Piras Romano, Internal migration in Italy: The role of migration networks, Italian Economic Journal, Vol. 6, n. 1, pp. 157-195.

Internal migration in Italy: The role of migration networks

Romano Piras
University of Cagliari
Department of Political and Social Sciences
Viale S. Ignazio, 17 - 09123 Cagliari (Italy)
Tel. +39 070 6753314
e-mail: pirasr@unica.it

This paper investigates the role of migration networks at internal level for the case of Italy and provides estimates of network elasticities which are found similar, in magnitude, to those estimated for international migration to the US. The empirical results, robust to different ways of dealing with unobserved heterogeneity, prove that network effects play a crucial role in explaining migration flows also at internal level. In addition, they also help to clarify the role of the other variables that have been found relevant in explaining internal migration flows in Italy during the last decades.

JEL Codes: J61, R23.
Key words: gravity model; internal migration; Italy; networks effects.

1 Introduction

In this paper we analyse the role of migration networks at the internal level for the case of Italy. From a theoretical point of view, gravity models of migration can be framed into a macroeconomic perspective looking at aggregate demand and supply of migrants (Karemera et. al., 2000), or can be derived from micro-founded economic models (Beine et al., 2016; Anderson, 2011).[1] Both approaches stress the role of migration networks (diasporas) as a key determinant of migration flows. As shown by a large amount of economic literature (Munshi, 2016; Beine et al., 2015; Beine et al., 2011) and also by sociological, geographical and demographic scholars (Bakewell et al., 2016; Garip and Asad, 2015; King, 2012a, 2012b; Haug, 2008; Amuedo-Dorantes and Mundra, 2007), moving costs (economic and non-economic) are lower when there is already an established network of previous migrants at the target destination.

Networks operate in various ways, for example, by giving information regarding job availability and first-aid help in looking for an accommodation to would-be migrants. It has been highlighted that once a network is established, a sort of self-perpetuating mechanism goes on (Massey et al., 1993). In addition, the recent empirical literature on international migration based on the gravity model has estimated that the share of explained variability due to migrant networks accounts for more than 70% of the observed variability of migration flows and more than 80% of the explained variability of the model (Beine et al., 2011, p.36). Hence, the existence of a diaspora is a key determinant to explain international migration.

Following Beine et al. (2015, p. 380), the network effects can be decomposed into a *policy* and an *assimilation* effect. The former 'is the overcoming of legal entry barriers imposed by the destination country' and while it has proved to be very important in the international context, it does not operate at the internal level, at least not in those countries in which internal mobility is not restricted. The second, the assimilation effect, 'operates through the lowering of private costs' and 'covers a wide range of hurdle faced by migrants in finding employment and deciphering foreign cultural norms.' At the macro level, Beine (2016) further splits the assimilation effect into two components: the first relates to a pure decrease in migration costs brought by migrant networks; the second relates to the benefits (for example in terms of information available to the network regarding social norms prevailing at destination, job opportunities and so on) that the existing networks provide to migrants. He explicitly recognises that these components, not related to immigration policy, apply to international as well as internal migration. It is through the

[1] Among others, the macroeconomic approach has recently been followed by Piras (2017) and Karemera et. al. (2000), the microeconomic one by Bertoli et al. (2016), Ortega and Peri (2013), Bertoli and Fernández-Huertas Moraga (2013) and Grogger and Hanson (2011).

assimilation effect that migrants find it easier to integrate in the community or society of the destination region. One could claim that the value of networks in migrating within a country that uses a single language and where its people share the same culture is limited, but this is just a speculation that needs to be tackled and empirically answered. Talking about Italy, for example, there is a well-known divide between Southern and Centre-Northern regions, which is not only economic in nature but it is also cultural and social, and that comes at least from the unification of Italy in 1861, and that has been extensively studied by other social sciences as well.[2] Add to this the fact that in the decades following World War II (WWII), Italy recorded millions of individuals flowing across its regions and, in particular, from Southern to Centre-Northern ones. All in all, an evaluation on the role of migrant networks at regional level is very important for a country like Italy that, in spite of its common language and culture, is still characterised by large social and economic disparities across its regions. Yet, and quite surprisingly, the empirical literature on internal migration in Italy completely neglects the role played by networks.

To fill this gap, we conducted our empirical investigation using internal migration flow data across Italian regions over the 1980–2013 time period. More precisely, we extended the work of Piras (2017) by introducing the networks of previously immigrated individuals showing that these networks serve a crucial role: on the one hand, to better explain internal migration in Italy, and, on the other hand, to clarify the role of the other variables involved in the migration phenomenon. As a matter of fact, Piras (2017) estimated a gravity model of internal migration in Italy during the time span from 1970 to 2005 and found macroeconomic variables to be the main drivers of internal migration. The author mainly found that at destination, human capital has had no role in explaining internal flows, whereas at origin, it has seemingly acted as a restraining factor. These results, however, were obtained without taking into account the role of networks that, as we will prove in this paper, have a paramount role in affecting migration.

The paper is organised into eight sections. Following the introduction, the different theoretical approaches regarding the gravity models of migration is presented are presented in Section 2. Section 3 provides a sort of road map that translates the theory into its empirical counterpart. Section 4 describes the variables and Section 5, the data. In Section 6, the empirical analysis and robustness checks are presented in Section 7 and Section 8, respectively. Finally, the conclusion is given in Section 8.

[2] For example, it is well known since the study of Putnam (1993) on social capital and civic tradition in Italy that wide variations do exists in the performance of regional governments and that these variations are intimately related to the vitality of associational life in each region.

2 Gravity models of migration

2.1 A macroeconomic gravity model of migration

Karemera et. al. (2000) assume that at time *t*, aggregate gross migration flows from the country/region of origin *i* to the country/region of destination *j* F_{ijt}, are driven by supply-push factors at home S_{it}, demand-pull factors at destination D_{jt}, and by other time-varying restraining and/or aiding factors associated with the specific origin-destination pair *i-j* R_{ijt}, such that:

$$(1) \qquad F_{ijt} = A_{ij}^{a_0}\left(\frac{S_{it}^{a_1} D_{jt}^{a_2}}{R_{ijt}^{a_3}}\right)$$

where A_{ij} catches time-invariant origin-destination pair effects, while the a_s (s = 0, 1, 2, 3) are structural elasticities linking migration flows to supply and demand factors. It is worth noticing that, although not specifically indicated by the authors, one of the most prominent of the specific origin-destination factors is the existence of a network of previously migrated individuals that acts as a cost-reducing factor. Going a step ahead, Karemera et. al. (2000) posit that a gravity model of migration can be derived from the demand and supply of migrants that depend on population (n_{it} and n_{jt}) as a mass variable, and expected income (y_{it} and y_{jt}) as the main economic explicative variable, both measured at origin and destination. Piras (2017) extends this approach and introduced human capital (h_{it} and h_{jt}) into the demand and supply of migrants which are assumed to be given by:

$$(2) \qquad S_{it} = b_i y_{it}^{b_1} n_{it}^{b_2} h_{it}^{b_3}$$

$$(3) \qquad D_{jt} = c_j y_{jt}^{c_1} n_{jt}^{c_2} h_{jt}^{c_3}$$

where the b_s and c_s (s = 1, 2, 3) are structural parameters, while b_i and c_j are, respectively, time-constant push factors and time-constant pull factors. Combining equations (1)-(3) and taking natural logarithms, we see that aggregate gross migration flows are:

$$(4) \qquad \begin{aligned} \ln F_{ijt} &= a_0 \ln A_{ij} + a_1 \ln b_i + a_2 \ln c_j + a_1 b_1 \ln y_{it} + a_1 b_2 \ln n_{it} + \\ &+ a_1 b_3 \ln h_{it} + a_2 c_1 \ln y_{jt} + a_2 c_2 \ln n_{jt} + a_2 c_3 \ln h_{jt} - a_3 \ln R_{ijt} \end{aligned}$$

As it will be discussed in Section 3, in order to properly estimate equation (4), in addition to the main macroeconomic and demographic variables, one has to consider origin *i*, destination *j* and origin-destination *i-j* dummies. In addition, in equation (4) as in all panel regressions, a time period dummy is usually inserted in order to control for common shocks that affect all units of the panel.

2.2 Micro-foundations of gravity models of migration

Beine et al. (2016) consider a random utility model (RUM) in which the utility that individual *k*, located in country/region *i* at time *t*-1, derives from moving to country/region *j* at time *t* is:

(5) $$u_{ijt}^{k} = w_{ijt} - c_{ijt} + \varepsilon_{ijt}^{k}$$

where w_{ijt} is the deterministic component of utility, c_{ijt} measures the moving costs from *i* to *j* and ε_{ijt}^{k} is an individual specific random term. If ε_{ijt}^{k} is assumed to follow an IID extreme value type I distribution, the results of McFadden (1974) can be exploited to obtain the expected probability that individual *i* moves to country/region *j* as:

(6) $$E(p_{ijt}) = \frac{e^{w_{ijt}-c_{ijt}}}{\sum_{l} e^{w_{ilt}-c_{ilt}}}$$

where *l* is any other country/region (including the origin) that individual *k* can choose. By definition, at time *t*, expected gross migration flows m_{ijt} from *i* to *j* are given by:

(7) $$E(m_{ijt}) = E(p_{ijt}) s_{it} = \left(\frac{e^{w_{ijt}-c_{ijt}}}{\sum_{l} e^{w_{ilt}-c_{ilt}}} \right) s_{it}$$

where s_{it} is the stock of population residing in country/region *i* at time *t*. Alternatively, if it is assumed that the deterministic component of utility does not vary with the country/region of origin *i*, equation (7) can be written as:

(8) $$E(m_{ijt}) = \left(\frac{e^{w_{jt}}}{e^{c_{ijt}}} \right) \left(\frac{1}{\sum_{l} e^{w_{ilt}-c_{ilt}}} \right) s_{it} = \frac{y_{jt}}{\phi_{ijt}} \frac{s_{it}}{\Omega_{it}}$$

Equation (8) is a gravity equation where expected gross migration flows positively depend on the attractiveness of destination (proxied by economic, social, cultural, institutional and other factors), y_{jt}, and on the stock of population at origin, s_{it}, whereas they negatively depend on the accessibility of destination, i.e. on the time-varying moving costs from *i* to *j*, ϕ_{ijt}, and on $\Omega_{it} = \sum_{l} e^{w_{ilt}-c_{ilt}}$, namely on the influence that a variation in the attractiveness of an alternative destinations *l* exerts on the bilateral migration flows from *i* to *j*. This influence is the multilateral resistance to migration (Bertoli and Fernández-Huertas Moraga, 2013).

In a similar RUM framework, and additionally making use of the market clearing conditions, Anderson (2011) developed a structural gravity model of migration in which two multilateral resistance to migration terms appear: the first, for *outward* migration from origin *i* to any other destination; the second, for *inward* migration to *j* from any other origin. In his model, expected gross migration flows m_{ijt} are given by:

(9) $$E(m_{ijt}) = \frac{s_{it} s_{jt}}{\sum_i s_{it}} \frac{(1/\delta_{ijt})}{\Omega_{jt}\Omega_{it}}$$

where, along with population at origin s_{it}, also population at destination s_{jt} exerts a positive effect on migration flows. In addition $\sum_i s_{it}$ is world (constant) labour supply, δ_{ijt} is the cost of migration from *i* to *j*, Ω_{jt} and Ω_{it} are, respectively, the inward and outward multilateral resistance to migration terms.

Going back to Beine et al. (2016), they show that introducing more general distributional assumptions regarding the correlation of the stochastic component of utility in equation (5) leads to far different expressions of expected gross migration flows. For example, they show that the approach followed by Ortega and Peri (2013) is coherent with the RUM and more general than the one that leads to equation (8). Indeed, by assuming that there is unobserved individual heterogeneity between migrants and non-migrants, Ortega and Peri (2013) show that expected gross migration flows are:

(10) $$E(m_{ijt}) = \left(\frac{y_{jt}}{\phi_{ijt}}\right)^{1/\tau} \frac{s_{it}}{\Omega_{ijt}}$$

where $0 < \tau < 1$ measures the degree of correlation of the random term across destinations and the multilateral resistance to migration Ω_{ijt} now depends also on the country/region of destination. Notwithstanding these undoubted improvements in the theoretical explanation of the migration decisions, Ortega and Peri (2013) model is not the more general conceivable framework to analyse them. As a matter of fact, Beine et al. (2016) show that when the migration decision is observed during a prolonged time period, namely when the potential migrant optimises her utility along all her life, then the corresponding expression for expected gross migration flows is:

(11) $$E(m_{ijt}) = \frac{y_{jt}}{\phi_{ijt}} \frac{e^{\beta V_{t+1}(j)}}{\Omega_{it}^{V}} s_{it}$$

where $\beta \le 1$ is a discount factor reflecting positive time preferences and $V_{t+1}(j)$ is the expected value of all future migration decisions, while the multilateral resistance term now becomes $\Omega_{it}^{V} = \sum_l e^{w_{lt} - c_{ilt} + \beta V_{t+1}(j)}$.

The estimates of these models raise some technical questions that will be discussed in the next section.

3 From theory to empirics

In the previous section, we have sketched some theoretical models of migration derived from different theoretical perspectives; in this section, we discuss how to estimate them coherently with the underlying theory. Let us start with the simplest way to write a general log-linearised gravity equation:

(12) $$m_{ijt} = \alpha_i' X_{it} + \gamma_j' X_{jt} + \varphi_{ij}' X_{ijt} + \varepsilon_{ijt}$$

to which an error term, ε_{ijt}, has been added. In equation (12), X_{it} and X_{jt} subsume, respectively, time-varying origin-specific and destination-specific (i.e. monadic) variables that are supposed to affect the migration decision. X_{ijt} represents origin-destination-specific time-varying variables, namely variables that measure the degree of accessibility (higher or lower) of the destination *j* for those living in *i* at time *t*. Finally, α_i', γ_j' and φ_{ij}' are vectors of parameters to be estimated.

In a panel data setting, a simple way to estimate equation (12) is the so-called traditional approach (Bertoli and Fernández-Huertas Moraga, 2013, p. 82) followed, among others, by Grogger and Hanson (2011), Mayda (2010), Clark et al. (2007) and Karemera et. al. (2000). This approach assumes a two-way (2-FE) error term structure such that $\varepsilon_{ijt} = \lambda_t + \nu_{ij} + \mu_{ijt}$, where λ_t are time dummies, ν_{ij} are origin-destination dummies and μ_{ijt} is a well-behaved error term, assumed to be orthogonal to the regressors and serially uncorrelated. It is worth noticing that the two-way structure is identical to an error structure with time (λ_t), origin (ς_i), destination (ς_j) and origin-destination (ν_{ij}) dummies, namely to $\varepsilon_{ijt} = \lambda_t + \nu_{ij} + \varsigma_i + \varsigma_j + \mu_{ijt}$ (Egger and Pfaffermayr, 2003). In this empirical specification, multilateral resistance to migration is not explicitly addressed; however, more general alternatives are available.

As a first one, in coherence with their theoretical analyses synthesised by equation (10), Ortega and Peri (2013) estimate various equations, including different sets of dummy variables. Their preferred empirical specification for the error term, beside time dummies, includes origin-year dummies (ϑ_{it}) along with origin and destination dummies such that $\varepsilon_{ijt} = \lambda_t + \varsigma_i + \varsigma_j + \vartheta_{it} + \mu_{ijt}$. Origin-year dummies ϑ_{it} capture the role of those factors that vary across origins and years but not across destinations,[3] ς_i control for time-invariant factors that vary only across origins, whereas ς_j catch time-invariant factors that vary only across destinations. Furthermore, Ortega and Peri (2013)

---

[3] The introduction of origin-year dummies creates a problem in estimating origin-specific time-varying variables. See below how such a problem can be dealt with.

estimated a second empirical specification, including also origin-destination dummies such that $\varepsilon_{ijt} = \lambda_t + \varsigma_i + \varsigma_j + \nu_{ij} + \vartheta_{it} + \mu_{ijt}$. The authors acknowledge that this is a rather demanding specification in that the inclusion of ν_{ij} absorbs all time-invariant bilateral variables, making estimates of their effects impossible.

In line with the study of Anderson (2011) and his structural gravity model, a second estimation strategy is to assume an error term structure where, along with origin-year dummies ϑ_{it}, also destination-year dummies θ_{jt} are included so that outward and inward multilateral resistance to migration is accounted for. In this case we have $\varepsilon_{ijt} = \lambda_t + \nu_{ij} + \vartheta_{it} + \theta_{jt} + \mu_{ijt}$.[4] A drawback of this specification is that it might be technically impossible to estimate because the contemporaneous inclusion of all these dummies implies a dramatic reduction in the degrees of freedom and an increase in multicollinearity problems. As noticed by Orefice (2015), who used the Anderson (2011) model to estimate the relationship between preferred trade agreements and bilateral migration flows, a researcher could afford to be less demanding and use origin-*period* and destination-*period* dummies, with *period* defined as multiple years. A second more severe drawback is that the contemporaneous inclusion of origin-year and destination-year dummies makes estimated coefficients of the time-varying origin- and destination-specific variables meaningless since the coefficients of these variables are identified by dropping one or more of these dummies (Head and Mayer, 2014). Notice that the same criticism also applies to the Ortega and Peri (2013) model with respect to the monadic variables at origin, given that they include origin-year dummies in the model. The solution proposed by Head and Mayer (2014) to overcome such a problem is a two-step procedure described in Appendix B. Briefly, in the first step of this procedure, migration flows are regressed on a set of origin-year, destination-year and origin-destination dummy variables. In the second step, the estimated origin-year and destination-year fixed effects are regressed on the monadic variables in order to retrieve, respectively, the outward (at origin) and the inward (at destination) effects of these variables on bilateral migration flows.

[Table 1]

As a third estimation strategy, one can apply the common correlated effects (CCE) estimator proposed by Pesaran (2006), which is the most appropriate to use when multilateral resistance to migration originates from an RUM, such as the one synthesised by equation (11). Furthermore, “it

[4] The set of fixed-effects considered in each model is summarised in Table 1.

has also the additional advantage of being robust even in the presence of residual cross-sectional dependence in the data" (Beine et al., 2016, p. 502). In order to control for multilateral resistance to migration, the CCE estimator introduces a linear combination of cross-sectional averages of the dependent and all the independent variables and estimates:

(13) $$m_{ijt} = \alpha_i' X_{it} + \gamma_j' X_{jt} + \varphi_{ij}' X_{ijt} + \xi_{ij}' \bar{Z}_t + \varepsilon_{ijt}$$

where ξ_{ij}' is the vector of origin-destination factor loading and $\bar{Z}_t$ is the vector of cross-sectional panel averages of all variables. In (13) the error term is assumed to be $\varepsilon_{ijt} = \lambda_t + \mu_{ijt}$

Finally, a complementary approach to the CCE that offers the possibility to tackle the presence of cross-sectional dependence in the data is given by the augmented mean group estimator (AMG) put forward by Eberhardt and Teil (2010). Compared with the traditional mean group estimator put forth by Pesaran and Smith (1995), the AMG estimator copes with cross-sectional dependence in the data by introducing a common dynamic effect in the unit specific regression, namely in the origin-destination pairs. The estimate is performed in two steps. Firstly, the following regression is estimated:

(14) $$\Delta m_{ijt} = b_i' \Delta X_{it} + b_j' \Delta X_{jt} + b_{ij} \Delta X_{ijt} + \sum_{t=2}^{T} c_t \Delta D_t + e_{ijt} \qquad \Rightarrow \hat{c}_t \equiv \hat{\pi}_t^{\bullet}$$

where e_{ijt} is an error term and D_t are time dummies. Secondly, the estimated coefficients of the time dummies $\hat{\pi}_t^{\bullet}$ are inserted into the origin-destination pairs regression along with origin-destination time trends that are intended to catch other omitted idiosyncratic processes:

(15) $$m_{ijt} = \alpha_i' X_{it} + \gamma_j' X_{jt} + \varphi_{ij} X_{ijt} + d_{ij} \hat{\pi}_t^{\bullet} + \tau_{ij} time\_trend + \varepsilon_{ijt}$$

To sum up, there are different estimation strategies that can be used in order to estimate bilateral migration flows and, as stated by Beine et al. (2016), determining which of them is more suited is an empirical question. In this respect, they claim that a minimum requirement for the estimates to be coherent with the RUM model is that residuals should not be affected by cross-sectional dependence. This can easily be checked, for example, using the CD test put forward by Pesaran (2004).

As regards endogeneity issues that typically plague econometric estimates, in this framework finding valid instruments is not easy. In the least square (fixed effects) estimates, a simple way adopted by many empirical studies (Ortega and Peri, 2013; Mayda, 2010; Pedersen et al., 2008), is to lag the dependent variables by one year, thus effectively also taking into account the information from the previous year that is available to potential migrants. This constitutes the approach we follow in this paper. It is worth noticing that, by controlling for multilateral resistance to migration through different types of fixed effects, the potential problems associated with omitted factors becomes much less severe. This as well holds true for the CCE and the AUG estimators because these also control for multilateral resistance to migration (Beine et al., 2016; Bertoli and Fernández-Huertas Moraga, 2013). Furthermore, panel data setting involving long time periods face other potential problems, namely the non-stationary and cointegration issues (Fidrmuc, 2009; Beine et al. 2016; Desbordes and Eberhardt, 2014) that are detectable by appropriate tests such as, for example, the Pesaran (2007) test for stationarity and the Kao (1999) residual test for cointegration.

4. Explanatory variables

Following the previous discussion, the most recent empirical literature on internal migration in Italy (see Piras, 2017 and the references therein) and the well-established literature on the determinants of migration (Bansak et al., 2015; Bodvarsson and Van den Berg, 2013), this sub-section will specify the variables to consider in equation (12).

First, we take into account per capita GDP and unemployment rate as proxies of expected income at origin and destination. Economic theory predicts that higher per capita GDP at origin (destination) deters (attracts) migration flows. Conversely, higher unemployment rate at origin (destination) spurs (pushes away) them. Second, population enters as the mass variable and its expected sign is positive with respect to both origin and destination. Third, we introduce the average years of schooling of the resident population as a proxy for human capital. The role of human capital as a determinant of migration flows has been largely discussed in the literature and summarised in Piras (2017). Briefly, human capital might have a positive role at origin and a negative effect at destination if, at destination, the skills and experiences of migrants are paid more with respect to the origin. However, the opposite might happen if, for example, agglomeration forces and economies of scale induce migrants, especially skilled ones, to leave their origin country/region and look for higher rewards to their accumulated human capital in another country/region. Finally, the variable to which we are most interested in, namely the migrant networks. We take into account network effects by inserting into the estimated equation the stock of migrants coming from origin-region i and living in the destination-region j. As claimed in the Introduction, the role of networks at the sub-national level as an explanatory variable for internal migration has not been studied so far, and while for some countries this lack of investigation might not be very important,[5] for the case of Italy – that has witnessed large flows of internal migrants from the poorer Southern regions towards the richer Centre-Northern regions – this is surprising.

[5] Another reason for the absence of empirical investigation at sub-national level could be the lack of data.

To accomplish this task, we compute the stock of internal migrants stk_{ijt} from origin-region i to destination-region j according to:

(16) $$stk_{ijt} = stk_{ijt-1} + inflow_{ijt} - d_{ijt} stk_{ijt-1} = inflow_{ijt} + (1 - d_{ijt}) stk_{ijt-1}$$

where $inflow_{ijt}$ is the flow of immigrants from region i to region j at time t and d_{ijt} is a 'depreciation' factor that adjusts for return migration and the mortality rate among immigrants. We calculate d_{ijt} using regional mortality and emigration rates as $d_{ijt} = (em_{jit} + mor_{jt})$, where, for any region i, em_{jit} is the emigration rate from region j to region i and mor_{jt} is the population mortality rate in region j, both expressed as a percentage of the resident population. The value of the regional stock at the beginning of the period under analysis, namely stk_{ij1980}, has been computed as follows:

(17) $$stk_{ij1980} = inflow_{ij1980} + (1 - d_{ij1979}) \times \sum_{t=1970}^{1979} inflow_{ijt}$$

We approximate the regional immigrants stock at year 1979 (the second term on the right-side of the equation) by summing up all previously available flows and correct it by taking into account the regional mortality and emigration rates registered in 1979.

To sum up, the equation to be estimated is:

(18) $$\ln m_{ijt} = \alpha_{1i} \ln pcy_{it} + \alpha_{2i} \ln u_{it} + \alpha_{3i} \ln n_{it} + \alpha_{4i} \ln h_{it} + \gamma_{1j} \ln pcy_{jt} + \gamma_{2j} \ln u_{jt} + \gamma_{3j} \ln n_{jt} + \gamma_{4j} \ln h_{jt} + \varphi_{ij} \ln stk_{ijt} + \varepsilon_{ijt}$$

where m_{ijt} represents gross migration flows from origin-region i to destination-region j, pcy_{it} (pcy_{jt}) is per capita GDP at origin (destination), u_{it} (u_{jt}) is unemployment rate at origin (destination), n_{it} (n_{jt}) is population at origin (destination), h_{it} (h_{jt}) is human capital at origin (destination) and, finally, stk_{ijt} is the network variable given by the stock of previous migrants from origin-region i residing in destination-region j.

5. Data sources and preliminary testing

The Italian National Institute for Statistics (ISTAT) provides data on unemployment (ISTAT, various years, b), interregional migration flows (ISTAT, various years, a) and, for the more recent years, on-line at http://dati.istat.it/. The average years of schooling for resident population are computed, exploiting the data available in ISTAT (various years, b). We considered four schooling levels and years of schooling as follows: 18 years for individuals with a university degree (ISCED 8, 7 and 6 of the OECD, 2015 classification), 13 years for individuals with an upper secondary school diploma (ISCED 5, 4 and 3), 8 years for those with a lower-secondary school attainment (ISCED 2) and 3 years for individuals with either a primary school educational level (ISCED 1) or without any formal schooling attainment (ISCED 0). Regional per capita GDP and population from 1970 to 2009 comes from SVIMEZ (2011) and have been updated with data available on-line at: http://dati.istat.it/.[6]

Our empirical investigation refers to the 1980–2013 time period. The data set contains 34 years, 380 bilateral flows and 12920 observations. We have only 31 observations (0.24%) with zero flows. Taking into account that the dependent variable is expressed in logs, in order to retain this piece of information we have set these zero flows to one. Overall, the average interregional migration flow between region of origin *i* and region of destination *j* is 747, however, such a figure is not at all informative of the phenomenon since there are huge variations across years and origin-destination pairs: from zero up to more than 10,000. A more detailed synthesis of all the variables at regional level is reported in Tables A1 and A2 in the Appendix A. Table A1 summarises interregional migration flows and it is divided into four panels corresponding to *intra* Centre-Northern regions flows (Panel A), Centre-North to South flows (Panel B), South to Centre-Northern regions flows (Panel C) and *intra* Southern regions flows (Panel D). As can be seen, on average during the time period under scrutiny, the largest flows are from Southern to Centre-Northern regions, particularly those from Sicilia (7339), Campania (6157), Puglia (5412), Calabria (4757) and Sardegna (1676) to Lombardia. Summary statistics for all the other variables are summarised in Table A2. Notice that regions in column (1) are taken as origin regions, accordingly columns from (2) to (5) report summary statistics for that region, whereas columns from (6) to (10) represent the average values of the 19 destination regions for the origin region in column (1).

As for cross-sectional dependence, unit roots and cointegration testing, we report the Pesaran (2004) CD tests to assess whether the series are affected by cross-sectional dependence, apply the Pesaran (2007) test for stationarity which is valid under cross-sectional dependence determined by unobserved common factors and look for cointegration by mean of the Kao (1999) residual cointegration test.[7] These tests are performed for the whole sample of the 20 Italian regions, with each one seen as a source and as a destination, then separately, for the two sub-samples of the 8 Southern (source) to the 12 Centre-Northern (destination) regions and from the 12 Centre-Northern (source) to the 8 Southern (destination) regions.

[Table 2]

[Tables 3a-3b-3c]

---

[6] The data set used in the empirical analysis is available from the author upon request.

[7] As an additional test for cointegration, we perform residual stationarity tests which can be interpreted as additional tests for cointegration (Eberhardt and Teil, 2013).

As expected, Table 2 shows that the null hypothesis of independence of the CD test is strongly rejected. For the whole sample of Italian regions (Table 3a), the CIPS test supports the hypothesis that non-stationary always holds for per capita GDP and population, and it holds for gross migration flows when more than one lag is considered and for the remaining variables when more than two lags are included. The same test for the South to Centre-North sample (Table 3b) gives strong evidence that for gross migration flows, per capita GDP (at origin and destination) and population (at origin and destination), the null of non-stationary cannot be rejected. Similarly, when more than one lag is considered, the stock of migrants at destination and the average years of schooling at destination show the same pattern. As for unemployment rate and average years of schooling at origin, the evidence in favour of non-stationarity seems less clear. Indeed, while non-stationarity seems to hold (with one lag for average years of schooling and two lags for unemployment rate) for these two variables at destination, at origin, for both of them non-stationarity shows up only if more than three lags are included. Finally, for the Centre-North to South sample (Table 3c), we obtain specular results compared to those of the South to Centre-North sample, with the difference that for gross migration flows, non-stationarity also holds when more than one lag is considered. Finally, the Kao (1999) residual cointegration test reported in Table 4 clearly rejects the null hypothesis of no cointegration in all panels. All in all, these results hint that the variables under scrutiny contain unit roots and are cointegrated, explaining why, as a robustness check in Section 7, along with the empirical specifications outlined in Section 3, we also needed to apply the dynamic ordinary least squares (DOLS) advocated by Mark and Sul (2003).

[Table 4]

6. Empirical results

6.1 Results for the whole sample of 20 Italian regions

Table 5 shows the results for the whole sample of Italian regions. All models are estimated first without the stock of migrants at destination, then included it.[8] Diagnostic tests reported below each estimate reveal on the one hand that variables are cointegrated (in all regressions the CIPS tests suggest that residuals are stationary) and on the other hand, that the inclusion of the stock of

[8] In columns (3)-(6) the reported coefficients for variables at origin are those obtained in the second-stage regressions. Similarly, in columns (7)-(8), the reported coefficients for variables at both origin and destination are those of the second-stage regressions. For more details on this two-step procedure, see Appendix B.

migrants at destinations contribute to making residuals cross-sectional independent in the 2-FE and in the two Ortega and Peri (2013) style estimates. In the Anderson (2011) model, the result for cross-sectional dependence is reverted: in column (7), the regression without the stock of immigrants shows that the CD test does not reject the null hypothesis of independence, while in column (8) the CD test refuse it. Somewhat surprisingly, none of the CCE and AUG regressions are able to cope with cross-sectional dependence and, furthermore, these regressions are generally much less satisfying in terms of estimated coefficients. This is the reason why, in what follows, we restricted our main comments on the fixed-effects estimates reported in columns (1)-(8). Taken together, these results suggest that the fixed-effects estimates are coherent with the RUM models sketched in Section 2. These results partially contrast the findings of Piras (2017) who finds that cross-section dependence is less severe for AUG estimates with respect to a simple two-way fixed effects estimate. However, that paper presents three important difference. First, the time period is different from the one considered in the present study. Second, the two Ortega and Peri (2013) style equations and the Anderson (2011) model are not estimated. Third, and more importantly, the stock of migrants at destination is not part of the regressors.

[Table 5]

Looking at the estimated coefficients, a number of interesting points need to be emphasised. Compared with the variables at origin, the variables at destination have a better explanatory power in terms of both statistical significance and number of significant variables. In particular, the stock of migrants at destination is always highly statistically significant and its point elasticity, in the fixed-effects estimates, is remarkable stable around one.[9] Very interestingly, this elasticity is close to that reported by Beine at al. (2015) with regard migration into the United States. For OECD countries, Beine at al. (2011) estimate a coefficient in the 0.62\0.77 range. Other estimates at international level (Beine, 2016) display lower elasticities (0.4). In addition, the introduction of the stock of migrants at destination has sizeable impacts on the other estimated parameters at both origin and destination. Let us consider these impacts in turns.

At destination, introducing migration networks lowers the estimated coefficient of population (from 2.24\2.09 to 0.61\0.74 depending on the specific model) and of per capita GDP (from 0.37\0.32 to 0.16\0.18), making it statistically insignificant. As regards the unemployment rate, it does not change across all fixed-effects specifications: estimated coefficients are very stable in the −0.18\−0.21 range. Finally, the introduction of the migrant stock increases (in absolute value, from

[9] Quite implausibly, it jumps up to five and seven in columns (12) and (10), respectively.

−0.24\−0.35 up to −0.48\−0.49) and, above all, makes the estimated coefficient of average years of schooling statistically significant.[10] Looking at variables at origin, the introduction of the migrants stock has no effects on per capita GDP (which is never significant) and lessens the impact of population (from 2.37\2.21 to 1.86\1.94) and of the unemployment rate, making it statistically insignificant. As far as the impact of the introduction of the migration stock variable on human capital is concerned, it is interesting to highlight that its role now becomes unambiguously positive and statistically significant across all fixed effects estimates, with an estimated elasticity ranging between 0.66 and 0.87 in columns (6) and (2), respectively. Compared with Piras (2017) who found basically no role at destination and a feeble negative impact at origin, the present results suggest a positive impact of human capital at origin and a negative role at destination. The positive role at origin and the negative role at destination are consistent with Borjas (1991) who claims that migration is lower (higher), the higher the mean educational level at destination (origin) due to the educational premium that occurs if human capital at destination is rewarded more than at origin.

6.2 Results for the South to Centre-North sample

In Italy, as it is well known, internal migration has been mainly driven by southerners moving to Centre-Northern regions so that any empirical investigation of Italian internal migration must tackle this peculiarity and separately analyse South to Centre-North from Centre-North to South flows. Table 6 reports the estimates from the 8 Southern regions towards the 12 Centre-Northern ones. Once again, estimates are first performed without the stock of migrants at destination and then including it. At the bottom of the Table, the CIPS test confirms that the variables are cointegrated whereas at standard 5% significance levels, the CD test does not allow to reject the null hypothesis of cross-sectional independence in the 2-FE estimates (columns 1 and 2), the two Ortega and Peri (2013) models in which the stock of migrants is considered (columns 4 and 6), in both CEE estimates (columns 9 and 10) and, finally, in the AUG estimate without the stock of migrants (column 11). Notice, however, that the CCE regressions performed very poorly in terms of significance, magnitude and sign of estimated coefficients. The AUG regression without the stock of migrants (column 11), on the contrary, seems to be roughly in accordance with the theory underlying the gravity model; however, when the stock of migrants is considered in column (12), the overall results worsen given that the CD test do not reject the null of cross-sectional dependence, with the estimated coefficient of population at destination having the wrong negative sign and the estimate coefficient of the stock of migrants that gets very high in magnitude. All in all,

[10] In the Anderson (2011) model, however, human capital is never statistically significant.

as regards the South to Centre-North flows, we find support for the RUM models sketched in Section 2 for the 2-FE and for the two Ortega and Peri (2013) style estimates.

Let us now have a look at estimated coefficients restricting, once more, our main comments to the fixed effects estimates. As regards the main variable under scrutiny, we can see that the elasticity of the stock of migrants is always highly statistically significant and slightly lower with respect to what are shown in Table 5, since it varies between 0.77 and 0.93. We also observe that the introduction of the stock of migrants has, again, a noteworthy impact on the other estimated coefficients. At destination, migration flows becomes slightly less reactive with respect to per capita GDP (from 1.13\1.07 to 0.86\0.92) and unemployment (from −0.31 to −0.25), while with respect to population, the estimated elasticities are halved (from around 4.7 down to 1.76\2.37). As for human capital, estimated coefficients are basically never significant in all regressions. With respect to the variables at origin, the introduction of the migrant stock has negligible impacts on per capita GDP and population (both variables are almost never statistically significant) while lowering (from 0.28\0.23 to 0.12\0.14) the elasticity of unemployment rate that now, contrary to what happens for the whole sample of Italian regions (see Table 5), is always statistically significant. At the same time, excluding the O-P (a) model, human capital is also affected by the introduction of the migrant stock. More in details, in the 2-FE estimates (column 1 and 2) it becomes positive and statistically significant while in the O-P (b) and Anderson (2011) models, it determines a sign inversion (from negative to positive) keeping statistical significance. Thus, when the migrant stock is included into the regressions, human capital at origin displays exactly the same pattern already seen for the whole sample of Italian regions in Table 5. On the contrary, at destination, human capital does not affect migration flows and this confirms the results already obtained by Piras (2017).

[Table 6]

6.3 Results for the Centre-North to South sample

Table 7 reports the estimates for migration flows from the 12 Centre-Northern regions towards the 8 Southern ones. The CIPS test confirms cointegration among the variables, whereas at standard 5% significance levels, the CD test always rejects the null of cross-sectional independence, with the exception of the 2-FE regression in column (2).

With the exception of CCE regression in column (10), the stock of migrants at destination turns out to be always highly statistically significant with a different magnitude, depending on the model. Compared with those of Tables 5 and 6, the estimated coefficients for variables at destination are never significant, with the noteworthy exception of human capital that, on the contrary, is always highly statistical significant across all fixed-effects specification and independently of whether the stock of migrants is included or not (estimated elasticities vary from −2.20 to −0.93). Looking at the variables at origin, it is seen that per capita GDP does not influence Centre-North to South migration flows, unemployment seems to display a restraining role,[11] population, on the contrary, exerts a strong positive effect on them. Finally, human capital is estimated to have a positive effect on migration flows and, while the magnitude of the coefficient decreases when the migration stock variable is included, its statistical significance is always very high in all fixed effects estimates. To conclude, it would seem that the role of human capital in the North to Centre-South direction is similar to its role in the overall interregional migration flows.

[Table 7]

7. Robustness checks

In this section we perform some robustness checks of the previous results. In particular, Table 8 reports the estimates of the Ortega-Peri (a) model with the inclusion of the (log) of the distance between origin and destination regions (measured as the kilometric distance between the two regional capital cities). As a matter of fact, while in all other models the distance effect is absorbed by the origin-destination dummy variable, this is not the case in the Ortega-Peri (a) model, allowing the role of geographical distance to be estimated. As already found in Table 5, also in Table 8 for the whole sample of the 20 Italian regions when the migration network variable is considered (see column 2), the CD test does not reject the null of cross-sectional independence at any statistical level. The same CD test for the South to Centre-North flows does not reject the null at the 5% of confidence (column 4). Furthermore, it is important to notice that the estimated coefficients for the whole sample of Italian regions (columns 1 and 2), for the South to Centre-North (columns 3 and 4) and for the Centre-North to South (columns 5 and 6) are very close to those previously estimated in Tables 5–7. As for distance, its estimated coefficient is expectedly always negative across all

[11] This result could be due to the fact that Centre-North to South flows have different motivations with respect to the South to Centre-North flows. As discussed in other studies (Piras, 2017; Mocetti and Porello, 2012; Etzo, 2011), one of the empirical explanations of Centre-North to South flows has been given in terms of return migration. During the 1950s and 1960s, millions of individuals moved from Southern to the Centre-Northern regions looking for a job and, more generally, for better living standard. In the following decades, many of them, after getting retired, came back towards their native regions, plumping Centre-North to South migration flows. Reasonably, such a return migration is not driven by economic factors such as per capita GDP or unemployment, but rather, other push and pull factors are likely to be important.

estimates, it turns out to be statistically significant in the regressions in which the network variable is excluded (columns 1, 3 and 5) and also in column (4) for the South to Centre-North flows in which the migration networks are also considered. Thus, while distance as a proxy of physical costs is an important determinant of South to Centre-North flows independent from the role played by migrant networks, the consideration of migrant networks for the whole sample on Italian regions and for the Centre-North to South flows shows that distance does not add any explicative power to the regressions.

Finally, given the evidence of non-stationarity and cointegration discussed in Section 5, in Table 9 we applied the DOLS estimator advocated by Mark and Sul (2003). Notice that across all regressions of Table 9, the null hypothesis of cross-sectional dependence is always rejected and this result comes not as a surprise given that this estimator is unable to deal with the cross-sectional dependence of the data. As far as the role of migrant networks is concerned, its estimated parameter is always highly statistically significant and roughly double with respect to the corresponding estimates in Tables 5–7. In addition, for the other estimated coefficients, there is much more variability when compared with those estimated in Tables 5–7.

[Tables 8 and 9]

8. Conclusions

This paper analysed the role of migration networks at the internal level for the Italian regions and briefly reviewed some issues regarding the link between the gravity models of migration and their empirical specifications. We have synthesised the different (micro and macro) approaches and tried to clarify how estimates should be performed in order to be coherent with the different theoretical perspectives. We then estimated various versions of the gravity model of migration across Italian regions inserting - and this is new in the literature regarding internal migration - a network variable (the stock of migrants at destination) inside an otherwise standard gravity model. We have shown that, by so doing, we can explain and interpret the determinants of internal migration in Italy more clearly than in previous empirical works. Specifically, when the stock of migrant is introduced, regression residuals are shown to be cross-sectional independent and, we have argued, the estimated model becomes coherent with the RUM put forward by the theoretical literature. According to our estimates, a 1% increase in the stock of internal migrants increases by about 1% the gross flows of internal migrants. Furthermore, the estimated impacts of the other variables are generally lower when the stock of migrants is considered. Finally, we have clarified the role of human capital, showing that it generally has a positive role at origin but exerts a negative impact at destination. These results regarding human capital partially conflict with those of the previous studies that, however, did not consider the stock of migrants at destination. Thus, we conclude that by introducing the network effects, we have come to a much better understanding of the determinants of migration flows across Italian regions.

More generally, it can be claimed that in order to explain migration flows (internal and international), the introduction of the network of migrants is simultaneously theoretically grounded and empirically relevant. As a matter of fact, it is likely that if network effects are not accounted for, the estimated parameters are biased, leading to a wrong assessment of their role in the migration process.

References

Amuedo-Dorantes C. and Mundra K. (2007) Social networks and their impact on the earnings of Mexican migrants. *Demography* 44: 849-863.

Anderson J. (2011) The gravity model. *Annual Review in Economics* 3: 133-160.

Bakewell O., Engbersen G., Fonseca M-L. and Horst C. (2016) *Beyond Networks. Feedback in International Migration*. Palgrave Macmillan, London.

Bansak C., Simpson N.B. and Zavodny M. (2015) *The Economics of Immigration*. Routledge, New York.

Beine M. (2016) The role of networks for migration flows: an update. *International Journal of Manpower* 37: 1154-1171.

Beine M., Docquier F. and Özden Ç. (2015) Dissecting network externalities in international migration. *Journal of Demographic Economics* 81: 379-408.

Beine M., Docquier F. and Özden Ç. (2011) Diasporas. *Journal of Development Economics* 95: 30-41.

Beine M., Bertoli S. and Fernández-Huertas Moraga J. (2016) A practitioners' guide to gravity models of international migration. *The World Economy* 39: 496-512.

Bertoli S. and Fernández-Huertas Moraga J. (2013) Multilateral resistance to migration, *Journal of Development Economics* 102: 79-100.

Bertoli S., Brücker H. and Fernández-Huertas Moraga J. (2016) The European crisis and migration to Germany. *Regional Science and Urban Economics* 60: 61-72.

Bodvarsson Ö. B. and Van den Berg H. (2013) *The Economics of Immigration: Theory and Policy*. Springer Science+Business Media, New York.

Borjas G. (1991) Immigration and self-selection. In Abowod J, Freeman R. (eds) *Immigration, Trade, and the Labor Market*. University of Chicago Press, Chicago.

Clark X., Hatton T. and Williamson J. (2007) Explaining U.S. immigration, 1971-1998. *The Review of Economics and Statistics* 89: 359-373.

Desbordes R. and Eberhardt M. (2014) Gravity models in the presence of slope heterogeneity and cross-section dependence. *Mimeo*.

Eberhardt M. (2012) Estimating panel time-series models with heterogeneous slopes. *The Stata Journal* 12: 61-71.

Eberhardt M. and Teil F. (2013) No mangoes in the tundra: Spatial heterogeneity in agricultural productivity analysis. *Oxford Economic Bulletin of Economics and Statistics* 75: 914-939.

Eberhardt M. and Teil F. (2010) Productivity analysis in global manufacturing production. Discussion Paper Series 515, *Department of Economics, University of Oxford, Mimeo*.

Egger P. and Pfaffermayr (2003) The proper panel econometric specification of the gravity equation: a three-way model with bilateral interaction effects. *Empirical Economics* 28: 571-580.

Etzo I (2011) The determinants of the recent interregional migration flows in Italy: A panel data analysis. *Journal of Regional Science* 51: 948-966.

Fidrmuc J. (2009) Gravity models in integrated panels. *Empirical Economics* 37: 435-446.

Garip F. and Asad L.A. (2015) Migrant Networks. In Scott R. and Kosslyn S., (ed.) *Emerging Trends in the Social and Behavioral Sciences*, John Wiley & Sons.

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

Hadri K. (2000) Testing for stationarity in heterogeneous panel data. *Econometrics Journal* 3: 148-161.

Haug S. (2008) Migration networks and migration decision-making. *Journal of Ethnic and Migration Studies* 34: 585-605.

Head K. and Mayer T. (2014) Gravity equations: Workhorse, toolkit, and cookbook, in Gopinath G., Helpman G. E. and Rogoff K. (Ed.) *Handbook of International Economics*, Vol. 4, 131-195, Elsevier, Oxford.

ISTAT (various years, a) Movimento migratorio della popolazione residente. Iscrizioni e cancellazioni anagrafiche, ISTAT, Roma.

ISTAT (various years, b) Forze di lavoro, ISTAT, Roma.

Karemera D., Iwuagwu. V. and Davis B. (2000) A gravity model analysis of international migration to North America. *Applied Economics* 32: 1745-1755.

Kao C. (1999) Spurious regression and residual-based tests for cointegration in panel data. *Journal of Econometrics* 90: 1-44.

King R. (2012a) Geography and migration studies: Retrospect and prospect. *Population Space and Place* 18: 134-153.

King R. (2012b) Theories and typologies of migration: An overview and a primer. Willy Brandt Series of Working Papers in International Migration and Ethnic Relations, 3/12.

Mark N. C. and Sul D. (2003) Cointegration vector estimation by panel DOLS and long-run money demand. *Oxford Bulletin of Economics and Statistics* 65: 655-680.

Mayda A.M. (2010) International migration: A panel data analysis of the determinants of bilateral flows. *Journal of Population Economics* 23: 1249-1274.

Massey D.S., Arango J., Hugo G., Kouaouci A., Pellegrino A. and Taylor J.E. (1993) Theories of international migration: A review and appraisal. *Population and Development Review* 19: 431-466.

McFadden D. (1974) Conditional logit analysis of qualitative choice behaviour. In Zarembka P. (ed.) *Frontiers in Econometrics*, New York, Academic Press.

Mocetti S, Porello C (2012) Le nuove migrazioni interne: tendenze nuove di un vecchio fenomeno. *Rivista di Politica Economica* CI: 275-310.

Munshi K. (2016) Community networks and migration. In Bramoullé Y., Galeotti A. and Rogers B. (ed.) *The Oxford Handbook of the Economics of Networks*, Oxford, Oxford University Press.

OECD/Eurostat/UNESCO Institute for Statistics (2015). ISCED 2011 Operational Manual: Guidelines for Classifying National Education Programmes and Related Qualifications, OECD Publishing, Paris.

Orefice G. (2015) International migration and trade agreements: the new role of PTAs. *Canadian Journal of Economics* 48: 310-334.

Ortega F. and Peri G. (2013) The role of income and immigration policies in attracting international migrants. *Migration Studies* 1: 47-74.

Pedersen P.J., Pytlikova M. and Smith N. (2008) Selection and network effects. Migration flows into OECD countries 1990-2000. *European Economic Review* 52: 1160-1186.

Pesaran M.H. (2007) A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics* 22: 265-312.

Pesaran M.H. (2006) Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74: 967-1012.

Pesaran M.H. (2004) General diagnostic tests for cross section dependence in panels. IZA Discussion Paper 1240.

Pesaran M.H. and Smith R.P. (1995) Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics* 68: 79-113.

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 Sciences* 96: 571-602.

Putnam D. (1993) *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, N.J.: Princeton University Press.

SVIMEZ (2011) *150 anni di statistiche Italiane: Nord e sud, 1861–2011*. Il Mulino, Bologna.

Table 1 – Fixed effects specifications.

	(2-FE)	Ortega-Peri (a)	Ortega-Peri (b)	Anderson
Year	Yes	Yes	Yes	Yes
....Origin-Dest.	Yes	No	Yes	Yes
Origin	No	Yes	Yes	No
Destination	No	Yes	Yes	No
Origin-Year	No	Yes	Yes	Yes
Dest.-Year	No	No	No	Yes

Table 2 - Pesaran (2004) CD test of cross section dependence.

Variable	All flows	South to Centre-North flows	Centre-North to South flows
Gross migration flows	430.66 [0.000]	58.44 [0.000]	98.44 [0.000]
Per capita GDP at origin	1447.35 [0.000]	384.66 [0.000]	353.45 [0.000]
Per capita GDP at destination		353.45 [0.000]	384.66 [0.000]
Unemployment rate at origin	961.01 [0.000]	288.23 [0.000]	314.34 [0.000]
Unemployment rate at destination		314.34 [0.000]	288.23 [0.000]
Population at origin	371.75 [0.000]	63.42 [0.000]	216.10 [0.000]
Population at destination		216.10 [0.000]	63.42 [0.000]
Average Years of schooling at origin	1556.82 [0.000]	391.58 [0.000]	392.14 [0.000]
Average Years of schooling at destination		392.14 [0.000]	391.58 [0.000]
Stock of previous migrants	1553.38 [0.000]	391.13 [0.000]	390.00 [0.000]

Notes: *p*-values in brackets. The Pesaran (2004) CD test is based on mean pair-wise correlation coefficients, it is normally distributed under the null hypothesis of no cross-sectional dependence, it is valid for N and T going to infinity in any order and it is robust to possible structural breaks.

Table 3a - Pesaran (2007) panel unit roots tests for the whole sample of 20 Italian regions.

	Gross migration flows	Per capita GDP	Unemployment rate	Population	Average years of schooling	Stock of migrants at destination
Lag 1	-3.06***	-2.18	-2.33***	-0.08	-3.46***	-2.84***
Lag 2	-2.48	-2.04	-2.46***	-0.47	-2.70***	-2.59**
Lag 3	-2.38	-1.68	-1.93	-0.34	-2.35	-2.47
Lag 4	-2.22	-1.66	-1.63	-0.27	-1.71	-2.26
Lag 5	-1.90	-1.68	-1.61	-0.14	-1.61	-1.98

Notes: Pesaran (2007) runs a test for unit roots in heterogeneous panels with cross-section dependence; the null hypothesis assumes that all series are non-stationary. The test has been run with constant and trend. 10%, 5% and 1% statistical levels of significance for the null hypothesis are indicated by *, ** and *** respectively.

Table 3b - Pesaran (2007) panel unit roots tests for the South to Centre-North flows.

	Gross migration flows	Per capita GDP		Unemployment rate		Population		Average years of schooling		Stock of previous migrants at dest.
		Origin	Dest.	Origin	Dest.	Origin	Dest.	Origin	Dest.	
Lag 1	-2.15	-2.51	-2.27	-3.37***	-3.05***	-2.18	-1.34	-3.79***	-3.46***	-2.76***
Lag 2	-1.60	-2.02	-2.26	-3.15***	-3.14***	-2.38	-1.61	-3.01***	-2.61	-2.43
Lag 3	-1.48	-1.55	-2.15	-3.16***	-2.40	-2.27	-1.69	-2.65**	-2.41	-1.95
Lag 4	-1.36	-1.77	-2.21	-2.57*	-2.31	-2.02	-1.58	-2.13	-1.84	-1.81
Lag 5	-1.07	-1.78	-1.94	-2.28	-1.90	-1.85	-1.14	-2.30	-1.57	-1.71

Notes: see Table 3a.

Table 3c - Pesaran (2007) panel unit roots tests for the Centre-North to South flows.

	Gross migration flows	Per capita GDP		Unemployment rate		Population		Average years of schooling		Stock of previous migrants at dest.
		Origin	Dest.	Origin	Dest.	Origin	Dest.	Origin	Dest.	
Lag 1	-3.27***	-2.27	-2.51	-3.05***	-3.37***	-1.34	-2.18	-3.46***	-3.79***	-2.64**
Lag 2	-2.62**	-2.26	-2.02	-3.14***	-3.15***	-1.61	-2.38	-2.61	-3.01***	-2.67**
Lag 3	-2.40	-2.15	-1.55	-2.40	-3.16***	-1.69	-2.27	-2.41	-2.65**	-2.46
Lag 4	-2.08	-2.21	-1.77	-2.31	-2.57*	-1.58	-2.02	-1.84	-2.13	-2.26
Lag 5	-1.83	-1.94	-1.78	-1.90	-2.28	-1.14	-1.85	-1.57	-2.30	-2.19

Notes: see Table 3a.

Table 4 - Kao (1999) residual cointegration tests.

	All flows	South to Centre-North flows	Centre-North to South flows
Modified DF *t*	-42.31	-26.93	-17.19
	[0.000]	[0.000]	[0.000]
DF *t*	-44.21	-24.40	-21.27
	[0.000]	[0.000]	[0.000]
Augmented DF *t*	2.82	1.37	1.45
	[0.002]	[0.085]	[0.074]

Notes: in all tests the null hypothesis is no cointegration. Number of lags for the specific autoregressive parameter selected with AIC criteria.

Table 5 – Estimation results for the whole sample of 20 Italian regions.

		2-FE		O-P (a)		O-P (b)		Anderson		CCE		AUG	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES AT ORIGIN	Per capita	-0.11	-0.02	-0.10	0.02	-0.07	0.05	-0.19	-0.00	-0.26	-0.17	-0.37***	-0.05
	GDP	[0.18]	[0.13]	[0.12]	[0.11]	[0.12]	[0.11]	[0.12]	[0.11]	[0.21]	[0.19]	[0.11]	[0.11]
	Unemp.	-0.02	-0.04	0.07**	0.01	0.07**	0.02	0.06**	0.01	0.08***	0.03	0.10***	0.04**
	rate	[0.03]	[0.03]	[0.03]	[0.02]	[0.03]	[0.02]	[0.03]	[0.02]	[0.03]	[0.03]	[0.02]	[0.02]
	Population	2.37***	1.90***	2.22***	1.86***	2.26***	1.94***	2.21***	1.86***	3.46**	2.98*	3.02***	1.44*
		[0.29]	[0.19]	[0.18]	[0.15]	[0.18]	[0.16]	[0.18]	[0.15]	[1.52]	[1.68]	[0.70]	[0.85]
	Av. years	0.40	0.87***	0.37	0.77***	0.29	0.66***	0.30	0.82***	0.27	0.40	0.10	-0.10
	of sch.	[0.32]	[0.24]	[0.25]	[0.21]	[0.25]	[0.22]	[0.25]	[0.21]	[0.32]	[0.34]	[0.21]	[0.20]
VARIABLES AT DESTINATION	Per capita	0.32*	0.16	0.32**	0.16	0.32**	0.17	0.37***	0.18	0.69***	-0.54***	0.50***	-0.05
	GDP	[0.17]	[0.12]	[0.16]	[0.11]	[0.16]	[0.12]	[0.13]	[0.14]	[0.20]	[0.20]	[0.11]	[0.11]
	Unemp.	-0.20***	-0.18***	-0.20***	-0.18***	-0.20***	-0.18***	-0.20***	-0.21***	-0.08**	0.04	-0.06***	0.01**
	rate	[0.03]	[0.03]	[0.03]	[0.02]	[0.03]	[0.02]	[0.03]	[0.03]	[0.03]	[0.03]	[0.02]	[0.02]
	Population	2.09***	0.65***	2.09***	0.61***	2.09***	0.74***	2.24***	0.67***	-3.05*	2.98*	-1.46*	-5.62***
		[0.27]	[0.20]	[0.26]	[0.19]	[0.26]	[0.20]	[0.15]	[0.19]	[1.64]	[1.68]	[0.78]	[0.88]
	Av. years	-0.35	-0.48**	-0.35	-0.49**	-0.35	-0.48**	-0.24	-0.23	0.27	0.50	0.10	0.11
	of sch.	[0.36]	[0.24]	[0.31]	[0.21]	[0.32]	[0.22]	[0.27]	[0.26]	[0.37]	[0.31]	[0.21]	[0.20]
	Stock of		0.94***		0.98***		0.89***		1.06***		6.97***		5.10***
	migrants		[0.05]		[0.01]		[0.05]		[0.06]		[0.20]		[0.10]
	Obs.	12540	12540	12540	12540	12540	12540	12540	12540	12920	12920	12920	12920
	Adj_R^2	0.98	0.98	0.97	0.97	0.98	0.98	0.98	0.98				
	CD test	6.26	0.53	3.73	-0.70	3.73	-0.73	1.02	-3.30	2.85	2.89	4.75	3.87
	[*p*-value]	[0.000]	[0.595]	[0.000]	[0.482]	[0.000]	[0.464]	[0.306]	[0.001]	[0.004]	[0.004]	[0.000]	[0.000]
	Stationarity	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)

Notes: unit 380; robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Constant included but not reported. In columns (3)-(6) the reported coefficients for monadic variables at origin are those of the second-stage regressions. In columns (7)-(8), the reported coefficients for monadic variables at both origin and destination are those of the second-stage regressions. In the second stage regressions, the number of observations depends on the number of origin-year (and destination-year) fixed effects estimated in the first-stage regressions. See Appendix B and the main text for more details. CD test is normally distributed under the null hypothesis of no cross-sectional dependence. Stationarity refers to the Pesaran (2007) CIPS tests for stationarity in the presence of cross-sectional dependence. The test is run up to five lags and in all estimates there is ample evidence that the null hypothesis of non stationary is rejected (full results are available upon request). For the CCE estimates, in column (9) the χ^2 test that the coefficients on cross-sectional averages are jointly zero is 528.76 (*p*-value=0.000), in column (10) it is 1859.8 (*p*-value=0.000). The user-written xtmg routine has been used (Eberhardt, 2012).

Table 6 – Estimation results for the South to Centre-North sample of Italian regions.

		2-FE		O-P (a)		O-P (b)		Anderson		CCE		AUG	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES AT ORIGIN	Per capita	-0.38	-0.32	-0.40	-0.31	-0.49**	-0.29	0.29	-0.41*	-0.11	-0.81**	-0.36**	-0.44**
	GDP	[0.32]	[0.23]	[0.25]	[0.23]	[0.24]	[0.23]	[0.24]	[0.24]	[0.31]	[0.36]	[0.18]	[0.18]
	Unemp.	0.28***	0.15***	0.23***	0.14***	0.25***	0.12***	0.26***	0.14***	0.00	-0.02	0.00	-0.08**
	rate	[0.06]	[0.05]	[0.06]	[0.05]	[0.05]	[0.04]	[0.05]	[0.05]	[0.06]	[0.05]	[0.04]	[0.03]
	Population	0.61	0.86*	0.42	0.67	0.20	0.84*	0.64	0.60	1.40	4.88	2.40*	0.40 [1.51]
		[0.62]	[0.49]	[0.47]	[0.44]	[0.44]	[0.44]	[0.44]	[0.42]	[3.57]	[3.30]	[1.42]	
	Av. years	-0.67	0.73**	-0.58	0.45	-0.76**	0.89***	-0.91***	0.54*	0.48	0.40	0.14	-0.01
	of sch.	[0.05]	[0.37]	[0.36]	[0.33]	[0.37]	[0.31]	[0.36]	[0.32]	[0.46]	[0.45]	[0.43]	[0.36]
VARIABLES AT DESTINATION	Per capita	1.13***	0.89***	1.13***	0.86***	1.13***	0.89***	1.07***	0.92***	-0.14	-0.75**	1.08***	0.71***
	GDP	[0.24]	[0.21]	[0.21]	[0.19]	[0.21]	[0.19]	[0.19]	[0.18]	[0.41]	[0.38]	[0.21]	[0.19]
	Unemp.	-0.31***	-0.25***	-0.31***	-0.24***	-0.31***	-0.25***	-0.31***	-0.25***	0.08	0.11**	-0.07**	-0.05**
	rate	[0.07]	[0.05]	[0.06]	[0.05]	[0.06]	[0.05]	[0.06]	[0.06]	[0.06]	[0.05]	[0.03]	[0.02]
	Population	4.73***	2.14***	4.73***	1.76***	4.73***	2.14***	4.75***	2.37***	-7.77***	-14.0***	2.55***	-5.53***
		[0.37]	[0.34]	[0.36]	[0.24]	[0.37]	[0.32]	[0.27]	[0.21]	[2.85]	[2.90]	[0.94]	[1.27]
	Av. years	0.62	-0.26	0.61	-0.39	0.61	-0.26	0.85*	0.48	0.16	-1.12*	0.29	-0.19
	of sch.	[0.61]	[0.49]	[0.59]	[0.44]	[0.60]	[0.47]	[0.52]	[0.48]	[0.60]	[0.65]	[0.40]	[0.35]
	Stock of		0.81***		0.93***		0.81***		0.77***		7.15***		3.97***
	migrants		[0.06]		[0.02]		[0.06]		[0.13]		[0.36]		[0.13]
	Obs.	3168	3168	3168	3168	3168	3168	3168	3168	3264	3264	3264	3264
	Adj_R^2	0.98	0.99	0.93	0.98	0.98	0.99	0.99	0.99				
	CD test	-1.89	-1.39	-2.27	-1.69	-2.27	-1.70	-3.29	-3.57	0.17	0.69	1.70	2.04
	[*p*-value]	[0.059]	[0.165]	[0.023]	[0.091]	[0.023]	[0.088]	[0.001]	[0.000]	[0.869]	[0.488]	[0.090]	[0.041]
	Stationarity	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)

Notes: units 96; robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Constant included but not reported. For the CCE estimates, in column (9) the χ^2 test that the coefficients on cross-sectional averages are jointly zero is 581.27 (p-value=0.000), in column (10) it is 700.83 (p-value=0.000). See Table 5 and the main text for more details.

Table 7 – Estimation results for the Centre-North to South sample of Italian regions

		2-FE		O-P (a)		O-P (b)		Anderson		CCE		AUG	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES AT ORIGIN	Per capita GDP	0.26	-0.04	0.32	0.11	0.30	0.07	0.23	-0.05	0.36	0.79*	-0.51**	-0.26
		[0.30]	[0.22]	[0.22]	[0.19]	[0.21]	[0.18]	[021]	[0.20]	[0.48]	[0.46]	[0.22]	[0.22]
	Unemp. rate	-0.17*	-0.21***	-0.08	-0.12**	-0.08	-0.13**	-0.10	-0.11*	0.04	0.02	0.02	-0.02
		[0.09]	[0.06]	[0.08]	[0.06]	[0.07]	[0.06]	[0.08]	[0.06]	[0.07	[0.06]	[0.03]	[0.03]
	Population	5.12***	2.97***	5.35***	3.77***	5.14***	3.50***	5.05***	3.69***	7.48***	5.03*	3.02**	-1.03
		[0.50]	[0.41]	[0.36]	[0.29]	[0.35]	[0.29]	[0.37]	[0.31]	[2.89]	[2.94]	[1.25]	[1.00]
	Av. years of sch.	1.77***	1.26***	2.05***	1.36***	1.86***	1.25***	2.01***	1.57***	-0.53	0.04	-0.00	0.48*
		[0.62]	[0.46]	[0.49]	[0.41]	[0.49]	[0.40]	[0.52]	[0.44]	[0.73]	[0.65]	[0.33]	[0.26]
VARIABLES AT DESTINATION	Per capita GDP	-0.01	-0.22	-0.01	-0.18	-0.01	-0.17	0.10	0.02	0.42	-0.15	0.29	0.37*
		[0.36]	[0.30]	[0.24]	[0.23]	[0.25]	[0.24]	[0.27]	[0.24]	[0.35]	[0.37]	[0.19]	[0.21]
	Unemp. rate	0.03	-0.00	0.03	0.00	0.03	0.01	0.02	-0.01	-0.05	-0.12	-0.00	-0.01
		[0.06]	[0.05]	[0.06]	[0.05]	[0.06]	[0.05]	[0.06]	[0.05]	[0.08]	[0.08]	[0.04]	[0.04]
	Population	0.29	0.50	0.29	0.46	0.29	0.44	0.14	0.59	-14.5***	-22.6***	-4.78***	-6.85***
		[0.91]	[0.63]	[0.75]	[0.54]	[0.76]	[0.56]	[0.66]	[0.57]	[4.47]	[4.97]	[1.54]	[1.53]
	Av. years of sch.	-2.19***	-0.93**	-2.20***	-1.17***	-2.20***	-1.27***	-2.05***	-1.17***	-0.08	-0.62	0.03	0.12
		[0.65]	[0.42]	[0.52]	[0.39]	[0.53]	[0.41]	[0.45]	[0.40]	[0.17]	[0.53]	[0.37]	[0.37]
	Stock of migrants		1.20***		0.97***		0.88***		0.77***		8.16		5.19***
			[0.12]		[0.03]		[0.15]		[0.16]		[0.43]		[0.16]
	Obs.	3168	3168	3168	3168	3168	3168	---	---	3264	3264	3264	3264
	Adj_R^2	0.97	0.98	0.92	0.98	0.98	0.98	0.98	0.98				
	CD test	-2.23	-0.62	-3.05	-2.86	-3.05	-2.88	-3.65	-3.74	2.22	2.06	2.06	4.09
	[*p*-value]	[0.025]	[0.536]	[0.002]	[0.004]	[0.002]	[0.004]	[0.000]	[0.000]	[0.027]	[0.039]	[0.039]	[0.000]
	Stationarity	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)

Notes: units 96; robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Constant included but not reported. For the CCE estimates, in column (9) the χ^2 test that the coefficients on cross-sectional averages are jointly zero is 202.38 (p-value=0.000), in column (10) it is 206.13 (p-value=0.000). See Table 5 and the main text for more details.

Table 8 – Estimation results for the model Ortega-Peri (a) with distance.

		Whole sample of 20 Italian regions		South to Centre-North sample		Centre-North to South sample	
		(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES AT ORIGIN	Per capita GDP	-1.10	0.02	-0.49	-0.31	0.32	0.08
		[0.12]	[0.11]	[0.24]	[0.23]	[0.21]	[0.18]
	Unemp. rate	0.07**	0.01	0.25**	0.13***	-0.09	-0.13**
		[0.03]	[0.02]	[0.05]	[0.04]	[0.07]	[0.06]
	Population	2.22***	1.86***	0.20	0.82*	5.25***	3.56***
		[0.18]	[0.15]	[0.44]	[0.43]	[0.36]	[0.29]
	Av. years of sch.	0.37	0.77***	-0.76**	0.79***	1.71***	1.27***
		[0.25]	[0.21]	[0.37]	[0.30]	[0.48]	[0.40]
	Distance	-0.48***	0.00	-1.80***	-0.20**	-1.76***	-0.13
		[0.07]	[0.02]	[0.25]	[0.09]	[0.26]	[0.10]
VARIABLES AT DESTINATION	Per capita GDP	0.32**	0.16	1.13***	0.87***	-0.01	-0.18
		[0.16]	[0.11]	[0.21]	[0.19]	[0.24]	[0.23]
	Unemp. rate	-0.20***	-0.18***	-0.31***	-0.24***	0.03	0.00
		[0.03]	[0.02]	[0.06]	[0.05]	[0.06]	[0.05]
	Population	2.09***	0.61***	4.73***	1.90***	0.29	0.45
		[0.26]	[0.19]	[0.36]	[0.25]	[0.75]	[0.54]
	Av. years of sch.	-0.35	-0.49***	0.61	-0.34	-2.20***	-1.20***
		[0.31]	[0.21]	[0.59]	[0.44]	[0.52]	[0.39]
	Stock of migrants		0.98***		0.88***		0.94***
			[0.01]		[0.02]		[0.03]
	Obs.	12540	12540	3168	3168	3168	3168
	Adj_R^2	0.86	0.97	0.95	0.98	0.95	0.98
	CD test	3.73	-0.70	-2.27	-1.69	-3.05	-2.87
	[*p*-value]	[0.000]	[0.483]	[0.023]	[0.091]	[0.002]	[0.004]
	Stationarity	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. See previous Tables and the main text for more details.

Table 9 – Estimation results with DOLS.

		Whole sample of 20 Italian regions		South to Centre-North sample		Centre-North to South sample	
		(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES AT ORIGIN	Per capita GDP	-0.39**	-0.49***	-1.28***	-0.75**	-0.46	-0.23*
		[0.16]	[0.15]	[0.39]	[0.36]	[0.37]	[0.30]
	Unemp. rate	0.16***	0.09**	0.53***	-0.00	-0.08	0.30***
		[0.04]	[0.03]	[0.08]	[0.06]	[0.11]	[0.08]
	Population	1.94***	0.61	1.23	1.54	4.27***	0.22
		[0.37]	[0.40]	[1.09]	[1.01]	[0.90]	[0.87]
	Av. years of sch.	-0.39	-1.66***	-1.55**	-1.30**	0.50	-4.19***
		[0.35]	[0.29]	[0.70]	[0.55]	[0.77]	[0.60]
VARIABLES AT DESTINATION	Per capita GDP	0.03	-0.77***	0.61*	-0.29	0.30	-0.80*
		[0.16]	[0.16]	[0.34]	[0.33]	[0.42]	[0.41]
	Unemp. rate	-0.09**	0.04	-0.29***	0.13*	0.05	-0.15**
		[0.04]	[0.04]	[0.10]	[0.07]	[0.09]	[0.07]
	Population	0.57	-0.41	4.23***	0.54	-4.88***	-4.61***
		[0.37]	[0.41]	[0.83]	[0.87]	[1.19]	[0.99]
	Av. years of sch.	-0.41	-1.09***	0.52	-1.30**	-1.59**	0.01
		[0.35]	[0.29]	[0.71]	[0.55]	[0.76]	[0.53]
	Stock of migrants		1.93***		1.79***		2.44***
			[0.07]		[0.14]		[0.17]
	Obs.	11780	11780	2976	2976	2976	2976
	Adj_R^2	0.98	0.99	0.99	0.99	0.98	0.99
	CD test	55.16	29.08	21.15	8.64	9.32	4.19
	[*p*-value]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	Stationarity	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)

Notes: *** p<0.01, ** p<0.05, * p<0.1. See previous Tables and the main text for more details.

Appendix A.

Table A1. Descriptive statistics of migration flows

	Dest	Pie	VdA	Lom	TAA	Ven	FVG	Lig	EmR	Tos	Umb	Mar	Laz	Abr	Mol	Cam	Pug	Bas	Cal	Sic	Sar
	Orig	*PANEL A: Origin CENTRE-NORTH – Destination: CENTRE-NORTH*												*PANEL B: Origin CENTRE-NORTH – Destination: SOUTH*							
mean	**Pie**		457	4304	155	1050	323	2992	1158	1019	183	361	1343	380	115	1667	1880	341	1832	2833	1163
s.d.			96.0	531.5	41.6	360.0	112.7	396.1	226.6	245.8	60.7	90.7	436.9	151.3	53.4	832.3	1121.1	235.2	987.1	1478.1	438.8
max			630	5446	272	1758	551	3913	1666	1648	348	551	2233	819	256	4014	4880	920	4661	6735	2328
min			302	3562	63	576	153	2226	868	705	96	222	827	202	51	858	763	122	903	1367	666
mean	**VdA**	331		103	8	34	13	49	70	39	6	13	43	13	3	28	35	5	81	42	38
s.d.		63.7		20.7	4.8	9.9	11.4	10.0	56.6	12.1	3.8	13.1	9.1	5.6	2.0	8.3	20.0	3.8	18.1	11.9	12.6
max		525		135	22	54	51	77	200	81	15	81	64	27	7	48	123	20	117	67	81
min		246		60	0	17	2	26	23	19	0	4	20	4	0	15	9	0	43	18	19
mean	**Lom**	4498	133		717	3055	781	2360	4200	2081	375	893	2407	741	176	3023	3366	434	2607	4557	1457
s.d.		377.8	21.0		116.0	573.1	209.2	278.8	519.2	311.2	101.7	158.9	483.6	226.3	76.2	815.8	1166.0	201.3	726.2	1422.8	283.5
max		5551	172		1052	4208	1217	3014	5747	3077	683	1241	3491	1309	431	5628	6102	936	4869	8314	2117
min		4003	94		504	2304	457	1852	3581	1673	246	655	1740	460	89	2204	2008	229	1735	3072	1067
mean	**TAA**	114	8	528		758	119	55	253	147	34	70	204	48	8	141	133	13	83	161	67
s.d.		31.4	5.0	129.4		186.3	34.8	12.2	42.1	37.6	13.8	23.9	59.3	20.8	3.7	28.9	34.1	4.7	22.3	39.2	14.4
max		182	22	876		1317	192	85	386	235	64	132	320	124	17	201	214	22	157	255	89
min		67	0	372		576	76	36	189	107	15	34	113	21	0	84	77	4	52	98	37
mean	**Ven**	643	26	2652	737		1725	252	1594	685	117	231	960	184	38	735	732	57	350	885	313
s.d.		147.0	6.7	365.0	88.6		215.5	56.4	146.9	202.6	25.9	41.0	193.4	57.7	11.6	119.6	92.0	11.3	55.9	187.7	82.9
max		955	41	3734	876		2400	406	2018	1231	196	336	1379	360	69	981	932	88	467	1286	466
min		476	15	1985	513		1356	175	1328	409	70	161	725	108	20	548	530	33	253	547	203
mean	**FVG**	202	11	628	122	1557		101	291	192	40	86	445	74	15	378	296	23	112	342	110
s.d.		50.4	4.6	137.6	23.6	235.2		37.5	39.5	58.1	12.0	24.2	139.8	24.3	6.9	85.1	80.7	9.1	34.1	85.2	26.2
max		305	19	870	173	2260		173	376	364	74	154	799	134	30	547	522	47	201	508	184
min		115	2	471	69	1222		53	221	133	21	55	275	36	5	210	172	9	63	200	64
mean	**Lig**	3065	67	2056	87	332	138		720	1450	80	128	653	119	22	441	412	52	436	769	546
s.d.		429.1	13.1	336.6	19.5	92.4	47.6		142.0	306.8	27.8	47.5	242.2	62.7	12.0	169.9	244.7	24.7	170	347.0	212.4
max		4018	88	2629	126	500	218		988	2292	143	227	1156	296	56	843	942	107	840	1548	1015
min		2469	44	1570	59	209	66		502	1108	35	63	372	44	5	258	174	15	209	416	273
mean	**Em**	636	29	3088	230	1452	238	512		1100	190	958	1010	386	98	1829	1384	181	722	1446	458
s.d.	**R**	89.6	8.2	315.3	40.6	133.0	31.4	83.2		164.5	36.6	110.4	166.7	73.9	25.3	541.4	267.4	47.4	151.5	306.3	85.6
max		893	49	3880	363	1828	316	785		1605	282	1271	1385	593	189	2657	1763	327	982	1981	627
min		502	17	2450	169	1171	191	411		887	142	751	706	302	50	1094	964	112	461	871	308
mean	**Tos**	619	30	1498	114	502	157	1113	1234		631	258	1889	198	51	1403	592	141	496	1221	514
s.d.		97.5	11.9	201.9	15.3	78.0	35.3	155.9	122.8		128.6	53.4	385.8	57.0	17.0	239.2	157.6	50.2	79.8	268.4	111.9
max		788	70	1833	139	639	272	1428	1489		1004	398	200	313	92	2048	926	263	671	1906	676
min		479	12	1101	90	325	118	823	988		432	185	1339	119	24	1030	409	89	389	799	297

Table A1 (cont.). Descriptive statistics of migration flows.

mean	**Umb**	113	6	295	28	109	39	53	217	535		262	1167	94	15	274	121	17	101	132	72
s.d.		16.5	4.2	43.8	8.5	17.5	10.4	11.7	32.4	111.6		57.1	263.2	25.6	6.0	117.3	11.0	7.1	33.4	18.5	13.8
max		164	18	419	52	141	62	82	283	816		419	1882	170	35	522	176	38	163	172	101
min		81	0	223	11	80	20	36	135	401		198	847	58	5	140	91	7	53	104	48
mean	**Mar**	189	7	591	48	235	74	78	1025	263	253		782	551	37	352	442	32	81	233	78
s.d.		36.9	3.4	95.9	11.3	39.7	20.3	23.8	137.3	59.5	70.4		226.9	142.1	8.6	172.1	78.2	8.2	13.6	69.9	14.6
max		255	15	856	76	322	114	144	1281	497	421		1285	960	62	741	591	48	112	405	107
min		120	0	409	25	175	42	47	700	199	116		526	378	24	161	307	13	55	137	49
mean	**Laz**	1110	52	2496	284	1144	511	579	1346	2367	1850	1152		2032	380	3164	1444	239	1241	1591	1151
s.d.		273.8	14.6	394.8	44.3	176.9	144.7	203.5	175.4	520.0	360.0	313.1		412.6	82.0	677.8	424.7	56.2	293.2	532.4	295.4
max		1676	81	3310	383	1584	844	1106	1644	3409	2661	1841		2784	545	4779	2235	326	1727	2618	1845
min		705	26	1822	205	880	363	357	1091	1721	1344	744		1466	248	2292	943	142	758	1049	739
		PANEL C: Origin SOUTH – Destination: CENTRE-NORTH												***PANEL C: Origin SOUTH – Destination: SOUTH***							
mean	**Abr**	250	11	688	53	232	79	88	561	250	114	601	1892		249	396	402	35	76	158	64
s.d.		68.2	4.6	120.1	12.2	35.2	19.8	38.0	76.4	57.4	26.1	118.3	483.7		98.9	59.1	76.9	10.3	25.3	36.2	16.1
max		468	23	1053	80	301	118	187	723	385	187	879	2946		478	531	545	56	176	210	100
min		169	2	523	35	165	44	44	434	179	83	452	1342		145	282	276	20	46	91	38
mean	**Mol**	101	4	219	13	59	20	22	211	98	30	68	495	381		340	218	18	30	40	19
s.d.		32.8	2.7	49.5	5.0	12.1	6.1	8.6	29.8	20.0	7.7	14.4	125.8	113.8		79.2	68.2	11.0	12.8	17.9	6.9
max		202	11	378	24	97	30	44	262	133	48	99	806	727		609	350	54	63	88	35
min		60	0	143	4	41	5	9	139	65	18	39	336	249		224	117	5	8	16	9
mean	**Cam**	2402	53	6157	345	1679	826	746	4781	3450	721	955	6693	808	581		1295	582	999	982	408
s.d.		593.7	10.6	707.7	95.5	282.3	196.8	194.1	1389.5	320.0	209.0	389.4	1177.7	183.5	136.2		420.5	149.1	170.2	255.4	67.9
max		3860	84	7815	610	2304	1208	1290	7551	4194	1007	1563	9420	1542	1041		2235	869	1444	1389	530
min		1584	37	5110	215	1182	478	533	2531	2862	317	416	4690	563	345		847	283	746	647	269
mean	**Pug**	2206	41	5412	303	1512	547	522	3487	1212	256	957	2529	736	322	1076		668	504	785	197
s.d.		774.8	10.1	947.0	90.5	251.1	92.1	234.1	849.4	199.7	42.6	254.7	472.9	93.5	62.9	318.5		216.4	155.8	340.3	58.2
max		3628	64	7308	434	2073	739	1030	4899	1787	377	1373	3396	918	479	1729		1156	796	1308	317
min		1251	17	4224	149	1129	340	261	1935	984	153	523	1873	570	230	693		398	291	393	118
mean	**Bas**	453	7	809	29	128	44	77	541	360	50	76	502	71	20	520	723		242	93	21
s.d.		186.0	3.3	216.1	9.1	17.5	12.8	35.0	75.2	81.1	13	13.8	98.1	15.9	7.8	163.5	284.4		92.1	34.8	7.4
max		937	16	1482	54	165	70	160	736	639	78	104	767	111	43	936	1554		473	159	35
min		228	0	583	14	82	22	35	371	267	30	45	330	50	6	308	438		123	45	7
mean	**Cal**	2457	175	4757	175	709	191	656	1864	1151	237	179	2392	136	40	771	570	230		1145	91
s.d.		915.4	64.3	1139.8	26.9	108.7	33.1	237.8	306.1	128.5	56.5	34.9	394.7	25.5	14.7	143.1	155.5	78.0		377.7	24.0
max		4164	301	7302	247	898	258	1099	2310	1445	336	255	3349	193	86	1071	926	388		1728	151
min		1314	81	3444	133	484	133	329	1147	955	116	113	1700	92	19	600	348	128		676	49
mean	**Sic**	3438	65	7339	324	1711	629	989	2981	2235	250	440	2581	221	55	811	809	97	1038		332
s.d.		1023.1	13.1	1218.7	86.6	408.5	157.3	294.9	620.4	392.0	47.1	136.3	634.5	41.8	18.9	194.7	325.5	35.3	374.4		91.7
max		5524	94	10165	491	2564	950	1566	4142	3148	356	704	3793	310	86	1108	1325	176	1736		485
min		2101	42	5756	171	1122	329	626	1855	1676	156	227	1885	141	28	539	424	39	589		212
mean	**Sar**	1009	61	1676	118	481	153	487	731	750	116	127	1221	78	18	244	173	18	83	282	
s.d.		258.7	15.0	237.0	36.5	148.5	37.6	201.6	128.7	195.8	25.7	27.8	389.2	20.5	7.9	42.1	59.0	6.5	26.7	87.5	
max		1507	100	2246	175	797	239	917	1064	1158	184	180	1944	128	39	371	307	30	140	464	
min		641	33	1334	47	296	79	280	509	493	64	70	783	49	8	176	102	7	40	176	

Table A2 – Summary statistics.

Origin Region		Variables at origin of migration flows: Per cap. GDP (€, 2000)	Un. rate (%)	Population (thousand)	Average Years School.	Variables at destination of migration flows: Per cap. GDP (€, 2000)	Un. Rate (%)	Population (thousand)	Average Years School.	Stock of migrants
(1)		(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
					Centre-Northern regions					
Pie	Mean	20754	7.14	4332845	7.44	18120	10.40	2795513	7.38	25150
	s. d.	2528	1.73	82593	1.23	5431	5.93	2301714	1.28	24963
	max	24066	10.53	4488250	9.62	29897	28.04	9897972	10.43	131025
	min	15926	4.08	4212943	5.39	8245	2.44	112262	5.00	497
VdA	Mean	27641	4.65	118408	7.30	17758	10.53	3017326	7.39	868
	s. d.	1357	1.23	5284	1.23	5019	5.84	2235119	1.28	1448
	max	29897	8.25	128003	9.53	28572	28.04	9897972	10.43	9944
	min	24409	2.93	112262	5.25	8245	2.44	314141	5.00	7
Lom	Mean	24716	5.37	9086978	7.67	17912	10.50	2545296	7.37	37292
	s. d.	3412	1.37	331471	1.28	5220	5.87	1807711	1.27	33231
	max	28572	8.02	9897972	9.87	29897	28.04	5852637	10.43	134005
	min	18011	3.39	8818278	5.54	8245	2.44	112262	5.00	276
TAA	Mean	24214	4.04	928423	7.47	17938	10.57	2974693	7.38	2989
	s. d.	2639	1.32	55631	1.23	5276	5.80	2281830	1.28	4110
	max	27349	7.14	1044950	9.40	29897	28.04	9897972	10.43	23522.
	min	19121	2.44	870611	5.33	8245	2.84	112262	5.00	29
Ven	Mean	21415	5.51	4516420	7.32	18085	10.49	2785852	7.38	12223
	s. d.	3355	1.63	196864	1.30	5388	5.87	2294468	1.28	14432
	max	25719	9.42	4914328	9.61	29897	28.04	9897972	10.43	83186
	min	15149	3.36	4326009	5.23	8245	2.44	112262	5.00	90
FVG	Mean	19709	6.10	1204501	7.75	18175	10.46	2960163	7.36	5066
	s. d.	3609	1.76	18751	1.35	5420	5.90	2293954	1.27	7384
	max	24386	9.10	1235243	10.35	29897	28.04	9897972	10.43	47714
	min	13586	3.40	1177057	5.69	8245	2.44	112262	5.00	38
Lig	Mean	19041	8.25	1656594	7.80	18210	10.34	2936369	7.36	11855
	s. d.	2381	2.07	81532	1.22	5463	5.95	2309595	1.28	15816
	max	22448	11.60	1823977	10.01	29897	28.04	9897972	10.43	88962
	min	14752	4.79	1568004	5.66	8245	2.44	112262	5.00	5.00
EmR	Mean	23191	5.45	4028728	7.51	17992	10.49	2811520	7.37	14285
	s. d.	3394	1.64	167976	1.24	5312	5.87	2311043	1.28	15745
	max	27580	8.55	4423051	9.67	29897	28.04	9897972	10.43	96071
	min	17255	2.84	3894993	5.48	8245	2.44	112262	5.00	52
Tos	Mean	20433	7.31	3567808	7.38	18137	10.39	2835778	7.38	12208
	s. d.	2796	1.69	74064	1.33	5431	5.94	2321524	1.27	12075
	max	24098	9.48	3747134	9.90	29897	28.04	9897972	10.43	58131
	min	15431	4.36	3490430	5.33	8245	2.44	112262	5.00	79
Umb	Mean	17754	8.32	830229	7.54	18278	10.34	2979862	7.37	3601
	s. d.	2218	2.34	29136	1.29	5468	5.95	2277092	1.28	5885
	max	20745	12.72	895116	9.78	29897	28.04	9897972	10.43	37129
	min	13856	4.61	803923	5.29	8245	2.44	112262	5.00	24
Mar	Mean	18345	6.32	1458759	7.32	18247	10.45	2946781	7.38	5191
	s. d.	3020	1.52	47902	1.35	5449	5.91	2303388	1.27	6060
	max	22671	10.94	1554261	9.86	29897	28.04	9897972	10.43	30851
	min	13563	3.78	1407608	5.17	8245	2.44	112262	5.00	19
Laz	Mean	22359	10.08	5194867	8.28	18036	10.25	2750144	7.33	23695
	s. d.	2775	1.72	187321	1.18	5370	5.98	2261737	1.26	19267
	max	26371	12.78	5758642	10.43	29897	28.04	9897972	10.35	96364
	min	16807	6.38	4972748	6.20	8245	2.44	112262	5.00	111

Table A2 (cont.). – Summary statistics.

Origin Region		Variables at origin of migration flows: Per cap. GDP (€, 2000)	Un. rate (%)	Population (thousand)	Average Years School.	Variables at destination of migration flows: Per cap. GDP (€, 2000)	Un. Rate (%)	Population (thousand)	Average Years School.	Stock of migrants
(1)		(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
					Southern regions					
Abr	Mean	16021	9.00	1263259	7.50	18369	10.30	2957071	7.38	6294
	s.d.	1892	1.75	35709	1.41	5451	5.97	2296271	1.27	9331
	max	18391	11.93	1334520	10.24	29897	28.04	9897972	10.43	59586
	min	12192	5.36	1212643	5.35	8245	2.44	112262	5.00	68
Mol	Mean	13544	12.46	324422	7.50	18500	10.12	3006483	7.40	2480
	s. d.	2021	2.81	5571	1.20	5360	5.93	2248637	1.28	3173
	max	16585	17.46	330614	9.15	29897	28.04	9897972	10.43	15654
	min	9819	8.06	314141	5.11	8245	2.44	112262	5.00	7
Cam	Mean	12277	18.82	5666106	7.18	18566	9.79	2725342	7.39	31554
	s. d.	1188	4.60	116065	1.00	5303	5.54	2232222	1.29	40476
	max	14129	25.81	5852637	8.91	29897	28.04	112262	10.43	202712
	min	9776	11.17	5421067	5.33	8245	2.44	9897972	5.00	116
Pug	Mean	12291	14.66	4014821	6.86	18566	10.01	2812251	7.41	22436
	s. d.	1460	2.91	68241	1.05	5300	5.86	2311695	1.28	28872
	max	14245	20.85	4103430	8.58	29897	28.04	9897972	10.43	168181
	min	9558	8.44	3843125	5.00	8245	2.44	112262	5.06	114
Bas	Mean	12536	15.74	601302	7.05	18553	9.95	2991910	7.40	5276
	s. d.	2176	3.53	11252	1.36	5302	5.79	2265038	1.27	6333
	max	15578	22.58	612229	9.84	29897	28.04	9897972	10.43	26910
	min	8842	9.43	579039	5.06	8245	2.44	112262	5.00	17
Cal	Mean	11639	19.60	2036059	7.07	18600	9.75	2916397	7.40	17662
	s. d.	1878	5.29	41746	1.11	5250	5.44	2318986	1.28	25620
	max	14321	28.04	2085397	8.90	29897	25.81	9897972	10.43	145007
	min	8245	11.12	1970735	5.17	8842	2.44	112262	5.00	102
Sic	Mean	12606	18.50	4982216	6.97	18549	9.81	2761336	7.40	25825
	s. d.	1138	4.48	44322	1.05	5323	5.57	2273653	1.28	37769
	max	14429	25.24	5080904	8.80	29897	28.04	9897972	10.43	225108
	min	10482	10.26	4893755	5.16	8245	2.44	112262	5.00	136
Sar	Mean	14553	17.47	1634851	7.11	18447	9.86	2937513	7.40	7989
	s. d.	1720	3.30	19840	1.19	5409	5.69	2309026	1.28	10302
	max	17004	21.53	1660431	9.20	29897	28.04	9897972	10.43	52134
	min	11628	9.81	1581274	5.17	8245	2.44	112262	5.00	60

Appendix B.

To retrieve the effects of monadic variables in the structural gravity specification derived from Anderson (2011) model and in the two versions of the Ortega and Peri (2013) model, we follow the two-step procedure suggested by Head and Mayer (2014).

With reference to trade determinants in a cross section gravity equation, Head and Mayer (2014; pages 157-158) highlight that when importer and exporter fixed effects are introduced into a trade equation, a number of country-specific (i.e. monadic variables) trade determinants cannot be identified. The solution proposed by Head and Mayer (2014) to overcome such a problem is a two-step estimator. In this Appendix, we sketch the procedure for estimating the structural gravity specification derived from Anderson (2011) model in which both origin-year and destination-year fixed effects are considered among the regressors. For the Ortega and Peri (2013) model, a similar method easily follows.

In the first step, we need to estimate two equations, the first without the stock of migrants:

(B1a) $\ln m_{ijt} = \vartheta_{it} + \theta_{jt} + \nu_{ij} + e^{1}_{ijt}$

the second including the stock of migrants:

(B1b) $\ln m_{ijt} = \vartheta^{stk}_{it} + \theta^{stk}_{jt} + \nu^{stk}_{ij} + \varphi_{ij} \ln stk_{ijt} + e^{2}_{ijt}$

in equations (B1a)-(B1b) ϑ_{it} and ϑ^{stk}_{it} represent origin-year dummies, θ_{jt} and θ^{stk}_{jt} are destination-year dummies, ν_{ij} and ν^{stk}_{ij} are origin-destination dummies, e^{1}_{ijt} and e^{2}_{ijt} are two error terms.

In the second step, we regress the estimated origin-year ($\hat{\vartheta}_{it}$ and $\hat{\vartheta}^{stk}_{it}$) and destination-year ($\hat{\theta}_{jt}$ and $\hat{\theta}^{stk}_{jt}$) fixed effects on the monadic variables (per capita GDP, unemployment, population and human capital) to recover, respectively, the outward (at origin) and the inward (at destination) effects of these monadic variables on bilateral migration flows. More in details, to recover the effects of monadic variables at origin for the case in which the stock of migrants at destination is excluded from the regression, we run the following regression:

(B2a) $\hat{\vartheta}_{it} = \lambda_{out,t} + \varsigma_{out,i} + \omega_{1,out} \ln pcy_{it} + \omega_{2,out} \ln u_{it} + \omega_{3,out} \ln n_{it} + \omega_{4,out} \ln h_{it} + e^{3}_{ijt}$

whereas, the effects of monadic variables at origin, when the stock of migrants is included in the regression, is retrieved from the following regression:

(B2b) $\hat{\vartheta}^{stk}_{it} = \lambda^{stk}_{out,t} + \varsigma^{stk}_{out,i} + \omega^{stk}_{1,out} \ln pcy_{it} + \omega^{stk}_{2,out} \ln u_{it} + \omega^{stk}_{3,out} \ln n_{it} \omega^{stk}_{4,out} \ln h_{it} + e^{4}_{ijt}$

Analogously, to estimate the coefficients of monadic variables at destination (inward effect), we run the following two regressions:

(B3a) $\hat{\theta}_{jt} = \lambda_{in,t} + \varsigma_{in,j} + \omega_{1,in} \ln pcy_{jt} + \omega_{2,in} \ln u_{jt} + \omega_{3,in} \ln n_{jt} + \omega_{4,in} \ln h_{jt} + e^{5}_{ijt}$

(B3b) $\hat{\theta}^{stk}_{jt} = \lambda^{stk}_{in,t} + \varsigma^{stk}_{in,j} + \omega^{stk}_{1,in} \ln pcy_{jt} + \omega^{stk}_{2,in} \ln u_{jt} + \omega^{stk}_{3,in} \ln n_{jt} \omega^{stk}_{4,in} \ln h_{jt} + e^{6}_{ijt}$

Notice that in these second-step estimates, time ($\lambda_{out,t}$, $\lambda_{out,t}^{stk}$, $\lambda_{in,t}$ and $\lambda_{in,t}^{stk}$) and origin ($\varsigma_{out,i}$, $\varsigma_{out,i}^{stk}$) or destination ($\varsigma_{in,j}$, $\varsigma_{in,j}^{stk}$) fixed effects are introduced. Moreover, e_{ijt}^{3}, e_{ijt}^{4}, e_{ijt}^{5} and e_{ijt}^{6} are error terms.

It should also be remarked that because of multicollinearity some of the first step origin-year and destination-year fixed effects cannot be estimated. Hence, in the second step, the number of observations is never the full across the different specifications (680 for the total sample, 272 for the South to Centre-North sample and 408 for the Centre-North to South sample). In our estimates of the Anderson (2011) model, the minimum number of observation in the second-step regressions is never lower than 623 for the whole sample, 255 for the South to Centre-North sample and 355 for the Centre-North to South sample. For the Ortega and Peri (2013) models, the figures are 643, 256 and 384, respectively.