Keywords**: migration; uncertainty; information; self-selection.
**JEL codes**: D84, F22, O15.

[1] *The author is grateful to an anonymous referee for the comments provided on the paper, and to Timo Baas, Ennio Bilancini, Nicola Coniglio, Marco Dardi and Andreas Hauptman for their remarks on earlier drafts; the usual disclaimers apply.*

## Introduction

The paper by Borjas (1987), which is based on Roy (1951), represents a seminal contribution in the literature on migrants' self-selection, and on their performance on the labor market at destination. Borjas (1987) identifies the relative dispersion of the earnings distributions in the sending and the receiving countries, and the correlation between the two, as the leading factors that shape the pattern of migrants' self-selection. A number of theoretical and empirical contributions has been produced to refine or test his theoretical predictions (e.g. Jasso and Rosenzweig, 1990; Borjas, 1995; Chiswick, 1999; Chiquiar and Hanson, 2005; McKenzie and Rapoport, forthcoming, Fernández-Huertas Moraga, forthcoming).

Interestingly, while Borjas (1987) assumes that domestic and foreign wages include a stochastic term, some of the most recent theoretical contributions in the literature, such as Chiquiar and Hanson (2005) and McKenzie and Rapoport (forthcoming), propose models where wages are described by deterministic functions.[2] Still, the difference between the two approaches is narrower than it appears, as the model by Borjas (1987) rules out uncertainty from the decision to migrate: an agent decides whether to migrate after he has observed the realizations of the stochastic component of both domestic and foreign wages. This informational structure portraits migration as a risk-free decision, and it entails that agents behave as if they were income maximizers.

This paper assesses the consequences of a slight modification in the informational structure of the decision problem that would-be migrants face, adopting the analytical framework originally proposed by Borjas (1987). Specifically, we introduce uncertainty in the migration decision, assuming that an agent decides whether to migrate after he has observed the realization of the domestic wage, but not that of the foreign one. This assumption reflects the idea that would-be migrants have a better knowledge of their domestic wage, though we maintain that the distributions of the two stochastic components are known to the economic agents. We show that the sets of model parameters that support the occurrence of the so-called refugee sorting and of negative migrants' self-selection change substantially, and that a greater dispersion of incomes at destination no longer suffices to predict that migrants will outperform natives in terms of income.

We first replicate the analysis in Borjas (1987), and then analyze the model under the alternative informational structure that we proposed.

[2] Chiquiar and Hanson (2005) argue that "implicitly, we imagine that there are random components to wage determination, but for simplicity we leave such features in the background" (p. 242).

**The self-selection model**

Let the subscript 0 denote the country of origin, and the subscript 1 denote the country of destination of the migrants. Domestic and foreign wages for the countries follow a log-normal distribution:

$$w_0 = (1+\pi)\, e^{\mu_0+\varepsilon_0}, \text{ with } \varepsilon_0 : \ N\left(0, \sigma_0^2\right) \qquad [1]$$

$$w_1 = e^{\mu_1+\varepsilon_1}, \text{ with } \varepsilon_1 : \ N\left(0, \sigma_1^2\right) \qquad [2]$$

The parameters $\mu_0$ and $\mu_1$ capture the influence of observable factors upon wages in the two countries, while the stochastic variables $\varepsilon_0$ and $\varepsilon_1$ describe the influence of unobserved characteristics upon wages. The correlation between the two distributions is represented by $\rho$, while the parameter $\pi>0$ describes time-equivalent migration costs.

*Migration with no uncertainty*

Borjas (1987) assumes that an agent decides whether to migrate after he has observed his realizations of the stochastic variables $\varepsilon_0$ and $\varepsilon_1$, so that he can compare his actual wage in country 0 with what he would earn in country 1. The decision to migrate can be characterized by the sign of the index function I:

$$I(\varepsilon_1, \varepsilon_0) = \ln \frac{w_1}{w_0} ; \ (\mu_1 - \mu_0 - \pi) + (\varepsilon_1 - \varepsilon_0) \qquad [3]$$

with migration occurring if $I>0$. Although Borjas (1987) portraits his model as describing the implications for migrants' self-selection that arise when "potential migrants are income maximizers", this is an unnecessary restriction to the scope of his model: the informational structure that underlies Eq. [3] leaves no role for uncertainty, so that agents behave as if they were income maximizers, independently of the possible introduction of a utility function which is non-linear in income. An agent would always simply maximize the argument of his utility function, as the decision to migrate boils down to the choice of the country which offers the highest wage. Observe that Eq. [3] is monotonically decreasing in $\varepsilon_0$, so that poorer realizations of the domestic stochastic component increase the probability that an agent will migrate.

The interest of the model resides in understanding how the average wage of the individuals who self-select themselves into migration relates to the average wage in both origin and destination countries. The two wage differentials can be defined as $Q_0$ and $Q_1$, and - following Heckman (1979) - Borjas (1987) shows that:

$$Q_0 = E\left(\ln\ w_0 \middle| I > 0\right) - \mu_0 = \frac{\sigma_0\sigma_1}{\sigma_\upsilon}\left(\rho - \frac{\sigma_0}{\sigma_1}\right)\lambda(z) \qquad [4]$$

$$Q_1 = E\left(\ln\ w_1 \middle| I > 0\right) - \mu_1 = \frac{\sigma_0\sigma_1}{\sigma_\upsilon}\left(\frac{\sigma_1}{\sigma_0} - \rho\right)\lambda(z) \qquad [5]$$

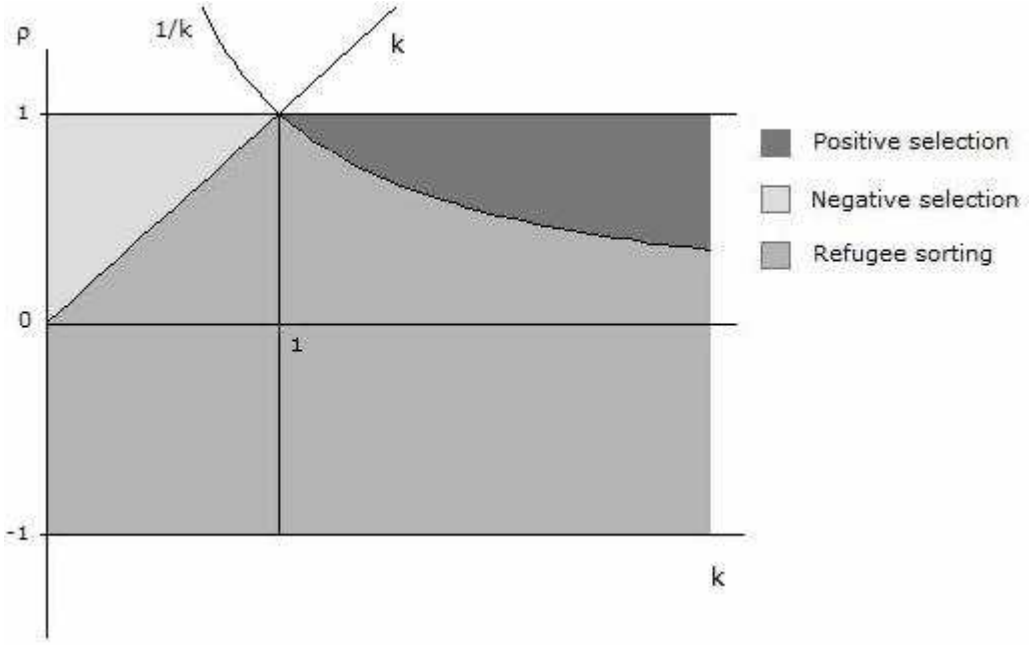
where $\upsilon = \varepsilon_1 - \varepsilon_0$, $\sigma_\upsilon{}^2 = \sigma_1{}^2 + \sigma_0{}^2 - 2\rho\sigma_1\sigma_0$, $z = (\mu_0 - \mu_1 + \pi)/\sigma_\upsilon$ and $\lambda(z) = \phi(z)/\Phi(-z)$, with $\phi(z)$ and $\Phi(z)$ denoting the probability and the cumulative density function of the standard normal respectively. As $\lambda(z)>0$ whenever z is finite, we can observe that:

$$\rho > \frac{1}{k}:\ Q_0 > 0 \qquad [6]$$

$$\rho < k:\ Q_1 > 0 \qquad [7]$$

where $k= \sigma_1/\sigma_0$. The combination of Eqs. [6]-[7] reveals that there are three possible scenarios with respect to migrants' self-selection in unobservables: positive selection, where migrants earn above average wages in both countries ($Q_0>0$ and $Q_1>0$); negative selection, where migrants are drawn from the left tail of the income distribution, and underperform the natives at destination ($Q_0<0$ and $Q_1<0$);[3] refugee sorting, where migrants are drawn from the left tail of the income distribution, but outperform natives at destination ($Q_0<0$ and $Q_1>0$). Figure 1 depicts the regions in the parameter space that correspond to the three scenarios.

**Figure 1.** Patterns of self-selection in unobservables with migration with no uncertainty




[3] Borjas (1987) first oberves that $\mu_1$ denotes "the mean income that residents from the home country would eam in the United States *if all* home country citizens were to migrate to the United States" (p. 532, emphasis in the original), but then it discusses the testable implications of his model as if $\mu_1$ denoted also the average income of natives; he writes that refugee sorting denotes a situation where "the United States draws below-average immigrants (in terms of the country of origin), but they outperform the *U.S. native born* upon arrival" (p. 534, emphasis added).

Observe, from Eq. [7], that whenever wages are more dispersed at destination than at origin, then migrants outperform the natives at destination. Moreover, Borjas (1987) predicts that migrants are drawn from the left-hand tail of the income distribution at origin but end up in the right-hand tail of the distribution at destination, i.e. refugee sorting occurs, when the distributions of $\varepsilon_0$ and $\varepsilon_1$ are independent.

From Eqs. [4]-[5], we can compute the average income of natives in country 0, w, accounting for their self-selection across destinations; we have that:

$$w = \Phi(-z)E(\ln\ w_1 | I > 0) + \Phi(z)E(\ln\ w_0 | I \le 0) = \Phi(-z)\left(\mu_1 + (\mu_0 + \pi)\frac{\Phi(z)}{\Phi(-z)}\right) + \sigma_\upsilon \Phi(-z)\lambda(z) \quad [8]$$

*Migration under uncertainty*

We now analyze the implications for migrants' self-selection of model described in Eqs. [1]-[2] under an alternative informational structure: we assume that an agent decides whether to migrate after he has observed the realization of $\varepsilon_0$, but not the realization of $\varepsilon_1$. For analytical convenience, we assume that agents are risk-neutral, as this limiting case suffices to show the implications of a reduction in the size of the information set on which would-be migrants take their decisions. This information set contains the realization of $\varepsilon_0$ and the expectation of $\varepsilon_1$ conditional on $\varepsilon_0$. Under the bivariate normality assumption, we have that:

$$(\varepsilon_1 \mid \varepsilon_0): \ N\left(\rho k \varepsilon_0, \sigma_1^2\left(1-\rho^2\right)\right) \quad [9]$$

Migration can be characterized by the index function I':

$$I'\left(E(\varepsilon_1|\varepsilon_0), \varepsilon_0\right) = \ln\ \frac{E(w_1)}{w_0}; \ \ (\mu_1 - \mu_0 - \pi) + (\rho k - 1)\varepsilon_0 \quad [10]$$

The parameters $\rho$ and k that characterize the bivariate normal distribution influence the decision to migrate, while they do not when migration entails no uncertainty.[4] Furthermore, we can observe that:

$$\rho > \frac{1}{k}: \ \frac{\partial I'\left(E(\varepsilon_1|\varepsilon_0), \varepsilon_0\right)}{\partial \varepsilon_0} > 0 \quad [11]$$

[4] Note that Eq. [10] is invariant in $\varepsilon_0$ when $\rho=1/k$, so that either no one migrates if $\mu_1 \le \mu_0+\pi$, or every one migrates; in this latter case, then average wage of the migrants at destination trivially coincides with that of the natives; this limiting case will be omitted from the discussion in the paper.

Differently from Borjas (1987), Eq. [11] reveals better realizations of the domestic stochastic component of the wage can increase the probability of migration, as the realization of $\varepsilon_0$ conveys information on $\varepsilon_1$ whenever the two distributions are not uncorrelated. If wages are more dispersed at destination than at origin, and the two distributions are positively and tightly correlated, then the signal about $\varepsilon_1$ that an agent obtains when observing a poor realization of $\varepsilon_0$ can more than offset the incentive to migrate that the observed realization creates.

We can asses the pattern of migrants' self-selection under the informational structure reflected in Eq. [10] along the lines of Borjas (1987):[5]

$$Q_0' = E\left(\ln\ w_0 \middle| I' > 0\right) - \mu_0 = \begin{cases} \sigma_0 \lambda\left(\frac{\sigma_\upsilon}{\sigma_\omega} z\right) & \text{if } \rho > 1/k \\ \left[-\sigma_0\, \Phi\left(-\frac{\sigma_\upsilon}{\sigma_\omega} z\right) \Big/ \Phi\left(\frac{\sigma_\upsilon}{\sigma_\omega} z\right)\right] \lambda\left(\frac{\sigma_\upsilon}{\sigma_\omega} z\right) & \text{if } \rho < 1/k \end{cases} \qquad [12]$$

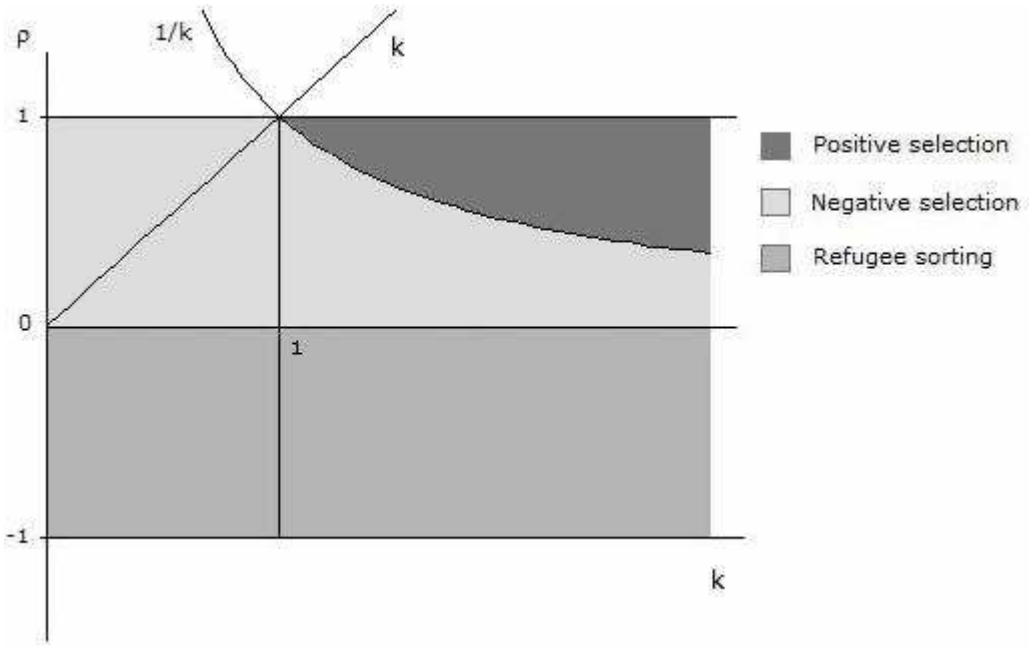
$$Q_1' = E\left(\ln\ w_1 \middle| I' > 0\right) - \mu_1 = \rho k Q_0' \qquad [13]$$

with $\omega = E\left(\varepsilon_1 \middle| \varepsilon_0\right) - \varepsilon_0, \sigma_\omega^2 = \left(\rho\sigma_1 - \sigma_0\right)^2$.[6] Eqs. [12]-[13] reveal that the migrants can obtain starkly different wage outcomes in the origin and in the destination countries only when the correlation coeffiecient $\rho$ is negative, while Borjas (1987) predicted the occurrence of the so-called refugee sorting even when $\rho \geq 0$. This can no longer happen when migration occurs before the realization of $\varepsilon_1$ is observed. A nonnegative correlation between the two distributions entails that migrants' wages are drawn from the same tail of the income distribution both at origin and at destination. Intuitively, the contraction in the size of the information set that we have introduced reduces migrants' ability to improve their relative income status when moving. Moreover, Eqs. [12]-[13] entail that a greater income dispersion at destination, i.e. k>1, no longer ensures that immigrants will outperform natives upon arrival, while Borjas argued that "in essence, the model says that prices matter: whether the immigrant flow is positively or negatively selected depends on the relative rewards to [unobservable] skills among countries" (Borjas, 1990; p. 306).

[5] Eq. [12] uses the fact that $E(x|x<y)=-\lambda(y)[\Phi(-y)/\Phi(y)]<0$, as $\lambda(y)=E(x|x\geq y)$, where x is the standard normal.

[6] Observe that the variance of $\omega$ is always smaller than the variance of $\upsilon$.

**Figure 2.** Patterns of self-selection in unobservables with migration under uncertainty




The implications of the proposed change in the informational structure of the migration decision are not limited to the reshaping of the parameter regions that correspond to the three possible patterns of self-selection, that can be observed from a comparison of Figure 1 and 2. As we did under the informational structure assumed in Borjas (1987), we can compute the average income of natives in country 0, w', under the alternative informational structure; specifically, we have that:

$$w' = \begin{cases} \Phi(-z')\left[\mu_1 + (\mu_0 + \pi)\dfrac{\Phi(z')}{\Phi(-z')}\right] + \Phi(-z')\lambda(z')(\rho\sigma_1 - \sigma) & \text{if } \rho > 1/k \\ \Phi(-z')\left[\mu_1 + (\mu_0 + \pi)\dfrac{\Phi(z')}{\Phi(-z')}\right] - \Phi(-z')\lambda(z')(\rho\sigma_1 - \sigma)\dfrac{\Phi(-z')}{\Phi(z')} & \text{if } \rho < 1/k \end{cases} \qquad [14]$$

The comparison of Eqs. [8] and [14] reveals that the exclusion of the realization of $\varepsilon_1$ from would-be migrants' information set reduces the average income of natives in country 0 compared to the information structure assumed by Borjas (1987). Specifically, we have that:

$$w - w' = \begin{cases} \sigma_\upsilon z\left[\Phi(z) - \Phi\left(\dfrac{\sigma_\upsilon}{\sigma_\omega}z\right)\right] + \sigma_\upsilon\phi(z) - \sigma_\omega\phi\left(\dfrac{\sigma_\upsilon}{\sigma_\omega}z\right) & \text{if } \rho > 1/k \\ \sigma_\upsilon z\left[\Phi(z) - \Phi\left(\dfrac{\sigma_\upsilon}{\sigma_\omega}z\right)\right] + \sigma_\upsilon\phi(z) - \sigma_\omega\phi\left(\dfrac{\sigma_\upsilon}{\sigma_\omega}z\right)\Phi\left(-\dfrac{\sigma_\upsilon}{\sigma_\omega}z\right) & \text{if } \rho < 1/k \end{cases} \qquad [15]$$

which is always positive, as $\sigma_\upsilon > \sigma_\omega$ entails that:

$$\sigma_{\upsilon} z\left[\Phi(z)-\Phi\left(\frac{\sigma_{\upsilon}}{\sigma_{\omega}} z\right)\right]>0,\ \frac{\sigma_{\upsilon}}{\sigma_{\omega}}\phi(z)>\phi\left(\frac{\sigma_{\upsilon}}{\sigma_{\omega}} z\right)>\phi\left(\frac{\sigma_{\upsilon}}{\sigma_{\omega}} z\right)\Phi\left(-\frac{\sigma_{\upsilon}}{\sigma_{\omega}} z\right) \qquad [16]$$

From Eq. [15], we can observe that when $\mu_1 = \mu_0 + \pi$, i.e. z=0, the income loss due to the contraction of the information set is proportional to the standard deviation the distribution of $\varepsilon_1$ conditional upon the realization of $\varepsilon_0$:

$$w - w' = (\sigma_{\upsilon} - \sigma_{\omega})\phi(0) = \phi(0)\sigma_1\left(1-\rho^2\right)^{1/2} \qquad [17]$$

Eq. [17] shows that the income loss is higher the lower the information about $\varepsilon_1$ that can be extracted from the realization of $\varepsilon_0$.

**Conclusions**

The informational structure that underlies the most theoretical representations of migration decision problems assumes – either explicitly or implicitly - a perfect knowledge of the foreign wages before migration occurs. We have shown that the introduction of an alternative informational framework, where only domestic wages are known, influences the theoretical predictions with respect to migrants' self-selection in unobservables. Adopting the framework proposed by Borjas (1987), the case for the so-called refugee sorting narrows down, while the case for migrants' negative self-selection symmetrically expands. Moreover, a destination country might fail to attract "the best and the brightest" even when its income distribution is more dispersed than the one prevailing in the country of origin of the migrants. These findings warn against the possible consequences of reducing the role that uncertainty plays in the decision to migrate. They also point to the additional insights that could be gained through the analysis of a further reduction in the size of the information set, assuming that would-be migrants do not know the objective distribution of incomes at destination, or from considering risk-averse would-be migrants.

**References**


Borjas, G.J.,1987, Self-Selection and the Earnings of Immigrants, American Economic Review 77 (4), 531-553.

Borjas, G.J.,1990, Self-Selection and the Earnings of Immigrants: Reply, American Economic Review 80 (1), 305-308.

Borjas, G.J., 1995, Assimilation and Changes in Cohort Quality Revisited: What Happened to Immigrant Earnings in the 1980s?, Journal of Labor Economics 13 (2), 201-245.

Chiquiar, D. and G.H. Hanson, 2005, International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States, Journal of Political Economy, 113 (2), 239-281.

Chiswick, B.R., 1999, Are Immigrants Favorably Self-Selected?, American Economic Review 89 (2), 181-185.

Fernández-Huertas Moraga, J., forthcoming, New Evidence on Emigrant Selection, The Review of Economics and Statistics.

Grogger, J. and G.H. Hanson, 2008, Income maximization and the selection and sorting of international migrants, NBER Working Paper No. 13821, Cambridge, MA.

Heckman, J.J., 1979, Sample Selection Bias as a Specification Error, Econometrica 47, 153-161.

Jasso, M. and M.R. Rosenzweig, 1990, Self-Selection and the Earnings of Immigrants: Comment, American Economic Review 80 (1), 298-304.

McKenzie, D.J. and H. Rapoport, forthcoming, Self-selection patterns in Mexico-U.S. migration: The role of migration networks, The Review of Economics and Statistics.

Roy, A.D., 1951, Some Thoughts on the Distribution of Earnings, Oxford Economic Papers 3 (2), 135-146.