A New Approach to the Estimation of Selective Migration with an Application to Italy
Very draft version. Please do not cite

Davide Fiaschi
University of Pisa

Angela Parenti
University of Pisa

Cristina Tealdi
Heriot-Watt University

June 10, 2018

Abstract

Migrants self-selection is one of the most fundamental issues in the economics of migration, with highly relevant policy implications (Borjas, 1994). Even though it has been widely recognized that migrants are not randomly selected from the populations of the origin countries, often both the theoretical and the empirical literature ignores the fact that immigrants are self-selected. Nevertheless, the research done on self-selection focuses on two main questions: first, whether and to what extent migrants and non-migrants differ in their characteristics and second, on the reasons why these differences exist. The theoretical foundation for studying the causes of self-selection is the Roy-model (Roy, 1951), which Borjas (1987a) has applied to migration. He developed a model of immigrant self-selection where the focus is on selectivity in unobserved characteristics. In his model, the fundamental driver of migration is the relative returns to skill in origin and destination countries. He claims that if earnings between the destination and the origin countries are positively correlated, whenever the destination country has more income inequality then the origin country, high-skilled migrants benefit the most for migrating, and hence a positive selection is observed. The opposite is true if earnings are more dispersed in the origin country. This result has important policy implications: since earnings inequality is lower in rich counties compared to poor countries on average, the Borjas model predicts that migrants from poor countries are unfavourably selected with regard to their skill levels. The empirical testing of this model, however, resulted in controversial results in the literature (Jasso and Rosenzweig, 1990; Cobb-Clark, 1993). For the case of Italy, the evidence is little and ambiguous. The only two papers on selective migration from the South of the country to the North in the period 1980s-early 2000s point at opposite results (Fratesi and Percoco, 2014a; Bartolucci et al., 2018a). In this paper, we aim at contributing to the debate on the selective migration in Italy. To achieve this goal, we use a random utility model where the decision to migrate from one region to another depends on the difference in the return of education between the two regions. Due to the lack of individual data, we use transition probabilities at regional level for different levels of human capital to estimate the sign of the self-selection of migrants among the Italian regions for the period 1992-2015. We find evidence of strong self-selection in the Italian interregional migration, and of its significant changes over time. We conclude drawing

some implications of our analysis for regional convergence and policy recommendations.

1 Introduction

Some general remarks on the starting point of our analysis

- Why migration is important? The general idea is that migration is a source of convergence and increase the overall efficiency Barro and Sala-i Martin, 2004.

- However, this conclusion is based on the hypothesis of no selection on the characteristics of migrants Fischer and Pfaffermayr, 2017. In presence of self-selection of migrants brain drain and brain gain can change this conclusion Dolado et al., 1994, Barro and Sala-i Martin, 2004, Kanbur and Rapoport, 2005 and Mountford and Rapoport, 2011. In particular, if emigration from a poor country is of high-skilled individuals (brain drain) then the speed of convergence can be decreased in the source country and increased in the destination country.

- Italy is especially interested in brain drain/gain phenomenon Becker et al. (2004). This happen in favour of foreign countries, but also from the South to the North Fratesi and Percoco (2014b) and it can be a candidate to explain the lack of convergence in regional GDP per capita Lagravinese (2015) and Faggian et al. (2017). Italy can be also subject to brain exchange, as explained in Milio et al. (2012).

- Some authors focus on the determinants of interregional migration in Italy adopting a push-pull approach Piras (2017) and Basile et al. (2017) Biagi et al., 2011

- The issue of overall efficiency is in the background, i.e. higher migration is considered overall a additional source of efficiency Decressin and Fatas, 1995 via decreasing regional unemployment as response to region-specific (persistent) shocks and by increasing the incentives in the accumulation of human capital Vidal, 1998.

Which are the main goals of the paper?

- To provide a methodology based on Borjas (1987b) and Markov matrices to identify self-selection in migration when no data on wages are available. We also provide a taxonomy of regions, which clarify the relationship between positive versus negative self-selection and brain drain versus brain gain.

- To apply the methodology to the Italian interregional migration in the period 1992-2015.

Which are the main finding of the paper?

- Italian regions can belong or to the class of positive self-selection-migration destination and negative self-selection-migration origin (brain gain) or to the class of negative self-selection-migration destination and positive self-selection-migration origin (brain drain). No evidence of regions with brain exchange is found. Compare results with Bartolucci et al. (2018b) (no education, nominal wages...), Fratesi and Percoco (2014b) (other methodology), Milio et al. (2012) (a survey...)
- regions shows a persistent membership to these two classes (Markov transition matrices).
- in the period 2006-2015 there is a clear distinction between South and North
- income of unemployed individuals plays a major role in the intensity of migration (sensitivity analysis)
- ...

2 The Model

Assume that the inter-temporal utility of individual q living in region i, with a working time-horizon before retirement of T_q, is given by:

$$U_{qi} = E_0 \left[\sum_{t=0}^{T_q} \beta(t) \, \delta^t u(c_{qi,t}) \right] + EF_i + \epsilon_{qi} \tag{1}$$

where $u(\cdot)$ is the instantaneous individual utility, $c_{q,t}$ the consumption of individual q in region i, $\delta \in (0,1)$ the intertemporal discount rate, and $\beta(t) = 1$ for $t = 0$ and $\beta(t) = \beta$ for $t > 0$ a time-varying parameter reflecting the possibility of quasi-hyperbolic discount rate as in Laibson 1997 . The term EF_i measures all the region i specific hedonic factors, while ϵ_{qi} is an idiosyncratic (unobserved) component of the utility of individual q living in region i; ϵ_{qi} is assumed to be not correlated with the other variables in Eq. (2).Assume that no saving is possible and that utility function is log-linear; then:

$$U_{qi} = \sum_{t=0}^{T_q} \beta(t)\,\delta^t \left[\left(1-\pi^u_{qi,t}\right)\log y^e_{qi,t} + \pi^u_{qi,t}\log b^e_{qi,t}\right] + EF_i + \epsilon_{qi}, \tag{2}$$

where $y^e_{qi,t}$ are the expected income at period t in region i when individual q is employed (e.g. the expected wages) and $b^e_{qi,t}$ when unemployed (e.g. the expected unemployment benefits); finally, $\pi^u_{qi,t}$ is the probability of individual q to be unemployed at period t in region i.

Assuming a log-linear utility model is tantamount to assume that migration flows are determined by the *relative* wages and migrants selection is a function of return to skill and fixed costs of migrating (Borjas, 1987a, Borjas, 1999). Differently a linear utility as in Grogger and Hanson (2011) would imply that *absolute* wage differences influence both the scale and the selectivity of migration flows, and migrants selection does not depend on fixed cost of migrating. The log-linear utility model has been criticised since it implies that the curvature of the utility function is relevant for comparison of all possible destinations, an assumption that might be problematic when looking at international migration given the extreme income differences that exist internationally (Grogger and Hanson, 2011). On the other hand, that in presence of a worldwide skill-neutral technological shift that increases wages in all countries and skill groups, the linear utility model would lead a substantial change in the rate of international migration (Borjas, 2014). "[...] there will be only one way of determining which framework is most useful for understanding the nature of immigrant selection - by looking at the data" (Borjas, 2014, p. 25). Since our aim is to provide evidence on selectivity of internal migration in Italy, we have decided to use the log-utility framework. Our empirical application is related to the literature on north-to-north migration where similar levels of productivity between the source and destination imply that low-skill wages are also similar. In this setting, both the log-linear and the linear model predict that individuals migrating form a source with high returns to skill to a destination with low returns should be negatively selected (Grogger and Hanson, 2011). This choice also allows us to avoid potential pitfalls deriving from the mis-measurement of skill-related wages, that most of the time also show within-country comparability issues.

By a simple manipulation of Eq. (2) we get:

$$U_{qi} = \sum_{t=0}^{T_q} \beta(t)\,\delta^t \left[\log y^e_{qi,t} + \pi^u_{qi,t}\log\left(b^e_{qi,t}/y^e_{qi,t}\right)\right] + EF_i + \epsilon_{qi}. \tag{3}$$

In order to make the model suitable to our scopes we make two assumptions. First, following Mincer (1974) (see also Rosen, 1992) we assume that the income of individual q living in region i in period t is the product of four components:

$$y^e_{qi,t} = y_{i,0} \ \exp\left(\sigma_i s_q + x_{i,t} + \eta_q\right) \tag{4}$$

- $y_{i,0}$ is a region-specific component related to the initial level of income of region i independent on the human capital level of individual q;
- $\sigma_i s_q$ represents the impact of the level of education on individual income, where s_q are the the years of education of individual q and σ_i is the return on human capital in region i;
- $x_{i,t}$ represents the impact of the post-schooling experience on individual income, which could be different among regions.
- η_q are possible other (unobservable) characteristics of individual q affecting her income.

Second, we assume that the income of employed and unemployed workers are expected to grow at the same rate but maintaining the possibility that their ratio is different between regions, that is:

$$\mathscr{b}_i = \log\left(\frac{b^e_{qi,t}}{y^e_{qi,t}}\right) \ \forall t. \tag{5}$$

Substituting Eqq. (4) and (5) into in Eq. (3) we get:

$$\begin{aligned} U_{qi} &= \sum_{t=0}^{T_q} \beta\left(t\right)\delta^t \left[\log y_{i,0} + \sigma_i s_q + x_{i,t} + \eta_q + \pi^u_{qi,t}\mathscr{b}_i\right] + EF_i + \epsilon_{qi} = \\ &= \gamma\left(\beta,\rho,T_q\right)\left(\log y_{i,0} + \sigma_i s_q + \eta_q\right) + \tilde{x}_i\left(\beta,\rho,T_q\right) + \mathscr{b}_i \widetilde{\pi}^u_{qi}\left(\beta,\rho,T_q\right) + EF_i + \epsilon_{qi}, \end{aligned}$$

where:

$$\begin{aligned} \gamma\left(\beta,\rho,T_q\right) &\equiv \frac{1+\delta\left(\beta-1\right)-\beta\delta^{1+T_q}}{1-\delta}; \\ \tilde{x}_i\left(\beta,\rho,T_q\right) &\equiv \sum_{t=0}^{T_q}\beta\left(t\right)\delta^t \log x_{i,t} = \log x_{i,0} + \sum_{t=1}^{T_q}\beta\delta^t\log x_{i,t}; \\ \widetilde{\pi}^u_{qi}\left(\beta,\rho,T_q\right) &\equiv \sum_{t=0}^{T_q}\beta\left(t\right)\delta^t\pi^u_{qi,t} = \pi^u_{qi,0} + \sum_{t=1}^{T_q}\beta\delta^t\pi^u_{qi,t}. \end{aligned}$$

Individual q who lives in region i with a working-time horizon T_q will then decide to migrate to region j at time 0 if the utility of living in region j, U_{qj,T_q}, minus the moving cost from region i to region j, CM_{qij}, is higher then the utility of living in region i, U_{qi,T_q}:

$$
\begin{aligned}
\Delta U_{qij} \equiv U_{qj} - U_{qi} - CM_{qij} &= \gamma\left(\beta,\rho,T_q\right)\log\left(\frac{y_{j,0}}{y_{i,0}}\right) + \gamma\left(\beta,\rho,T_q\right)\left(\sigma_j-\sigma_i\right)s_q + \\
&+ \tilde{x}_j\left(\beta,\rho,T_q\right) - \tilde{x}_i\left(\beta,\rho,T_q\right) + \\
&+ \mathscr{b}_j\widetilde{\pi}^u_{qj}\left(\beta,\rho,T_q\right) - \mathscr{b}_i\widetilde{\pi}^u_{qi}\left(\beta,\rho,T_q\right) + \\
&+ EF_j - EF_i - CM_{qij} + \epsilon_{qj} - \epsilon_{qi}
\end{aligned} \tag{6}
$$

Let R the number of regions in the economy; the probability that individual q living in region i with a working time horizon T_q, P_{qij}, will migrate to region j is given by:

$$
P_{qij} = \text{Prob}\left(\Delta U_{qij} - \Delta U_{qik} > 0,\ \forall k \neq j\right),
$$

that is:

$$
\begin{aligned}
P_{qij} = \text{Prob}\left(\epsilon_{qk} - \epsilon_{qj}\right. &< \gamma\left(\beta,\rho,T_q\right)\log\left(\frac{y_{j,0}}{y_{k,0}}\right) + \gamma\left(\beta,\rho,T_q\right)\left(\sigma_j-\sigma_k\right)s_q + \\
&+ \tilde{x}_j\left(\beta,\rho,T_q\right) - \tilde{x}_k\left(\beta,\rho,T_q\right) + \\
&+ \mathscr{b}_j\widetilde{\pi}^u_{qj}\left(\beta,\rho,T_q\right) - \mathscr{b}_k\widetilde{\pi}^u_{qk}\left(\beta,\rho,T_q\right) + \\
&+ \left. EF_j - EF_k - CM_{qij} + CM_{qik},\ \forall k \neq j\right).
\end{aligned} \tag{7}
$$

P_{qij} is therefore the probability that all the $R-1$ random terms $\epsilon_{qk} - \epsilon_{qj}$ for $k = 1,\ldots,R, k \neq j$, are below the value reported on the right-hand side of Eq. (7). Let $f(\boldsymbol{\epsilon_q})$ the joint density function of the random vector $\boldsymbol{\epsilon_q} = (\epsilon_{q1},\ldots,\epsilon_{qR})$; then:

$$
\begin{aligned}
P_{qij} = \int_{\epsilon} I\left(\epsilon_{qk} - \epsilon_{qj}\right. &< \gamma\left(\beta,\rho,T_q\right)\log\left(\frac{y_{j,0}}{y_{k,0}}\right) + \gamma\left(\beta,\rho,T_q\right)\left(\sigma_j-\sigma_k\right)s_q + \\
&+ \tilde{x}_j\left(\beta,\rho,T_q\right) - \tilde{x}_k\left(\beta,\rho,T_q\right) + \\
&+ \mathscr{b}_j\widetilde{\pi}^u_{qj}\left(\beta,\rho,T_q\right) - \mathscr{b}_k\widetilde{\pi}^u_{qk}\left(\beta,\rho,T_q\right) + \\
&+ \left. EF_j - EF_k - CM_{qij} + CM_{qik},\ \forall k \neq j\right) f(\boldsymbol{\epsilon_q})d\boldsymbol{\epsilon_q}.
\end{aligned} \tag{8}
$$

where $I(\cdot)$ is the indicator function, equalling 1 when the expression in parenthesis is true and 0 otherwise.

If $\epsilon_{qk},\ \ k = 1,\ldots,R$, are independent and identically distributed and follow a Gumbel distribution,the difference $\epsilon_{qk} - \epsilon_{qj}$ follow the logistic distribution (Gumbel,

1958) and, consequently, the probability of individual q to move from region i to region j is given by (see McFadden, 1973):

$$P_{qij} = \frac{\exp\left(\boldsymbol{\beta}'\mathbf{X}_{qij}\right)}{\sum_{r=1}^{R}\exp\left(\boldsymbol{\beta}'\mathbf{X}_{qir}\right)}, \tag{9}$$

where:

$$\begin{aligned}
\boldsymbol{\beta}'\mathbf{X}_{qij} \equiv\ & \gamma\left(\beta,\rho,T_q\right)\log\left(\frac{y_{j,0}}{y_{i,0}}\right) + \gamma\left(\beta,\rho,T_q\right)\left(\sigma_j - \sigma_i\right)s_q + \\
+\ & \tilde{x}_j\left(\beta,\rho,T_q\right) - \tilde{x}_i\left(\beta,\rho,T_q\right) + \mathbb{b}_j\widetilde{\pi}_{qj}^{u}\left(\beta,\rho,T_q\right) - \mathbb{b}_i\widetilde{\pi}_{qi}^{u}\left(\beta,\rho,T_q\right) + \\
+\ & EF_j - EF_i + CM_{qij}.
\end{aligned}$$

In order to clarify the meaning of Eq. (9), it is worth remarking that the probability not to migrate of individual q is given by:

$$P_{qii} = \frac{\exp\left(\boldsymbol{\beta}'\mathbf{X}_{qii}\right)}{\sum_{r=1}^{R}\exp\left(\boldsymbol{\beta}'\mathbf{X}_{qir}\right)} = \frac{1}{\sum_{r=1}^{R}\exp\left(\boldsymbol{\beta}'\mathbf{X}_{qir}\right)}, \tag{10}$$

($\exp\left(\boldsymbol{\beta}'\mathbf{X}_{qii}\right) = 1$ by definition), which substituted into Eq. 9, leads to:

$$P_{qij} = \exp\left(\boldsymbol{\beta}'\mathbf{X}_{qij}\right)P_{qii}. \tag{11}$$

Summing for all $j \neq i$ we have:

$$\sum_{j=1,j\neq i}^{R} P_{qij} = P_{qii}\sum_{j=1,j\neq i}^{R}\exp\left(\boldsymbol{\beta}'\mathbf{X}_{qij}\right) = 1 - P_{qii}, \tag{12}$$

i.e.

$$P_{qii} = \frac{1 - P_{qii}}{\sum_{j=1,j\neq i}^{R}\exp\left(\boldsymbol{\beta}'\mathbf{X}_{qij}\right)}, \tag{13}$$

and therefore substituting in Eq. (11):

$$P_{qij} = \left[\frac{\exp\left(\boldsymbol{\beta}'\mathbf{X}_{qij}\right)}{\sum_{j=1,j\neq i}^{R}\exp\left(\boldsymbol{\beta}'\mathbf{X}_{qij}\right)}\right]\left(1 - P_{qii}\right), \tag{14}$$

Eq. (14) shows that the probability to migrate from region i to region j can be decomposed into the probability to leave region i, i.e. $1 - P_{qii}$, and another term, $\exp\left(\boldsymbol{\beta}'\mathbf{X}_{qij}\right)/\sum_{j=1,j\neq i}^{R}\exp\left(\boldsymbol{\beta}'\mathbf{X}_{qij}\right)$, which represents the conditioned probability to migrate to region j once individual q has decided to migrate from region i, i.e.:

$$P_{qij}|q \text{ is migrating} = \frac{P_{qij}}{1 - P_{qii}} = \frac{\exp\left(\boldsymbol{\beta}'\mathbf{X}_{qij}\right)}{\sum_{j=1,j\neq i}^{R} \exp\left(\boldsymbol{\beta}'\mathbf{X}_{qij}\right)}. \tag{15}$$

Eq. (15) shows that the estimate of the determinants of migration could be done only considering migrating individuals and ignoring the not migrating individuals.

The main obstacle to estimate Eq. (9) and/or Eq. (15) by using the standard logit estimation technique proposed by McFadden (1973) is the lack of publicly available micro-data on the choice to migrate, an issue plaguing the most of EU countries. Additional minor difficulties are that i) the econometric model is non-linear in T_q; ii) we should specify as individuals formulate their expectations on the future growth rates of income (earnings) and unemployment rates), and iii) endogeneity can be present. In Section 3 we will present an empirical strategy to circumvent these difficulties when the interest is only focused on the estimate of self-selection in migration, i.e. in the estimate of the sign and magnitude of the regional difference in the returns to education represented by $\sigma_j - \sigma_i$.

3 Empirical strategy

In this section we propose an empirical strategy to directly estimate the difference in the return of education, i.e. $\sigma_j - \sigma_i$, without any information but the observed flows of migrants between regions divided by age and education. In Section 3.3 ee also elaborate a strategy to calculate the counterfactual migration probability under this limitation in the data.

From Eqq. (10) and (11) we have:

$$\begin{aligned} \log\left(\frac{P_{qij}}{P_{qii}}\right) = \boldsymbol{\beta}'\mathbf{X}_{qij} &= \gamma\left(\beta,\rho,T_q\right)\log\left(\frac{y_{j,0}}{y_{i,0}}\right) + \gamma\left(\beta,\rho,T_q\right)\left(\sigma_j - \sigma_i\right)s_q + \\ &+ \tilde{x}_j\left(\beta,\rho,T_q\right) - \tilde{x}_i\left(\beta,\rho,T_q\right) + \mathbb{b}_j\widetilde{\pi}_{qj}^{u}\left(\beta,\rho,T_q\right) - \mathbb{b}_i\widetilde{\pi}_{qi}^{u}\left(\beta,\rho,T_q\right) + \\ &+ EF_j - EF_i + CM_{qij}, \end{aligned} \tag{16}$$

where P_{qij}/P_{qii} is the migration probability from region i (j) to region j (i) normalized to the mass of individuals who decide not to move.Taking the difference between these normalized migration probability from region i to region j for two individuals 1 and 2 having the same working-time horizon $T_1 = T_2 = \bar{T}$ but two different levels of education $s_2 > s_1$, we have:

$$
\begin{aligned}
\log\left(\frac{P_{2ij}}{P_{2ii}}\right) - \log\left(\frac{P_{1ij}}{P_{1ii}}\right) &= \gamma\left(\beta,\rho,\bar{T}\right)(\sigma_j-\sigma_i)(s_2-s_1) + \\
&+ \mathbb{b}_j\left[\tilde{\pi}^u_{2j}\left(\beta,\rho,\bar{T}\right) - \tilde{\pi}^u_{1j}\left(\beta,\rho,\bar{T}\right)\right] - \mathbb{b}_i\left[\tilde{\pi}^u_{2i}\left(\beta,\rho,\bar{T}\right) - \tilde{\pi}^u_{1i}\left(\beta,\rho,\bar{T}\right)\right] \\
&+ CM_{2ij} - CM_{1ij}; \qquad (17)
\end{aligned}
$$

consequently, the differential return to education between region i and region j is given by:

$$
\begin{aligned}
\Delta\sigma_{ij} &\equiv \sigma_j - \sigma_i = \\
&= \left[\frac{1}{\gamma\left(\beta,\rho,\bar{T}\right)(s_2-s_1)}\right]\left\{\log\left(\frac{P_{2ij}}{P_{2ii}}\right) - \log\left(\frac{P_{1ij}}{P_{1ii}}\right) + \right. \\
&- \mathbb{b}_j\left[\tilde{\pi}^u_{2j}\left(\beta,\rho,\bar{T}\right) - \tilde{\pi}^u_{1j}\left(\beta,\rho,\bar{T}\right)\right] - \mathbb{b}_i\left[\tilde{\pi}^u_{2i}\left(\beta,\rho,\bar{T}\right) - \tilde{\pi}^u_{1i}\left(\beta,\rho,\bar{T}\right)\right] \\
&+ \left. CM_{2ij} - CM_{1ij}\right\} \qquad (18)
\end{aligned}
$$

Eq. (18) points out that the difference in the returns to education between region j versus region i, $\Delta\sigma_{ij}$, determining the self-selection of migrants between the two regions, can be calculated by taking:

1. a scaling factor depending on the work-time horizon and discount rate, $\gamma\left(\beta,\rho,\bar{T}\right)$, such that the difference between the normalized migration probabilities calculated for the inter-temporal utility with time horizon $\bar{T}$ are reported to yearly timing;

2. the difference between the levels of education, $s_2 - s_1$, which report the differences in the return to yearly timing;

3. the difference between the normalized migration probabilities from region i to region j for different levels of education;

4. the different regional unemployment benefits $\mathbb{b}_j$ and $\mathbb{b}_i$ weighted by the difference in the probability to be unemployed for different levels of education $\tilde{\pi}^u_{qi}\left(\beta,\rho,\bar{T}\right)$; and

5. the differences in cost of moving of individuals with different level of education $CM_{2ij} - CM_{1ij}$.

While for the first four points we have some information to exploit, point 5 is not easily computed. However, we observe that according to Eq. (18) the difference

$\sigma_i - \sigma_j$ can be expressed as:

$$
\begin{aligned}
\Delta\sigma_{ji} &\equiv \sigma_i - \sigma_j = \\
&= \left[\frac{1}{\gamma\left(\beta,\rho,\bar{T}\right)(s_2-s_1)}\right]\left\{\log\left(\frac{P_{2ji}}{P_{2jj}}\right) - \log\left(\frac{P_{1ji}}{P_{1jj}}\right) + \right. \qquad (19)\\
&- \mathbb{b}_i\left[\widetilde{\pi}^u_{2i}\left(\beta,\rho,\bar{T}\right) - \widetilde{\pi}^u_{1i}\left(\beta,\rho,\bar{T}\right)\right] - \mathbb{b}_j\left[\widetilde{\pi}^u_{2j}\left(\beta,\rho,\bar{T}\right) - \widetilde{\pi}^u_{1j}\left(\beta,\rho,\bar{T}\right)\right] \\
&+ \left. CM_{2ji} - CM_{1ji}\right\} \qquad (20)
\end{aligned}
$$

and since:

$$\Delta\sigma_{ij} = -\Delta\sigma_{ji}, \tag{21}$$

we have that:

$$CM_{2ij} - CM_{1ij} + CM_{2ji} - CM_{1ji} = -\log\left(\frac{P_{2ij}}{P_{2ii}}\right) + \log\left(\frac{P_{1ij}}{P_{1ii}}\right) - \log\left(\frac{P_{2ji}}{P_{2jj}}\right) + \log\left(\frac{P_{1ji}}{P_{1jj}}\right). \tag{22}$$

3.1 Regional Indexes of Self-Selection

Borjas (1987b) discusses how a positive sign of $\Delta\sigma_{12,ij}$ implies a positive self-selection of migrants for region of destination j, i.e. the more educated workers should tends to migrate from region i to region j attracted by the higher return on education. Conversely, a negative sign of $\Delta\sigma_{12,ij}$ implies a negative self-selection of migrants for region of origin i. In a multiregional framework the type of self-selection for every region can be represented by *local synthetic indexes of self-selection*. In particular, on the base of Eq. (18) a natural local synthetic index of self-selection of destination region is given by the weighted sum of individual local synthetic indexes of self-selection related to region of destination j:

$$SSM^D_j\left(\bar{T}, s_1, s_2\right) = \sum_{i=1}^{R}\frac{m_{ij}}{\sum_{i=1}^{R} m_{ij}}\Delta\sigma_{ij} = \sum_{i=1}^{R}\tilde{m}^D_{ij}\Delta\sigma_{ij}, \tag{23}$$

where $\tilde{m}^D_{ij}$ is the share of migrants going to region j from region i on total migrants going to region j.In the same manner, a local synthetic index of self-selection for region of origin is given by the weighted sum of individual local synthetic indexes of self-selection related to region of origin i:

$$SSM^O_i\left(\bar{T}, s_1, s_2\right) = \sum_{j=1}^{R}\frac{m_{ij}}{\sum_{j=1}^{R} m_{ij}}\Delta\sigma_{ij} = \sum_{j=1}^{R}\tilde{m}^O_{ij}\Delta\sigma_{ij} \tag{24}$$

where $\tilde{m}^O_{ij}$ is the share of migrants coming from region i to region j on total migrants coming from region i.

Consider the region of destination j. If we observe a positive $\Delta\sigma_{ij}\left(\bar{T}, s_1, s_2\right)$, this means that the return to education in region j is higher than the return to education in region i, so that region j will be a destination for high-skilled migrants living in region i. Summing on all the origin regions, a positive $SSM^D_j\left(\bar{T}, s_1, s_2\right)$ implies that, on average, people migrating to region j should be high-skilled. On the contrary, a negative $\Delta\sigma_{ij}\left(\bar{T}, s_1, s_2\right)$, implies that the return to education in region j is lower than the return to education in region i and, consequently, region j should be a destination for low-skilled migrants living in region i. Therefore, a negative $SSM^D_j\left(\bar{T}, s_1, s_2\right)$ means that, on average, people migrating to region j are low-skilled. With the same reasoning we can analyse the index of self-selection for the region of origin i. When $\Delta\sigma_{ij}\left(\bar{T}, s_1, s_2\right)$ is positive, high-skilled individual living in region i should migrate to region j. Summing on all the destination regions, a positive $SSM^O_i\left(\bar{T}, s_1, s_2\right)$ implies that, on average, people emigrating from region i are high-skilled. On the contrary, when $\Delta\sigma_{ij}\left(\bar{T}, s_1, s_2\right)$ is negative region i is an origin region of low-skilled emigrants. Then, a negative $SSM^O_i\left(\bar{T}, s_1, s_2\right)$ means that, on average, people emigrating from region i are low-skilled. Figure 1 reports the four possible cases.

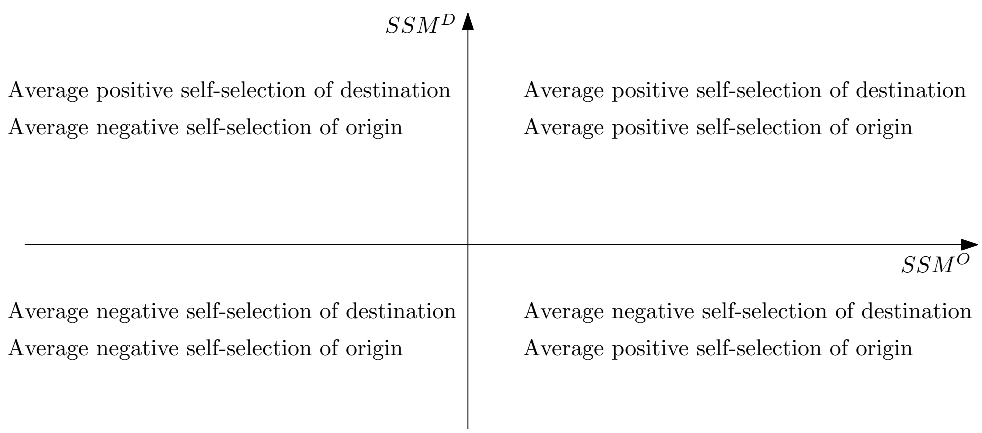

Figure 1. A taxonomy of regions based on local synthetic indexes of self-selection in migration.

3.2 An Example with Two-regions Country

Consider an economy with only two regions, N and S and two levels of education $s_2 > s_2$ for a sample of migrants with the same working-time horizon T. Suppose that the two transition matrices representing the interregional migration for the two levels of education are reported in Table 1.

s_1	N	S
N	P_{1NN}	$1-P_{1NN}$
S	$1-P_{1SS}$	P_{1SS}

s_2	N	S
N	P_{2NN}	$1-P_{2NN}$
S	$1-P_{2SS}$	P_{2SS}

Table 1. Transition matrix of interregional migration between Regions N and S, where migrants are divided according their two possible levels of education $s_2 > s_1$.

From Eq. (18) we have that:

$$\Delta\sigma_{SN} = \left(\frac{1}{s_2-s_1}\right)\left\{\frac{1}{\gamma\left(\rho,\bar{T}\right)}\log\left(\frac{P_{1SS}}{P_{2SS}}\times\frac{1-P_{2SS}}{1-P_{1SS}}\right)-\tilde{b}_N\left(Pr^u_{2N}-Pr^u_{1N}\right)+\tilde{b}_S\left(Pr^u_{2S}-Pr^u_{1S}\right)\right\}, \tag{25}$$

i.e. the nature of self-selection of migrants should depend on P_{1SS}, P_{2SS}, on differential expected unemployment rates and unemployment benefits between the two regions.

Assuming that $\tilde{b}_S\left(Pr^u_{2S}-Pr^u_{1S}\right)-\tilde{b}_N\left(Pr^u_{2N}-Pr^u_{1N}\right)=0$, then Eq. (25) simplifies to:

$$\Delta\sigma_{SN} = \frac{1}{(s_2-s_1)\,\gamma\left(\rho,\bar{T}\right)}\log\left(\frac{P_{1SS}}{P_{2SS}}\times\frac{1-P_{2SS}}{1-P_{1SS}}\right); \tag{26}$$

thus, if $P_{1SS} > P_{2SS}$, i.e. the high-skilled workers show a higher migration rate from Region S to Region N, $\Delta\sigma_{SN} > 0$, i.e. we should observe a positive self-selection toward the Region N.However, if $\tilde{b}_S\left(Pr^u_{2S}-Pr^u_{1S}\right)-\tilde{b}_N\left(Pr^u_{2N}-Pr^u_{1N}\right)<0$, e.g. because Region N is more generous as unemployment benefits and the difference between the unemployment rates of low- and high-skilled workers are similar for the two regions, such conclusion could be reverted. The presence of more than two regions makes more fuzzy any conclusion on the way of self-selection.

3.3 Counterfactual Migration Probabilities

We start defining the counterfactual probability as:

$$P^{CF}_{qij} = \frac{\exp\left(\boldsymbol{\beta}'\mathbf{X}^{CF}_{qij}\right)}{\sum_{r=1}^{R}\exp\left(\boldsymbol{\beta}'\mathbf{X}^{CF}_{qir}\right)}, \tag{27}$$

where $\mathbf{X}^{CF}_{qij,0}$ represents the counterfactual level of determinants (e.g. the difference in the return to schooling σ). Therefore:

$$\log P^{CF}_{qij} = \boldsymbol{\beta}'\mathbf{X}^{CF}_{qij} - \log\left(\sum_{r=1}^{R}\exp\left(\boldsymbol{\beta}'\mathbf{X}^{CF}_{qir}\right)\right) \tag{28}$$

and

$$\log P_{qij} = \boldsymbol{\beta}'\mathbf{X}_{qij} - \log\left(\sum_{r=1}^{R}\exp\left(\boldsymbol{\beta}'\mathbf{X}_{qir}\right)\right) \tag{29}$$

Hence:

$$\log P^{CF}_{qij} = \log P_{qij} + \boldsymbol{\beta}'\mathbf{X}^{CF}_{qij} - \boldsymbol{\beta}'\mathbf{X}_{qij} - \log\left(\frac{\sum_{r=1}^{R}\exp\left(\boldsymbol{\beta}'\mathbf{X}^{CF}_{qir}\right)}{\sum_{r=1}^{R}\exp\left(\boldsymbol{\beta}'\mathbf{X}_{qir}\right)}\right) \tag{30}$$

and therefore:

$$\log P^{CF}_{qij} = \log P_{qij} + \boldsymbol{\beta}'\mathbf{X}^{CF}_{qij} - \boldsymbol{\beta}'\mathbf{X}_{qij} - \log\left(\frac{P_{qii}}{P^{CF}_{qii}}\right), \tag{31}$$

i.e.

$$\log P^{CF}_{qij} - \log P_{qij} = \boldsymbol{\beta}'\Delta\mathbf{X}^{CF}_{qij} - \log\left(P_{qii}\right) + \log\left(P^{CF}_{qii}\right) \tag{32}$$

Eq. (32) shows that the counterfactural effect of a determinant can be decomposed into two components: the first related to the change in the destination, i.e. $\boldsymbol{\beta}'\Delta\mathbf{X}^{CF}_{qij}$, and the second related to the change in the probability to migrate, i.e. $\log\left(P_{qii}\right) - \log\left(P^{CF}_{qii}\right)$.

Rearranging:

$$\log\left(\frac{P^{CF}_{qij}}{P^{CF}_{qii}}\right) = \log\left(\frac{P_{qij}}{P_{qii}}\right) + \boldsymbol{\beta}'\Delta\mathbf{X}^{CF}_{qij}. \tag{33}$$

Since $P^{CF}_{qii} \equiv 1 - \sum_{j,j\neq i} P^{CF}_{qij}$, then

$$\log\left(\frac{P^{CF}_{qij}}{1 - \sum_{j,j\neq i} P^{CF}_{qij}}\right) = \log\left(\frac{P_{qij}}{P_{qii}}\right) + \boldsymbol{\beta}'\Delta\mathbf{X}^{CF}_{qij} \tag{34}$$

Taking the exponent of both sides and summing in j:

$$\frac{\sum_{j} P^{CF}_{qij}}{1 - \sum_{j,j\neq i} P^{CF}_{qij}} = \sum_{j,j\neq i}\exp\left(\log\left(\frac{P_{qij}}{P_{qii}}\right) + \boldsymbol{\beta}'\Delta\mathbf{X}^{CF}_{qij}\right) = \sum_{j,j\neq i} A_{qij} \tag{35}$$

from which:

$$\sum_{j,j\neq i} P^{CF}_{qij} = \frac{\sum_{j,j\neq i} A_{qij}}{1 + \sum_{j,j\neq i} A_{qij}}. \tag{36}$$

and therefore:

$$P_{qii}^{CF}=1-\sum_{j,j\neq i}P_{qij}^{CF}=\frac{1}{1+\sum_{j,j\neq i}\exp\left(\log\left(\frac{P_{qij}}{P_{qii}}\right)+\boldsymbol{\beta}'\Delta\mathbf{X}_{qij}^{CF}\right)} \tag{37}$$

Substituting Eq. (37) into Eq. (32) leads to:

$$\log P_{qij}^{CF}=\log P_{qij}+\boldsymbol{\beta}'\Delta\mathbf{X}_{qij}^{CF}-\log\left(P_{qii}\right)-\log\left(1+\sum_{j,j\neq i}\exp\left(\log\left(\frac{P_{qij}}{P_{qii}}\right)+\boldsymbol{\beta}'\Delta\mathbf{X}_{qij}^{CF}\right)\right) \tag{38}$$

i.e.

$$\log P_{qij}^{CF}=\log\left(\frac{\exp\left(\boldsymbol{\beta}'\Delta\mathbf{X}_{qij}^{CF}\right)P_{qij}}{P_{qii}+\sum_{j,j\neq i}\exp\left(\boldsymbol{\beta}'\Delta\mathbf{X}_{qij}^{CF}\right)P_{qij}}\right) \tag{39}$$

and therefore (using $\Delta X_{qii}^{CF}=0$):

$$P_{qij}^{CF}=\left[\frac{\exp\left(\boldsymbol{\beta}'\Delta\mathbf{X}_{qij}^{CF}\right)}{\sum_{r=1}^{R}\exp\left(\boldsymbol{\beta}'\Delta\mathbf{X}_{qir}^{CF}\right)P_{qir}}\right]P_{qij}. \tag{40}$$

4 Empirical Evidence

...

4.1 Counterfactual Analysis of Self Selection Migration in Italy

...

5 Concluding Remarks

...

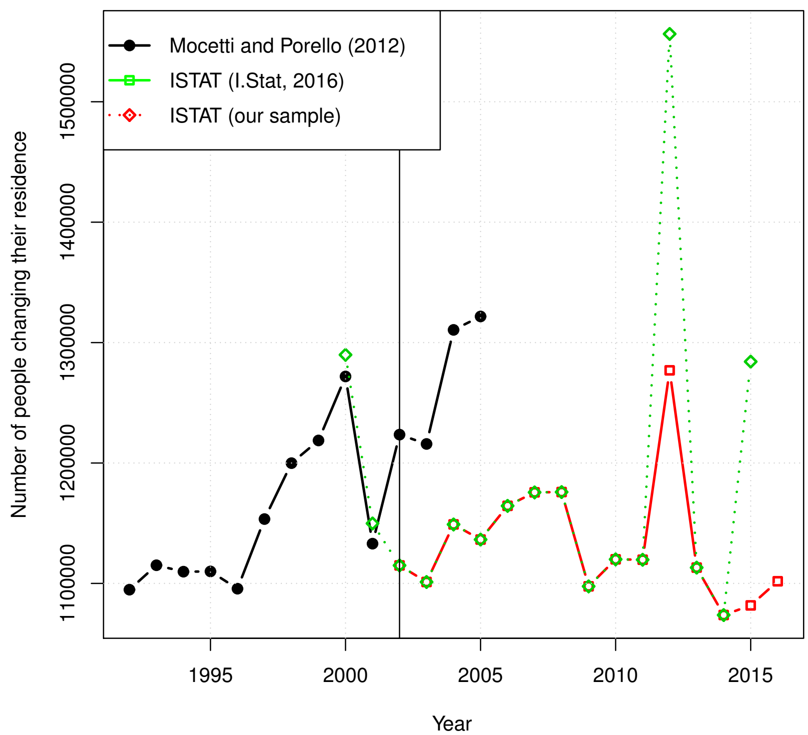

Figure 2. Comparison of different datasets on interregional Italian migration

A Comparison of different datasets on interregional Italian migration

References

Ambrosini, J. W. and G. Peri (2012). The determinants and the selection of mexico-us migrants. *The World Economy 35*(2), 111–151.

Barro, R. J. and X. Sala-i Martin (2004). Economic growth: Mit press. *Cambridge, Massachusettes*.

Bartolucci, C., C. Villosio, and M. Wagner (2018a). Who migrates and why? Evidence from Italian administrative data. *Journal of Labor Economics (forthcoming)*.

Bartolucci, C., C. Villosio, and M. Wagner (2018b). Who migrates and why? evidence from italian administrative data. *Journal of Labor Economics 36*(2), 000–000.

Basile, R., A. Girardi, M. Mantuano, and G. Russo (2017). Interregional migration, human capital externalities and unemployment dynamics: Evidence from italian provinces.

Becker, S. O., A. Ichino, and G. Peri (2004). How large is the" brain drain" from italy? *Giornale degli Economisti e Annali di Economia*, 1–32.

Belot, M. V. K. and T. J. Hatton (2012). Immigrant selection in the oecd. *The Scandinavian Journal of Economics 114*(4), 1105–1128.

Bertoli, S. (2010). The informational structure of migration decision and migrants selection. *Economics Letters 108*, 89–92.

Biagi, B., A. Faggian, and P. McCann (2011). Long and short distance migration in italy: the role of economic, social and environmental characteristics. *Spatial Economic Analysis 6*(1), 111–131.

Borjas, G. J. (1987a). Self-selection and the earnings of immigrants. *American Economic Review 77*(4), 531–551.

Borjas, G. J. (1987b). Self-selection and the earnings of immigrants. *The American Economic Review*.

Borjas, G. J. (1994). The economics of immigration. *Journal of Economic Literature XXXII*, 1667–1717.

Borjas, G. J. (1999). The economic analysis of immigration. In *Handbook of labor economics*, Volume 3, pp. 1697–1760. Elsevier.

Borjas, G. J. (2014). *Immigration Economics*. Harvard University Press.

Borjas, G. J. and B. Bratsberg (1996). Who leaves? the outmigration of the foreign-born. *The Review of Economics and Statistics 78*(1), 165–176.

Card, D. (2001). Immigrant inflows, native outflows, and the local market impacts of higher immigration. *Journal of Labor Economics 19*(1), 22–64.

Carrington, William J., D. E. . V. T. (1996). Migration with endogenous moving costs. *The American Economic Review 86*(4), 909–930.

Chiquiar, Daniel, . H. G. H. (2005). International migration, self selection and the distribution of wages: Evidence from mexico and the united states. *Journal of Political Economy 113*(2), 1105–1128.

Chiswick, B. R. (1999). Are immigrants favorably self-selected? *The American Economic Review Papers and Proceedings 89*(2), 181–185.

Cobb-Clark, D. A. (1993). Immigrant selectivity and wages: The evidence for women. *The American Economic Review 83*(4), 986–993.

de Coulon, A. and M. Piracha (2005). Self-selection and the performance of return migrants: the source country perspective. journal of population economics. *Economics Letters 18*, 779–807.

Decressin, J. and A. Fatas (1995). Regional labor market dynamics in europe. *European Economic Review 39*(9), 1627–1655.

Dolado, J., A. Goria, and A. Ichino (1994). Immigration, human capital and growth in the host country. *Journal of population economics 7*(2), 193–215.

Faggian, A., I. Rajbhandari, and K. R. Dotzel (2017). The interregional migration of human capital and its regional consequences: a review. *Regional Studies 51*(1), 128–143.

Fernández-Huertas Moraga, J. (2011). New evidence on emigrant selection. *The Review of Economics and Statistics 93*(1), 72–96.

Fischer, L. B. and M. Pfaffermayr (2017). The more the merrier? migration and convergence among european regions. *Regional Science and Urban Economics*.

Fratesi, U. and M. Percoco (2014a, August). Selective migration, regional growth and convergence: Evidence from italy. *Regional Studies 48*(10), 1650–1668.

Fratesi, U. and M. Percoco (2014b). Selective migration, regional growth and convergence: evidence from italy. *Regional Studies 48*(10), 1650–1668.

Grogger, J. and G. H. Hanson (2011). Income maximization and the selection and sorting of international migrants. *Journal of Development Economics 95*(1), 42–57.

Gumbel, E. J. (1958). *Statistics of Extremes*. New York, Columbia University Press.

Ibarraran, P. and D. Lubotsky (2007). Mexican immigration and self-selection: New evidence from the 2000 mexican census. In G. J. Borjas (Ed.), *Mexican Immigration to the United States*, pp. 201 – 234. The University of Chicago Press.

Jasso, G. and M. R. Rosenzweig (1990). Self-selection and the earnings of immigrants: Comment. *The American Economic Review 80*(1), 298–304.

Kaestner, R. and O. Malamud (2014). Self-selection and international migration: New evidence from mexico. *The Review of Economics and Statistics 96*(1), 78–91.

Kanbur, Ravi, . R. H. (2005). Migration selectivity and the evolution of spatial inequality. *Journal of Economic Geography 5*, 43–57.

Kanbur, R. and H. Rapoport (2005). Migration selectivity and the evolution of spatial inequality. *Journal of economic geography 5*(1), 43–57.

Lagravinese, R. (2015). Economic crisis and rising gaps north–south: evidence from the italian regions. *Cambridge Journal of Regions, Economy and Society 8*(2), 331–342.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*. Wiley, New York.

McKenzie, David, S. S. and J. Gibson (2010). How important is selection? experimental vs. non-experimental measures of the income gains from migration. *Journal of the European Economic Association 8*(4), 913–945.

Milio, S., R. Lattanzi, F. Casadio, N. Crosta, M. Raviglione, P. Ricci, and F. Scano (2012). Brain drain, brain exchange and brain circulation. the case of italy viewed from a global perspective. *National interest*.

Mincer, J. (1974). Schooling, experience, and earnings. human behavior & social institutions no. 2.

Mountford, A. and H. Rapoport (2011). The brain drain and the world distribution of income. *Journal of Development Economics 95*(1), 4–17.

Pedersen, P. J., M. Pytlikova, and N. Smith (2008). Selection or network effects? migration flows into 27 oecd countries 1990-2000. *European Economic Review 52*, 1160–1186.

Piras, R. (2017). A long-run analysis of push and pull factors of internal migration in italy. estimation of a gravity model with human capital using homogeneous and heterogeneous approaches. *Papers in Regional Science 96*(3), 571–602.

Rooth, D.-O. and J. Saarela (2007). Selection in migration and return migration: Evidence from micro data. *Economics Letters 94*(1), 90–95.

Rosen, S. (1992). Distinguished fellow: Mincering labor economics. *Journal of Economic Perspectives 6*(2), 157–170.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers 3*, 135–146.

Vidal, J.-P. (1998). The effect of emigration on human capital formation. *Journal of Population Economics 11*(4), 589–600.