doi:10.1111/pirs.12211

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*

Romano Piras

University of Cagliari, Department of Economics and Business, Viale S. Ignazio, 1709123 Cagliari, Italy
(e-mail: pirasr@unica.it)

Received: 3 February 2015 / Accepted: 12 November 2015

Abstract. We estimate a gravity model of internal migration with human capital across Italian regions during the 1970–2005 time period. The estimates confirm that the macroeconomic variables are the main drivers of migration flows. As for human capital, while at destination it has had no role, at origin it has worked as a restraining factor. Such a restraining role has mainly worked in the Centre-North to South direction. We interpret this result in terms of agglomeration economies that makes the Centre-North as the core and the South as the periphery of Italy. We have framed our analysis inside a cointegration setup and applied both homogeneous and heterogeneous estimators, proving that heterogeneous estimators are more appropriate.

JEL classification: J61, R23

Key words: Gravity model, Italy, internal migration, human capital

1 Introduction

During the last decades the empirical economic literature on internal migration across Italian regions has highlighted the fundamental role of the macroeconomic variables as the main determinants of interregional flow. Such a literature has been growing in the recent years (Biagi et al. 2011; Etzo 2011; Mocetti and Porello 2012; Piras 2012a, 2012b; Lamonica and Zagaglia 2013; Fratesi and Percoco 2014), yet two main voids need to be filled. First, a systematic investigation of the push and pull factors in the long-run has never been pursued. As a matter of fact, the great majority of the recent empirical works is interested in the recent upsurge of internal mobility starting from the middle of the 1990s and ignores the long-run perspective of the phenomenon. In addition, notwithstanding the empirical evidence at both internal and international level (Hunt 2006; Mayda 2010) that pull and push factors operate asymmetrically, virtually none of the available evidence for Italy looks at push and pull factors separately. Second and more important, the role of regional human capital has not been investigated.

* The author gratefully acknowledge the editor Alessandra Faggian and two anonymous reviewers for their helpful comments on an earlier version of the paper. The usual disclaims apply.

As a driver of bilateral migration flows, human capital is expected to have opposite effects in the sending and the receiving economies (countries or regions) whenever they are structural different, namely whenever countries or regions are heterogeneous. According to Borjas (1991) migration is expected to be higher, when the mean level of education in the source country is higher and lower when the mean level of education in the destination country is higher. Yet, it is conceivable there is an opposite effect of human capital on migration flows on the basis of the new economic geography models popularized in the literature by Krugman (1991) and Fujita and Thiesse (2002). In these models agglomeration forces, led by economies of scale, attract individuals and firms into a core region. The core region accumulates human capital that further attracts more individuals and firms and, at the same time, prevents individuals from leaving the region. Given these two alternative theoretical predictions, the answer as to how human capital impacts on migration flows have to rely on empirical studies that, quite surprisingly, are scanty. At the international level, Clark et al. (2007), Pedersen et al. (2008), Hatton and Williamson (2010) and Mayda (2010) introduce human capital for different samples of countries and distinct time periods finding mixed results, not robust to alternative specifications of estimated regressions. Altogether, their results seem to indicate that the average level of human capital might affect migration negatively at destination and positively at origin. However, the few studies that investigate the role of human capital at internal level – Mitze and Reinkowski (2011) for Germany, Maza (2006) and Maza and Villaverde (2004) for Spain – do not confirm these feeble results. Indeed, none of these papers find any significant effect for regional human capital in explaining the migration patterns neither in Germany, nor in Spain. Contrary to the aforementioned empirical results, in this paper while we do not find statistically significant links between migration flows and human capital at destination, we detect a negative role for human capital at origin. In line with the results of Fratesi and Percoco (2014), who find that human capital gains from migration in Italy are larger in Centre-Northern regions where the human capital stock is large than in Southern ones, we interpret this result in terms of agglomeration economies that make the Centre-Northern regions the core and the Southern regions the periphery of Italy.

We tackle all these issues from three different geographical perspectives. First, we consider the 20 Italian regions, contemporaneously, as source and as destination regions. Second, we take the eight Southern regions as a source and the 12 Centre-Northern regions as destination. Third, we reverse the perspective: the Centre-Northern regions are considered as sources and the Southern ones as destinations. By so doing, along with the overall internal migration flows, as well as the traditional South to Centre-North flows, on the one hand we are able to explore the still under-investigated and less-understood Centre-North to South migration phenomenon; on the other hand we can untangle the dissimilar and asymmetric role played by human capital in these two geographical areas. To accomplish this task, we extend the gravity model of migration proposed by Karemera et al. (2000) by including human capital among the determinants of internal migration flows and investigate one of the longest time periods ever scrutinized, starting from 1970 up to 2005.

From an econometric point of view, two main questions need to be properly tackled. The first one is the order of integration of the variables under scrutiny. The panels under study are not short and stationary cannot be assumed to hold, rather a careful analysis of the dynamic characteristics of the series must be carried out as a preliminary step. The second question is whether a homogeneous or a heterogeneous estimator is more appropriate. When T is large, heterogeneous estimators provide consistent estimates of long-run coefficients, but these will be inefficient if slope homogeneity holds. Yet, a full homogeneity assumption 'can produce inconsistent and potentially very misleading estimates of the average values of the parameters unless the slope coefficients are identical' (Pesaran et al. 1999, p. 622). In order to tackle this second question, along with standard homogeneous panel data estimator – two way fixed effect (2FE), fully modified ordinary least squares (FMOLS) and dynamic ordinary least squares (DOLS) – we apply also three heterogeneous panel data estimators, namely the mean group (MG) estimator advocated by Pesaran and Smith (1995), the common correlated effects mean group estimator (CCEMG) proposed by Pesaran (2006) and the augmented mean group estimator (AMG) introduced by Eberhardt and Teal (2010). In particular, the CCEMG and the AMG estimators are very useful in our context since it is very likely that our data is plagued by the presence of cross-section dependence, common factors along with non-stationary of the series. Quite surprisingly, cross-section dependence is almost always ignored in empirical research on migration. Given that Italian regions form part of a single country and are deeply integrated among themselves, cross-section dependence is very likely to be present.

To the best of our knowledge, there are only two empirical studies on migration that use heterogeneous panel data methods inside a gravity model framework with common factors: Bertoli and Fernández-Huertas Moraga (2013) and Bertoli et al. (2013). The former studies international migration flows from 61 origin countries to Spain, the latter uses data from 28 origin countries to Germany. Thus, in the present study, for the first time these novel estimators are applied in the field of internal migration and, more generally, for the first time they are applied to the general case of multiple origin and multiple destination units (regions).

The paper is organized as follows. Section 2 provides an updated review of the empirical literature on internal migration in Italy. In Section 3 we present the extended version of the Karemera et al. (2000) model in which human capital is included among the determinants of migration flows. In this Section a detailed discussion on how human capital might affect migration flows is done. Section 4 illustrates the econometric approach. The empirical analysis is conducted in Section 5 while Section 6 concludes.

2 A review of empirical literature on internal migration in Italy

Internal migration in Italy has principally been from the South (the *Mezzogiorno*) to the Central-Northern regions (Figure 1). Descriptive data reported by Pugliese (2002) shows that 3,708,392 individuals moved from the Southern towards the Central-Northern regions during the period from 1951 to 1975, and that 1,363,553 individuals went the other way. Piras and Melis (2007) update to the 1971–2002 time period the migration data and report that 3,881,437 individuals emigrated from the South to the Centre-North, compared with 2,337,996 individuals that travelled in the opposite direction. Very recently, SVIMEZ (2015) highlights that the South to Centre-North flows have still been taking place between 2001 and 2014, with 1,667,000 individuals that left the *Mezzogiorno* regions, compared with 923,000 that moved the other way round.

A detailed review of the empirical works on internal migration across Italian regions goes beyond the scope of the present paper. A digest of these works along with their main results is summarized in Table 1. The first empirical studies are those of Salvatore (1977) and Attanasio and Padoa-Schioppa (1991), followed by Faini et al. (1997), Daveri and Faini (1999), Cannari et al. (2000), Brunello et al. (2001) and Murat and Paba (2002). More recently, due to the upsurge of internal migration recorded from the middle of 1990s, a new wave of empirical studies has been conducted. Recent investigations include Furceri (2006), Basile and Causi (2007), Fachin (2007), Etzo (2011), Biagi et al. (2011), Bonasia and Napolitano (2012), Mocetti and Porello (2012), Piras (2012a, 2012b), Lamonica and Zagaglia (2013) and Fratesi and Percoco (2014).

As one can see from Table 1, the time period, methodology and variables used in these analyses greatly differ. What clearly emerges from the majority of them, particularly from the most recent ones, is the role played by macroeconomic variables, *per capita* GDP and unemployment rates, whereas other variables used in the empirical research have not been found as robust as the main macroeconomic ones.

As far as the territorial disaggregation of the data is concerned, with the exception of Faini et al. (1997) who do not estimate an econometric model, the approach is quite differentiated. There are a number of studies that focus on migration flows from the *Mezzogiorno* as a whole,

Fig. 1. Italian regions

or from each of the eight Southern regions, towards the Centre-Northern regions (Salvatore 1977; Daveri and Faini 1999; Cannari et al. 2000; Brunello et al. 2001). Others divide the country into some more or less arbitrary macro-areas and focus on migration flows across these macro-areas (Attanasio and Padoa-Schioppa 1991; Fachin 2007). Basile and Causi (2007) and Biagi et al. (2011) disaggregate migration flows at the provincial level. All the other papers cited in Table 1 use regional data, however there is no a common approach. Furceri (2006) study the determinants of total net migration rates and find that interregional migration responds to both short and long-run market forces (*per capita* GDP and unemployment rates). Piras (2012a) analyses net migration rates from each Italian region towards all other regions. He defines the dependent variables in relative terms with respect to nationwide average and provides empirical evidence that *per capita* GDP, unemployment rate and migrants' human capital are the main determinants of migration flows across Italian regions from 1970 to 2002. In a similar manner, Piras (2012b) brings the analysis back to 1964 and splits migration rates into high and low-skilled migration rates. He finds that high-skilled individuals react more promptly to regional unbalances than low-skilled ones. Also Bonasia and Napolitano (2012) split net migration rates into low and high-skilled migration rates and run separate regressions for the 1985–1995 and the 1996–2006 time periods. Their results suggest that during the latter time interval income and unemployment differentials are important for the skilled component of migration flows but they turn out to be irrelevant for the unskilled. Lamonica and Zagaglia (2013) preliminary perform a factor analysis in order to detect the principal variables to consider into their regression analysis. They conduct an investigation by means of a factor variance analysis and point out that together with the size of regional population and distance, economic factors were always highly significant in explaining internal migration of Italians between 1995 and 2006. Finally, Fratesi and Percoco (2014)

Table 1. Empirical papers on internal migration in Italy

Author(s)	Time Period	Estimation method	Dependent variable	Main independent variables	Human capital	Notes	Main results
Salvatore (1977)	1958–1974	OLS	Net and gross migration rates	Unemployment and wages	NO	He analyses migration flows from the *Mezzogiorno* as a whole and from each one of the Southern regions to both the North-western and the Northern regions	South-North labour migration is viewed as a response to interregional unemployment and earnings differentials
Attanasio and Padoa-Schioppa (1991)	1960–1986	OLS	Net and gross migration rates	Local and national wages in public and private sector, local and national male unemployment	NO	They study migration flows across six macro areas of the country	Gross migration rates are strongly correlated with unemployment differentials
Faini et al. (1997)	1995	Multinomial logit	Mobility choice	Environmental factors and individual characteristics	NO	They define an empirical puzzle the falling of migration flows with growing unemployment differentials during the second half of ’70 and the ’80 of last century	The pattern of low mobility is attributed to inefficiencies in job-matching and high mobility costs
Daveri and Faini (1999)	1970–1989	Panel fixed effects and SUR	Gross outward migration	Unemployment and wage differentials. Two variables aimed to measure risk are also included	NO	They analyze gross outward migration from southern regions to the rest of the country and to international destinations	Risk is a significant determinant in the migration decision
Cannari et al. (2000)	1967–1992	Logit	Transfers from South to Centre-North	Relative unemployment, relative per capita consumption and housing market variables	NO	A non-negligible share of South-North migration flows remains unexplained	Housing price differentials are important to explain the decline in migration
Brunello et al. (2001)	1970–1993	IV	Gross migration rates	Relative wages and unemployment rate	NO	They estimate migration outflows from each of the eight Southern regions to the rest of the country	South-North labour mobility has declined with the reduction in earnings differentials

(Continues)

Table 1. (Continued)

Author(s)	Time Period	Estimation method	Dependent variable	Main independent variables	Human capital	Notes	Main results
							and the increase in social transfer
Murat and Paba (2002)	Four decennial sub-periods: 1951–61 1961–71 1971–81 1981–91	OLS	Net migration rate	Per capita income, the share of workers employed by firms with more than 500 employees	NO	They use provincial data and run separate regressions for four decennial periods	Industrial structure is important to explain inter-provincial migration flows.
Furceri (2006)	1985–2001	OLS, Panel fixed effect and Panel GMM	Net migration	Per capita GDP and unemployment rate	NO	His aim is to ascertain whether net migration responds to GDP regional cyclical components	He finds support to the conjecture
Basile and Causi (2007)	1991–2000	SUR	Net migration rate	Per capita income and unemployment	NO	They use provincial data and run separate regressions for the two periods: 1991–1995 and 1996–2000.	During the first period the effect of economic variables on net migration flows were negligible or nil; during the second they reacted to them
Fachin (2007)	1970–2006	Panel cointegration analysis (FM-OLS, PMG)	Net migration rate	Per capita GDP and unemployment rate differentials	NO	He considers seven macro-areas and concentrates his analysis on male migration from the two Southern towards the other macro-areas	Income growth in the sending regions is the main factor that explains the decline in migration rates. Income and unemployment differentials play a minor role
Etzo (2011)	1996–2005	FE, RE, Fixed-effect vector decomposition, GMM	Gross migration flows	Per capita GDP and unemployment rate	NO	He applies a gravity model of bilateral flows	He finds a significant impact of the main macroeconomic variables
Biagi et al. (2011)	2001 and 2002	Negative binomial, GMM	Gross migration flows	Per capita GDP and unemployment rate	YES, secondary school attainment per 10,000 inhabitants	They use provincial data in order to disentangle different patterns for	Long-distance reflects a disequilibrium model of migration; short-run distance an equilibrium

(Continues)

Table 1. (Continued)

Author(s)	Time Period	Estimation method	Dependent variable	Main independent variables	Human capital	Notes	Main results
						short- and long-distance migration	model of migration. Human capital (does not) affects (short) long-distance migration
Mocetti and Porello (2012)	1995–2005	FE, GMM	Net migration	Per capita GDP and unemployment rate. All variables are expressed as differences between origin and destination	NO	They disaggregate by other socio-economic characteristics of origin and destination of flows	Employment opportunities in the Centre-Northern regions pull Southern migrants. Laureates are sensitive to per capita GDP but less sensitive to unemployment with respect to lower-secondary school migrants
Piras (2012a)	1970–2002	Panel cointegration analysis (OLS, FE, FGLS, GMM)	Net migration rate	Per capita GDP and unemployment rate expressed in relative terms with respect to national average	YES, average years of schooling of migrants. Relative human capital negatively affects net migration	He studies the accommodating potential of internal migration to regional unbalances	A significant role of the macroeconomic variables and of the share of young population on regional migration is detected
Piras (2012b)	1964–2002	Panel cointegration analysis (FE, FGLS, LSDV, GMM)	Net migration rate	Per capita GDP and unemployment rate expressed in relative terms with respect to national average	NO	He splits migration rates for high and low skilled	Separate regressions shows that high skilled individuals react more promptly to regional unbalances than low-skilled ones.
Bonasia and Napolitano (2012)	1985–2006	GMM	Net migration rate	Per capita GDP and unemployment rate expressed in relative terms with respect to national average	NO	They split migration rates for high and low skilled and run separate regressions for the two periods: 1985–1995 and 1996–2006	In the first period the traditional variables neither explain migration flows for skilled, nor for unskilled. In the second period income and unemployment differentials seem to be

(Continues)

Table 1. (Continued)

Author(s)	Time Period	Estimation method	Dependent variable	Main independent variables	Human capital	Notes	Main results
							important for skilled migrants but not for unskilled ones
Lamonica and Zagaglia (2013)	1995–2006.	OLS	Gross migration flows	Economic and demographic factors synthesized by means of factor analysis	NO	They estimate year by year internal migration flows of both Italians and foreigners	For the Italians, together with the size of regional population and the distance, economic factors were always highly significant.
Fratesi and Percoco (2014)	1980–2001	FE	Net migration rate; net human capital gain	Per capita GDP and unemployment rate at origin	NO/YES. In the net human capital gain regression, average years of schooling of resident population has a positive role	The focus is on selective migration, regional growth and convergence	The loss of human capital in the South has been detrimental to growth

Notes: In all these works other explanatory variables are included.

concentrate on selective migration, growth and convergence at regional level and claim that selective migration has been a diverging force in the Italian convergence process.[1] In their work they look for the determinants of both net migration rate and human capital gain. As for the net migration rate, they find a role for *per capita* GDP and unemployment rate. As regards human capital gain, they regress net human capital gain, defined as the difference of average years of schooling of immigrants minus average years of schooling of emigrants over regional population, with respect to *per capita* GDP, unemployment rate, housing price, human capital and human capital squared. The main result of their analysis is that regional human capital net gains are positively linked with the stock of regional human capital, though possibly non-linearly, given that the squared term is also negative and statistical significant.

None of the aforementioned papers estimate a gravity model with bilateral flows as we do in this paper. Two noteworthy exceptions are Etzo (2011) and Mocetti and Porello (2012). Etzo (2011) estimates a gravity model for the 1996–2005 time period and reports that *per capita* GDP and unemployment rates are important factors that explain internal flows. However, since he is interested in investigating the upsurge of internal migration recorded from the middle of 1990s, the long-run perspective of the phenomenon is ignored. Moreover, regional human capital is not included among the regressors. Mocetti and Porello (2012) define all the variables as the difference between the value at origin and that at destination. Their analysis is executed for the overall population and for specific groups of individuals, differentiated on the basis of educational attainment, age and gender. Overall, their results show that migration from the *Mezzogiorno* is principally driven by better employment opportunities in the Centre-Northern regions. This paper has three main drawbacks: two of them, namely the short time period investigated and the absence of human capital among the regressor, are common with Etzo (2011); the third is that, given the way in which the variables are defined, push and pull factors for origin and destination regions are not separately reported. As a matter of fact, treating push and pull factors symmetrically imposes arbitrarily restrictions on structural parameters and does not allow us to properly disentangle the different impact that the same variable could have as a push rather than as a pull factor.

As for human capital, there are just two works that introduce it among the explicative variables.[2] Biagi et al. (2011) consider a proxy for provincial human capital given by resident population with a secondary school diploma per 10,000 inhabitants in 1991. In their estimates, they find that human capital positively affects long-distance movements, but that it does not influence short-distance migration. However, it is not possible to untangle the role of human capital separately as a push and as a pull factor. Piras (2012a) computes regional human capital as the average years of schooling thus taking into account all educational levels. In the regressions he presents, the relative human capital of migrants (defined as the average human capital of immigrant minus the average human capital of emigrants over the average human capital of resident population) exerts a negative influence on net regional migration. Again, given the way in which the model is defined, the role of regional human capital as a push or as a pull factor is not disentangled.

3 A gravity model of internal migration and the role of human capital

3.1 A gravity model of internal migration with human capital

The gravity model of internal migration proposed by Karemera et al. (2000) is our starting point. In their work they assume that both supply and demand factors together with other restraining and/or aiding factors determine migration flows. More precisely, they consider:

[1] A similar result has been found by Piras (2013).

[2] We do not discuss further the paper by Fratesi and Percoco (2014) which considers human capital among the regressors for the net human capital gains equation, but does not introduce it into the net migration rate equation. We will come back to their results below.

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

where F_{ijt} are the gross migration flows from the region of origin i to the region of destination j at time t. F_{ijt} is a function of supply-push factors at home S_{it}, demand-pull factors at destination D_{jt} and of other variables R_{ijt}, restraining and/or aiding migration, associated with the specific origin-destination region-pair i-j. Note that the constant A_{ij} catches all time-invariant region-pair effects linked with bilateral migration flows from i to j, whereas the variable R_{ijt} captures time-variant effects of bilateral flows. In the traditional gravity model R_{ijt} typically represents the costs associated with moving from i to j at time t. Interestingly, inside the framework of the random utility maximization problem used by Bertoli and Fernández-Huertas Moraga (2013), the variable R_{ijt} is the multilateral resistance to migration. In our framework, R_{ijt} grasps the influence of supply-push and demand-pull factors of origin regions other than i and destination regions other than j. Finally, the a_k (k = 1, 2, 3) are structural elasticities linking migration flows to supply and demand factors.

Karemera et al. (2000) claim that a gravity model is conceivable as a reduced form equation that can be derived from a demand and a supply relationship, and that supply and demand of migrants are linked to population and income in both countries. The rationale behind such an assumption is that the supply function of migrants S_{it} depends on the capacity to migrate from i, summarized by expected income y_{it}, on the stock of population n_{it} living in i and on other time-constant push factors b_i. Analogously, the demand function of migrants D_{jt} depends on the capacity of the host economy to attract migrants captured by expected income y_{jt}, on the stock of population living in j, n_{jt}, and on other time-constant pull factors of the receiving country/region j, c_j.

In the context of internal migration in an advanced economy such as the Italian one, it is reasonable to think that also other variables could be added and human capital is the first candidate to the list. Therefore, we extend the Karemera et al. (2000) approach and assume that, at time t, the supply and demand functions of migrants depends also on the average level of human capital at origin h_{it} and destination h_{jt}. Shortly, we assume:

$$S_{it} = b_i y_{it}^{b_1} n_{it}^{b_2} h_{it}^{b_3}, \tag{2}$$

$$D_{jt} = c_j y_{jt}^{c_1} n_{jt}^{c_2} h_{jt}^{c_3}. \tag{3}$$

After substituting Equations (2) and (3) into (1) and taking natural logarithms, we get:

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

Equation (4) posits that bilateral migration flows from origin i to destination j are spurred by the ‘mass’ of the two regions, namely by their respective population sizes, n_{it} and n_{jt}: the bigger their mass, the bigger the bilateral flows. Expected incomes y_{it} and y_{jt} also play a crucial role. Following Harris and Todaro (1970) and in line with the macroeconomic literature on migration, we assume that expected income be a function of *per capita* income and employment opportunities. In order to disentangle the role of these two factors, in the empirical implementation of Equation (4) *per capita* income and unemployment rates at origin, pcy_{it} and u_{it} respectively, and destination, pcy_{jt} and u_{jt}, will simultaneously be considered. Thus, everything else being equal, the higher level of *per capita* income in the sending (receiving) region leads to lower (higher) migration flows from i to j. On the contrary, higher levels of unemployment rates in the sending (receiving) region leads to higher (lower) migration flows from i to j. As regards human capital, some theoretical reasons lead us to believe that it might operate with opposite effects at origin and at destination.

It is worth noting that many empirical studies on migration that do not use the gravity model assume, arbitrarily and implicitly, perfect symmetry for push and pull factors. This happens, for example, if one considers the ratio of the variables at origin and destination. Making such an assumption in Equation (4) would imply $a_1b_1 = -a_2c_1$ and $a_1b_2 = a_2c_2$ for expected income and population, respectively, and $\pm a_1b_3 = \mp a_2c_3$ for human capital. However, there are no clear theoretic underpinnings why these restrictions should be imposed rather than tested. On the contrary, there is evidence that asymmetric impacts of pull and push factors are the norm and not the exception at both the internal and international level.[3] Hence, in our empirical estimates, instead of imposing them we formally test whether they are valid or not.

3.2 The role of human capital

It should be re-emphasized that in this paper we aim to study the role that the average level of regional human capital could have as a driver of interregional migration in the source as well as in the destination region. Hence, we do not deal with the human capital endowment of interregional migrants (i.e., their skill level) that has already been investigated, though not extensively as discussed in Section 2, in the Italian context and widely discussed at international level by a large body of research (see the recent work of Docquier et al. 2014 and references therein). The main finding of this strand of literature, and this is a quite robust result across empirical studies, is that individuals with higher levels of human capital are more willing to migrate. Furthermore, we do not look for the impact of migration prospects on regional human capital formation and, possibly, on regional growth. On this point, starting from the seminal theoretical papers of Mountford (1997) and Beine et al. (2001), a huge amount of research has also been devoted showing, however, mixed results.[4]

Having said that, let us now discuss the role of human capital as a driver of interregional migration and try to elucidate its effects as a push and as a pull factor. In the literature on the economics of migration, among the many authors that contributed to understand the linkages between education and migration, George Borjas deserves particular attention. In a series of papers, he addresses many issues related to the role of human capital. In particular Borjas (1991) predicts that migration is higher (lower) the higher the mean level of education in the source (destination) country. His result is explained on the basis of the ‘educational premium’ which occurs if, in the destination country, the characteristics of the migrants (skills and experience) are rewarded more than they are at origin. As a matter of fact, whenever *per capita* GDP is introduced into a migration equation such as (4) at both origin and destination, the average skill levels should be controlled for. Indeed, *per capita* GDP is just a proxy for the wages that a potential migrant takes into account in order to decide whether or not to move, the reason being that *per capita* GDP depends also on the rates of returns of capital and labour. As clearly pointed out by Mayda (2010) a higher *per capita* GDP at destination does not necessarily imply that a migrant would obtain it. A higher *per capita* GDP at destination might be due to higher *per capita* capital or to higher levels of average human capital.

Empirical evidence on the positive role at origin and on the negative role at destination is not robust however, and seems also to depend on whether international or internal migration flows are analysed.

To the best of our knowledge, at the international level, Clark et al. (2007), Pedersen et al. (2008), Hatton and Williamson (2010) and Mayda (2010) are the only few empirical works on migration that

[3] See, for example, Hunt (2006) and Mayda (2010). The former studies internal migration in German states and explains the asymmetric impact of push and pull factors in terms of differential responses of young and old migrants. The latter analyses international migration flows and claims that the asymmetry emerges from the demand side of migration flows, namely from the migration policies pursued in the host countries.

[4] For a recent empirical investigation, see Brunow et al. (2015).

introduce a proxy for human capital, finding mixed results. Hatton and Williamson (2010) estimate emigration rates from 62 developing countries to the US over five-year intervals starting from 1970 up to 2004. Among the regressors, they introduce the ratio of average years of education in the US with respect to the source country and find that the estimated elasticity has the expected negative sign, but is not significant. Mayda (2010) considers international migration flows into 14 OECD countries from 70 origin countries between 1980 and 1995. When she controls for the average schooling level alone, she finds that the average skill level of population at destination (origin), negatively (positively) affect the emigration rate. However, when the endowment of physical capital per worker is added into the regression, both coefficients of physical and human capital turn out to be statistically insignificant. In the same line, Pedersen et al. (2008) study migration flows into 22 OECD countries from 129 origin countries. They introduce the illiteracy rate at the origin and find that countries with a high illiteracy rate tend to have lower emigration flows. This result, too, is not robust across all their regressions. Finally, Clark et al. (2007) consider immigration into US from a panel of 81 countries during the 1971–1998 time period. They introduce the ratio of average years of schooling at origin relative to the US in order to deflate the effects of the human capital stock and of the average returns on human capital on the relative income variable. Following this approach, relative schooling years is expected to have a positive effect on immigration into US. As a matter of fact, in the fixed effects regressions they do find it but the estimated coefficient is not statistically significant, while in the between effects estimates the estimated coefficient is positive and statistically significant.

Conversely, at internal level the available empirical evidence delivers no role at all for human capital. Mitze and Reinkowski (2011) study German internal net migration rates for 97 planning regions in the decade 1996–2006. They regress the net migration rate (inflows minus outflows over regional population) against the explanatory variables defined in relative terms with respect to the country aggregate and find that different proxies that they use to measure the regional endowment of human capital are never significant. Their results mirror those of Maza and Villaverde (2004) and Maza (2006) for Spain. Maza (2006) investigates the factors influencing regional migration and regional convergence across Spanish regions during the 1995–2002 time span. For each region, the variables are defined in relative terms, as performed by Mitze and Reinkowski (2011), and when he looks at interregional migration he finds that the net migration rate is not influenced by regional human capital. Maza and Villaverde (2004) apply semi-parametric techniques to estimate internal migration flows and confirm that human capital does not appear to exert any effect on them. The authors give a possible explanation of these results affirming that human capital might affect inflows and outflows in opposite directions (though they do not give any explanation of the mechanisms behind this hypothesized opposite relationship), therefore analysing net migration might compensate the two counterbalancing effects. In our opinion, another possible explanation common to these three papers is that the dependent variables, hence also regional human capital, are defined as ratios with respect to the country aggregate, therefore they cannot properly catch the push and pull factors affecting regional migration flows.

The predictions of Borjas (1991) theoretical approach are reversed and human capital could exert opposite effects on migration whenever it works as a magnet, particularly for skilled migrants who look at higher rewards to their individual human capital. This is what happens in the new economic geography models, pioneered by Krugman (1991) and well described in the Fujita and Thiesse (2002) book that stress the role of agglomeration economies in attracting individuals.[5] Indeed, labour migration is essential to the new economic geography models in that it responds to market signals in a counterintuitive way with respect to the traditional classical theory of migration that sees migration as an equilibrating mechanism with respect to regional unbalances in the labour market (Blanchard and Katz 1992; Decressin and Fatas 1995). When agglomeration forces

[5] On this point and, more generally, for a thorough discussion of the intertwined relationship of regional development, migration and human capital, see also Faggian and McCann (2009).

are at work, the resulting concentration processes triggers a self-reinforcing mechanism of circular causation and firms and worker concentrates where demand is larger in order to benefit from economies of scale. In this framework, if a region is highly attractive because of its role as a centre – and as such it has high levels of human capital – not only does it attract individuals from other regions (positive role of human capital at destination) but it also withholds individuals from moving towards other regions (negative role of human capital at origin).

In addition, as for human capital at origin, it could also be the case that rewards to skill are higher at origin rather than at destination. Then for better educated individuals the migration choice is inversely linked with the educational level at home. As a matter of fact, human capital required at destination might be country specific and different from that owned by would-be migrants (Friedberg 2000). The simplest form of country specific human capital in the international context is language proficiency. The case of Italian regions is an internationally known case study because of their long-lasting social and economic divide that has characterized and still characterizes them.

4 Estimation methods, empirical specification, data and testing

4.1 Estimation methods

The estimation of a gravity model faces several econometric challenges that have been recently discussed by Fidrmuc (2009), Hanson (2010), Anderson (2011), Bertoli and Fernández-Huertas Moraga (2013), Desbordes and Eberhardt (2014), Beine et al. (2015), to mention just a few. The great majority of these papers concentrate on the gravity model applied to international trade; yet, given the similarities between trade and migration, most of the results also apply to the gravity model of migration. As claimed by Fidrmuc (2009), standard gravity models of international trade include non-stationary variables and are characterized by cross-sectional correlation across units (country-pair). If these issues are not properly tackled, spurious regressions and misspecification problems are likely to be present. In the case of non-stationary variables, a well-known result is that the standard fixed effects (FE) estimator is biased. Among the various estimators put forward to cope with this problem, the fully modified ordinary least squares (FMOLS) and the dynamic ordinary least squares (DOLS) estimators have been extensively used in the applied economic literature. Kao and Chiang (2000) have demonstrated that both estimators are asymptotically equivalent, but that they perform differently in finite sample. In particular, the DOLS estimator seems to be better suited in such a case.

In addition to the non-stationary and spurious regressions issues, Desbordes and Eberhardt (2014) uphold that a fundamental concern in estimating the structural gravity model is the modelling of time-varying unobserved multilateral resistance terms. In the international trade literature, trade between two countries not only depends on bilateral trade costs, but also depends on trade costs with other potential trading partners. Similarly, when applying the structural gravity model to migration, it is very likely that migration flows between one origin and one destination do not depend on their respective push and pull factors alone, but that they also depend on the opportunities of moving to other destinations. Thus, migration flows from, say region i to region j, will be influenced by the attractiveness of alternative destinations, other than j. Bertoli and Fernández-Huertas Moraga (2013) refer to the influence exerted by the attractiveness of other destinations (third-region effects) as multilateral resistance to migration. Since these influences make the error term of standard panel data estimators both serially and spatially correlated, more general econometric approaches that cope with them are required.

Following the very recent contributions of Eberhardt and co-authors (see, among others, Desbordes and Eberhardt 2014; Eberhardt and Teal 2013a, 2013b) as well as those of Bertoli and Fernández-Huertas Moraga (2013) and Bertoli et al. (2013), in this paper we apply the common correlated effects mean group estimator (CCEMG) proposed by Pesaran (2006) and the augmented mean group estimator (AMG) introduced by Eberhardt and Teal (2010). These novel heterogeneous estimators, which will be briefly explained below, are compared with the mean group (MG) estimator of Pesaran and Smith (1995), that does not cope with cross-sectional dependence in the data, and with three standard homogeneous panel data estimators, namely the two ways (time and region-pairs) fixed effect (2FE),[6] the FMOLS and the DOLS estimators. These three homogeneous estimators neither account for the potential bias introduced by the imposed parameters homogeneity, nor do they control for unobserved common factors. Thus, they should be viewed as *prima facie* evidence for our gravity model.

4.2 Empirical specification

In light of the previous discussion, the empirical specification of bilateral migration with human capital expressed by Equation (4) can be estimated in the following general form:

$$\begin{aligned}\ln F_{ijt} &= \gamma_{1ij} \ln pcy_{it} + \gamma_{2ij} \ln pcy_{jt} + \gamma_{3ij} \ln u_{it} + \gamma_{4ij} \ln u_{jt} + \gamma_{5ij} \ln n_{it} + \gamma_{6ij} \ln n_{jt} + \\ &\quad + \gamma_{7ij} \ln h_{it} + \gamma_{8ij} \ln h_{jt} + \lambda_{ij} + \varepsilon_{ijt},\end{aligned} \tag{5}$$

where λ_{ij} is a dummy for the region-pair *i-j* (clearly $\lambda_{ij} \neq \lambda_{ji}$) and the error term $\varepsilon_{ijt} = \xi'_{ij} f_t + \eta_{ijt}$ has a multifactor error structure given by the inner product of a vector of panel specific factor loadings ξ_{ij} and a vector of time-varying factors f_t and an *i.i.d.* error term η_{ijt}. Note that λ_{ij} is intended to capture all region-pair specific fixed factor, such as distance or common borders, that will not be included as additional explanatory variables.

The CCEMG estimator advocated by Pesaran (2006) expresses the term $\xi'_{ij} f_t$ as a linear combination of cross-sectional panel averages of both the dependent and the independent variables, namely, as $\delta'_{ij} \overline{Z}_t$, where δ'_{ij} is the vector of region-pair *i-j* factor loading and $\overline{Z}_t$ is the vector of cross-sectional panel averages of the dependent and of all independent variables; and estimates:

$$\begin{aligned}\ln F_{ijt} &= \gamma_{1ij} \ln pcy_{it} + \gamma_{2ij} \ln pcy_{jt} + \gamma_{3ij} \ln u_{it} + \gamma_{4ij} \ln u_{jt} + \gamma_{5ij} \ln n_{it} + \gamma_{6ij} \ln n_{jt} + \\ &\quad + \gamma_{7ij} \ln h_{it} + \gamma_{8ij} \ln h_{jt} + \lambda_{ij} + \delta'_{ij} \overline{Z}_t + \tau_i time + \eta_{ijt}.\end{aligned} \tag{6}$$

In this equation the coefficients of the extra regressor $\overline{Z}_t$ cannot be interpreted in any meaningful way, since they just serve to consistently estimate the model in the presence of unobserved common factors. Nevertheless, a formal test to check whether cross-sectional averages are jointly different from zero, which is required for the CCEMG estimator to be valid, will be shown in the empirical section. In addition, a deterministic group-specific time trend, labelled *time* with coefficient τ_i, is introduced in Equation (6) in order to capture omitted idiosyncratic processes evolving over time. Pesaran (2006) shows that the CCEMG estimator is able to control for serially and spatially correlated error terms and that it has good finite sample properties. Various simulation studies (Coakley et al. 2006; Kapetanios et al. 2011) have shown that the CCEMG estimator perform quite well also when the variables are non-stationary, cointegrated or not, in the presence of local spillovers and global/local business

[6] In the 2FE regression we apply the Driscoll and Kraay (1998) approach in order to obtain a covariance matrix estimator that yields heteroscedasticity consistent standard errors robust to general forms of both temporal and spatial dependence. As we will see in the empirical section, this does not suffice to eliminate residual cross-sectional dependence; hence more general estimators are needed.

cycles and when the relationship is subject to structural breaks (Eberhardt and Teal 2013a, 2013b).[7]

As for the AMG estimator, a common dynamic effect in the group specific regression is introduced in order to account for cross-sectional dependence in the presence of slope heterogeneity and non stationary variables. This common dynamic effect variable is constructed by taking the coefficients of the $T-1$ time dummies in a first stage pooled regression run in first differences. In a second stage these coefficients are included in the group specific regressions along with group specific linear time trends which are intended to catch omitted idiosyncratic processes. Bond and Eberhardt (2013) show the AMG estimator performs similarly well to the CCEMG estimator in many empirical setups. The AMG is thus implemented in two steps. In the first step the following regression is estimated:

$$\begin{aligned}\Delta\ln F_{ijt} &= b_1\Delta \ln pcy_{it} + b_2\Delta \ln pcy_{jt} + b_3\Delta \ln u_{it} + b_4\Delta \ln u_{jt} + b_5\Delta \ln n_{it} + b_6\Delta \ln n_{jt}+ \\ &+ b_7\Delta \ln h_{it} + b_8\Delta \ln h_{jt} + \sum_{t=2}^{T} c_t\Delta D_t + e_{ijt} \qquad \Rightarrow \hat{c}_t \equiv \hat{\mu}^{\bullet}_t,\end{aligned} \tag{7}$$

where e_{ijt} is a well-behaved error term and D_t are time dummies. In the second step, the estimated coefficients of year dummies $\hat{\mu}^{\bullet}_t$ – which represent an estimated average of the evolution of unobservable over time, namely the common dynamic process – are plugged into each group-specific regression possibly along with group-specific linear time trends:

$$\begin{aligned}\ln F_{ijt} &= \gamma_{1ij} \ln pcy_{it} + \gamma_{2ij} \ln pcy_{jt} + \gamma_{3ij} \ln u_{it} + \gamma_{4ij} \ln u_{jt} + \gamma_{5ij} \ln n_{it} + \gamma_{6ij} \ln n_{jt}+ \\ &+ {}_{\gamma 7ij}\ln h_{it} + \gamma_{8ij} \ln h_{jt} + d_{ij}\hat{\mu}^{\bullet}_t + \tau_i time + e_{ijt}.\end{aligned} \tag{8}$$

Alternatively, the common dynamic process can be imposed on each group member with unit coefficient by subtracting it from the dependent variable (AMGC):

$$\begin{aligned}\ln F_{ijt} - \hat{\mu}^{\bullet}_t &= \gamma_{1ij} \ln pcy_{it} + \gamma_{2ij} \ln pcy_{jt} + \gamma_{3ij} \ln u_{it} + \gamma_{4ij} \ln u_{jt} + \gamma_{5ij} \ln n_{it} + \gamma_{6ij} \ln n_{jt}+ \\ &+ \gamma_{7ij} \ln h_{it} + \gamma_{8ij} \ln h_{jt} + \tau_i time + e_{ijt}.\end{aligned} \tag{9}$$

By controlling for multilateral resistance to migration both the CCEMG and the AMG estimators account for endogeneity problems due to common factors that drives interregional migration flows and that contemporaneously affect the dependent variables (Bertoli and Fernández-Huertas Moraga 2013). In addition, as long as the resistance terms catch out the effect of omitted variables, then this source of endogeneity is accounted for (Beine et al. 2015). As for endogeneity due to reverse causality, instrumentation of some variables might be needed. Unfortunately, finding valid and informative instruments in gravity models of migration, as it is more generally the case in most macro panel econometric analysis (Bazzi and Clemens 2013), is not an easy task, since viable instruments are very difficult to discover. Furthermore, according to Beine et al. (2015) instrumentation in gravity models of migration would need to be carried out in a Poisson regression framework like the Poisson pseudo-maximum likelihood (PPML) advocated by Santos Silva and Tenreyro (2006) and extended by Tenreyro (2007). Resorting to such estimator, however, would induce further complications. On the one hand it may face problems of convergence; on the other, contrary to CCEMG and AMG estimators, it is not able to cope with the presence of cross-section dependence and common factors. All in all, weighing the pros and cons, we deem that heterogeneous panel data estimators are to be preferred to pseudo-maximum likelihood estimators.

[7] In the presence of common factors, as an alternative of the CCEMG approach, Bai et al. (2009) have advocated the constantly updated and fully modified bias corrected estimators. In this regard, the recent contributions of Bailey et al. (2012) and Westerlund and Urbain (2015) suggest that the CCEMG approach has to be preferred on the basis of both theoretical and computational easiness.

Finally, the MG estimator either ignores unobserved common factors, or it just captures them with group-specific linear time trends:

$$\ln F_{ijt} = \gamma_{1ij} \ln pcy_{it} + \gamma_{2ij} \ln pcy_{jt} + \gamma_{3ij} \ln u_{it} + \gamma_{4ij} \ln u_{jt} + \gamma_{5ij} \ln n_{it} + \gamma_{6ij} \ln n_{jt} + \\ + \gamma_{7ij} \ln h_{it} + \gamma_{8ij} \ln h_{jt} + \lambda_{ij} + \tau_i time + \eta_{ijt}. \quad (10)$$

All these heterogeneous estimators yield estimates of the parameters as averages of the group-specific region-pair *i-j* coefficients, namely:[8]

$$\gamma_{k,ij} = N^{-1} \sum_{ij}^{N} \gamma_{k,ij} (k = 1, \ldots 8). \quad (11)$$

4.3 Data

We gathered data principally from the Italian national institute for statistics (ISTAT). Unemployment is taken from ISTAT (various years b), interregional migration flows derives from ISTAT (various years a).[9] ISTAT takes regional data from municipal registers' offices and provides origin-destination matrices from 1970 to 2005, and this is why our analysis is bounded by this time period. The presence of zero interregional flows is encountered in less than 8 per cent of observations. Although this percentage is quite low and could be safely ignored without compromising the empirical estimates, we have dealt with the presence of zeroes by imputing '1' instead of '0'. By so doing, we are able to maintain the double logarithmic format and to keep all the information contained in the data, including zeros. The average years of schooling for resident population are computed exploiting the detailed information available in ISTAT (various years, b). In computing this proxy for human capital we considered four schooling levels (s = 1, … 4) and years of schooling as follows: 18 years for individuals with a university degree (ISCED 6), 13 for individuals with an upper secondary school diploma (ISCED 3a–3c), eight years for those with a lower-secondary school attainment (ISCED 2) and, finally, three years for individuals with either a primary school educational level (ISCED 2) or without any formal schooling attainment.[10] To be more precise average years of schooling of resident population are defined as:

$$aysp op_t = \frac{\sum_{s=1}^{4} as_s P_{st}}{P_t}, \quad (12)$$

where as_s is the number of years of schooling for schooling level s, P_{st} is the number of individuals with schooling level s and P_t is total population. The reason why we attached three years to the last group is that starting from 1993, ISTAT merges these two groups into a single group and we are forced to do the same also for the previous period, namely from 1970 to 1992. Regional *per capita* GDP and population comes from SVIMEZ (2011). Summary statistics and a more detailed description of all variables are presented in Appendix A.

[8] In the empirical section we apply the robust option of the user-written STATA routine xtmg which allows for weighted mean coefficients in such a way that the averages of the group-specific coefficients are robust to outliers. Basically, this is the procedure devised by Hamilton (1992), which attributes less weight to outliers in computing the estimates. See Eberhardt (2012) for more details.

[9] For the more recent years, the data is available on-line at http://dati.istat.it/?lang=en.

[10] The United Nations Educational, Scientific and Cultural Organisation (UNESCO) has developed the International Standard Classification of Education (ISCED) in order to facilitate the comparisons of education statistics across countries on the basis of uniform and internationally agreed definitions. For more details, see http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx.

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	592.06 [0.000]	81.94 [0.000]	119.59 [0.000]
Per capita GDP at origin	1576.56 [0.000]	395.78 [0.000]	398.92 [0.000]
Per capita GDP at destination		398.92 [0.000]	395.78 [0.000]
Unemployment rate at origin	1363.16 [0.000]	373.36 [0.000]	358.95 [0.000]
Unemployment rate at destination		358.95 [0.000]	373.36 [0.000]
Population at origin	525.84 [0.000]	281.84 [0.000]	116.42 [0.000]
Population at destination		116.42 [0.000]	281.84 [0.000]
Average Years of schooling at origin	1604.07 [0.000]	404.49 [0.000]	403.32 [0.000]
Average Years of schooling at destination		403.32 [0.000]	404.49 [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 and is valid for N and T going to infinity in any order and it is robust to possible structural breaks.

4.4 Cross-sectional dependence, panel unit root and stationary tests

In this sub-section first we report the Pesaran (2004) CD tests to assess whether the series are affected by cross-sectional dependence, second we give details on three different panel unit root tests – Hadri (2000); Im et al. (2003); Pesaran (2007) – with different assumption regarding the null and the alternative hypothesis, third we look for cointegration by means of the Kao (1999) residual cointegration test.

Table 2 shows that on the basis of Pesaran (2004) CD test, the null hypothesis of independence is strongly rejected, so that cross-sectional dependence is pervasive and must be considered when computing diagnostic tests and, more importantly, when performing the regression analysis and post-estimation checking.[11]

Several procedures to test for unit roots in panels are available. In this paper we apply three of them. The Hadri (2000) test that accounts for heterogeneous and serially correlated error terms and assumes a null hypothesis of stationary against the alternative that at least one unit of the panel contains a unit root. The Im et al. (2003) test (IPS) has a null that all panels contain a unit root, against the alternative that some units of the panels are stationary, thus rejection of the null does not imply that the unit root is rejected for all units but only for a positive share of the sample. Neither the Hadri (2000) nor the IPS tests account for cross-sectional dependence, thus we also run the Pesaran (2007) test (CIPS) that allows for the presence of a single unobserved common factor with heterogeneous factor loadings as a cause of cross-sectional dependence. In addition, it also allows for heterogeneous autoregressive coefficients in the unit specific Dickey-Fuller regression. The results of these tests are presented in Table 3 for the whole sample of Italian regions and Tables 4 and 5 for the sample of South to Centre-North and Centre-North to South regions, respectively.

The Hadri (2000) test always rejects the null of stationary for the whole sample (Table 3) as well as for the two sub-samples (Tables 4 and 5). The IPS test rejects the null for migration flows and for the average years of schooling of resident population in the whole sample. As far as the South to Centre-North sample is concerned (Table 4), it rejects the null of unit root for migration flows, *per capita* GDP at origin and average years of schooling both at origin and at destination. For the Centre-North to South sample (Table 5), it refuses the null for migration flows, *per capita* GDP at destination and, again, for average years of schooling both at origin and at destination. Finally, the CIPS test is in favour of non-stationary in the great majority of cases, particularly whenever more than one lag is considered. Still,

[11] It is worth noting that, for the whole panel of twenty Italian regions, each region is at the same time a sending and receiving region, the test is identical for the variables measured at origin or at destination.

Table 3. 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
IPS (2003)	−4.06***	−1.84	−1.24	−1.19	−2.41***
Pesaran (2007)					
Lag 1	−15.86***	−1.51*	−5.92***	40.24	−28.19***
Lag 2	−0.84	4.20	−0.93	35.41	−3.52***
Lag 3	2.92	5.87	−5.53***	40.51	−5.43***
Lag 4	4.66	11.12	−1.72**	41.83	11.92
Lag 5	8.81	2.35	8.67	44.44	13.66
Hadri (2000)					
Homoscedastic disturbances across the panel	53.25***	239.54***	275.98***	275.73***	181.58***
Heteroskedastic disturbances across units	70.15***	212.06***	270.65***	221.25***	184.44***
Serial dependence in errors	27.01***	47.73***	49.39***	48.76***	42.07***

Notes: IPS (Im et al. 2003) test has a null that all panel contain a unit root, against the alternative that some panels are stationary. 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. Hadri (2000) test has a null of stationarity against the alternative that at least one unit of the panel contains a unit root. All tests have been run with constant and trend. 10%, 5% and 1% statistical levels of significance for the null hypothesis are indicated by *, ** and *** respectively.

for the whole sample (Table 3) unemployment rate and average years of schooling of resident population present three out of five tests against non-stationary, that is, they reject the null hypothesis of unit root, whereas for the South to Centre-North sample (Table 4) and for the Centre-North to South sample (Table 5) it suggests that, at least for some units, unemployment rate at destination (origin) and average years of schooling at origin (destination) are possibly stationary in four cases out of five.

Overall, the results presented in Tables 3–5 suggest that the variables contain unit roots, with the Hadri (2000) test always in favour and the IPS and CIPS tests in the great majority of cases supporting the same result.

Finally, the Kao (1999) residual cointegration test reported in Table 6 clearly rejects the null hypothesis of no cointegration in all panels. As an additional test for cointegration, as suggested by Eberhardt and Teal (2013b), in the next Section in which we present the regression results, we perform residual stationarity tests which can be interpreted as additional tests for cointegration.

5 Empirical results

5.1 Empirical results for the whole sample of 20 Italian regions

We present the estimation results for the whole panel of twenty Italian regions in Table 7. Residual diagnostic tests strongly reject the null hypothesis of non stationary for both homogeneous and heterogeneous estimates (CIPS test), thus proving that the variables are cointegrated. Cross-sectional independence is rejected in all homogeneous estimates as well as in the MG and in the CCEMG estimates at standard levels of statistical confidence (CD test). Cross-section dependence is less severe for AMG and AMGC estimates (columns 6 and 7 with a p-value of 0.059 and 0.053, respectively): both of them account more satisfactory for cross-sectional dependence with respect to all other estimators.

The AMG and AMGC regressions have the lowest mean absolute correlation coefficient – below 0.16 – whereas in all other regressions it is always higher than that. The root mean square error is higher in the homogeneous estimates with respect to the heterogeneous ones. All in all, regression results favour the AMG and AMGC estimators and this is the reason why, hereafter,

Table 4. Panel unit roots tests for the South to Centre-North flows

	Gross migration flows	*Per capita* GDP at origin	*Per capita* GDP at destination	Unempl. rate at origin	Unempl. rate at destination	Population at origin	Population at destination	Average years schooling at origin	Average years of schooling at destination
IPS (2003)	−3.83***	−2.41***	−1.46	−0.92	−1.45	−1.24	−1.16	−2.55***	−2.31*
Pesaran (2007)									
Lag 1	−7.04***	−3.85***	0.56	−3.99***	−5.80***	7.48	20.10	−23.51***	−10.80***
Lag 2	−0.16	−0.02	1.06	−0.11	−1.62*	4.99	11.15	−9.91***	1.32
Lag 3	1.80	4.52	1.07	−5.09***	−1.61*	10.82	11.84	−9.70***	2.67
Lag 4	3.26	2.90	4.26	−1.05	−1.71**	10.43	12.15	−5.25***	8.75
Lag 5	5.45	3.58	−1.13	0.51	6.24	9.47	12.75	0.97	8.98
Hadri (2000)									
Homoscedastic disturbances across the panel	35.27***	120.57***	120.31***	134.96***	140.51***	162.47***	87.08***	84.41***	94.58***
Heteroscedastic disturbances across units	42.54***	98.89***	111.72***	134.54***	137.03***	158.64***	79.58***	84.92***	97.90***
Serial dependence in errors	17.04***	33.53***	34.20***	37.41***	36.03***	41.31***	23.04***	25.66***	29.70***

Notes: See notes for Table 3.

Table 5. Panel unit roots tests for the Centre-North to South flows

	Gross migration flows	*Per capita* GDP at origin	*Per capita* GDP at destination	Unempl. rate at origin	Unempl. rate at destination	Population at origin	Population at destination	Average years schooling at origin	Average years of schooling at destination
IPS (2003)	−4.00***	−1.46	−2.41***	−1.45	−0.92	−1.16	−1.24	−2.31*	−2.55***
Pesaran (2007)									
Lag 1	−9.20***	0.56	−3.85***	−5.80***	−3.99***	20.10	7.48	−10.80***	−23.51***
Lag 2	−3.82***	1.06	−0.02	−1.62*	−0.11	11.15	4.99	1.32	−9.91***
Lag 3	−2.73***	1.07	4.52	−1.61*	−5.09***	11.84	10.82	2.67	−9.70***
Lag 4	1.34	4.26	2.90	−1.71**	−1.05	12.15	10.43	8.75	−5.25***
Lag 5	3.42	−1.13	3.58	6.24	0.51	12.75	9.47	8.98	0.97
Hadri (2000)									
Homoscedastic disturbances across the panel	25.67***	120.31***	120.57***	140.51***	134.96***	87.08***	162.47***	94.58***	84.41***
Heteroscedastic disturbances across units	37.18***	111.72***	98.89***	137.03***	134.54***	79.58***	158.64***	97.90***	84.92***
Serial dependence in errors	13.13***	34.20***	33.53***	36.03***	37.41***	23.04***	41.31***	29.70***	25.66***

Notes: See notes for Table 3.

Table 6. Kao (1999) residual cointegration test

	All flows	South to Centre-North flows	Centre-North to South flows
SIC criteria	6.56 [0.000]	12.53 [0.000]	18.05 [0.000]
AIC criteria	17.15 [0.000]	21.41 [0.000]	19.08 [0.000]

Notes: Null hypothesis is no cointegration; both Schwarz (SIC) and Akaike (AIC) information criteria have been used.

when discussing in general the estimates, we report them in parentheses as a point elasticity range (AMG\AMGC estimated coefficient).

As a general result, it emerges that almost all estimated elasticities are highly statistically significant and have the expected sign. Moreover, *per capita* GDP at origin (–0.189\–0.137) and unemployment rate at destination (–0.073\–0.076) are always negatively linked with interregional migration flows, whereas *per capita* GDP at destination (0.744\0.660), unemployment rate at origin (0.106\0.112) and population at origin (1.429\1.534) are always positively related with them. Population at destination (1.211\1.205) is found to be positively associated with migration flows in all but in the DOLS estimate (column 3) where it is not significant. As for human capital at origin, leaving aside the 2FE estimator, the estimated coefficient is always negative, it is higher (in absolute terms) in the homogeneous estimates with respect to heterogeneous ones and, most of the times, it is statistically significant. If we consider the AMGC estimator, a one per cent increase of human capital at origin lowers migration flows of 0.319 per cent. The results for human capital at destination, however, deliver mixed results. Indeed, only in the 2FE regression is the estimated elasticity is positive and statistically significant, but it turns out insignificant in all other regressions. Note, however, that in our preferred AMG and AMGC regressions, the estimated coefficient is positive in line, as discussed in subsection 3.2, with the hypothesized role that human capital plays in attracting migrants according to the new economic geography models. We have also performed a series of robustness checks using the quota of graduate over population instead of the average years of schooling. The results, available upon request, on the one hand validate the role of *per capita* GDP, unemployment and population that have all the expected signs and, most of the time, are statistically highly significant. On the other hand, while they confirm the negative sign for human capital at origin across all regressions (though the estimated elasticity turns out statistically significant in the MG regression alone), they deliver mixed results for human capital at destination.

This first set of results is very important from a general migration modelling perspective as well as for the specific case of Italian internal migration and can be summarized as follows. From the point of view of general migration modelling, they agree with the results of Bertoli and Fernández-Huertas Moraga (2013) and Bertoli et al. (2013) and prove that heterogeneity matters and that unobserved common factors are relevant and cannot be ignored. As a consequence, future studies on internal as well as international migration should tackle these two issues properly, otherwise severely distorted estimates might be found. In addition, as the coefficient diagnostics section of Table 7 demonstrates, in the great majority of circumstances the same variable has a statistically different impact according to whether it operates as a push or as a pull variable. Therefore, the gravity model that we estimate warns empirical researchers, whenever it is possible, to include separately (as push and pull factors) the variables of interest, rather than using them as ratios or as differences. As for the specific case of Italy, these results prove that our extended gravity model fits well the long-run interregional migration flows across Italian regions and confirm that the macroeconomic variables are the main drivers of internal migration as highlighted in the previous empirical works. In addition, and this is a novel and important result, they suggest that while the role of human capital is virtually nil at destination, it works as a restraining factor at origin. According to the previous discussion, there could be two

Table 7. Estimation results for the whole sample of 20 Italian regions

	Homogeneous estimates			Heterogeneous estimates			
	(1) 2FE	(2) FMOLS	(3) DOLS	(4) MG	(5) CCEMG	(6) AMG	(7) AMGC
Per capita GDP at origin	−0.352*** (0.005)	−0.332*** (0.000)	−0.072 (0.633)	−0.490*** (0.000)	−0.248 (0.110)	−0.189** (0.033)	−0.137* (0.100)
Per capita GDP at destination	0.299* (0.097)	0.134** (0.046)	0.414*** (0.006)	0.345*** (0.000)	0.852*** (0.000)	0.744*** (0.000)	0.660*** (0.000)
Unempl. rate at origin	0.086*** (0.007)	0.107*** (0.000)	0.181*** (0.000)	0.074*** (0.000)	0.071*** (0.009)	0.106*** (0.000)	0.112*** (0.000)
Unempl. rate at destination	−0.138*** (0.005)	−0.171*** (0.000)	−0.084** (0.030)	−0.113*** (0.000)	−0.063** (0.016)	−0.073*** (0.000)	−0.076*** (0.000)
Population at origin	1.780*** (0.000)	1.167*** (0.000)	2.026*** (0.000)	1.045** (0.045)	1.966** (0.041)	1.429*** (0.002)	1.534*** (0.002)
Population at destination	2.366*** (0.000)	1.516*** (0.000)	−0.631 (0.232)	0.277 (0.597)	0.867 (0.377)	1.211*** (0.009)	1.205** (0.011)
Av. years sch. at origin	0.159 (0.251)	−0.918*** (0.000)	−2.174*** (0.000)	−0.742*** (0.000)	−0.234 (0.390)	−0.242 (0.144)	−0.319** (0.050)
Av. years sch. at destination	0.638*** (0.003)	−0.055 (0.677)	−0.511 (0.254)	−0.230 (0.162)	0.126 (0.659)	0.177 (0.245)	0.156 (0.306)
Regression diagnostics							
RMSE	0.274	0.226	0.464	0.180	0.153	0.169	0.172
Share trends (n. trends)				0.24 (93)	0.18 (69)	0.27 (101)	0.25 (96)
Cross−sec. aver. (*p*-value)					913.73 (0.000)		
Residuals diagnostics							
Mean $\lvert\rho_{ij}\rvert$	0.300	0.225	0.177	0.203	0.164	0.157	0.154
CD test (*p*-value)	6.80 (0.000)	235.28 (0.000)	63.23 (0.000)	239.86 (0.000)	3.06 (0.002)	1.89 (0.059)	1.94 (0.053)
Stationarity	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)
Coefficients diagnostics							
Per capita GDP	0.06 (0.810)	7.30 (0.007)	3.11 (0.078)	1.09 (0.296)	7.47 (0.006)	18.97 (0.000)	18.74 (0.000)
Unempl. rate	0.97 (0.325)	−43.32 (0.000)	7.54 (0.006)	3.31 (0.069)	0.04 (0.846)	2.87 (0.090)	3.47 (0.062)
Population	1.85 (0.175)	1.18 (0.277)	9.71 (0.002)	1.08 (0.299)	0.58 (0.447)	0.11 (0.742)	0.23 (0.628)
Av. years sch.	8.31 (0.004)	50.755 (0.000)	28.28 (0.000)	15.95 (0.000)	0.07 (0.787)	0.08 (0.773)	0.53 (0.466)

Notes: Obs 13,680; unit 380; robust *p*-values in parentheses: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Constant included but not reported. All heterogeneous estimates are outlier-robust. In columns (2) and (3) a common trend is introduced in the cointegrating equation. In columns (4), (5), (6) and (7) a group-specific linear time trend is introduced in the regressions (coefficients not reported). RMSE is the root mean square error. Share trends (n. trends) reports the share (number) of region-pair specific time trend statistically significant at 10 per cent level significance. Cross-sec. aver. is a χ^2 test that the coefficients on cross-sectional averages are jointly zero. Mean $|\rho_{ij}|$ is the mean absolute correlation coefficient of residuals. Pesaran (2004) 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). Coefficients diagnostics are χ^2 tests for the following restrictions on coefficient of equation (5): $\gamma_1 = -\gamma_2$, $\gamma_3 = -\gamma_4$, $\gamma_5 = \gamma_6$, $\gamma_7 = -\gamma_8$. See the main text for more details.

Table 8. Estimation results for the South to Centre–North sample of Italian regions

	(1) MG	(2) CCEMG	(3) AMG	(4) AMGC
Per capita GDP at origin	-0.506^{***} (0.001)	-0.484^{**} (0.030)	-0.313^{**} (0.035)	-0.268^{*} (0.065)
Per capita GDP at destination	0.681^{***} (0.000)	0.875^{***} (0.002)	0.849^{***} (0.000)	0.854^{***} (0.000)
Unempl. rate at origin	0.159^{***} (0.000)	0.019 (0.701)	0.034 (0.223)	0.013 (0.639)
Unempl. rate at destination	-0.209^{***} (0.000)	−0.037 (0.340)	-0.054^{**} (0.012)	-0.046^{**} (0.020)
Population at origin	-2.253^{***} (0.001)	4.120^{**} (0.026)	−0.670 (0.268)	−0.488 (0.421)
Population at destination	1.117^{*} (0.092)	1.101 (0.591)	3.058^{***} (0.000)	2.933^{***} (0.000)
Av. years sch. at origin	−0.455 (0.185)	0.320 (0.453)	0.127 (0.693)	0.101 (0.764)
Av. years sch. at destination	−0.366 (0.203)	−0.689 (0.178)	0.244 (0.416)	0.355 (0.232)
Regression diagnostics				
RMSE	0.156	0.116	0.140	0.143
Share trends (n. trends)	0.29 (28)	0.07 (7)	0.33 (32)	0.30 (29)
Cross-sec. averages		435.94 (0.000)		
Residuals diagnostics				
Mean $\lvert\rho_{ij}\rvert$	0.236	0.180	0.165	0.163
CD test (*p*-value)	80.89 (0.000)	1.13 (0.258)	0.46 (0.643)	0.75 (0.451)
Stationarity	I(0)	I(0)	I(0)	I(0)
Coefficients diagnostics				
Per capita GDP	0.65 (0.419)	1.20 (0.272)	6.36 (0.012)	8.09 (0.004)
Unempl. rate	1.59 (0.207)	0.07 (0.785)	0.32 (0.572)	0.95 (0.331)
Population	12.84 (0.000)	1.20 (0.274)	16.53 (0.000)	13.84 (0.000)
Av. years sch.	3.37 (0.066)	0.31 (0.580)	0.71 (0.399)	1.03 (0.310)

Notes: Obs 3,456; unit 96; robust *p*-values in parentheses: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. See Table 7 and the main text for more details.

possible explanations for this result. The first calls for agglomeration economies linked to human capital at origin that prevent individuals from moving; the second suggests that rewards to skill are higher at origin rather than at destination and that human capital at destination might be region specific and different from that owned by would-be migrants. However, these results, particularly those regarding the role of human capital, are valid for overall flows and they need to be carefully and separately verified for South to Centre-North and Centre-North to South flows.

5.2 Empirical results for the South to Centre-North sample

Table 8 reports the results of the heterogeneous estimators for the South to Centre-North sample.[12] The CIPS test strongly rejects the null hypothesis of non stationary for all estimates, proving that the variables are cointegrated. The CD test refuses the null of cross-sectional independence in the MG estimate, whereas it does not reject it in the heterogeneous estimates which, in addition, also have lower levels of both RMSE and mean absolute correlation coefficient. Thus, regression and residual diagnostics tests favour CCEMG and AMG estimates; yet, in the light of these tests alone, it is not easy to compare the CCEMG vis-à-vis AMG estimates. The CCEMG estimate has a lower RMSE but higher mean absolute residuals correlation coefficients. It is possible that these estimates, along with common elements, capture different facets of South to Centre-North bilateral flows. As a general result, CCEMG, AMG and AMGC estimates suggest that *per capita* GDP in the *Mezzogiorno* regions operates

[12] Results for homogeneous estimators, analogously to those of Table 7 for the whole sample, are plagued by cross-sectional dependence. Though not reported here to save space, they are available upon request.

as a restraining factor – estimated coefficients vary from –0.268 in column (4) to –0.484 in column (2) – whereas *per capita* GDP in the Centre-Northern regions works as a pull factor with an estimated elasticity very robust and stable around 0.85\0.87. The CCEMG regression highlights also the role as a push factor of population at origin, with a very high estimated elasticity, above 4. In addition, both the AMG and the AMGC regressions find strong support for the unemployment rate at destination (coefficients around –0.05) and population at destination (near or above 3). Overall these results prove to be consistent with theoretical expectations, nevertheless none of the estimates find a statistically significant impact for human capital, neither at origin, nor at destination. Note, however, that in the AMG and AMGC regressions, that we believe are to be trusted more, the sign of human capital is positive at both origin and destination which is coherent with the idea that agglomeration forces in the Centre-North attract individuals from the South and that, at the same time, higher levels of human capital in the *Mezzogiorno* spur southerners to go to the Centre-North.[13]

In general, the results regarding human capital might appear a little surprising: given that the recent South to Centre-North migration flows have been characterized by individuals with high level of human capital (Piras 2005; Etzo 2011; Fratesi and Percoco 2014), one would expect they should be significantly affected by the level of human capital at origin and/or at destination. According to Murat and Paba (2002) in the case of Italy, human capital required in the Centre-North is region specific and different from that owned by would-be migrants from the South. Italian productive structure is largely based on the industrial districts, mainly localized in the Centre-North. In these districts human capital belongs to native workers and can be acquired only at a positive cost by southern migrants who, *ceteris paribus*, are discouraged from moving because of such a positive cost. It might be the case that this peculiarity of the Italian productive system has curbed the role of human capital as a driver of internal migration. To further deepen this point, we have run regressions with heterogeneous estimators for the period from 1994 onwards (during which an increasing wave of interregional migration has occurred) and the results (not reported here but available upon request) confirm a positive role for human capital at origin in all regressions[14] and mixed effects, though never statistically significant, for human capital at destination.

How come, then, can the significant negative impact of human capital at origin for the whole sample of Italian regions shown in Table 7 be explained?

5.3 *Empirical results for the Centre-North to South sample*

The answer to the previous question, quite unexpected, is found looking at the results for the Centre-North to South sample reported in Table 9. Let us first comment the overall performance of the regressions and the estimated coefficients other than those of human capital. The CIPS test rejects the null of non stationary in all regressions, while the CD test on the one hand rejects the null of cross-sectional independence in the MG estimates, on the other hand it does not reject it for all heterogeneous regressions. In the case of Centre-North to South flows the CCEMG performs very poorly and only the AMG and AMGC estimates deliver good estimates. From the results shown in columns (3) and (4), one can infer that Centre-North to South migration flows are negative linked with *per capita* GDP at origin and positively related with both *per capita* GDP at destination and population at origin. In addition, also unemployment rate at both origin and destination play a significant role in explaining them. On the contrary, population at destination does not affect them.

As for human capital, there is no statistically significant link between migration flows and human capital at destination, while a negative and statistically significant role for human capital at origin is

[13] We find the same results using the quota of graduates as an alternative measure of human capital.

[14] In the AMG regression the estimated elasticity is highly statistically significant, whereas in the AMGC regression the significance level is slightly above 10 per cent.

Table 9. Estimation results for the Centre-North to South sample of Italian regions

	(1) MG	(2) CCEMG	(3) AMG	(4) AMGC
Per capita GDP at origin	-0.710^{***} (0.000)	-0.399 (0.216)	-0.325^{**} (0.040)	-0.196 (0.204)
Per capita GDP at destination	0.421^{**} (0.045)	-0.007 (0.976)	0.403^{***} (0.006)	0.286^{**} (0.045)
Unempl. rate at origin	0.119^{***} (0.000)	0.170^{***} (0.001)	0.045^{*} (0.070)	0.057^{**} (0.028)
Unempl. rate at destination	-0.072^{**} (0.026)	-0.030 (0.561)	-0.055^{**} (0.047)	-0.059^{*} (0.054)
Population at origin	4.580^{***} (0.000)	4.238^{*} (0.087)	3.726^{***} (0.000)	3.746^{***} (0.000)
Population at destination	-1.244^{*} (0.087)	-2.976 (0.164)	-0.305 (0.668)	-0.662 (0.360)
Av. years sch. at origin	-1.483^{***} (0.000)	-0.151 (0.815)	-0.647^{**} (0.048)	-0.679^{**} (0.039)
Av. years sch. at destination	0.341 (0.299)	0.186 (0.770)	0.348 (0.225)	0.364 (0.201)
Regression diagnostics				
RMSE	0.194	0.147	0.179	0.183
Share trends (n. trends)	0.18 (17)	0.08 (8)	0.18 (17)	0.20 (19)
Cross-sec. averages		343.34 (0.000)		
Residuals diagnostics				
Mean $\|\rho_{ij}\|$	0.206	0.173	0.162	0.163
CD test (*p*-value)	64.62 (0.000)	0.60 (0.549)	0.90 (0.366)	1.63 (0.102)
Stationarity	I(0)	I(0)	I(0)	I(0)
Coefficients diagnostics				
Per capita GDP	0.99 (0.320)	1.01 (0.315)	0.13 (0.717)	0.18 (0.669)
Unempl. rate	1.14 (0.285)	3.62 (0.057)	0.07 (0.785)	0.00 (0.955)
Population	30.06 (0.000)	4.87 (0.027)	13.10 (0.000)	15.25 (0.000)
Av. years sch.	5.26 (0.022)	0.00 (0.969)	0.47 (0.493)	0.53 (0.468)

Notes: Obs 3,456; unit 96; robust *p*-values in parentheses: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. See Table 7 and the main text for more details.

detected in columns (3) and (4), with an estimated elasticity stable around a value of –0.65\–0.68.[15] This result conforms and strengthens what previously hypothesized in relation to what found for the whole sample of Italian regions. Indeed, it is conceivable that since the Centre-Northern area of Italy is the most advanced, the higher levels of human capital in these regions favour the creation of agglomeration economies that deter individuals from migrating towards the *Mezzogiorno* regions. At the same time, human capital in Southern regions (at destination) has no capacity to attract people from the Centre-North. In this respect, the results of Fratesi and Percoco (2014) who find that human capital gains from migration in Italy are larger in regions where the human capital stock is large, may be seen as an indirect proof of agglomeration economies taking place in the Centre-Northern regions of the country. Furthermore, since it has been shown (Guagnini and Mussida 2009) that rewards to skill are higher in the Centre-Northern regions, rather than in the *Mezzogiorno*, it is possible that people from the Centre-North are less willing to migrate to the South for this reason as well. To sum up and to answer the question posed above: the negative impact of human capital at origin for the whole sample of Italian regions is not explained by South to Centre-North flows, rather it is attributable to Centre-North to South flow.

It is worth noticing that earlier studies (Etzo 2011; Mocetti and Porello 2012; Piras 2012b) did not find any satisfactory explanation of Centre-North to South flows when using standard econometric techniques (fixed effects, generalized methods of moments and fixed effects vector decomposition) without the regional average human capital and attributed the lack of explicative power of their estimated models to return migration. Many of those individuals who emigrated

[15] Using the quota of graduates we find similar results, although the general performances of the estimated equations are poorer with respect to those of Table 9.

from the *Mezzogiorno* to the Centre-North during the 1950s and 1960s decided to come back to their regions of origin after they retired in the following decades, thus partially fuelling the Centre-North to South flows. These individuals are not motivated by economic factors when making such a decision, rather other cultural, social and ageing factors come into play. The results presented in Table 9 partially overturn this explanation and answer the failure of previous empirical findings, calling into question both the theoretical model and the empirical methodology used in the preceding works. We are not saying that return migration is not relevant and that other factors do not come into play when a retiree decide to come back towards her/his native region. We claim that appropriate econometric techniques are needed in order to tackle the dynamic characteristics of the series, regional heterogeneity, presence of cross-sectional correlation and of common factors. By so doing, Centre-North to South migration flows can also be framed inside the gravity model with human capital.

6 Conclusions

In this paper we have investigated bilateral migration flows across Italy during the 1970–2005 time period, thus encompassing 36 years of internal mobility. We have estimated a gravity model of internal migration with human capital in which bilateral flows have been regressed on *per capita* GDP, unemployment rate, population and average years of schooling.

Given that the main results of the paper have been stressed throughout the text, here they can be shortly summarized. The gravity model augmented with human capital explains quite well the long-run patterns of Italian internal migration flows, confirming that the macroeconomic variables (*per capita* GDP and unemployment) are the main drivers of migration flows in the long run as previously found in the literature (Etzo 2011; Mocetti and Porello 2012; Piras 2012a, 2012b). Novel and interesting results have been found regarding the role of human capital. While at destination it has had no role, at origin it has worked as a restraining factor. Maybe quite surprising at first sight, such a restraining role has been operating in the Centre-North to South direction rather than vice versa, as one might have expected given the long-run trends of internal migration flows across Italian regions. Two main explanations have been provided for this result: the first in terms of agglomeration economies, the second calls for higher rewards to skill. Both of them benefit the Centre-North and disadvantage the South.

As far as the econometric methodology is concerned, it has been proved that homogeneous estimators are not able to cope with heterogeneity of the data and that more general approaches need to be used. Therefore, in this paper for the first time the CCEMG and the AMG estimators have been applied for the case of internal migration and for the general case of multiple origin and multiple destination flows. These novel heterogeneous estimators deal with cross-sectional dependence, presence of unobserved common factors, non-stationary of variables, presence of global/local business cycles and structural breaks. In addition and differently from alternative empirical approaches, they are also easy to implement. Preliminary to this, a careful analysis of the dynamic characteristics of the series has been carried out, showing that stationary cannot be assumed to hold and that such an issue should be a concern in applied research on this topic.

Appendix A

For each region, Table A1 reports migration data (in-flows and out-flows) along with the data of the other explicative variables. Recalling that we analyse gross bilateral migration flows from each region towards any other region, the migration data reported in Table A1 provide just a synthetic piece of information regarding the regions that, on average during the 1970 to 2005 time period, have been losing or gaining population due to internal migration. Explicative variables at origin of migration flows refer to the single region identified by row in the table. By

Table A1. Summary statistics

Region		Migration in-flows (units)	Migration out-flows (units)	Variables at origin of migration flows				Variables at destination of migration flows			
				Per capita GDP (€ 1995)	Unemp. rate (%)	Population (thousand)	Average Years Schooling	Per capita GDP (€ 1995)	Unemp. Rate (%)	Population (thousand)	Average Years Schooling
Centre-Northern regions											
	Mean	1,827	1,646	16,257	6.16	4,375	6.35	13,630	9.30	2,758	6.29
Piemonte	s. d.	2,265	1,588	3,198	2.43	84	1.28	4,572	6.10	2,254	1.37
	c. v.	1.24	0.96	0.20	0.39	0.02	0.20	0.34	0.66	0.82	0.22
	Mean	74	55	20,935	4.18	115	6.15	13,383	9.40	2,982	6.30
Valle d'Aosta	s. d.	126	93	2,758	1.16	4	1.43	4305	6.05	2,191	1.36
	c. v.	1.69	1.69	0.13	0.28	0.04	0.23	0.32	0.64	0.73	0.22
	Mean	2,991	2,382	17,951	4.66	8,908	6.50	13,541	9.38	2,520	6.28
Lombardia	s. d.	3,047	1,887	3,856	1.79	192	1.42	4476	6.06	1,779	1.36
	c. v.	1.02	0.79	0.21	0.38	0.02	0.22	0.33	0.65	0.71	0.22
	Mean	217	197	18,443	3.66	895	6.23	13,515	9.43	2,941	6.29
Trentino Alto Adige	s. d.	249	278	4,095	1.49	36	1.33	4437	6.02	2,237	1.37
	c. v.	1.15	1.41	0.22	0.41	0.04	0.21	0.33	0.64	0.76	0.22
	Mean	1,013	793	15,616	5.03	4,387	6.20	13,663	9.36	2,757	6.29
Veneto	s. d.	1,041	923	3,827	1.93	140	1.55	4564	6.07	2,253	1.36
	c. v.	1.03	1.16	0.25	0.38	0.03	0.25	0.33	0.65	0.82	0.22
	Mean	418	326	15,072	5.52	1,210	6.56	13,692	9.33	2,925	6.27
Friuli Venezia Giulia	s. d.	533	461	3,825	2.14	18	1.33	4574	6.08	2,250	1.37
	c. v.	1.27	1.41	0.25	0.39	0.01	0.20	0.33	0.65	0.77	0.22
	Mean	801	747	14,983	7.59	1,731	6.75	13,697	9.22	2,897	6.26
Liguria	s. d.	1,008	911	3,307	2.77	96	1.28	4597	6.12	2,268	1.37
	c. v.	1.26	1.22	0.22	0.36	0.06	0.19	0.34	0.66	0.78	0.22
	Mean	1,457	885	17,125	5.08	3,950	6.39	13,584	9.35	2,780	6.28
Emilia Romagna	s. d.	1,418	889	3,956	1.73	67	1.33	4,510	6.08	2,268	1.37
	c. v.	0.97	1.00	0.23	0.34	0.02	0.21	0.33	0.65	0.82	0.22
	Mean	,1173	762	15,,300	6.61	3,546	6.28	13,680	9.27	2,802	6.29
Toscana	s. d.	1,043	672	3075	2.39	34.37	1.48	4600	6.11	2,277	1.36
	c. v.	0.89	0.88	0.20	0.36	0.01	0.24	0.34	0.66	0.81	0.22
	Mean	306	232	13,359	7.92	815	6.39	13,782	9.20	2,945	6.28
Umbria	s. d.	455	388	3,027	2.63	21	1.38	4615	6.13	2,233	1.37
	c. v.	1.49	1.68	0.23	0.33	0.03	0.22	0.33	0.67	0.76	0.22

(Continues)

Table A1. (Continued)

Region		Migration in-flows (units)	Migration out-flows (units)	Variables at origin of migration flows				Variables at destination of migration flows			
				Per capita GDP (€ 1995)	Unemp. rate (%)	Population (thousand)	Average Years Schooling	Per capita GDP (€ 1995)	Unemp. Rate (%)	Population (thousand)	Average Years Schooling
	Mean	440	338	13,877	5.57	1,429	6.13	13,755	9.33	2,913	6.30
Marche	s. d.	426	403	2,949	1.71	39	1.39	4619	6.09	2,259	1.37
	c. v.	0.97	1.19	0.21	0.31	0.03	0.23	0.34	0.65	0.78	0.22
	Mean	1,820	1,489	15,507	9.25	5,068	7.24	13669	9.13	2,722	6.24
Lazio	s. d.	1,671	1,043	3,294	2.76	173	1.29	4,588	6.13	2,221	1.35
	c. v.	0.92	0.70	0.21	0.30	0.03	0.18	0.34	0.67	0.82	0.22
Southern regions											
	Mean	429	410	11,990	8.31	1,242	6.28	13,854	9.18	2,923	6.29
Abruzzo	s.d.	527	595	2,716	2.30	38	1.38	4,607	6.14	2,252	1.37
	c. v.	1.23	1.45	0.2333	0.28	0.03	0.22	0.33	0.67	0.77	0.22
	Mean	139	162	10,690	11.04	328.2	6.05	13,923	9.04	2,971	6.30
Molise	s. d.	182	201	2,734	4.27	5	1.33	4569	6.07	2,205	1.37
	c. v.	1.31	1.24	0.26	0.39	0.01	0.22	0.33	0.67	0.74	0.22
	Mean	1,100	2,012	9,382	16.51	5,574	6.27	13,992	8.75	2,695	6.29
Campania	s. d.	1,205	2,413	1,630	6.86	238	1.20	4537	5.70	2,189	1.38
	c. v.	1.10	1.20	0.17	0.42	0.04	0.19	0.32	0.65	0.81	0.22
	Mean	969	1,446	9,484	12.41	3,942	5.90	13,986	8.97	2,781	6.31
Puglia	s. d.	1,139	1,946	1,758	4.92	162	1.25	4540	6.01	2,268	1.37
	c. v.	1.17	1.35	0.19	0.40	0.04	0.21	0.32	0.67	0.82	0.22
	Mean	217	357	9,683	14.23	610	5.92	13,976	8.87	2,956	6.31
Basilicata	s. d.	291	511	2,045	5.24	7	1.23	4544	5.93	2,222	1.37
	c. v.	1.34	1.43	0.21	0.37	0.01	0.21	0.33	0.67	0.75	0.22
	Mean	673	1,151	8,695	17.72	2,063	6.10	14,028	8.70	2,880	6.30
Calabria	s. d.	855	1,683	1,764	7.03	48	1.28	4495	5.60	2,276	1.37
	c. v.	1.27	1.46	0.20	0.40	0.02	0.21	0.32	0.64	0.79	0.22
	Mean	1,111	1,696	9,850	15.71	4,978	6.04	13,967	8.79	2,726	6.30
Sicilia	s. d.	1,511	2,546	1,763	7.33	139	1.19	4558	5.72	2,226	1.37
	c. v.	1.36	1.50	0.18	0.47	0.03	0.20	0.33	0.65	0.82	0.22
	Mean	437	686	1,809	5.79	58	1.26	13,905	8.80	2,904	6.30
Sardegna	s. d.	538	7	8,054	3.85	1,,467	3.96	4604	5.82	2,265	1.37
	c. v.	1.23	1.30	0.16	0.37	0.04	0.21	0.33	0.66	0.78	0.22

(Continues)

Table A1. (Continued)

Region		Migration in-flows (units)	Migration out-flows (units)	Variables at origin of migration flows				Variables at destination of migration flows			
				Per capita GDP (€ 1995)	Unemp. rate (%)	Population (thousand)	Average Years Schooling	Per capita GDP (€ 1995)	Unemp. Rate (%)	Population (thousand)	Average Years Schooling
					Italy						
Italy	Mean	880.6	880.6	13,761	9.14	2839	6.29	13,761	9.14	2,839	6.29
	s. d.	1407	1407	4,547	6.00	2224	1.37	4,547	6.00	2,224	1.37
	c. v.	1.60	1.60	0.33	0.66	0.78	0.22	0.33	0.66	0.78	0.22

contrast, variables at destination of migration flows represent their average values in all 19 destination region. Let us consider Piemonte in the first row of Table A1. On the one hand, this region has recorded positive net migration flows (in-flows are higher than out-flows), *per capita* GDP was 16,257 € (constant price 1995), unemployment rate 6.16 per cent, population 4,375,000 and average years of schooling (our measure of regional human capital) 6.35 years. On the other hand, the average per capita GDP in all other 19 regions towards which Piemonte has sent migration flows was 13,630 €, unemployment rate 9.30 per cent, population 2,758,000 and average years of schooling 6.29 years.

References

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

Attanasio O, Padoa-Schioppa F (1991) Regional inequalities, migration and mismatch in Italy 1960–1986. In: Padoa-Schioppa F (ed) *Mismatch and labour mobility*. Cambridge University Press, Cambridge

Bai J, Kao C, Ng S (2009) Panel cointegration with global stochastic trends. *Journal of Econometrics* 149: 82–99

Bailey N, Kapetanios G, Pesaran MH (2012) *Exponent of cross-sectional dependence: Estimation and inference* (CESIFO Working Paper 3722). CESifo GmbH (Munich Society for the Promotion of Economic Research), München

Basile R, Causi M (2007) Le determinanti dei flussi migratori nelle province italiane: 1991–2001. *Economia & Lavoro* 41: 139–159

Bazzi S, Clemens M (2013) Blunt instruments: Avoiding common pitfalls in identifying the causes of economic growth. *American Economic Journal, Macroeconomics* 5: 152–186

Beine M, Bertoli S, Fernández-Huertas Moraga J (2015) A practitioners' guide to gravity models of international migration. *The World Economy*. doi:10.1111/twec.12265

Beine M, Docquier F, Rapoport H (2001) Brain drain and economic growth: Theory and evidence. *Journal of Development Economics* 64: 275–289

Bertoli S, Brücker H, Fernández-Huertas Moraga J (2013) *The European crisis and migration to Germany: Expectations and the diversion of migrants flows* (FEDEA Working Paper 3). Fundación de Estudios de Economía Aplicada, Madrid

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

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

Blanchard O, Katz L (1992) Regional evolutions. *Brookings Papers on Economic Activity* 1: 1–61

Bonasia M, Napolitano O (2012) Determinants of interregional migration flows: The role of environmental factors in the Italian case. *The Manchester School* 80: 525–544

Bond S, Eberhardt M (2013) *Accounting for unobserved heterogeneity in panel time series models*. Nuffield College, University of Oxford, Mimeo

Borjas G (1991) Immigration and self-selection. In: Abowod J, Freeman R (eds) *Immigration, trade, and the labor market*, University of Chicago Press, IL, Chicago.

Brunello G, Lupi C, Ordine P (2001) Widening differences in Italian regional unemployment. *Labour Economics* 8: 103–129

Brunow S, Nijkamp P, Poot J (2015) The impact of international migration on economic growth in the global economy. In: Chiswick BR, Miller PW (eds) *Handbook of the economics of international migration* (Vol. 1B). North-Holland, Amsterdam

Cannari L, Nucci F, Sestito P (2000) Geographic labour mobility and the cost of housing: Evidence from Italy. *Applied Economics* 32: 1899–1906

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

Coakley J, Fuertes AM, Smith R (2006) Unobserved heterogeneity in panel time series models. *Computational Statistics & Data Analysis* 50: 2361–238

Daveri F, Faini R (1999) Where do migrants go? *Oxford Economic Papers* 51: 595–622

Decressin J, Fatas A (1995) Regional labour market dynamics in Europe. *European Economic Review* 39: 1627–1655

Desbordes R, Eberhardt M (2014) Theory and Estimation of Gravity Equations. *Gravity models in the presence of slope heterogeneity and cross-section dependence*. Paper presented at the Scottish Institute for Research Economics workshop on 'Theory and Estimation of Gravity Equations' at Strathclyde University, Mimeo

Docquier F, Peri G, Ruyssen I (2014) The cross-country determinants of potential and actual migration. *International Migration Review* 48: S37–S99

Driscoll JC, Kraay AC (1998) Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics* 80: 549–560

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

Eberhardt M, Teal F (2013a) Structural change and cross-country growth empirics. *The World Bank Economic Review* 27: 229–271

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

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

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

Faggian A, McCann P (2009) Human capital and regional development. In: Capello R, Nijkamp P (eds) *Handbook of regional growth and development theories*. Edward Elgar, Cheltenham.

Fachin S (2007) Long-run trends in internal migration in Italy: A study in panel cointegration with dependent units. *Journal of Applied Econometrics* 22: 401–428

Faini R, Galli G, Gennari P, Rossi F (1997) An empirical puzzle: Falling migration and growing unemployment differentials among Italian regions. *European Economic Review Papers and Proceedings* 41: 571–579

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

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

Friedberg RM (2000) You can't take it with you? Immigrants assimilation and the portability of human capital. *Journal of Labor Economics* 18: 221–151

Fujita M, Thiesse JF (2002) *Economics of agglomeration*. Cambridge University Press, Cambridge

Furceri D (2006) Does labour respond to cyclical fluctuations? The case of Italy. *Applied Economics Letters* 13: 135–139

Guagnini M, Mussida C (2009) Il rendimento dell'istruzione nelle regioni italiane. AIEL Conference Paper, XXIV National Conference of Labour Economics (AIEL), Sassari (Italy)

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

Hamilton LC (1992) How robust is robust regression? *Stata Technical Bulletin* 2: 21–26

Hanson GH (2010) International migration and the developing world. In: Rodrik D, Rosenzweig M (eds) *Handbook of Development Economics* (Vol. 5). North-Holland, Amsterdam

Harris J, Todaro M (1970) Migration, unemployment, and development: A two sector analysis. *American Economic Review* 60: 126–142

Hatton T, Williamson J (2010) Are third world emigration forces abating? *World Development* 39: 20–32

Hunt J (2006) Staunching emigration from East Germany: Age and the determinants of migration. *Journal of the European Economic Association* 4: 1014–1037

Im KS, Pesaran MH, Shin Y (2003) Testing for unit roots in heterogeneous panels. *Journal of Econometrics* 115: 53–74

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

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

Kapetanios G, Pesaran MH, Yamagata T (2011) Panels with nonstationary multifactor error structures. *Journal of Econometrics* 160: 326–348

Karemera D, Iwuagwu V, 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

Kao C, Chiang MH (2000) On the estimation and inference of a cointegrated regression in panel data. In: Baltagi BH (ed) *Advances in econometrics: Nonstationary panels, panel cointegration and dynamic panels* (Vol. 15). Elsevier, New York

Krugman P (1991) Increasing returns and economic geography. *Journal of Political Economy* 99: 483–499

Lamonica G, Zagaglia B (2013) The determinants of internal mobility in Italy 1995–2006: A comparison of Italians and resident foreigners. *Demographic Research* 29: 407–440

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

Maza A (2006) Migration and regional convergence: The case of Spain. *Review of Regional Research/Jahrbuch für Regionalwissenschaft* 26: 191–202

Maza A, Villaverde J (2004) Interregional migration in Spain: A semiparametric analysis. *Review of Regional Studies* 34: 156–171

Mitze T, Reinkowski J (2011) Testing the neoclassical migration model: Overall and age-group specific results for German regions. *Journal for Labour Market Research* 43: 277–297

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

Mountford A (1997) Can a brain drain be good for growth in the source economy? *Journal of Development Economics* 53: 287–303

Murat M, Paba S (2002) Capitale umano specifico e flussi migratori. *Rivista di Politica Economica* 92: 63–108

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

Pesaran MH (2004) *General diagnostic tests for cross section dependence in panels* (IZA Discussion Paper 1240). Institute for the Study of Labor, Bonn

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

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

Pesaran MH, Shin Y, Smith RP (1999) Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American Statistical Association* 94: 621–634

Pesaran MH, Smith RP (1995) Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics* 68: 79–113

Piras R (2005) Il contenuto di capitale umano dei flussi migratori interregionali: 1980–2002. *Politica Economica* 21: 461–491

Piras R (2012a) Internal migration across Italian regions: Macroeconomic determinants and accommodating potential for a dualistic economy. *The Manchester School* 80: 499–524

Piras R (2012b) Disentangling the role of migrants' educational level in the long-run Italian internal migration trends. *Studies in Regional Science* 42: 377–396

Piras R (2013) Can the augmented Solow model with migration explain the Italian internal brain drain? *Labour* 27: 140–163

Piras R, Melis S (2007) Evoluzione e tendenze delle migrazioni interne. *Economia Italiana* 2: 437–469

Pugliese E (2002) *L'Italia tra migrazioni internazionali e migrazioni interne*. Il Mulino, Bologna

Salvatore D (1977) An econometric analysis of internal migration in Italy. *Journal of Regional Science* 17: 395–408

Santos Silva JMC, Tenreyro S (2006) The log of gravity. *Review of Economics and Statistics* 88: 641–658

SVIMEZ (2015) *Rapporto 2014 sull'economia del mezzogiorno*. Il Mulino, Bologna

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

Tenreyro S (2007) On the trade impact of nominal exchange rate volatility. *Journal of Development Economics* 82: 485–508

Westerlund J, Urbain JP (2015) Cross sectional averages versus principal components. *Journal of Econometrics* 185: 372–377