Università degli Studi di Cagliari

Essays on Internal Migration Determinants
From a macro to a micro approach

Dottorato di Ricerca in Economia

Scuola di Dottorato in Scienze Economiche ed Aziendali

XXVII Ciclo

Settore scientifico-disciplinare: SECS-P/06

Presentata da: Elisa Gagliardini

Coordinatore: Prof. Romano Piras

Tutor: Prof. Raffaele Paci e Prof.ssa Emanuela Marrocu

Esame finale anno accademico 2014 – 2015

“La presente tesi è stata prodotta durante la frequenza del corso di dottorato in Economia dell’Università degli Studi di Cagliari,- XXVII ciclo, con il supporto di una borsa di studio finanziata con le risorse del P.O.R. SARDEGNA F.S.E. 2007-2013 - Obiettivo competitività regionale e occupazione, Asse IV Capitale umano, Linea di Attività l.3.1 “Finanziamento di corsi di dottorato finalizzati alla formazione di capitale umano altamente specializzato, in particolare per i settori dell’ICT, delle nanotecnologie e delle biotecnologie, dell'energia e dello sviluppo sostenibile, dell'agroalimentare e dei materiali tradizionali”.

Elisa Gagliardini gratefully acknowledges Sardinia Regional Government for the financial support of her PhD scholarship (P.O.R. Sardegna F.S.E. Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2007-2013 - Axis IV Human Resources, Objective l.3, Line of Activity l.3.1.)”.

Unione europea
Fondo sociale europeo

FSE 2007-2013
POR SARDEGNA
obiettivo competitività regionale e occupazione

REGIONE AUTÒNOMA DE SARDIGNA
REGIONE AUTONOMA DELLA SARDEGNA

Abstract

Italy has a long history of internal migration. The arguments addressed by the researchers in the last century cover two main questions: *where* migrants go, and *who* they are. This thesis focuses on these aspects using two different approaches.

The first study is based on a macro-approach. By means of a spatial gravity model, we investigate the determinants of internal migration using bilateral flows across Italian regions in the period 2000-2013. We address the issue of cross-regional dependence arising from the existence of regional spillovers by including spatial lags of the explanatory variables. The main results indicate the importance of spatial dependency induced by neighbouring regions at origin, and at destination. Interesting results are found for two different sub-sample of population: foreigners and Italians.

The second study focuses on individuals' behaviour. Weighted logit models of the probability that an individual changes his or her region of residence from one year to the next over the 2011–2012 periods are estimated using Labour Force Survey data. Our results show that alongside strictly economic determinants, migration choices are driven by a large set of personal, professional, and family characteristics.

Contents

List of Tables

List of Figures

Acknowledgments

First and foremost I want to thank my supervisor Prof.ssa Emanuela Marrocu for guiding me in developing this thesis, and for her inestimable contribution of time and ideas. I would also like to thank Prof. Raffaele Paci for encouraging my research since the beginning.

I have benefited from valuable comments and suggestions by Prof. Romano Piras, Prof. Stefano Usai, Dott. Giovanni Sulis and all the participants at the internal seminars at University of Cagliari.

Furthermore, I am grateful to my fellow colleagues, Tiziana, Vania, Daniela and Marco for their constant support and especially to Benedetta and Federica for their precious friendship.

Lastly, I would like to thank my wonderful parents, all my family and Antonio for all their love, their patience and encouragement during these years.

Summary

As many other European countries characterized by large sub-national disparities, Italy has a long and complex history of population flows from the Southern backward regions to the rest of the country. Italian migration has changed over time, from being sizeable and unqualified (in particular during the so called *miracolo economico*, i.e. between 1950s and 1970s) to a more 'selective' process.

For its intensity and particular features, we consider Italy as our specific case-study of this thesis. The aim is to analyse this phenomenon covering two important and related aspects. The first aspect concerns the role played by push and pull factors in determining the directions of the migration flows. The second concerns the role played by personal features on the individual decision to migrate.

In the first Chapter we provide a review of the main empirical literature on internal migration giving particular attention on the different migration modelling approaches. The large stream of literature developed so far, has been mainly based on a macro perspective, namely researchers have focused on migration with respect to the spatial context and the related aggregate variables in order to find the determinants of migration or to study the consequences of migration. Less attention has been devoted to the decision to migrate as micro-economic behaviour.

In the second Chapter, we employ different econometric approaches to estimate the parameters of a spatial gravity model using panel data from 2000-2013 on interregional migration flows in Italy.

The application of a spatially lagged explanatory variables (SLX) model allows us to examine the issues related to spatial dependence patterns in a panel migration context.

Results confirm that in this last decade, migrants continue to respond to economic push and pull factors in their home region and at their potential destinations. Moreover, we show that other location-specific characteristics (like quality of life and infrastructure) may act as potential push and pull factors. The SLX model provides significant evidence of the existence of both origin and destination regional spillovers. Furthermore, using two different subsamples of population and three different migration patterns (total flows, South to Centre- North flows, Centre-North to South flows) we provide a new and better understanding of the multiple aspects of this complex phenomenon.

Differently, in the third and last Chapter we move to a micro approach of the migration decision process. The object of the analysis is indeed the single individual behaviour and the factors that influence his\her decision of whether to migrate or not. Using Labour Force Survey data we estimate weighted logit models of the probability that an individual changes region of residence

from one year to the next over the 2011–2012 period. Consistent with the life-cycle model, our findings suggests that age and family ties are negatively related to mobility: family ties tend to discourage migration. Conversely, having a degree or post-degree qualification is found to increase the probability of migrating. Moreover, the empirical evidence shows that migration decision is also strongly influenced by the previous labour status.

Chapter 1

Literature review on internal migration

1.1 Introduction

Migration is one of the most important social and demographic phenomena. Nowadays, like in the past, a significant number of individuals move internally to the same country, or even to a different one. People may move more than once during their life for different reasons: education, job opportunities, family reunification, retirement, better quality of life and so on.

Since the first scientific work of Ravenstein[1] (1885), the causes and consequences of the spatial relocation of people have become a matter of interest among sociologists, economists, demographers, historians and urban planners although from different perspectives. For more than a century, economists and other social scientists have been interested in three migration questions: *where* migrants go, *why* they go and *who* they are.

The majority of existing studies address the determinants of migration empirically by the estimation of gravity equations[2] (see Anderson 2011 for a review of the gravity model), which allow to identify the effects of potential push and pull factors. According with the standard gravity model, migration is directly correlated with the population size[3] of both origin and destination and inversely correlated with the spatial distance. Geographical distance (a proxy for general migration costs) is identified as the main impedance factor. It has long been central in the migration literature and alternative approaches to proximity have been deeply investigated. In particular, the empirical literature on international migration patterns, shows that other unobserved transaction costs, mainly related to cultural and institutional barriers, are important in explaining bilateral migration patterns (Belot and Ederveen 2012; Caragliu et al. 2013).

The majority of existing studies have augmented the basic gravity model with variables that reflect additional potential determinants. In particular, macroeconomic variables (like gross domestic product, unemployment rate, price index and so on) are generally found to be the most influential factors in explaining the variations in migration flows. However, the role of amenities (and disamenities), immigration policy regimes, or other social and political conditions in the source and destination countries for migration rates are also acknowledged. Recently, the impact of measures of immigration policy has been studied by Ortega and Peri (2009) and Mayda (2010). Others, like Clark et al. (2007), focus on the age distribution of population. American researchers instead, give a

[1] Ravenstein in his famous article on migration, define the so called 'laws of migration' which formed the basis for modern migration theory.

[2] The name gravity model reflects the analogy to the law of universal gravitation developed by Newtown in 1687 to describe the gravitational interaction between two objects, for example planets. This model has been applied for the first time in 1962 by Jan Tinbergen who estimated a gravity equation of international trade flows.

[3] Population is defined a 'mass' variable: a larger population at destination is expected to generate higher emigration volumes while the population size of the destination reflects the size of the labor market that which is supposed to be the target of potential migrants .

special emphasis to the role of natural amenities (such as warm winters, proximity to oceans and lakes, or pleasant landscapes) as important push factors (Glaeser et al., 2001; Partridge and Rickman, 2003).

The important role of personal characteristics in migration research has been considerably emphasized by the past and more recent literature (Greenwood, 1975, 1985, 1997; Plane and Bitter, 1997; Cushing and Poot, 2004). A general conclusion is that migration phenomenon is a *selective* process (Moraga, 2011).

Micro data analysis allow researchers to account for the central role of personal characteristics (e.g. age, education, family structure, labour status) and the extent to which such factors may modify the propensity to move and thus, select migrants. Due to the high availability of macro aggregate data (flows, rates), the empirical literature on migration as an 'aggregate' and macro phenomenon is always much more copious than micro data based migration research. However, during the past three decades, thanks to the rapid advances in computer technologies and availabilities of microdata sets, the use of discrete choice models (logit and probit) have become fairly standard to study the underlying factors which influence the personal decision to migrate.

It has been stressed that these two approaches (macro and micro) are complementary rather than alternative. As Etzo (2008) pointed out 'decision-making process of the single individual affects the aggregate utility function, which in turn determines the aggregate migration flows'. Moreover, some models that use aggregate data are essentially derived from micro theoretical principles (Champion et al., 1998; Ortega and Peri, 2009).

In this thesis we will focus on internal migration which involves the reallocation of people within the national borders. The US is by far, the country where the literature on internal migration patterns is most influent (Borjas, et al. 1992; Davies and Greenwood, 2001; Molloy et al., 2011 are only some examples). Internal migration rates in US are indeed many times greater than those of Europe. However, the empirical literature for European countries like UK (Simpson and Finney, 2009), Spain (Maza and Villaverde, 2001), Netherlands (Venhorst et al., 2011), Finland (Ritsilä and Haapanen, 2003) but also for many Eastern countries (Kulu and Billari, 2004; Paci et al. 2007) is quite abundant. Each country is a specific case study; each of which has a different history and particular features. As a result, it is still difficult to draw general conclusions.

Within the European context, the Italian migration case continues to attract the interest of many researchers for its intensity and particular features. Italy has a long tradition of internal emigration from the less developed Southern regions to the wealthy part of the country (the Centre-North).

Moreover, it is a well known fact that this phenomenon (and especially the migrations of the most skilled part of the population) has contributed to further impoverishing the South and exacerbating

the persistent dualism between the two areas of the country (Piras, 2005). Thus, for its peculiarities and interesting features, we have chosen Italy as object of analysis in this thesis.
In the next Section of this chapter we provide a brief excursus of the main phases of Italian migration history comparing the main empirical contributions in this field of study. Therefore, in the third Section we move our attention at the micro level and we shortly discuss few of the main empirical results.

1.2 A long history of migration

The empirical contributions on interregional migration patterns differ in methodologies adopted, data sample, geographical areas (macro-areas, regions or provinces) and time periods. In what follows we summarize the principal results reached by some of the most relevant Italian migration studies. The aim is to have a general overview of how the explanation of migration determinants has changed during the time. In order to facilitate comparisons across the reviewed contributions, for each of them we report the main features of the analyses in Table 1.1.
The literature identifies four main phases of Italian interregional migration. The first and most important wave concerned the period between the end of the Second World War and late 1960s. Those two decades were characterized by a massive migration flow from the South (so called *Mezzogiorno*) to the main industrial cities of the North: Torino, Milano and Genova (defined *industrial triangle*) and to the capital city, Roma. Indeed, it was during this historical period that Italy experienced the *miracolo economico* (economic miracle) which enabled the country to become one of the most developed economies of the world. It is interesting to note that the phenomenon of outflow migration was not limited to the South, but it also involved some areas of the North. Pugliese (2002) shows that during the period from 1951 to 1975, more than 3 million individuals moved from the Southern towards the Central-Northern regions, while more than one million migrated in the opposite direction.
As far as we know, Salvatore (1977) is the first empirical study of internal migration across Italian regions. Using OLS, he analyses the considerable migration flows from the Mezzogiorno regions to the Northern regions between the 1958 and the 1974. He finds that South-North labour migration responds to interregional unemployment and earnings differentials, thus reflecting a so called disequilibrium model of migration[4].

[4] The theoretical assumption underlying this model states that migration is an adjustment mechanism. People react to initial disequilibria in wages and unemployment by moving to areas where the level of wages is higher and unemployment is lower. This movement would lead to an equalization of labour productivity and regional income per capita, restoring the equilibrium across space (Muth, 1971). The second and alternative view of interregional migration is called the '*equilibrium model*'. According to this view, differences in wages are partially compensating for spatial

The second phase, between the early 70s until the first half of the 90s, was marked by a progressive decline of migratory flows and an increasing economic gap between the two parts of the country. For this reason, researchers started to look at the potential obstacles that held back internal mobility and had widened the secular mismatch between North-South.

Attanasio and Padoa- Schioppa (1991) examine the determinants of the internal migration flows across six macro regions from 1960 to 1986 within a linear regression model. Besides their main conclusion that gross migration rates are strongly correlated with unemployment differentials, they put forward the fact that the financial support from the family (informal network) and the substantial public transfers for the Southern regions have made the unemployment status more affordable, discouraging in this way the migration propensity. This argument has been more deeply studied by Faini et al. (1997). Using a multinomial logit specification, they identify as main causes of this empirical puzzle[5], 'a combination of demographic factors, high mobility costs and inefficiencies in the job matching process' (p. 531). More precisely, they results contradict the idea that family and government support has been used to finance unemployment and avoid migration (as suggested by Attanasio and Padoa- Schioppa, 1991). On the opposite, they find that higher household income (proxied by the family's employment rate) is associated with greater mobility.

Through a logit model for the period 1967-1992, Cannari et al. (2000) provide empirical evidence that the North-South housing price differential was a notable factor in explaining the falling pattern of mobility. As the authors suggest 'the positive impact on migration from the South to the North (..) has been offset by the housing price differential, which has steadily risen at least from the mid-1980s onwards' (p.189). Brunello et al. (2001) use an instrumental variables (IV) estimator, in order to give evidence that the reduction in earnings differentials and the rapid increase in social transfers (mainly during the 1970s and the 1980s), were the main causes to explain this empirical puzzle. Others, like Murat and Paba (2002) focus their attention in the industrial structure. Running separate ordinary least squared (OLS) regressions for four decennial periods (from 1951 to 1991) they show that the rapid increase in technological progress occurred on the last decades studied, had a great influence in labour demand in Centre-North industries. In particular, the high demand of more qualified and specialized workers were mainly satisfied by native workers instead of by potential migrants from the South.

Emigration from the backward regions of the South to the North regained momentum in the second half of the nineties following the phase of stagnation characterizing the previous decade. This is

variations in non-tradable and non-economic factors. Graves (1980) focused on location-specific natural amenities such as climate and temperature.

[5] The term was coined by the authors to define the falling of migration flows with growing unemployment differentials during the second half of '70 and the '80 of last century.

often identify as beginning of a new (the third) phase of internal migration. One of the main distinctive elements of this phase has been the increased number of graduates which moved from the South to the Centre-North regions. Due to the undisputed economic impact of this qualitatively significant group of migrants on regional development, many empirical studies have focused on the analysis and assessment of pull and push factors of highly skilled individuals (Piras, 2012b; Nifo and Vecchione, 2014).

Basile and Causi (2007), propose an accurate analysis of internal migration flows, based on a seemingly unrelated regressions (SUR) model in two distinct periods: 1991-1995 and 1996-2000. By using iterative feasible generalized least squares (FGLS) estimator, they show that, during the period 1996-2000, the net migration flow between Italian provinces tuned to be consistent with the traditional theories in which economic variables (such as unemployment and GDP per capita) play a crucial role in explaining internal migration. As they suggest, this result is mainly due to the increase in labour demand in the Centre-North and the drastic reduction of public support for the South. Conversely, the authors demonstrate that economic fundamentals have no effect during the 1991-1995.

More recently, these results have been confirmed by Etzo (2011) and Piras (2012a) using fixed error vector decomposition (FEVD) and error correction model (ECM) respectively. Etzo (2011) estimates a gravity model to bilateral migration flows across Italian regions for the 1996-2005 time periods including some non-economic variables. Empirical results show that internal migration flows are significantly influenced by the main macroeconomic variables[6]. Additionally, he shows that internal movement has also been favoured by the existence of hard infrastructures (like airports) and that migrants have a certain preference for warmer regions. He finds that Northern migrants (i.e. migrants which move from Centre-Northern to Southern regions) respond differently to the push and pull forces with respect to Southern migrants (i.e. migrants which move from Southern regions to Centre-Northern). Namely, the Northern seem to give more importance to location-specific amenities while the Southern better react to economic incentives. Using a panel cointegration framework, Piras (2012a) provides empirical evidence that per capita GDP, unemployment rate and migrants' human capital are the main determinants of net migration rates across Italian regions from 1970 to 2002[7].

Piras (2015) has recently proposed an accurate analysis of bilateral migration flows across Italy during the 1970-2005 time periods. Besides confirming the role of macroeconomic variables (per

---

[6] More precisely, per capita GDP seems to play a strong role both at origin and destination, whereas the effect of unemployment appear to be significant only in the sending regions.

[7] Similar results are found in Piras (2012b) where he runs separate regressions for high and low skilled migrants. He finds that the former react more promptly to regional unbalances.

capita GDP and unemployment) as main drivers of migration flows; novel and interesting results regarding the role of human capital have been found. Namely, by means of heterogeneous panel data estimators (mean group, common correlated effects mean group, and augmented mean group), the author provide evidence that while there is no significant role of human capital at destination, it acts as a restraining factor at origin. This result is confirmed only for the Centre-North to South direction, the main explanation give by the author is in terms of agglomeration economies which seem to deter individuals from migrating towards the Southern regions.

Biagi et al. (2011) give an important contribution to migration literature decomposing for the first time, mobility flows into short (between provinces within the same region) and long (between provinces of different regions) distance. They estimate an extended version of the gravity model for the years 2001 and 2002, using a negative binomial (NB) model, augmented with instruments to control for potential endogeneity issues. Their main findings demonstrate that long distance migration mainly reflects a disequilibrium model of migration where economic/labour market variables play a dominant role. Short distance movements are rather similar to equilibrium model of migration where, quality of life and amenities differences seem to play a dominant role.

The years after 2000 have been characterized by a new progressive decline of internal migratory flows. Mocetti and Porello (2012), besides giving descriptive evidence on the socio-demographic characteristics and work history of Italian migrants, provide an interesting panel analysis for the period 1995-2005. By adopting standard fixed-effect (FE) estimator, they find that the strong growth of house prices in the Centre-North, the spread of temporary contracts and immigration from abroad in the first decade of 2000's, are the main factors which have contributed to reduce the migration phenomenon. Moreover, they underline that short-time contracts may have discouraged 'official' migration, and favoured a temporary mobility that is not officially captured (defined long commuting).

In more recent years, the increased number of foreign immigrants leads many economists to study the internal movements of population in conjunction with international movements. The high mobility among foreign residents has contributed substantially to the overall internal mobility trend (ISTAT 2010, Casacchia et al., 2010). Indeed, the spatial relocation of (foreign) labour force may affects the skills composition of the local markets, the housing price (Saiz, 2007) and mostly, influence native migration choice for personal attitudes toward immigrants (Mocetti and Porello 2010, 2012).

Up to now, the empirical evidence on the relationship between native internal mobility and immigration has produced contrasting results and they are often referred to the US context.

For Italy, some relevant contributions are worth mentioning: Brűcker et al. (2009) use panel cointegration approach to exploit the variance of international and internal migration over time identifying significant displacement effects. Conditional on unemployment and wage differentials, the share of foreign workers in the labour force discourages internal labour mobility of Italian natives. Mocetti and Porello (2010) investigate the impact of immigration inflows on native internal mobility but distinguish between low- and high educated individuals. The authors face endogeneity issues related to the location choices of immigrants exploiting both the existence of previous enclaves and the proximity to 'gateways' as instruments for immigrant geographical distribution. Using IV panel regression (for the period 1996-2003) they show that a displacement effect of immigration acts on less skilled natives; in particular, immigrant concentration in the Northern regions has partially substituted the South-North flows of low-educated natives. By contrast, immigration seems to be positively associated to high-educated native inflows; especially in more urbanized areas. Therefore, they conclude that the foreign labour force contributes to attract younger natives from other regions and positively acts on the ageing of the population of the destination region.

More recently, following preliminary analysis by Casacchia et al. (2010)[8], Lamonica and Zagaglia (2013) examine the determinants of the internal migration flows across the Italian provinces over the period 1996-2005 by means of a factor analysis. They estimate a single cross-sectional regression for Italians and the foreign migrants separately. Significant evidence of economic factors as main drivers of Italian and foreign migration is found. This paper is one of the first which considers the two subsample of population (foreigners and Italians). Mostly, the main novelty of this work is that it is the first which consider the issue of spatial of spatial autocorrelation within the flows by means of the Griffith's eigenvector spatial filtering method.

1.3 From a macro to a micro approach

The majority of the studies cited so far are essentially based on the macro modelling approach since they attempt to explain the observed flows and the causes of net inward and outward migration of specific provinces\regions. In other words, interregional migration is mostly studied with respect to the spatial context and the related aggregate variables. Economic considerations, such as such as unemployment rate, available income, economic structure, population density or index of living costs have been identified as the main features driving migration choice. Theory predicts that the decision to migrate of a rational individual is based on a cost–benefit assessment of the economic

[8] They adopt a basic form of gravitational model which allows to quantify the effects on internal mobility due to the dimension of the population at origin and destination and the distance between macro-areas.

benefits of moving to the area of destination, net of transfer costs (Sjastaad, 1962; Harris and Todaro, 1970). However, as the Sjastaad's (1962) model[9] suggests, it is not likely that the sample of migrants that is observed would be a random sample of individuals. More precisely, the fact that one individual migrates, whereas another does not, implies that between them there may be some observable (and/or unobservable) characteristics, which create a relative wider spread between expected earnings and costs, and therefore ensure a different propensity to migrate. Todaro (1980) wrote that migrants 'tend to be disproportionately young, better educated, less risk-averse, more achievement-oriented and to have better personal contacts in destination areas than the general population in the region of out-migration'. Moreover, an outdated paper by Mincer (1977) defines family ties relevant to migration decisions and explains their effects on the probability of migration In particular, he finds very low migration rate for singles living with parents and relatives.

As already stated in the Introduction, the empirical research which has addressed the 'who moves' issue econometrically has been limited by the unavailability of longitudinal micro database. However, this has been changing in recent years. More and better micro data are available, although still mostly focus on some subgroups of population (like graduates).

In this Section we present a selected review of the background literature which has focused on the importance of personal characteristics among the basic determinant of migration decision. As before, a brief recap of these empirical studies, along with their main relevant features, is summarised in a table (Table 1.2).

The seminal contributions by Pissarides and Wadsworth (1989) and Antolin and Bover (1997) are worth mentioning. They both use data from the Labour Force Survey (LFS) respectively for UK and Spain in order examine the relation between unemployment and the inter-regional migration of labour force. Pissarides and Wadsworth (1989) argue that unemployment might affect mobility at three levels. First, unemployed workers are more likely to move than employed ones. Second, regional unemployment differentials encourage mobility: the probability that workers change region of residence is higher if they live in a high-unemployment region and the bigger the region's unemployment differential the higher the migration probability. Third, at higher overall unemployment rates the probability of migration is less. However, only the first and the third effects are confirmed by their empirical analysis. Antolin and Bover (1997) for Spain, have shown that the positive effect of personal unemployment status on the propensity to migrate becomes lower the larger the unemployment insurance given. In line with previous studies, they find that the probability of migration is higher for young people, and higher educated people. They also confirm

[9] Sjastaad (1962) developed a micro model where migration decision is mainly defined and modeled as an investment in human capital.

Mincer (1977) hypothesis' concerning the negative effects of family ties in the propensity to move. In particular, one of the main contributions of the authors is the emphasis on the importance of interactions between individual characteristics and regional variables. Their results demonstrate that personal characteristics not only have an important direct effect on migration but also alter the effect of regional economic variables on migration.

Paci et al. (2007) using micro data from the EU- LFS try to explain why internal movement (commuting or migration) has not played a bigger role in mitigating regional unemployment disparities in seven countries of Central and East Europe (Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland and Slovak Republic). By means of a logit model they provide empirical evidence on socioeconomic and demographic determinants to commuting and migration[10]. They find that the probability of migration is statistically associated with various regional economic indicators (employment and per capita income differentials). Gender, age, family status, and education attainment are the most relevant personal features which influence migration decision. More specifically: young, single and more educated men are the individuals with the higher propensities to move (migrate or commute equally). Moreover, even with mixed results for the different countries, they find that, workers in selected occupations (or in selected sectors of employment) are more mobile than others. For example, in general, construction workers seem to be the most mobile while, education and health workers are the less mobile.

Finnie (2004) provides another interesting contribution which. He addresses the ‘who moves’ issue using a broad-based longitudinal database over an extended period of time (1982-1995). The empirical analysis of interprovincial migration in Canada is carry out by estimating a logit model where the probability that a person moves from one province to another in a given year is taken to be a function of various ‘environmental’ factors (current province of residence, the provincial unemployment rate, the area size of residence), personal characteristics (language, age, marital status, the presence of children) and some key labour market attributes (earnings level, the receipt of unemployment insurance and social assistance). Mostly, the role of age and gender in migration patterns has been deeply analyzed. They run separate models by age groups (Entry, Younger, Prime-Younger and Prime-Older) and for each sex.

A different approach has been followed by Ritsilä and Haapanen (2003). The study focuses on actual Finnish migrants and it examines the direct effect of personal characteristics on destination

[10] In a recent contribution, Parenti and Tealdi (2015) use the Italian LFS from 1992 to 2008 to estimate a model where the probability of commuting is regressed on a wide set of individual, job, firm and regional characteristics. They also find that the increased utilization of temporary contracts did not have a strong impact on the commuting decisions of Italian workers.

choices. The main results of the ordered probability model indicate that the highly educated migrants are likely to move to urban municipalities. Conversely, large households, females and unemployed are more attracted to peripheral and less densely populated municipalities.

To the best of our knowledge, there are no in-depth studies focused on Italian migration decision at individual level. However, there are four interesting contributions in the field of study of internal migration using micro data which are worth noticing. The first is the paper by Faini et al. (1997) already cited above. They rely on the subjective attitudes toward mobility for unemployed individuals, using the information provided by the LFS[11]. Their results, based on a logit model, suggests that males and more educated individuals are the most likely to be willing to move and to take a job anywhere in Italy. Similarly, age has a positive impact on the attitude towards mobility, although this effect declines with age itself. Moreover, their results indicate that, a higher percentage of employed (or retired) members in the household are associated with more mobility. The main explanation of this latter finding is that those components (employed or retired) are able to finance migration costs of other family members.

The second is the contribution by Coniglio and Prota (2008). The data set used in this analysis has been generated through a survey, designed and conducted by the authors. Appling a logit model they identify the characteristics that differentiate migrants from non-migrants among highly skilled individuals residing in Basilicata. Their analysis shows that even among a group of highly educated individuals, the probability to move is higher for most talented ones especially with business or engineering studies. Among other things, they find that individuals with a previous migration experience are more likely to migrate. Differently from the pioneers (i.e. Pissarides and Wadsworth, 1989), they find that unemployed individuals are more likely to move than the employed. This result is explained by the strong family networks in the Southern regions that work like a kind of social security benefits, which restrain young southerners to leave their regions.

Similar results have been found by Nifo and Vecchione (2014). Using data on a sample of more than 47 thousand Italian graduates, they study the impact of provincial institution quality on the probability of migration controlling for individual and macroeconomic variables, both in the province of origin and destination. The Heckman probit estimation indicates that institutions do matter for migration decisions and their importance is comparable with that of per capita income provincial differences. Moreover, they provide strong evidence of a large set of individual characteristics. Among the sample of graduated, the probability to migrate is higher for female,

[11] They exploit the question 'where you would be willing to take a job?' (a) only in their own town, (b) in a neighboring town, (c) anywhere.

married, most talented graduates and the ones which belong to a family context with a high educational level.

More recently, Bartolucci et al. (2014) give an interesting contribution to the migration literature based on micro-data. They use a panel dataset from the Italian Social Security Administration (INPS) archives in order to identify the importance of unobserved worker characteristics (often defined 'ability') for the selection of migrants and returns to migration. They propose a complex analysis of a sample of about 1 percent of Italian workers from 1985 to 2004 using a novel iterative estimation method. They report evidence that the returns to ability are lower in the Northern than in the Southern regions, thus migrants tend to be drawn from the lower-end of the ability distribution. Whit this novel result they provide evidence against the conventional 'brain drain' from the Southern regions. In other words, they find evidence of a negative selection; lower ability workers are more likely to migrate from South to North while the 'best and brightest' are found to be more likely to stay in the South. Differential returns to observable characteristics are indeed found to be far less important.

Table 1.1 Econometric studies on the determinants of internal migration flows in Italy.

Paper	Period	Coverage	Sub-Samples	Estimation method	Main dependent variable	Main indipendet variables	Other control variables
Attanasio and Padoa-Schioppa (1991)	1960-1986	Macro areas		OLS	Net and gross migration rate	Wages in public and private sectors, male unemployment rate	
Basile and Causi (2007)	1991-1995, 1996-1000	Provinces		SUR (seemingly unrelated regression)	Net migration rate	Per capita income, unemployment rate	Population density, active population, share of employed in industry sector, price index
Biagi et. Al (2011)	2000 - 2001	Provinces	Short and long distance	NB, GMM	Gross migration flows	GDP per capita, unemployment rate	Distance, population (by sub-grups) , robberies, airports, universities, natural amenities, social capital
Brűcker et al. (2009)	1978-2001	Regions	South-North	Panel cointegration analysis (FM-OLS)	Gross migration flows (male)	Unemployment rate, wage, share of foreigners on labour force of destination area	Home per capita household consumption, past migration, housing price
Brunello et al. (2001)	1970-1993	Regions	South-North	IV	Gross migration rate	Wages and unemployment rate	Government social trasfer per capita
Cannari (2000)	1967-1992	Regions	South- CentreNorth	Logit	Gross migration flows	Unemployment rate, housing price	
Etzo (2011)	1996- 2005	Regions	Total flows; South to North North to South	FEVD[1] , GMM	Gross migration flows	Per capita income, unemployment rate	Distance, temperature, airports
Faini et. Al (1997)	1995	Italy		Multinomial logit	Mobility attitude	Gender, age, education, unemployment rate, employment rate	
Lamonica and Zagaglia (2013)	1995-2006	Provinces	Italians and foreigners	OLS	Gross migration flows	Population, distance	Socio-economic and demographic variables synthesized by means of factor analysis
Mocetti and Porello (2012)	1995-2005	Regions	Age, gender, education	FE, GMM	Net migration rate	Per capita GDP, unemployment rate, housing price	Population, share of immigrants
Mocetti and Porello (2010)	1995-2005	Regions	High skilled and low skilled	IV	Net migration rate	Per capita GDP, unemployment rate, housing price, share of immigrants	Population
Murat and Paba (2002)	1951-61, 1961-71, 1971-81, 1981-91	Provinces		OLS	Net migration rate	Per capita GDP, share of workers emplyed by firms with more than 500 emplyees	Population
Piras (2012)	1964-2002	Regions	High and low skilled	Panel cointegration analysis (OLS, FE, FGLS, GMM)	Net migration rate	Per capita GDP , unemployment rate, human capital	
Piras (2015)	1970-2005	Regions	Total flows; South to North North to South	FE, FMOLS, DOLS, MG, CCMG, AMG[2]	Gross migration flows	Per capita GDP , unemployment rate, human capital	Population
Salvatore (1977)	1958-1974	Regions	South-North	OLS	Net and gross migration rate	Wages, unemployment rate	

[1]Fixed effects vector decomposition

[2] Two way fixed effects (2FE); modified ordinary least squares (FMOLS) , dynamic ordinary least squares (DOLS); mean group (MG) ; common correlated effects mean group estimator (CCEMG); augmented mean group estimator (AMG)

Table 1.1. (cont.) Econometric studies on the determinants of internal migration flows in Italy.

Paper	Period	Country	Method	Data source	Sample	Dependent variable	Main Individual charactestichs	Regional characteristics
Antolin and Bover (1997)	1987-1991	Spain	Logit and probit	Spanish Labour Force Survey	Labour force	Mobility choice	Age, family status, labour status (ex-ante), economic sector (ex-ante)	Wage differentials, housing price differentials, unemployment rate differentials
Bartolucci et al. (2015)	1985- 2004	Italy	Iterative estimation	Work Histories Italian Panel (INPS)	Employed	Wage	Experience, tenure, years in the North , indicators for occupation (blue collar, white collar and managerial occupation), part time job	-
Coniglio and Prota (2011)	2002	Basilicata	Logit and conditional logit	Survey designed by the authors	Graduates	Mobility choice	Gender, education, experience, marks	GDP, unemployment rate, distance, past migration stock, population density, amenities, disamenities.
Faini et al. (1997)	1995	Italy	Multinomial logit	Italian Labour Force Survey	Unemployed	Mobility attitude	Gender, age, education	Unemployment, employment rate
Finnie (2004)	1982-1995	Canada	Logit	Longitudinal Administrative Database	Population (20-54 years old)	Mobility choice	Gender, age, family status, working status, wage, language	Distance, area size, unemployment rate
Nifo and Vecchione (2013)	2004	Italy	Heckman probit	Survey on the professional recruitment of graduates (ISTAT)	Graduates	Mobility choice	Age, gender, education, experienze, marital staus, family education	Wage, GDP per capita, quality of institutions proxies (both at origin and destination)
Paci et al . (2007)	2004	Czech Republic, Estonia,Hungary, Latvia, Lithuania, Poland, Slovak Republic, and Slovenia.	Probit	European Union Labour Force Survey (EUROSTAT)	Employed	Mobility choice	Age, family status, labour status (ex-ante), economic sector (ex-ante)	Unemployment rate, long term unemployment, population density and GDP (both at origin and destination)
Pissarides and Wadsworth (1989)	1976 and 1983	UK	Logit	British Labour Force Survey	Labour force	Mobility choice	Gender, age, education, family status, work sector	Relative wage, unemployment differential, cost of unemployment, relative vacancy rate
RitsilÃa and Haapanen (2010)	1994-1995	Finland	Ordered probit model	Finnish Longitudinal Census	Total population (17 - 64 years old)	Destination choices of migration	Gender, education , experience, age, working status, family composition, house howners	Urban area and population density (at origin)

Table 1.2 Econometric studies on the determinants of internal migration flows based on micro data.

Paper	**Period**	**Country**	**Method**	**Data source**	**Sample**	**Dependent variable**	**Main Individual charactestichs**	**Regional characteristics**
Antolin and Bover (1997)	1987-1991	Spain	Logit and probit	Spanish Labour Force Survey	Labour force	Mobility choice	Age, family status, labour status (ex-ante), economic sector (ex-ante)	Wage differentials, housing price differentials, unemployment rate differentials
Bartolucci et al. (2015)	1985- 2004	Italy	Iterative estimation	Work Histories Italian Panel (INPS)	Employed	Wage	Experience, tenure, years in the North , indicators for occupation (blue collar, white collar and managerial occupation), part time job	-
Coniglio and Prota (2011)	2002	Basilicata	Logit and conditional logit	Survey designed by the authors	Graduates	Mobility choice	Gender, education, experience, marks	GDP, unemployment rate, distance, past migration stock, population density, amenities, disamenities.
Faini et al. (1997)	1995	Italy	Multinomial logit	Italian Labour Force Survey	Unemployed	Mobility attitude	Gender, age, education	Unemployment, employment rate
Finnie (2004)	1982-1995	Canada	Logit	Longitudinal Administrative Database	Population (20-54 years old)	Mobility choice	Gender, age, family status, working status, wage, language	Distance, area size, unemployment rate
Nifo and Vecchione (2013)	2004	Italy	Heckman probit	Survey on the professional recruitment of graduates (ISTAT)	Graduates	Mobility choice	Age, gender, education, experienze, marital staus, family education	Wage, GDP per capita, quality of institutions proxies (both at origin and destination)
Paci et al . (2007)	2004	Czech Republic, Estonia,Hungary, Latvia, Lithuania, Poland, Slovak Republic, and Slovenia.	Probit	European Union Labour Force Survey (EUROSTAT)	Employed	Mobility choice	Age, family status, labour status (ex-ante), economic sector (ex-ante)	Unemployment rate, long term unemployment, population density and GDP (both at origin and destination)
Pissarides and Wadsworth (1989)	1976 and 1983	UK	Logit	British Labour Force Survey	Labour force	Mobility choice	Gender, age, education, family status, work sector	Relative wage, unemployment differential, cost of unemployment, relative vacancy rate
RitsilÄa and Haapanen (2010)	1994-1995	Finland	Ordered probit model	Finnish Longitudinal Census	Total population (17 - 64 years old)	Destination choices of migration	Gender, education , experience, age, working status, family composition, house howners	Urban area and population density (at origin)

Chapter 2

What drives natives and foreign migrants? A spatial gravity analysis of interregional flows in Italy.

2.1 Introduction

The literature addressing internal migration of foreign population has become increasingly popular in recent decades, especially for countries that have experienced large population gains as a result of international migration. The large amount of studies refers to United States (Bartel and Koch 1991; Belanger and Rogers 1992; Rogers and Henning 1999; Molloy et al., 2011) and Canada (Nogle 1994; Newbold, 2001), where the peculiarity of the internal migration phenomenon has always attracted the attention of a number of researches. Only more recently, some empirical studies have been written also for European countries like Germany (Schündeln, 2014), Spain (Hierro and Maza, 2010; Maza, et al, 2013) and Italy (Lamonica and Zagaglia, 2013).

Italy is a country with a long history of emigration but a shorter experience of immigration. After the Italian unification it became one of the leading European emigration countries. Del Boca and Venturini (2005) report that over that period more than 26 million people[12] left the country towards other European and American countries, in search of better job opportunities. After the World War II a greatest internal migration flow across the country began. The rapid development of the industrial Northern regions stimulated millions of low-skilled workers to move from the backward Southern (or Mezzogiorno regions) and North-eastern towards the Central and North-Western ones contributing to the so-called *miracolo economico* (economic miracle) which enabled the country to become one of the most industrialised of the world. In the second half of the 1970s emigration sharply declined and Italy changed from being a sender into a host country, receiving a great number of immigrants mainly from developing countries and later from Eastern Europe[13]. Nowadays, the migration phenomenon (both international and interregional) continues to be a distinctive feature of the Italian economy, although with dimensions and characteristics quite different from the past decades. Two main features are worth mentioning: firstly, it has lost the connotation of 'mass' phenomenon taking indeed a more selective character and secondly, the foreign population substantially contribute to the internal migration trend.

For different reasons foreigners constitute a self-selected group of individuals with specific features which presumably differ with respect to natives individuals[14]. For example, foreign born individuals who have already chosen to leave their home country toward a new host country may be more

[12] Two fifths of all these emigrations originated from the regions of the Italian Mezzogiorno.

[13] Discussing the large wave of immigration during the 2000s, Bratti and Conti (2014) write: 'the characteristics of immigration in Italy are such that immigrants mainly appear as a source of low-skilled or cheap labour force, which is employed in traditional (i.e. low value added) economic sectors'

[14] A foreigner is defined as a person who lacks Italian citizenship. Italian citizenship is mainly acquired by *ius sanguinis.* Under restrictive and particular conditions it can be granted on request to foreign citizens married to Italians and to foreign citizens who reside in Italy.

inclined with regards to internal migration and, in particular, they may react differently to push and pull factors.

The main aim of this Chapter is to assess the factors explaining the internal mobility of the natives (Italian citizens) and the foreigners across Italian regions, emphasizing the different nature (economic or non-economic factors) of the variables determining the internal migration movements. This study contributes to the existing literature addressing, in a panel framework, the issue of cross-regional dependence arising from the existence of regional spillovers. In fact, bilateral migration flows do not depend only on the specific characteristics of the origin and the destination regions and on their distance, but are also potentially influenced by neighbouring characteristics from origin and destination. Although this issue has been long acknowledged (Curry, 1972), it has been often overlooked in empirical research on Italian migration. The only exception is the recent paper by Lamonica and Zagaglia (2013). They estimate year by year internal migration flows (from 1995 to 2006) using the spatial filtering method in order to remove spatial autocorrelation within migration flows. Differently from Lamonica and Zagaglia (2013), in order to investigate whether spillovers effects may affect migration patterns, we estimate an augmented gravity model by including spatial lags of push and pull factors variables. Furthermore, along with the overall internal migration flows we consider also two additional subsamples. Firstly, we take the eight Southern regions as origin and the twelve Centre-Northern regions as destination, and secondly, we consider the reverse flow, namely, from the Centre-Northern regions to the Southern ones. Indeed, very limited evidence on this latter migration pattern has been offered so far (Etzo 2011, Piras 2015).

We base our econometric setting considering two different ways of modelling the dependent variable, namely the migration flows from region i to region j. The first is to consider the variable as purely count data, thus the natural starting point is to consider the Poisson model or the Negative Binominal (NB) variant, which is generally the most adequate model to capture overdispersion in migration data. Alternatively, when the dependent variable takes sufficiently large values, a normal approximation can be exploited and thus standard econometric estimation methods for continuous variables can be applied. One of the most common approaches is to consider the log-linear formulation of the gravity equation and use OLS estimators. However, as largely discussed by Santos Silva and Tenreyro (2006), under heteroskedasticity, the parameters of log-linearized models estimated by OLS, may be biased because of Jensen’s inequality. As suggested by Santos Silva and Tenreyro (2006), in order to address the various estimation problems arising from the log-linearization, we use Poisson pseudo-maximum-likelihood (PPML) estimator.

Chapter 2 is organised as follows. In Section 2 we briefly outline some theoretical arguments on the determinants of internal migration with a particular emphasis on the relationship between foreigners and natives' migration choices. Section 3 presents the main features of internal migration flows across Italian regions and a brief discussion about the selection of destination and origin determinants. The description of the methodology adopted to carry out the empirical analysis follows in Section 4. The main econometric results are presented in Section 5 and 6. Concluding remarks are offered in Section 7.

2.2 Literature review

Since the seminar contribution of Salvatore (1977) who for the first time studied internal migration across Italian regions, a great number of studies have been published reaching varying conclusions, given the wide range of methodologies adopted and the differences in time periods considered and the coverage of geographical areas[15].

As discussed in Chapter 1, internal migration in Italy has experienced different migration trends. The first and most intense migration flow, from the rural Southern regions toward the more industrialised Northern regions, characterized the fifties and the sixties. From the first half of seventies internal flows were marked by a dramatic decrease which lasted until the first half of nineties. The mismatch between internal migration and regional disparities which characterized these years and the failure of traditional theory to explain such phenomenon (known as 'the empirical puzzle'), attracted the interest of many researchers (Faini et. al, 1997). After a long period of mobility stagnation, lasted for more than two decades, internal migration flows started to grow again in 1996.

It is worth mentioning, that while the main direction of the flows has not changed during those decades (i.e., from South to Centre-North), its composition has revealed some relevant changes in terms of age structure and educational attainment of migrants[16]. Today's migrants are older and more skilled than past migrants.

In more recent years, internal movements of population are being considered in conjunction with international movements, rather than as separate from. As a matter of fact, from the second half of the 90s, Italy has become one of the prime European destinations for foreign immigrants. It is well

[15] A comprehensive review of the empirical contribution on internal migration across Italian regions goes beyond the scope of the present Chapter. This aspect has been reviewed in the first Chapter of this thesis.

[16] Piras (2005) analyses the human capital endowment of migration flows across Italian regions during the period from 1980 to 2002 detecting evidence of human capital losses for almost all Southern regions. More precisely, this brain drain seems to have reduced their growth potential.

known that the high mobility among resident foreigners has contributed substantially to the overall internal mobility trend (ISTAT 2010)[17].

As we already said in the first Chapter, the empirical evidence on the relationship between native internal mobility and immigration is often referred to the US context where the migration phenomenon (both internal and international) is particularly relevant and persistent. For example, Card and DiNardo (2000) using IV regression methods find that an increase in immigrant population in specific skill groups lead to small increases in the migration of native-born individuals of the same skill group. In 2001, Card (2001) shows that inflows of new immigrants did not generate large offsetting mobility flows by natives. On the contrary, Borjas et al. (1997) report a strong negative correlation between native net migration and immigration by states. Borjas (2006) finds that immigration is associated with lower in-migration rates, higher outmigration rates, and a decline in the growth rate of the native workforce.

The work published by Schündeln (2014) for the period 1996-2003, addresses jointly the mobility behaviour of natives and foreigners for the case of Germany. By using probit and conditional logit regressions, he proves, after taking into account a set of individual characteristics, that immigrants are more likely than natives to internally migrate within Germany and moreover, seem to be more responsive to labour market differentials than natives.

As previously stated one of the most contribution is the paper of Lamonica and Zagaglia (2013). The authors apply an augmented version of the gravity model to the internal migrant flows of Italians and resident foreigners. To account for the existence of spatial autocorrelation they use Griffith's eigenvector spatial filtering method and estimate the resulting model with OLS. Estimating a single cross-sectional regression for each year (1996-2005), they found significative effects of the two main gravity variables (i.e. the population sizes and the distance) for Italians and foreigners. For native migrants, economic conditions act as influential push and pull factors; by contrast, for non-natives both, economic and demographic conditions attracted them to other regions, but they do not seem to be significant as push factors. Differently from Lamonica and Zagaglia's analysis, we empirically assess the role of spatial spillovers and hence, distinguish between the effects due to regional internal determinants from those generated by interactions among neighbouring regions. As far as we know, this model has never been used before in the Italian migration context.

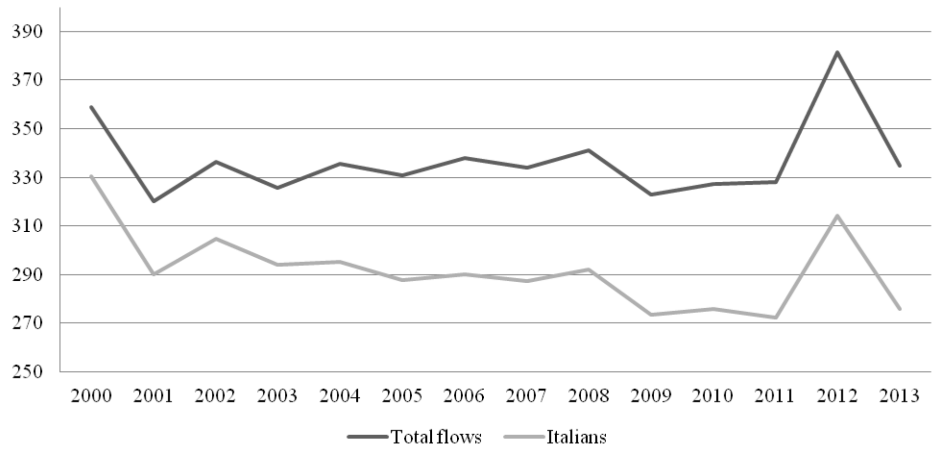
[17] Some studies carried out in Canada, the United States have arrived to the conclusions that foreign born population tend to be more mobile than natives because of their demographic and social characteristics, the situation of the labour market and their academic attainment (Bartel 1989; Bartel and Koch 1991; Nogle 1994).

2.3 Data and variables descriptions

2.3.1 Interregional and immigration flows

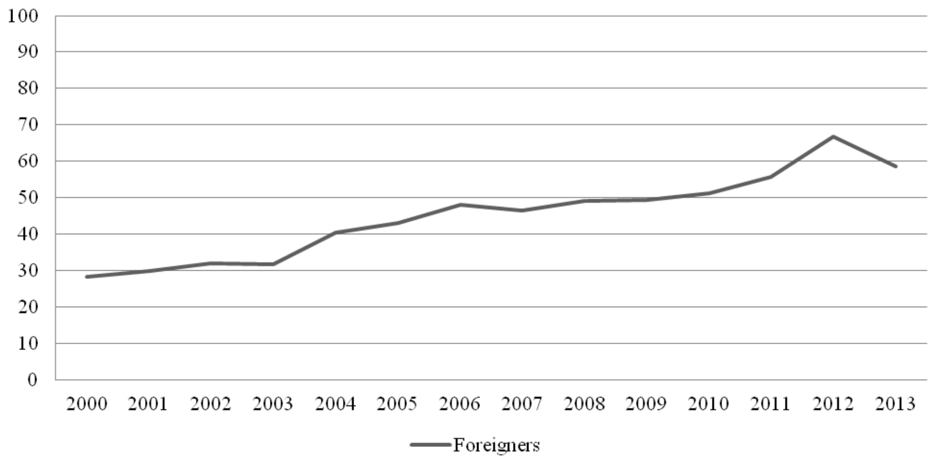
Annual data collections on changes of residence between the Italian municipalities are reported in the population register (ISCAN – Rilevazione sugli iscritti e cancellati per trasferimento di residenza). The data collection is based on individual forms for each registration/deregistration (one form for each change of residence) sent from each municipality to the Italian National Statistical Office (ISTAT). Personal information about (gender, place and date of birth, address, citizenship, place of origin/destination) for each recorded migrant are collected by ISTAT and are normally available some nine to twelve months later. Changes of residence from one municipality to another are effective as of the day of application for registration in the new municipality register, but are recorded when the migration process, returning from the municipality of deregistration, is completed. We use interregional flows occurred over the period 2000-2013 among two of the 20 Italian regions. Clearly, this measure of residential mobility only refers to formal change of residence. Non official transfers to other regions without administrative formalization are totally missed. For each year, we construct a 20x20 Origin-Destination (OD) matrix which describes migration flows from each region of origin (region *i*) to each possible region of destination (region *j*). The main diagonal of each matrix is set to zero to exclude intra-regional flows. This leaves us with 380 bilateral OD individual flows per year. Figures 2.1 and 2.2 report the trend of gross interregional migration flow for the period under consideration in our analysis (2000-2013) distinguishing between Italians and foreigners respectively.

Figure 2.1 Gross interregional migration flows (thousand). Total migrants and Italian migrants.



Source: our elaboration on Istat data.

Figure 2.2. Gross interregional migration flows (thousand). Foreign migrants.



Source: our elaboration on Istat data.

As we can see, in the early 2000s the outflow was quite stable at more than 320 thousand migrants per year, a level higher compared to that of the 90s when the internal migration phenomenon drastically lost intensity. In 2009, the migration flow considerably decreased by around 20 thousand. It is unlikely to be just a coincidence that this low point corresponds to the advent of the economic crisis. For the object of our analysis it is interesting to analyze the contribution of the foreign population to this considerable phenomenon.

As we can see from Table 2.1 migrant flow is constantly increasing over time, i.e. from 8% in the early 2000s to 18% in 2013. By contrast, Italian citizen flows tend to be relatively more stable over time. Their contribution to total flow falls from 92% to 82% over the last 13 years. As suggested by the literature, foreigners seem to have a higher propensity to be mobile than Italians (ISTAT, 2010). Therefore, the less deep roots in the territory, as well as the effort in search of better jobs and adequate social protection, determine in 2013 a rate of internal mobility of 13 per thousand foreign residents, about four times higher than the Italian's one (3 per thousand). Clearly, there is a considerable heterogeneity among foreign communities: people differ in their propensity to migrate. As confirmed by ISTAT, in 2013 the Chinese community is the most mobile across regions: 44 individuals over 1000 resident Chinese moved to another region. As reported in Table 2.1, is the Asian community that gives the major contribution to foreign internal flows (5%, with more than 18 thousand Chinese), followed by European member citizens (only the Romanian represent 80% of total internal movers from the EU members) and finally other non-EU member citizens.

Table 2.1 Interregional migration flows by citizens, absolute and percentage values.

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Total flows	359008	320133	336461	325856	335643	330969	338068	333815	341154	323015	327258	327866	381251	334689
Italians (abs. values)	330572	290279	304566	294073	295177	287803	290042	287428	292113	273534	276026	272210	314384	275933
Italians (percentage values)	*92.08*	*90.67*	*90.52*	*90.25*	*87.94*	*86.96*	*85.79*	*86.10*	*85.62*	*84.68*	*84.35*	*83.02*	*82.46*	*82.44*
Foreigners (abs. values)	28436	29854	31895	31783	40466	43166	48026	46387	49041	49481	51232	55656	66867	58756
Foreigners (percentage values)	*7.92*	*9.33*	*9.48*	*9.75*	*12.06*	*13.04*	*14.21*	*13.90*	*14.38*	*15.32*	*15.65*	*16.98*	*17.54*	*17.56*
Subgrups of inter-regional flow by citizenship:														
EU28 (abs. values)	3393	3724	4063	4261	6471	7506	8307	9575	12801	12504	13132	14297	17471	13859
EU28 (percentage values)	*0.95*	*1.16*	*1.21*	*1.31*	*1.93*	*2.27*	*2.46*	*2.87*	*3.75*	*3.87*	*4.01*	*4.36*	*4.58*	*4.14*
Non EU members (abs. values)	6882	7465	8561	8545	11956	12760	14155	12906	12346	11676	11563	12307	13272	11712
Non EU members (percentage values)	*1.92*	*2.33*	*2.54*	*2.62*	*3.56*	*3.86*	*4.19*	*3.87*	*3.62*	*3.61*	*3.53*	*3.75*	*3.48*	*3.50*
African (abs. values)	9355	9576	8896	8173	9994	9209	9913	9204	9258	9304	9157	9898	11666	10844
African (percentage values)	*2.61*	*2.67*	*2.48*	*2.28*	*2.78*	*2.57*	*2.76*	*2.56*	*2.58*	*2.59*	*2.55*	*2.76*	*3.25*	*3.02*
Central and Southern	182	187	209	195	267	266	301	306	286	327	365	392	454	412
Western	2807	2704	2402	2162	2638	2468	2711	2370	2400	2498	2387	2785	3439	3458
Eastern	561	629	564	531	532	496	541	601	821	822	982	882	965	874
Northern	5805	6056	5721	5285	6557	5979	6360	5927	5751	5657	5423	5839	6808	6100
Asian (abs. values)	6853	7091	8195	8522	9172	10482	12409	11694	11238	12230	13593	15424	20089	18494
Asian (percentage values)	*1.91*	*1.98*	*2.28*	*2.37*	*2.55*	*2.92*	*3.46*	*3.26*	*3.13*	*3.41*	*3.79*	*4.30*	*5.60*	*5.15*
Central and Southern	3059	3265	3599	3443	3741	3946	4478	4311	4359	4344	4837	5420	7367	7172
Western	254	242	215	246	287	257	260	251	255	338	349	491	696	721
Eastern	3540	3584	4381	4833	5144	6279	7671	7132	6624	7548	8407	9513	12026	10601
American (abs. values)	1922	1973	2135	2250	2840	3178	3208	2983	3349	3735	3758	3709	4333	3824
American (percentage values)	*0.54*	*0.55*	*0.59*	*0.63*	*0.79*	*0.89*	*0.89*	*0.83*	*0.93*	*1.04*	*1.05*	*1.03*	*1.21*	*1.07*
Central and Southern	1785	1852	1975	2106	2714	3059	3085	2867	3194	3602	3610	3566	4178	3701
Northen	137	121	160	144	126	119	123	116	155	133	148	143	155	123
Oceania (abs. values)	25	19	36	22	33	24	23	19	37	28	23	18	19	16
Oceania (percentage values)	*0.09*	*0.06*	*0.11*	*0.07*	*0.08*	*0.06*	*0.05*	*0.04*	*0.08*	*0.06*	*0.04*	*0.03*	*0.03*	*0.03*
Stateless (abs. values)	6	6	9	10		7	11	6	12	4	6	3	17	7
Stateless (percentage values)	*0.02*	*0.02*	*0.03*	*0.03*	*0.00*	*0.02*	*0.02*	*0.01*	*0.02*	*0.01*	*0.01*	*0.01*	*0.03*	*0.01*

Source: our elaboration in ISTAT data.

Table 2.2 describes the geography of migration flows. It reports the change of residence in the period 2000-2013 distinguishing between Italian and foreign citizens with different nationalities and from three macro-regions (namely North, Centre and South[18]) of origin and destination. Moreover, data are reported for four subgroups of population: European, African, Asian, American and other countries. Overall, in the time interval considered, more than 4 million of Italians have changed region of residence. Table 2.2 shows that natives tend to move mainly from Southern to Central-Northern regions and that, approximately, 45% per cent of their total mobility is generated by Southern regions. On the contrary, foreign internal migrants mainly originate from the Northern regions of the country. The North of Italy is the most preferred destination for the two subgroups. Short-range mobility (across the same macro-region) is relatively high for all macro nationalities, while is limited for Italian natives. With regards to Italians, we can observe an opposite flow from Northern and Central regions toward the less wealthy Southern regions, which might be originated by a kind of return migration, driven by non economic reasons (e.g. retired people going back to

[18] The three areas include the following regions: Piemonte, Val D'Aosta, Lombardia and Liguria, Trentino Alto Adige, Veneto, Friuli Venezia Giulia and Emilia Romagna (North); Toscana, Umbria, Marche and Lazio (Centre); Abruzzo, Molise, Campania, Aquila, Basilicata, Calabria, Sicilia and Sardegna (South or Mezzogiorno).

their native regions). This flow is indeed quite modest for foreigners suggesting the existence of some spatial pattern. These results are confirmed by inflow and outflow indicators described in Table 2.3, for three different years (2000, 2006 and 2013). On average, the outflow rate is higher in the Central-Northern regions than in the Southern ones. As described by the coefficient of variation for Italian movers, regional disparities are slightly decreasing over time. By contrast, they are slowly rising (the coefficient of variation moves from 0.83 to 0.86) for foreign internal migrants. Over the three years considered, regions that create more interregional mobility among Italian citizens are those most densely populated (i.e., Campania and Sicilia in the South, Lombardia, in the North, and Lazio, among the Central regions). Conversely, smallest and least populated regions (like Valle d'Aosta, Bolzano, Trento, Friuli Venezia-Giulia, Molise) show very low outflow rates, mostly below 1%. Sardegna, most likely due to insularity and related problem of depopulation, generates less than 2% of total flows, a figure that is declining over time. Similar results are found for the foreign population. However, in 2006 and 2013 Sicilia leaves the group of the four regions that create more mobility leaving room to Emilia Romagna and Toscana. Moving to the inflow rate, we can see that the regions that over the three years attract more migrants are Lombardia and Emilia-Romagna. Among Central regions, Lazio is also particularly attractive for Italians while foreigners (especially in 2000 and 2006) strongly prefer the Northern regions (such as Veneto). Toscana as well, shows a modest and quite stable attraction rate (between 7-8%). We can also see that while the attraction rate for Southern regions is essentially negligible for foreign population (between 1-2% on average), some regions like Sicilia and especially Calabria constantly exhibit higher inflows rate (above 5%) reflecting, as already observed a potential return migration to native regions. These facts are clearly in line with the theory behind the gravity model where population size (both at origin and destination) is intended to capture the relative force of mass. The population of the region of origin is referred to as the population 'at risk of migration', while the population size of the region of destination is assumed to be proportional to the number of migrants that can be accommodated (in terms of housing and jobs, for example).

Table 2.2 Interregional migration flows by citizenships and by origin and destination macro-areas.

	Italians		*Foreigners*									
			European		African		Asian		American		Other	
From Southern origins	1833421	45%	79819	28%	43653	32%	37082	22%	7912	18%	104	23%
From Northern origins	1584837	39%	127570	44%	65765	49%	84295	51%	24605	57%	205	46%
From Central origins	665882	16%	80081	28%	25029	19%	44109	27%	10680	25%	137	31%
total	4084140	100%	287470	100%	134447	100%	165486	100%	43197	100%	446	100%
From Southern origins to Southern destinations	280496	15%	8299	10%	4020	9%	6235	17%	1001	13%	19	18%
From Southern origins to Northern destinations	1035259	56%	50749	64%	31543	72%	20543	55%	4680	59%	65	63%
From Southern origins to Central destinations	517666	28%	20771	26%	8090	19%	10304	28%	2231	28%	20	19%
total	1833421	100%	79819	100%	43653	100%	37082	100%	7912	100%	104	100%
From Northern origins to Southern destinations	587409	37%	13821	11%	9737	15%	8811	10%	2586	11%	25	12%
From Northern origins to Northern destinations	744413	47%	88844	70%	46422	71%	57986	69%	17933	73%	130	63%
From Northern origins to Central destinations	253015	16%	24905	20%	9606	15%	17498	21%	4086	17%	50	24%
total	1584837	100%	127570	100%	65765	100%	84295	100%	24605	100%	205	100%
From Central origins to Southern destinations	273903	41%	9868	12%	4138	17%	8122	18%	1873	18%	35	26%
From Central origins to Northern destinations	235474	35%	49351	62%	15537	62%	26169	59%	5948	56%	58	42%
From Central origins to Central destinations	156505	24%	20862	26%	5354	21%	9818	22%	2859	27%	44	32%
total	665882	100%	80081	100%	25029	100%	44109	100%	10680	100%	137	100%

Source: our elaboration on ISTAT data.

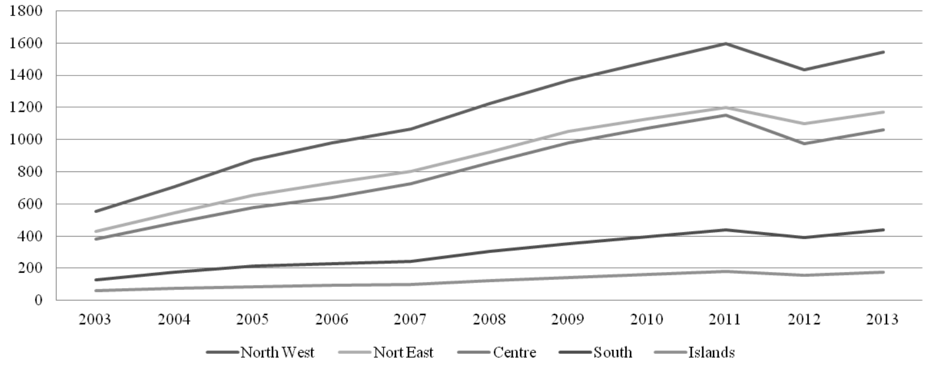
Table 2.3 Patterns of interregional migration by citizenships (percentages).

	2000				**2006**				**2013**			
	Inflow rate	**Outflow rate**	**Inflow rate**	**Outflow rate**	**Inflow rate**	**Outflow rate**	**Inflow rate**	**Outflow rate**	**Inflow rate**	**Outflow rate**	**Inflow rate**	**Outflow rate**
Regions/Citizenship	*Italians*		*Foreigners*		*Italians*		*Foreigners*		*Italians*		*Foreigners*	
Piemonte	7.69	7.04	6.71	6.60	7.23	7.30	8.26	7.58	7.54	6.63	9.15	8.23
Lombardia	15.49	12.23	22.41	11.24	15.44	13.56	22.42	14.92	17.38	12.65	19.15	16.62
Valle d'Aosta	0.42	0.32	0.26	0.33	0.43	0.43	0.39	0.35	0.42	0.33	0.36	0.25
Trentino-Alto Adige	1.35	0.92	2.66	1.30	1.55	1.09	2.55	1.66	1.91	1.17	2.70	1.15
Veneto	6.59	4.67	15.63	6.43	6.28	5.32	12.67	9.64	5.66	5.19	9.07	10.60
Friuli-Venezia Giulia	2.78	1.68	4.74	1.58	2.36	1.90	3.94	2.41	2.31	1.72	3.16	2.55
Liguria	3.78	3.51	2.94	2.60	4.03	3.59	3.58	2.69	3.75	3.72	4.07	3.25
Emilia-Romagna	11.83	5.62	16.56	6.64	11.95	7.05	15.25	8.70	10.01	7.28	12.03	11.07
Toscana	7.37	4.12	7.47	6.37	7.17	4.73	8.32	7.68	6.75	5.26	8.33	7.42
Umbria	2.17	1.27	2.82	2.30	2.05	1.44	2.45	2.49	1.77	1.68	2.22	2.98
Marche	3.16	1.91	6.07	2.52	3.13	2.17	4.44	3.58	2.41	2.48	3.27	4.54
Lazio	11.39	8.27	4.25	15.69	10.26	7.41	4.95	11.19	11.18	8.28	9.57	9.45
Abruzzo	2.34	2.11	1.54	2.88	2.55	2.20	2.19	2.35	2.80	2.64	2.44	2.91
Basilicata	1.00	1.54	0.27	1.11	0.86	1.52	0.36	0.88	1.03	1.47	0.61	0.77
Calabria	3.88	6.54	0.84	3.77	3.45	5.73	1.08	3.44	3.45	5.55	1.65	2.94
Campania	6.05	14.67	1.70	9.85	6.85	13.69	2.33	9.36	7.50	13.69	4.62	6.03
Molise	0.76	0.81	0.25	0.39	0.75	0.79	0.28	0.43	0.82	0.92	0.40	0.49
Puglia	4.83	8.77	1.24	7.06	4.81	8.35	1.73	4.60	4.96	7.96	2.92	3.51
Sicilia	5.01	10.98	1.16	9.41	6.13	9.25	1.86	5.06	5.99	8.86	3.04	3.93
Sardegna	2.09	3.04	0.48	1.92	2.72	2.48	0.95	0.98	2.36	2.50	1.23	1.30
standard deviation	4.08	4.17	6.21	4.15	3.99	4.02	5.79	4.11	4.21	3.87	4.80	4.32
mean	5.00	5.00	5.00	5.00 #	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
coefficient of variation	0.82	0.83	1.24	0.83	0.80	0.80	1.16	0.82	0.84	0.77	0.96	0.86

Note: Outflow rate: ratio between cancellations from region *i* and total migration flows in Italy; *Inflow rate*: ratio between registrations in region *j* and the total migration flows in Italy. *Source: our elaboration on ISTAT data.*

Due to the relative increasing weight of foreigners with respect to internal migration, it is also interesting to shed some light on the recent evolution of Italian immigration phenomenon. Over the last decades, Italy went from being an origin country of immigration to one of the major destination country for international migration flows. Figure 2.3 reports ISTAT data on foreign resident population from 2003 to 2013. It is evident that during the first decade of 2000s Italy, and especially its North-Western regions, had faced a rapid evolution of the migration phenomenon. According to data from the population register, in 2011 there were 4.57 million foreigners residing in Italy. However, this stock of registered foreign population faced a considerably decline in 2012 (-11.34% compared to 2011) but started to increase again in 2013 (+8.28%) reaching 4.3 million or, 7.5% of the total population.

Figure 2.3 Trend of foreign-born population by macro areas (thousand).



Year	Italy	% var
2003	1549.37	-
2004	1990.16	28.45
2005	2402.16	20.70
2006	2670.51	11.17
2007	2938.92	10.05
2008	3432.65	16.80
2009	3891.30	13.36
2010	4235.06	8.83
2011	4570.32	7.92
2012	4052.04	-11.34
2013	4387.69	8.28

Source: our elaboration on ISTAT data.

2.3.2 Characteristics of regions at origin and destination

In this paragraph we illustrate the variables that are expected to influence migration choices as well as the ability of regions to attract inflows and restrain outflows of Italians and foreigners citizens, distinguishing between pull factors (at Origin) and push (at Destination) factors[19].

The first standard gravity variable is *population* size of both the region of origin and destination. The former variable reflects the fact that a larger source region may generate higher emigration volumes, while at the same time, highly populated regions are expected to exploit economies of scale, provide better services and thus, be more attractive for potential migrants. In the robustness analysis, for empirical generality, we also include another demographic indicator: the share of economically active population aged 15-64, which aim to effectively capture the effect of the labour force population as push and/or pull factor.

The second variable is *per capita GDP*. The income level in the destination regions represents an adequate indicator of the economic development in the origin and receiving area. It is expected to influence positively the incoming migration flows at destination and negatively the outflows at origin. As suggested by empirical evidence (Basile and Causi, 2007; Furceri, 2006; Etzo, 2011) low

[19] A complete description of all variables used in the econometric analysis in Table A2.1 in the Appendix.

levels of economic development are able to push flows of migrants away from their regions and to attract them to better destinations.

For robustness we also control for *unemployment rate*[20] which, measures interregional differences in employment opportunities. According with economic theory, we expect that the higher the unemployment rate in the origin the higher the out-migration (push factor) and vice versa (DaVanzo, 1978). Along with strict economic variables, recent literature has also highlighted the importance of indicators of amenities (or disamenities) supply at the regional level (Biagi et al. 2011, Etzo 2011). The availability of hard (like road, airports) and soft (such as universities) infrastructures are often used to proxy urban agglomeration economies. In our analysis we focus on public transport. We try to assess the relative importance of the availability of *public transport* as a push and pull factor using the urban networks of public transport in the provincial capitals for 100 square km. As a matter of fact, a good provision of public transport makes not only faster and cheaper movements of people across urban areas, but mainly, it reduces considerably negative environmental externalities (congestion, gas emissions, pollution and so on) related to private transports. Taking into account these considerations, we expect that an adequate endowment of public infrastructures deter the propensity to leave a region. Conversely, at the destination, it may attract new migrant flows.

We also include the *tourism attraction* index (defined by per capita nights[21]), which is intended to be a proxy for a more ample and interrelated set of locally provided services and amenities (accommodation, restaurants, bars, cultural attractions, historical and natural amenities and so on). As suggested by the empirical literature (Marrocu and Paci, 2013), natural and other tourism-related amenities are directly (and indirectly) linked to tourism inflows and are also particularly important on migration decision (Graves 1980, Glaeser et al., 2001; Florida, 2002). Indeed, research suggests that high-amenity regions experience more rapid population growth than low-amenity regions do.

Regional productivity and development may also be influenced by the level of social capital (Marrocu et al., 2012), which can be defined as a complex mixture of shared norms, ties and trust. Due to the difficulty in measuring such a complex and informal phenomenon, several indicators have been employed in empirical studies[22]. In this paper, following Putnam (1993, 1995), we proxy social capital with the membership in voluntary associations, assuming that such groups and

[20] Due to the high correlation between the two economic variables we choice to including the two variables separately.

[21] This variable refers to the formal accommodation sector only. Non official accommodations, such as those provided by friends and relatives, are totally missed.

[22] Some example of indicators for social capital which have been used are: trust level, blood donations (Guiso et al., 2004), voluntary organisation density (Paldam and Svendsen, 2000), associational activity (Beugelsdijk and van Schaik, 2005; Dettori et al., 2012) and so on.

associations function as 'schools of democracy', in which cooperative values and trust are easily socialized. More precisely, we use the share of the population aged 14 and over that have taken part at least once in the last 12 months in social activities such as voluntary service, unions and cultural associations meetings or have worked free for voluntary organizations. A better social endowment is expected to restrain population outflows and increase the inflows.

We also control for ‘negative’ social capital measured as the number of robberies and thefts per 10,000 inhabitants (defined *crime rate* from now). As they occur where people reside, they might affect inhabitants’ decision on where to live. Property crime level is considered an important indicator for local liveability and quality of life. The sign in the sending regions is expected to be positive and negative in the destination regions. Indeed, it is reasonable to assume that people move from regions with high level of crime toward safer and more liveable regions.

Human capital represents another important intangible asset just like other regional physical endowments. In this analysis we employ the standard and most used indicator for human capital namely, the share of population who attained at least a university degree. Besides the large consensus among social researchers on the positive role played by highly skilled population on local economic performances, other important factors, like the creation of new ideas and technological innovations are strongly reliant on human capital (Lucas, 1988). According to the literature in new economic geography (Krugman, 1991; Fujita and Thiesse, 2002), high accumulation of human capital may triggers a systemic process of agglomeration forces, which not only discourage individuals from leaving the region but also attract more and more talents and innovative firm.

Following this reasoning, we would expect a negative sign at origin and positive one at destination.

However, empirical evidence for Italy (Piras, 2015) shows that this result seems to depend on whether Centre-North to South or South to Centre-North flows are analysed.

Finally, geographical *distance* may be defined as a proxy for general transportation and information costs and is generally found to be the most influential factor (together with average wages differentials) in explaining the variations in migration flows. It is expected to be negatively correlated with internal mobility and it should also account for spatial correlation among the observational units. In this paper the geographical distance between each origin and each destination regions is measured in kilometres.

As already stated, we should clearly consider that some of the variables included (like tourism attraction rate or social capital), which cannot be directly observed and measured, are proxies which aim to account for regional push and pull factors that are not captured by standard gravity variables. For this reason, we should take a reasonable cautious approach when we derive implications from this empirical study.

2.4 Methodology

2.4.1 Estimation issues and model selection

From a methodological point of view, the gravity model[23] is the most common theoretical framework used in empirical analyses to study the spatial determinants of migrations. This model has been applied for the first time in 1962 by Jan Tinbergen who estimated a gravity equation of international trade flows[24].

The general gravity law between any origin location *i* and destination *j* may be expressed as follows:

$$M_{ij} = k^{\alpha_0} \frac{P_i^{\alpha_1} P_j^{\alpha_2}}{d_{ij}^{\alpha_3}} \quad [1]$$

In its simplest form, the gravity equation for migration states that the flow of people from place *i* to place *j*, denoted by M_{ij}, is proportional to the product of the two masses, denoted by P_i and P_j, and inversely proportional to their geographical distance, d_{ij}, broadly construed to include all factors that might create migration resistance. In empirical studies, stochastic version of the equation is used and it may be written as:

$$M_{ij} = \alpha_0 X_i{}^{\alpha_1} X_j{}^{\alpha_2} f\left(dist_{ij}, k_{ij}\right) \varepsilon_{ij} \quad [2]$$

where the matrices X_i and X_j include the variables describing the most relevant features of the regions at origin and destination, respectively. $dist_{ij}$ represents the geographical distance between origin *i* and destination *i*, and k_{ij} are other pair-specific impeding factors; ε_{ij} is an error term assumed to be statistically independent of the regressors.

At this point, a brief digression about the nature of the dependent variable (M_{ij}) and the most appropriate model is needed. In our case M_{ij} represents the gross bilateral migration flows between Italian regions. Namely, it is the number of persons cancelled for change of residence from region *i* and registered in region *j* at time *t.* It is therefore represented by integer numbers, and for this reason the natural starting point is to consider the use of count data models.

The simplest distribution used for modelling count data is the Poisson distribution with probability density function defined as:

[23] The name gravity model reflects the analogy to the law of universal gravitation developed by Newtown in 1687 to describe the gravitational interaction between two objects, for example planets.

[24] Ravenstein (1885) has been the pioneer of the use of gravity to model migration patterns.

$$f_y(y;\mu) = e^{-\mu} * \frac{\mu}{y!} \quad [3]$$

where y is the response variable, a strictly non-negative number representing (the dependent variable) and μ is the expected number of occurrences often called 'intensity parameter'. One important implication of the Poisson distribution is that it assumes, by construction, that the variance of the response variable is equal to its mean (denoted μ_i). This is called equidispersion property.

$$\mu_i = \exp(x_i\beta) = E[y_i|x_i] = Var[y_i|x_i] \quad [4]$$

In empirical analysis, this property has often been found to be quite restrictive, as the data are usually overdispersed (i.e. the conditional variance is most often higher than the conditional mean) and this may result in consistent, yet inefficient, estimation of the dependent variable (Cameron and Trivedi, 2005). According to Cameron and Trivedi (2005) one of the most common causes of overdispersion is neglected unobserved heterogeneity which originates from omitted variables. In order to deal with overdispersion, mixture models are most frequently employed. These models model heterogeneity among observations by adding an extra parameter, which is a function of unobserved heterogeneity. The negative binomial (NB) model is a specific example of a continuous mixture model. The use of the NB to model migration flows has been frequently applied in recent migration literature (see Devillanova and Garcìa Fontes, 2004, Biagi et al, 2011; Balia et al., 2015).
The NB model can be considered as a generalization of Poisson regression since it has the same mean structure as the Poisson regression ($E[y_i|x_i] = \mu_i$) but a different variance, defined as: $Var[y_i|x_i] = \mu_i + \alpha\mu_i^2$. The parameter α model the over-dispersion and has to be estimated by the model. NB parameters are estimated using maximum likelihood (ML) estimators.
Another common approach in migration literature is to consider a linear approximation of the dependent variable. More precisely, for sufficiently large values of M_{ij} the normal distribution may be an adequate approximation to the Poisson distribution. This is clearly the case of trade or migration flows which often assume quite large values. For this reason, a most commonly strategy in empirical analyses requires to take the log-normal formulation of the equation [2] and to use OLS estimator to estimate the gravity model coefficients. However, as emphasized by Flowerdew and Aitkin (1982) in the early '80s there are some serious methodological issues that must be taken into account with the log-normal formulation of the gravity model. The first concerns the fact that logarithmic transformation has an intrinsic effect on the nature of the estimation process. This problem arises from the well known *Jensen's inequality* which states that the expected value of the logarithm of a random variable is different from the logarithm of its expected value. Secondly, the

log-normal model is based on the assumption of homoskedasticity of the error term which is clearly unlikely to be met. Therefore, if the errors are heteroskedastic, the transformed errors will be generally correlated with the covariates violating OLS assumption and leading to inconsistent estimators. The third key problem with the OLS specification is the need to discard non positive (zero) values since the logarithm of zero is undefined.

These limitations have been remarked by Santos Silva and Tenreyro (2006, 2011). The solution proposed is to estimate the model in levels (instead of taking logarithms), thorough Pseudo Poisson Maximum Likelihood (PPML) estimators. It does not require the data to follow a Poisson distribution (that is why it is a pseudo-maximum likelihood estimator and not a maximum likelihood estimator).

2.4.2 Modelling migration flows

Our empirical analysis is conducted adopting the following functional form:

$$M_{ijt} = \alpha_0+\alpha_1 ln\,(X_{it-2}) + \alpha_2 ln\,(X_{j,t-2}) + \alpha_3\, ln\,(dist_{ij}) + years_t + d_i + d_j + \varepsilon_{ij} \qquad [5]$$

or alternatively, if we consider the log-linear formulation we can write the functional form as:

$$ln\,(M_{ijt}) = \alpha_0+\alpha_1 ln\,(X_{it-2}) + \alpha_2 ln\,(X_{j,t-2}) + \alpha_3\, ln\,(dist_{ij}) + years_t + d_i + d_j + \varepsilon_{ij} \qquad [6]$$

where the subscript *i* refers to the region of origin, *j* to the region of destination and *t* to time, with *t*= 2000, 2013. The observations in each year refer to pairs of OD regions, *ij* = 1, 2, ..., N =380. M_{ijt} is the number of persons cancelled for change of residence from region *i* and registered in region *j* at time *t*. The matrices $X_{i,t}$ and $X_{i,t}$ include the most relevant push and pull factors of the regions at origin and destination, previously discussed[25]. The variable $dist_{ij}$ captures the geographical distance between regions in each OD pair. Following the specification of Anderson and van Wincoop (2003) we also include fixed effects for origin and destination (d_i and d_j), while $years$ are time dummies which are supposed to capture the effect of macro shocks common to all region pairs. Moreover, we tackle potential endogeneity problem arising from reversal causality and/or

[25] Since there is not clear consensus among scholars whether variables should enter the model as a regional individual characteristics or in relative terms, in our analysis we follow the traditional specification, so variables are included as regional characteristics both at origin and destination.

from the correlation of any of the regressors with the error term by including all the explanatory variables lagged two years[26].

Additionally, as our sample observations refer to spatial units, one of the most econometric issues is related to the existence of cross-regional association, due to the existence of spatial spillovers (Griffith and Jones, 1980; Le Sage and Pace, 2009). More precisely, it is clear that since spatial units are intrinsically related to each other, the assumption of cross-sectional independence is excessively restrictive, whatever the methodological approach we consider (i.e. consider M_{ijt} as pure count variable or its linear approximation).

In other words, it is not reasonable to assume that flows of a given origin are independent from the features of the neighbouring regions, and analogously, that flows towards a specific destination respond only to specific features of itself. As emphasized in LeSage and Pace (2009) and LeSage and Dominguez (2012), if cross-regional dependence is not properly modelled, the estimated coefficients are likely to be biased and inconsistent. Recent developments in spatial econometrics (Halleck Vega and Elhorst, 2015) recognize the spatially lagged explanatory variable (SLX) model as the most flexible approach for modelling spillover effects. In particular, this model overcomes relevant drawbacks of other 'traditional' spatial models like the spatial autoregressive model (SAR) and the spatial error model (SEM). Indeed, despite, their popularity the SAR and SEM models have given evidence of many serious limitations, largely discussed by applied researchers (Elhorst (2010), Pace and Zhu (2012), Corrado and Fingleton, 2012). Moreover, as recently pointed out by LeSage and Thomas-Agnan (2015), a limitation of the SAR model is that Poisson or NB estimation procedure have not been developed for panel data models. Conversely, the major drawback of the SEM model is that it removers, by construction, any information about spatial spillovers.

Taking into account these considerations, and following Halleck Vega and Elhorst (2015) we tackle the issue of spatial dependence by estimating a SLX model and thus, including spatial lags of the explanatory variables in our previous specifications. The resulting models may be defined as:

$$M_{ijt} = \alpha_0 + \alpha_1 ln\,(X_{it-2}) + \alpha_2 ln\,(X_{j,t-2}) + \alpha_3\,ln\,(dist_{ij}) + \beta_1 ln(WX_{it}) + \beta_2 ln(WX_{jt}) + \\ + years_t + d_i + d_j + \varepsilon_{ij} \qquad [7]$$

or:

$$ln\,(M_{ijt}) = \alpha_0 + \alpha_1 ln\,(X_{it-2}) + \alpha_2 ln\,(X_{j,t-2}) + \alpha_3\,ln\,(dist_{ij}) + \beta_1 ln(WX_{it}) + \beta_2 ln(WX_{jt}) + \\ + years_t + d_i + d_j + \varepsilon_{ij} \qquad [8]$$

[26] We also consider longer lag structures; since the main results do not change appreciably we keep the second lag in order to maintain a larger number of observations.

where WX_{it} and WX_{jt} are the weighted average matrices of the neighbouring regions. Thus, for each variable of interest we compute a spatial lag by pre-multiplying each value by a spatial weighted matrix (W). The latter is constructed by the inverse of the distance between all possible pairs of regions and normalized with its maximum-eigenvalue, which has the main advantage of preserving the importance of the absolute distance between each region pair[27.] In the SLX model, compared to alternative spatial models, the direct and spillover effects do not require further calculation. More precisely, the total effect of a given variable may easily be decomposed into a direct component, due to changes occurred in a region's own variable, and an indirect effect, caused by changes in the same variable that take place in neighbouring regions at origin or destination. The direct effects are the coefficient estimates of the non-spatial variables (α_1, α_2) and the spillover effects are those associated with the spatially lagged explanatory variables (β_1, β_2). Finally, the term $\varepsilon_{i,j}$ is a pure random error, uncorrelated with the regressors. The other terms are the same as in [5] and [6].

2.5 Estimation results

2.5.1 Basic and Spatial gravity model

Columns (1), (3) and (5) of Table 2.4. display results obtained from estimation of our baseline specification using as dependent variable the interregional total flows. As we can see, we compare three different models: the NB, the log-linear and the PPML. Estimated coefficients are obtained using maximum-likelihood (ML), OLS and Pseudo-ML respectively. However, OLS estimates are reported only for comparing purposes with the NB and PPML model, thus, they will not be discussed. As a matter of fact, as previously highlighted, the estimated coefficients of the log-linear model are likely to be biased.

As discussed in the methodological Section, the main difference among NB and PPML models is how the dependent variable is considered. More precisely, the NB model consider M_{ijt} as pure count process while, the log-linear model and the PPML model support its normal approximation and thus, favour its approximation to a continuous variable.

The interpretation of reported coefficients is straightforward because all covariates are log-transformed, thus, the estimated parameters measure elasticities.

[27] Alternatively, many applied studies use the *W* matrix is row-standardized. In this case the impact of all other regions on a particular region is given by the weighted (relative) average of all regions' impacts.

From Table 2.4 we can see that most of the estimated effects using the three different estimators are significant and exhibit the expected signs. Leaving aside the PPML estimator, the mass variable, population, is positive and highly significant both at origin and destination. GDP per capita seems to play an important role as restraining factor indicating that richer regions significantly reduce people outflows. Conversely, no significant effect is detected at destination.

At destination, the estimated coefficients indicate that better endowment of public transports is important attractive features. A negative effect at origin prevails only using PPML.

In line with our expectations, positive evidence as push factor is found for crime rate. People leave regions with high crime rates. The NB model however, provides evidence of significant and positive effects at destination. This sign at destination is quite surprising. We would have logically expected to have a negative effect of crime rate on inflows. This result may due to the fact that that our proxy is not capturing adequately the social disamenities related to crime. Relative regional safety may be indeed related with other objective aspects (like crime associations, homicides or other criminal behaviour).

Our expectations are confirmed by the proxy for tourism amenities. The negative coefficient at origin suggests that tourism-oriented regions face lower outflows. Conversely, this variable turns out insignificant at destination.

As a general result, we find that human capital is relevant to explain the outflow from a given region; the positive sign suggests that an increase in the regional human capital level encourage out migration. On the other side, only in the NB model, human capital works as attracting factor at destination.

No significant effect (neither at origin and destination) is found for the proxy for social capital.

Focusing on the determinant at the region-pair level, we found that geographical distance has the expected adverse effect on migration flows.

As we have remarked in the methodological Section, the results obtained from this standard gravity specification are likely to be unreliable since spatial interactions among neighbouring regions are completely neglected. For this reason, these results may be interpreted at ‘face value’ only for comparing purpose with those obtained by an 'augmented' specification which deals with the issue of spatial dependence.

In columns (2), (4) and (6) of Table 2.4 we extend the basic specification including spatially lagged variables at origin and destination. This specification (SLX model) allows to fully account to cross-region dependence. As stated before, the coefficients of the spatial lag variables can be interpreted as indirect effects, which arise as the result of a change in a given variable occurring in the focal region’s neighbours.

As we can see, signs and relative magnitudes of the estimated coefficients which capture the direct effects remain roughly stable. The only two exceptions concern the NB regression: the estimated coefficient for public transport turns out significant and with a negative sign, whereas the human capital variable loses its explanatory power.

The spatial lag of per capita GDP at origin is significant only using PPML estimator. A negative and highly significant effect is detected. This result could suggest that, *ceteris paribus*, outflows from a certain region are discouraged if it is surrounded by prosperous regions. The negative sign of this coefficient, may be explained by the fact that richer neighbouring regions may trigger positive spillovers which restrain people to leave their home region since, due to scale economies and network effects they may benefit from positive externalities without incurring the cost of migration. Moreover, we should also consider that income opportunities are often counterbalanced by higher housing price and cost of living, discouraging migration flow (Mocetti and Porello 2010, 2012). Conversely, at destination this coefficient is positive and significant only for the NB model. This could indicate that prosperous neighbouring regions put into effect positive spillovers which increase the attractiveness of the focal region. Whit respect to the spatial lags of the public transport at origin, the negative and significant sign of the coefficient indicates that a good provision of public transport in neighbouring regions makes the focal region relatively more attractive, thus, migrant outflows from that region will be restrain. At destination, the NB regression shows that the spatial lag of public transport does not play any significant role, while it turns out significant and with a negative sign using the PPML estimator.

Estimation results for the NB model show symmetric (positive and significant) indirect effects, at origin and destination, for crime rate. These results further reinforce the positive direct effects previously discussed. These coefficients suggest from one side (at origin), that high levels of crime in the surrounded regions increase the outflows in the focal region while, from the other side (at destination), a given destination is relatively more attractive when its surrounding areas have a higher level of crime. In the PPML regression, indirect effects are detected only a destination.

Symmetric results (negative and significant coefficients) are also found for the proxy of tourism amenities. The results at origin suggest that outflows in a specific region will be lower when its neighbouring regions are relatively more 'touristic', thus, it can take advantage of positive spillovers related to tourism amenities. This effect seems to be not confirmed in the PPML regression. At the same time, all three models confirm an unexpected result at destination. The negative coefficient suggests that the inflow in a given destination is negatively correlated with the tourism attraction rate in neighbouring regions. More specifically, as those regions become more touristic the lower will be the inflows toward the focal region.

As for human capital, in the NB model the estimated coefficient of the spatial lag variable at origin is negative. This result may be linked to knowledge and innovation spillovers which arise from high level of human capital in neighbouring regions and that restrain the outflows from the focal region. Moreover, if we consider the PPML estimator, we find that the spatial lag of human capital at destination is significant and it exhibits a negative coefficient. This result is coherent with the idea that agglomeration forces in neighbouring regions attract fewer migrants in the focal region and make surrounding regions more favourite destinations.

Finally, any type of spatial dependence is found for social capital.

These first sets of results are quite important from the point of view of general migration modelling. In fact we find that neighbouring spatial effects are relevant in explaining migration flows and they cannot be ignored. Moreover, we have shown that in the great majority of circumstances the same lagged variable has a statistically different impact at origin and at destination. Finally we also provide evidence on the fact that, depending on how the nature of M_{ijt} is modelled, we can use different estimators and thus have different estimates.

Table 2.4 Panel gravity models for total interregional flows (2000-2013). NB, OLS and PPML regressions.

	NB		OLS		PPML	
Dependent variable: Migrants flows from Origin i to Destination j	Total flow		Total flow (ln)		Total flow	
	(1)	(2)	(3)	(4)	(5)	(6)
Origin characteristics						
Population	0.931***	1.142***	0.912***	1.288***	1.216***	1.112***
	(0.307)	(0.343)	(0.313)	(0.374)	(0.374)	(0.332)
GDP pc	-1.437***	-0.927***	-1.439***	-0.739**	-0.888***	-1.015***
	(0.330)	(0.293)	(0.358)	(0.358)	(0.263)	(0.222)
Spatial lag - GDP pc	-	0.256	-	1.420	-	-3.168**
		(1.668)		(1.739)		(1.436)
Public transports	-0.042	-0.085**	0.005	-0.081	-0.129***	-0.097**
	(0.042)	(0.042)	(0.056)	(0.055)	(0.046)	(0.040)
Spatial lag - Public transports	-	-0.619***	-	-1.068***	-	0.274
		(0.239)		(0.319)		(0.211)
Crime	0.230***	0.217***	0.244***	0.225***	0.120**	0.091**
	(0.043)	(0.042)	(0.050)	(0.048)	(0.050)	(0.042)
Spatial lag - Crime	-	0.535*	-	0.717**	-	-0.041
		(0.325)		(0.318)		(0.351)
Tourism attraction	-0.350***	-0.367***	-0.319***	-0.362***	-0.283***	-0.248***
	(0.061)	(0.059)	(0.061)	(0.061)	(0.080)	(0.066)
Spatial lag - Tourism attraction	-	-0.937**	-	-0.906**	-	-0.103
		(0.465)		(0.456)		(0.394)
Human capital	0.149**	0.099	0.064	0.027	0.255***	0.231***
	(0.074)	(0.070)	(0.081)	(0.079)	(0.070)	(0.072)
Spatial lag - Human capital	-	-1.470***	-	-1.624***	-	-0.246
		(0.491)		(0.601)		(0.326)
Social capital	0.032	0.008	0.022	-0.015	0.015	0.015
	(0.031)	(0.030)	(0.036)	(0.036)	(0.027)	(0.025)
Spatial lag - Social capital	-	0.195	-	0.230	-	0.140
		(0.200)		(0.214)		(0.165)
Destination characteristics						
Population	1.007***	0.805**	1.019***	0.899**	0.523	0.644*
	(0.320)	(0.360)	(0.317)	(0.360)	(0.402)	(0.368)
GDP pc	-0.089	0.324	0.037	0.459	0.062	0.547**
	(0.257)	(0.298)	(0.320)	(0.338)	(0.233)	(0.253)
Spatial lag - GDP pc	-	2.546*	-	2.293	-	1.745
		(1.346)		(1.697)		(1.495)
Public transports	0.124***	0.157***	0.176***	0.200***	0.233***	0.182***
	(0.040)	(0.046)	(0.051)	(0.052)	(0.049)	(0.047)
Spatial lag - Public transports	-	0.237	-	0.311	-	-0.412*
		(0.254)		(0.296)		(0.219)
Crime	0.087**	0.090**	0.047	0.042	0.019	0.056
	(0.044)	(0.043)	(0.048)	(0.047)	(0.057)	(0.053)
Spatial lag - Crime	-	0.601**	-	0.267	-	0.954***
		(0.262)		(0.305)		(0.273)
Tourism attraction	0.088	0.031	0.161**	0.087	0.113	0.095
	(0.064)	(0.062)	(0.067)	(0.066)	(0.075)	(0.078)
Spatial lag - Tourism attraction	-	-1.116***	-	-1.342***	-	-1.191***
		(0.336)		(0.425)		(0.350)
Human capital	0.177***	0.177**	0.091	0.069	0.098	0.106
	(0.066)	(0.069)	(0.076)	(0.082)	(0.080)	(0.075)
Spatial lag - Human capital	-	0.224	-	0.032	-	-0.662**
		(0.421)		(0.507)		(0.300)
Social capital	-0.022	-0.038	-0.029	-0.032	0.047	0.027
	(0.032)	(0.032)	(0.033)	(0.034)	(0.034)	(0.033)
Spatial lag - Social capital	-	-0.111	-	0.122	-	0.260
		(0.180)		(0.205)		(0.187)
Origin- Destination characteristics						
Distance (km)	-0.604***	-0.604***	-0.510***	-0.510***	-0.247***	-0.247***
	(0.058)	(0.058)	(0.061)	(0.061)	(0.074)	(0.074)
Observations	4,557	4,557	4,557	4,557	4,557	4,557
R-squared	-	-	0.59	0.89	0.79	0.80
Ln(alpha)	-1.260***	-1.262***	-	-	-	
	(0.062)	(0.062)				0.1713
Log-likelihood	-30269	-30265	-	-	-490938	-490104

Note: All models include a constant time dummies and fixed effects at origin level and destination level.
All variables (excluding time dumminies, origin dummies and destination dummies) are log-transformed and lagged two years
Cluster-robust standard errors at region-pair level are reported in pharentsis.
Level of significance: *** 1%, ** 5%, * 10%

2.5.2 Foreigners and Italians.

Previous results are related to the overall flows; however, it is interesting to verify how those results may change depending on the sample of migrants. More precisely, in this sub-section we present the results of the empirical investigation on the determinants of internal migration flows across Italian regions by distinguishing between native and foreign migrants.

Estimated results are reported in Table 2.5. As a general result, it is worth noting that the magnitude of the estimated coefficients, both at origin and destination, are steadily quite higher for foreigners compared to Italians.

The NB and PPML regressions highlights the role as a push factor of total population at origin (only for Italians in the NB model) while, at destination it is found to be negatively associated with foreign migration inflows. The fact that foreigners seem to have preferences for less crowded regions may be interpreted as a sort of crowding-out effect. Differently, Italians continue to be attracted by most populated regions.

The two models show that per capita GDP at origin is always negatively linked with interregional migration flows. The positive role of per capita GDP at destination is confirmed for foreign migrants using both PPML and NB, while for Italians, we find significant coefficients only using the PPML model.

Leaving aside the second column of Table 2.5, all regressions confirm a negative and significant coefficient for the spatial lag of per capita GDP at origin and positive and highly significant effects of per capita GDP at destination. Mostly, the estimated elasticities of per capita GDP are quite higher for foreigners than for Italians suggesting that the former are more sensitive to macroeconomic conditions compared to Italians..

The two different estimators report mixed results for public transport at origin. Using PPML, foreigners and Italians migration outflows decrease as long as the endowment of public infrastructure in the source region increases. However, this effect turns out insignificant in NB regressions. The spatial lag coefficient at origin is positive for foreign migrants and negative for Italians (only for the NB model). At destination, estimates confirm the fact that migrants, both Italians and foreigners, are attracted toward region which may provide a better endowment of public transport. However, this effect is partially restrained for foreigners. The negative sign of the spatial lag of this variable suggests that neighbouring regions offering a better public transport provision are view (by foreigners) as more attractive alternative destinations.

As for crime, NB results are quite similar for the two subgroups of migrants. Crime rate acts as pushing factor for both, Italians and foreigners. However, we find opposite signs for the spatial lag of crime rate at origin. The negative sign for foreigners suggests that, *ceteris paribus*, neighbouring effects due to diffused criminality may slightly decrease the outflows from a certain region, while a positive sign is found for Italians. This result suggests that negative spillovers effects from neighbouring regions with high criminality, push natives toward other 'safer' regions. The only difference with the PPML model is that no significant and direct effect is found for Italians.

The positive (and unexpected) direct effect of crime on migrants' inflows, previously found for total flows, is confirmed only for foreigners in the NB regression. However, the spatial lag is confirmed positive and highly significant for all four regressions.

In the PPML and NB regressions, the estimated coefficients for tourism related amenities at origin are broadly the same (in sign and magnitude) for the two groups, confirming again, the results previously found for total flow. For spatial lag variables (at origin and destination) results are quite mixed. NB regression provides evidences of negative indirect effects (both at origin and destination) on Italian migration flows. Conversely, PPML estimates provide the same results for the group of foreign migrants. Moreover, the negative indirect effect at destination is confirmed for Italians also using PPML estimator.

There are substantial differences in terms of human capital that are worth mentioning. PPML estimates for Italians show a positive and highly significant sign for human capital at origin. Apparently, this result appears consistent with the return migration hypothesis, discussed in recent empirical studies (Piras, 2015). More specifically, those involved in this type of migration are in general natives and the motivation beyond this decision is often driven by non economic factors (cultural, social and ageing). However, it may be also consistent with the idea that since high skilled individuals are the fringe of the population most involved in migration phenomena (and they are mainly Italians), regions with high level of human capital may be also regions which generate higher outflows. At destination, while the human capital coefficient is insignificant (for the two sub-groups), its spatial lag is negative and highly significant only for foreign population. According to the new economic geography models, this result suggests that the availability of well-educated population in neighboring regions represents an advantage for the localization of innovative firms that further attracts more and more individuals (especially high skilled) and firms. In other words, neighboring regions are relatively more attractive than the focal one.

Estimated results for human capital are quite different for the NB model. At origin we do not find any direct effect of human capital, while the spatial lag for Italians is negative. This result suggests that the potential human capital spillovers coming from neighbouring regions are able to make a certain region of origin more attractive for potential migrants.

On the other hand, at destination, human capital plays a positive role in attracting Italian migrants. Conversely, the negative sign of the spatial lag indicates, again, that neighbouring regions are relatively more attractive for foreigners.

The role of social capital seems quite ambiguous. PPML estimates lead us to conclude that increasing social participation both, in the focal region and the neighbouring regions, encourage (although the estimated elasticity is quite low and slightly significant) Italian outflows. Social capital theory assumes that active participation in local services and voluntary associations may better identify and support collective goals that reinforce civil norms. However, the role and importance of social capital may depend on social, economic, and political contexts. In addition, our proxy of the social capital concept may fail to measure the type of social capital relevant to Italians migrants and how it operates in migrant populations.

Opposite effects, and more in line with our expectations are found for foreign migrants. At destination, foreign inflows are positively related to the social involvement in a specific destination region and to that in its close regions. Conversely, we do not find any significant effects for natives.

NB regression does not provide evidence for direct effect of social capital at origin while, the negative (positive) sign for foreigners (Italians) are confirmed. At destination, leaving aside the spatial lag for Italians which turns out to be significant and with negative sign, results are in line with those provided with the PPML estimator.

As far as the role of space is concerned, the distance between the region of origin and the region of destination matters for the two subgroups. As expected the coefficient is negative. It should be also highlighted that this effect is much higher for foreigners than for natives. The former have in fact a higher propensity to move but as suggested by descriptive statistics, they manly relocate between closer regions.

The estimation results provided so far, are quite interesting, however, they have to be thought and interpreted as 'general' in term of directions of migration flows. More specifically, we are not able to discern which the regions of origin are and which those of destination. Due to the persistent dualism between Mezzogiorno and Centre-North regions, it can be reasonable to assume that migrants leaving the regions in the South may respond differently to the push and pull forces with respect to Northern migrants. Analyze interregional determinants for the two main directions of migration (i.e. from South to Centre-North and vice versa) may help to shed some light on results find so far, especially with respect to some key factors, like human capital.

Table 2.5 Panel spatial gravity models for interregional flows (2000-2013) by migrants' citizenship. NB, and PPML regressions.

	NB		PPML	
Dependent variable: Migrants flows from Origin i to Destination j	Foreigners	Italians	Foreigners	Italians
	(1)	(2)	(3)	(4)
Origin characteristics				
Population	0.773	0.853***	1.938**	0.669**
	(0.739)	(0.327)	(0.949)	(0.306)
GDP pc	-2.005***	-0.639**	-2.333***	-0.685***
	(0.574)	(0.298)	(0.617)	(0.220)
Spatial lag - GDP pc	-8.960***	0.919	-8.472**	-2.239**
	(2.808)	(1.671)	(3.681)	(1.070)
Public transports	-0.071	-0.065	-0.175**	-0.076*
	(0.085)	(0.044)	(0.087)	(0.039)
Spatial lag - Public transports	0.740*	-0.615**	1.072**	0.138
	(0.386)	(0.240)	(0.464)	(0.200)
Crime	0.164*	0.195***	0.345***	0.046
	(0.090)	(0.043)	(0.098)	(0.040)
Spatial lag - Crime	-2.551***	0.987***	-4.522***	0.659*
	(0.592)	(0.316)	(0.664)	(0.347)
Tourism attraction	-0.225*	-0.324***	-0.249**	-0.218***
	(0.123)	(0.058)	(0.125)	(0.066)
Spatial lag - Tourism attraction	-1.033	-0.797*	-1.332**	0.078
	(0.740)	(0.463)	(0.646)	(0.335)
Human capital	-0.025	0.113	0.150	0.211***
	(0.154)	(0.072)	(0.183)	(0.070)
Spatial lag - Human capital	-0.359	-1.587***	0.711	-0.449
	(0.854)	(0.494)	(0.920)	(0.302)
Social capital	0.023	0.017	-0.194***	0.048*
	(0.064)	(0.031)	(0.070)	(0.025)
Spatial lag - Social capital	-1.839***	0.478**	-3.010***	0.620***
	(0.394)	(0.201)	(0.350)	(0.147)
Destination characteristics				
Population	-3.593***	0.817**	-3.523***	0.672*
	(0.833)	(0.357)	-0.874	(0.359)
GDP pc	1.574***	0.487	1.618***	0.666***
	(0.570)	(0.298)	(0.461)	(0.246)
Spatial lag - GDP pc	4.752*	3.969***	8.449***	2.069
	(2.675)	(1.345)	(2.824)	(1.439)
Public transports	0.182**	0.167***	0.458***	0.167***
	(0.089)	(0.047)	(0.098)	(0.044)
Spatial lag - Public transports	-2.305***	0.422	-1.929***	-0.311
	(0.487)	(0.262)	(0.427)	(0.215)
Crime	-0.134	0.089**	-0.191	0.068
	(0.091)	(0.043)	(0.134)	(0.049)
Spatial lag - Crime	4.566***	0.607**	6.398***	0.824***
	(0.581)	(0.253)	(0.591)	(0.269)
Tourism attraction	0.364***	0.053	0.446***	0.074
	(0.125)	(0.060)	(0.152)	(0.073)
Spatial lag - Tourism attraction	-0.879	-0.714**	-2.376***	-0.732**
	(0.726)	(0.317)	(0.654)	(0.347)
Human capital	0.100	0.152**	0.122	0.081
	(0.150)	(0.071)	(0.152)	(0.074)
Spatial lag - Human capital	-1.918**	0.547	-4.074***	-0.186
	(0.835)	(0.427)	(0.771)	(0.297)
Social capital	0.158**	-0.040	0.306***	0.017
	(0.065)	(0.034)	(0.074)	(0.032)
Spatial lag - Social capital	3.249***	-0.302*	3.322***	0.078
	(0.358)	(0.180)	(0.328)	(0.189)
Origin- Destination characteristics				
Distance (km)	-0.662***	-0.598***	-0.618***	-0.175**
	(0.041)	(0.063)	(0.054)	(0.079)
Observations	4,557	4,557	4,557	4,555
Ln(alpha)	-1.749***	-1.132***	-	-
	(0.060)	(0.060)		
Log-likelihood	-19852	-29821	-50163	-486776

Note: All models include a constant time dummies and fixed effects at origin level and destination level.
All variables (excluding time dummines, origin dummies and destination dummies) are log-transformed and lagged two years
Cluster-robust standard errors at region-pair level are reported in pharentsis.
Level of significance: *** 1%, ** 5%, * 10%

2.5.3 South to Centre-North and Centre-North to South migration flows

The direction of migration flows deserves to be discussed in more detail in order to better understand the role of push and pull factors on internal migration. For this reason, we extend our analysis by considering separately, two different directions of internal movements.

The first direction is from the eight Southern regions to the twelve regions in the Centre-North. Clearly, this represents the main direction of interregional flows in Italy and it encompasses the largest volume of migrants. The second direction goes from the Central and Northern regions to the Mezzogiorno (Southern regions and the two main islands). This less sizeable flow is often defined 'return migration' flow. It is supposed to be characterized mainly by past migrants (who often emigrated from during the 1950s and 1960s) who have decided to come back towards their regions, mainly for cultural, social and family factors rather than to pure economic one.

While the analysis of South toward the Centre-Northern flows is not new in migration literature (Salvatore, 1977; Daveri and Faini, 1999; Cannari et al., 2000; Brunello et al., 2001), the number of studies which focus on Centre-North to South migration flows are quite new for Italian migration literature (Etzo 2011, Piras 2015). In particular, to the best of our knowledge this type of analysis has never been carried out separately for two different subgroup of population (natives and foreigners).

Estimation results are reported in Table 2.6. As we can observe, the most populated regions of the South generate large outflows of Italians and foreigners (except using the PPML model). At destination, only the NB model shows that Southern Italians are slightly more attracted toward high populated Northern regions. Conversely, the population size is not relevant to explain Centre-North to South flows, neither at origin nor at destination.

As we expected, there are noticeable differences with regard to the economic determinants of migrations flows in the two opposite directions but also, among Italian and foreigners. Per capita GDP at origin is significant only for Italians Southern migrants (i.e. individuals which move from South to Centre-North) suggesting that natives prefer to stay in their region of origin if the economic development level is sufficiently high. The same reasoning may be applied for the foreign counterpart. In fact, even if any direct effect is detected, the negative coefficient of the spatially lagged term for per capita GDP suggests that, the foreign outflows from a given Southern region is lower if it is surrounded by more prosperous regions. The high elasticity for destination per capita GDP signals that migrants' inflows (Italians and foreigners) toward Northern regions are enhanced by high levels of economic development.

By contrast, the role of per capita GDP for Northern migrants is quite different. While the well-begin in the Northern regions has a considerable effect on restraining foreign outflows, the economic prosperity of Southern regions does not work as attracting factor for migrants.

For Northern Italians migrants, we find significant and positive evidence only for the spatial lag of per capita GDP at destination. These results may suggest, as we would expect, that this 'opposite' migration does not seem to be directly driven by pure economic factors.

Only using PPML, public transport endowment negatively affects Italian outflows from Southern regions, while it does not influence migration outflows from the Northern ones. At destination, Northern regions with better endowment of public transport are preferred by Italian and foreigners, any direct effects is found for the South to North migration. For the South to Centre North direction, the spatially lagged variable (at destination) shows an opposite sign for the two groups of migrants suggesting that (as we found for total flows), foreigner inflows in a certain region decrease if neighbouring regions provide better infrastructures, while the Italian ones slightly increase (this latter result is not supported by the PPML model where it turns out not significant).

For the spatial lag at origin we find a positive sign for foreigners (using NB model) and for Italians (sing the PPML). In the opposite direction (namely form Centre-North to South), we do not find any significant indirect effects of public transports at origin. Conversely, at destination a negative sign of the spatial lag is confirmed for three models over four.

Quality of citizens' life (proxied by the crime rate) seems to be an important determinant for Italian migration. We find that an increase in crime rate pushes Italian outflows from both, Southern and Northern regions. While no significant effects at origin is found for foreigners.

Conversely at destination, the negative sign suggests that the level of crime in Northern region discourages foreign inflows while, it does not affect Italians (neither with the NB nor with the PPML model). The level of crime in Southern destination regions does not have any direct effects on inflows migration of both, Italians and foreigners.

Moreover, the positive and significant coefficient of the spatial lag of crime at destination, suggests that the inflow in a certain region is positively correlated with the crime rate in the surrounding regions, meaning that neighbouring regions are not viewed as attractive alternative destinations.

This result, it is confirmed for all models, excluding columns (2) and (8).

Tourism amenities represent a restraining factor for foreigners which move from South to Centre-North. In fact, as already stated, foreigners are often employed in the tourism sector (in restaurants and hotels) which constitutes a relevant part of Southern regions' economy. Conversely, when Italians are involved in such migration a negative sign (using the NB model) at origin is detected. Thus, using the NB model, tourism amenities seems to be a push factor for Italian flow, both from

South to Centre- North and from Centre-North to South. For Italians migrants involved in North to South migration, we also find a negative and significant sign for tourism attraction at destination (while no significant effects is found for foreigners) probably because the most touristic regions are also the less developed and isolated, like the two main islands (Sardinia and Sicilia). Conversely, more touristic regions among Northern destination regions seem to be preferred by both Italians and foreigners (expect using the PPML estimator). However, this direct effect seems to be partially offsets by the tourism attraction coming from the neighbouring regions, as suggested by the negative coefficients of the spatially lagged variables. In the North to South direction, we only find a significant and positive term in the PPML model for foreign migrants.

At origin, the spatial lag is positive for Northern foreign migrants while, only in the PPML model, we find a negative sign for the Southern one. As far as Italians are concerned, the estimated models report a positive sign only for Southern Italians.

As for human capital, there are considerable differences that are worth mentioning. Estimated coefficient of human capital at origin for South to Centre-North migration flows is significant and negative as we found in Table 2.4. This result confirms the hypothesis advanced before. In fact, it indicates that among Southern regions, those with the highest level of human capital are also the most able to generate migrants' outflows, probably as a results of the positive selection on migration phenomena (i.e. high skilled people have a higher propensity to migrate). At the same time (except in column 6), Northern regions with a high level of human capital significantly attract more migrants from the South Italy. Comparing the magnitude we also notice that the impact is slightly stronger at destination rather than at origin. When migration occurs in the opposite direction, namely from Centre-North to South, we can see that migration outflows is restrained when human capital increases in the focal region (for foreigners) or when it increases in the neighbouring regions (for Italians in the NB model). As suggested by Piras (2015) '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. Only the spatial lag term is significant (and only for Italians) and it exhibits a positive coefficient. This result suggests that being surrounded by regions with high human capital contributes to increase regional attractiveness. For foreigners (in the PPML model), however, we find a negative sign of this spatial lag.

The role of social capital is quite ambiguous and heterogeneous. The positive sign (find in columns 2, 3 and 7) suggests that foreign outflows migration from Centre-North regions and Italian outflows from Southern regions, are enhanced by the high level of social capital. Similarly, looking at columns 1, 2, 4 and 8, we can see that, on one hand foreigners are attracted toward Northern regions with high social capital, while, on the other hand, evidence of discouraging and adverse effects, is found for Italians, as inflows are expected to be lower in Sothern region with high social capital. As already said, those results for social capital may appear surprising and quite difficult to interpret. The main explanation that we can provide for this result is that social capital (which in this case we proxy with the membership in voluntary associations) is a particular feature of our society which is often very hard to measure.

Spatial effects are quite mixed as well. For the South to Centro-North direction we can see that that the social capital in surrounding regions enhances the attraction of foreign and Italian migrates (only with PPML). Conversely, in the opposite direction we find that high levels of social capital in neighbouring makes deter the inflows in the focal destination region.

At the destination side, a positive sign is confirmed for Northern Italian migrants and a negative one for Northern foreign migrants (only for the NB model). In the South to North direction PPML regressions report a negative sign for the two sub-sample of migrants (foreign and Italians).

Finally, we can see that geographical distance exhibits a stronger impact on migration flows compared with previous results (Table 2.4). However, this result is not surprising considering that intra-areas flows (i.e. among same macro areas) are excluded in those models and thus, only 'long distance' migration is considered.

Table 2.6 Panel spatial gravity models for interregional flows (2000-2013) by migrants' citizenship and different directions. NB and PPML regressions.

	NB				PPML			
	From South to Centre-North		From Centre-North to South		From South to Centre-North		From Centre-North to South	
Dependent variable: Migrants flows from Origin i to Destination j	Foreigners	Italians	Foreigners	Italians	Foreigners	Italians	Foreigners	Italians
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Origin characteristics								
Population	4.668**	3.319***	0.349	0.478	3.251	3.544***	1.670	-0.198
	(1.904)	(0.828)	(1.630)	(0.773)	(2.051)	(0.906)	(1.671)	(0.859)
GDP pc	-0.476	-1.256***	-4.637***	-0.581	-0.156	-0.978**	-3.875***	-0.713
	(0.908)	(0.434)	(1.067)	(0.592)	(1.102)	(0.494)	(0.936)	(0.505)
Spatial lag - GDP pc	-20.829***	-4.099	-8.324*	0.136	-25.237***	-2.382	-8.691**	-1.873
	(6.790)	(3.425)	(4.900)	(3.301)	(6.463)	(4.334)	(4.261)	(1.947)
Public transports	-0.002	-0.154**	-0.117	0.053	0.049	-0.094	-0.010	0.046
	(0.165)	(0.063)	(0.190)	(0.087)	(0.161)	(0.079)	(0.195)	(0.081)
Spatial lag - Public transports	1.521**	0.256	-0.011	-0.212	0.838	0.681**	0.839	0.169
	(0.750)	(0.373)	(0.697)	(0.450)	(0.584)	(0.304)	(0.669)	(0.258)
Crime	0.154	0.176**	0.023	0.255**	0.221	0.138**	0.010	0.086
	(0.146)	(0.070)	(0.216)	(0.109)	(0.145)	(0.064)	(0.196)	(0.087)
Spatial lag - Crime	-1.443	1.190*	-1.735	0.907*	-0.611	1.381*	-3.105***	1.013*
	(1.613)	(0.682)	(1.150)	(0.508)	(1.779)	(0.758)	(1.096)	(0.548)
Tourism attraction	0.391*	-0.190*	-0.068	-0.372***	0.390*	0.092	-0.021	-0.307***
	(0.205)	(0.099)	(0.163)	(0.120)	(0.210)	(0.099)	(0.162)	(0.096)
Spatial lag - Tourism attraction	-1.969	1.577*	3.968***	-0.867	-3.833*	1.445	3.863***	0.410
	(2.221)	(0.894)	(1.351)	(0.706)	(2.274)	(0.909)	(1.457)	(0.456)
Human capital	0.578**	0.209*	-1.018***	-0.115	0.733**	0.258**	-0.826***	-0.062
	(0.233)	(0.110)	(0.299)	(0.145)	(0.296)	(0.105)	(0.301)	(0.122)
Spatial lag - Human capital	-0.871	-1.395**	0.944	-1.790**	1.695	0.233	2.623*	0.406
	(1.710)	(0.699)	(1.439)	(0.792)	(1.981)	(0.574)	(1.447)	(0.710)
Social capital	0.030	0.133***	0.236*	-0.122	0.000	0.033	0.272**	-0.022
	(0.097)	(0.037)	(0.131)	(0.078)	(0.083)	(0.038)	(0.134)	(0.058)
Spatial lag - Social capital	-1.155	-0.352	-1.194*	0.580*	-2.040**	-0.753*	-1.109	0.980***
	(1.013)	(0.412)	(0.708)	(0.340)	(0.872)	(0.395)	(0.812)	(0.319)
Destination characteristics								
Population	-1.384	1.539*	-1.672	0.973	-1.614	1.065	-2.843	0.918
	(1.825)	(0.886)	(1.798)	(0.895)	(2.086)	(1.064)	(1.871)	(0.965)
GDP pc	2.830***	1.361***	0.992	0.302	3.367***	1.799***	-0.768	0.203
	(0.973)	(0.497)	(1.101)	(0.524)	(0.884)	(0.313)	(0.997)	(0.377)
Spatial lag - GDP pc	-6.955	0.627	4.768	10.833***	6.345	0.730	-4.423	12.431***
	(5.119)	(2.172)	(7.967)	(3.755)	(6.060)	(2.403)	(8.340)	(2.968)
Public transports	0.519***	0.241***	0.080	0.016	0.842***	0.392***	0.152	0.059
	(0.180)	(0.074)	(0.155)	(0.085)	(0.173)	(0.111)	(0.142)	(0.069)
Spatial lag - Public transports	-1.825**	0.625*	-1.560*	-0.136	-2.784***	0.007	-1.519**	-0.535**
	(0.785)	(0.369)	(0.902)	(0.449)	(0.770)	(0.308)	(0.701)	(0.229)
Crime	-0.785***	-0.095	-0.090	-0.000	-0.436**	-0.011	0.026	0.076
	(0.174)	(0.087)	(0.130)	(0.069)	(0.205)	(0.102)	(0.142)	(0.070)
Spatial lag - Crime	4.755***	0.648	3.190**	1.548**	7.598***	1.601**	4.129**	0.547
	(1.292)	(0.502)	(1.607)	(0.687)	(1.408)	(0.625)	(1.803)	(0.648)
Tourism attraction	0.604***	0.386***	-0.247	-0.162**	0.228	0.273**	-0.134	-0.244***
	(0.220)	(0.100)	(0.161)	(0.080)	(0.237)	(0.124)	(0.165)	(0.077)
Spatial lag - Tourism attraction	-3.606***	-1.109***	3.979	1.291	-6.188***	-2.005***	4.644*	-0.056
	(1.216)	(0.424)	(2.466)	(0.988)	(1.180)	(0.506)	(2.483)	(0.819)
Human capital	0.594**	0.281**	0.084	0.065	0.652***	0.144	0.241	-0.148
	(0.257)	(0.132)	(0.228)	(0.114)	(0.241)	(0.146)	(0.173)	(0.092)
Spatial lag - Human capital	-5.596***	0.847	4.127**	0.170	-6.648***	-0.603	4.310**	-1.627**
	(1.298)	(0.754)	(2.016)	(0.942)	(1.221)	(0.622)	(2.090)	(0.694)
Social capital	0.478***	0.018	-0.141	-0.160***	0.538***	0.010	-0.185*	-0.102***
	(0.151)	(0.063)	(0.087)	(0.050)	(0.158)	(0.063)	(0.096)	(0.035)
Spatial lag - Social capital	4.495***	0.275	1.439	-0.998*	4.653***	0.482*	1.168	-0.588
	(0.627)	(0.285)	(0.958)	(0.550)	(0.657)	(0.263)	(0.862)	(0.437)
Origin- Destination characteristics Distance (km)	-1.058***	-1.711***	-1.265***	-1.710***	-1.251***	-1.791***	-1.352***	-1.826***
	(0.231)	(0.197)	(0.203)	(0.211)	(0.312)	(0.301)	(0.244)	(0.420)
Observations	1,149	1,149	1,152	1,152	1,149	1,149	1,152	1,152
ln(alpha)	-2.534***	-2.401***	-2.634***	-2.336***	0.947	0.959	0.930	0.944
	(0.142)	(0.133)	(0.136)	(0.136)	-	-	-	-
Log-likelihood	-4871	-7177	-3959	-6671	-7356	-37266	-4901	-27805

Note: All models include a constant time dummies and fixed effects at origin level and destination level. All variables (excluding time dummines, origin dummies and destination dummies) are log-transformed and lagged two years.
All variables (excluding time dummines, origin dummies and destination dummies) are log-transformed and lagged two years. Cluster-robust standard errors at region-pair level are reported in pharentsis. 10%

2.6 Robustness analysis

In order to test the strength of the results discussed so far, we conduct additional robustness analysis with respect to alternative explicative variables and to the inclusion of region-pairs fixed effects.

2.6.1 Robustness analysis with alternative explanatory variables

In Table 2.6 we check whether the results of the basic model [7] are robust with respect to the inclusion of alternative measures for the mass variable (population) and the economic variable (per capita GDP), both at origin and destination. Specifically, the total population is replaced by the population aged 15-64 (defined as economically active population) and the per capita GDP by the unemployment rate. The former variable aims to capture the size of the labour market while unemployment rate is a more accurate indicator for labour market conditions and job opportunities. All other variables are unchanged. Although with lower estimated elasticities compared to those found in Table 2.4, economically active population affects migration flows at origin and destination. As before, the former effect, appear to be stronger than the latter. This finding is confirmed by the two regressions (NB and PPML) and it suggests a negative *net* effect for the most populated Italian regions (like Campania, Sicilia, Puglia).

For the NB model, the estimated coefficient of the unemployment rate turns out to be significant only at destination. Conversely, using PPML estimator, the outcomes for the unemployment rate reports the results predicted by economic theory: it has a positive (push) effect in the sending and a negative (pull) effect in the destination region. However, it seems to have less impact on migration than the per capita GDP probably because, as already stated, the per capita GDP better reflects the overall economic conditions. All other results are sufficiently stable, especially for the NB model.

2.6.2 Robustness analysis on unobservable heterogeneity

The spatial gravity specifications estimated so far distinguish between unobservable regional specific characteristics at origin and destination including origin and destination fixed effects. However, there might be unobservable heterogeneity between region pairs. For this reason, we repeat the analysis, including pairwise effects in order to take into account those unobservable characteristics at both locations. In other words, this heterogeneity may be seen as the unobservable propensity of the origin *i* migrants to move in a given destination *j* [28]. In linear panel data models, this is typically done by using the standard fixed-effect (FE) estimator. In the second column of Table 2.8 we report the results of the fixed-effects Poisson (Quasi-ML) regression model[29].

[28] Our model fails to converge including simultaneously, fixed effects at origin and destination a pairwise fixed effects.
[29] Since *ppml* Stata command does not have an option to include country-pair fixed effects, as suggest by Santos Silva and Tenreyro (2006) we consider Timothy Simcoe's *xtpqml* Stata command.

In NB models the inclusion of region pairs FE is more problematic because neither a conditional or unconditional FE estimator can be used[30]. In fact, as demonstrated by Allison and Waterman (2002), the conditional FE estimator for NB model (proposed by Hausman et al., 1984) is not a true FE method. Similarly, unconditional FE estimator for short panel is not feasible due to the incidental parameter problem (IPP). As discussed by Cameron and Trivedi (2013), the conditionally correlated random effects (CCRE) model seems to be useful substitute when researchers are not able to deal with FE specification and the RE assumption appear too restrictive. The basic assumption of this approach is to assume that exogenous regressors and the unobservable effect are conditionally correlated. More specifically, unobserved effects are specified as a function of the time-averages of all time-varying exogenous regressors. Thus, this model may be thought of as an intermediate between the FE and the RE model. Therefore, the final specification of the CCRE-NB model may be defined as:

$$M_{ijt} = \gamma_{ij} + \alpha_0 + \alpha_1 ln\ (X_{it-2}) + \alpha_2 ln\ (X_{j,t-2}) + \alpha_3\ ln\ (dist_{ij}) + \beta_1 ln(WX_{it}) + \beta_2 ln(WX_{jt}) + \\ + years_t + d_i + d_j + \varepsilon_{ij}$$

$$\text{with } \gamma_{ij} = \lambda_o\ ln(\overline{x_\imath}) + \lambda_d\ ln(\overline{x_\jmath}) + \phi_{ow} ln(W\overline{x_\imath}) + \phi_{dw} ln\ (W\overline{x_\jmath} + \varepsilon_{ij}) \qquad [9]$$

where, γ_{ij} are the unobservable effects which are assumed to be correlated with the time-averages of, origin and destination variables $\left(\overline{x_\imath} = \frac{1}{T\sum_{t=1}^{T} x_{it}}\ ;\ \overline{x_\jmath} = \frac{1}{T\sum_{t=1}^{T} x_{jt}}\right)$, and spatial lags of the same variables. The first column of Table 2.8 displays econometric results of model (9). As we can see estimates of these two models compare favourably with the results, both in terms of significance level and magnitude, provided in Table 2.4. The only noteworthy difference refers to three variables (public transport, spatial lag of crime rate at origin and population, at destination), which turn out to be insignificant. The geographical distance exhibits a slightly lower impact on migration flows compared to the basic model results[31]. This highlights the fact that part of the explanatory power is captured by the region-pair effects. The results discussed so far confer additional robustness to our results.

[30] The idea has been originally developed by Mundlak (1978) and Chamberlain (1982) in the context of linear panel models.

[31] The CCRE-NB model allows to estimate the coefficients of the time-invariant regressors (e.g. geographical distance), which by construction are removed in a standard FE model.

Table 2.7 Panel spatial gravity models for total interregional flows (2000-2013) with alternative variables. NB and PPML regressions.

	NB	PPML
Dependent variable: Migrants flows from Origin i to Destination j	Total Flow	Total Flow
Origin characteristics		
Population aged 15-64	0.887***	0.611***
	(0.179)	(0.231)
Unemployment rate	0.010	0.083**
	(0.036)	(0.039)
Spatial lag - Unemployment rate	-0.270	-0.154
	(0.232)	(0.164)
Public transports	-0.073*	-0.045
	(0.042)	(0.043)
Spatial lag - Public transports	-0.621***	0.192
	(0.238)	(0.204)
Crime	0.223***	0.083*
	(0.047)	(0.046)
Spatial lag - Crime	0.761**	0.473
	(0.352)	(0.336)
Tourism attraction	-0.379***	-0.308***
	(0.061)	(0.064)
Spatial lag - Tourism attraction	-1.172***	-0.360
	(0.427)	(0.397)
Human capital	0.057	0.202***
	(0.072)	(0.072)
Spatial lag - Human capital	-1.203***	-0.480*
	(0.374)	(0.258)
Social capital	-0.009	0.016
	(0.032)	(0.025)
Spatial lag - Social capital	0.214	0.059
	(0.198)	(0.175)
Destination characteristics		
Population aged 15-64	0.450**	0.358*
	(0.182)	(0.204)
Unemployment rate	-0.072**	-0.069*
	(0.029)	(0.039)
Spatial lag - Unemployment rate	0.248	0.240
	(0.170)	(0.181)
Public transports	0.203***	0.232***
	(0.046)	(0.053)
Spatial lag - Public transports	0.346	-0.272
	(0.243)	(0.229)
Crime	0.122***	0.097*
	(0.042)	(0.053)
Spatial lag - Crime	0.829***	1.163***
	(0.270)	(0.296)
Tourism attraction	0.091	0.128
	(0.062)	(0.079)
Spatial lag - Tourism attraction	-0.770**	-0.890***
	(0.329)	(0.297)
Human capital	0.148**	0.058
	(0.069)	(0.071)
Spatial lag - Human capital	0.449	-0.366
	(0.393)	(0.291)
Social capital	-0.031	0.024
	(0.032)	(0.032)
Spatial lag - Social capital	0.063	0.295*
	(0.181)	(0.160)
Origin- Destination characteristics		
Distance (km)	-0.604***	-0.247***
	(0.058)	(0.074)
Observations	4,557	4,557
ln(alpha)	-1.262***	-
	(0.062)	
Log-likelihood	-30264	-490108

Note: All models include a constant time dummies and fixed effects at origin level and destination level.
All variables (excluding time dummines, origin dummies and destination dummies) are log-transformed and lagged two years
Cluster-robust standard errors at region-pair level are reported in pharentsis.
Level of significance: *** 1%, ** 5%, * 10%

Table 2.8 Panel spatial gravity models for total interregional flows (2000-2013) with region pairs fixed effects. CCRE-NB and PQML regressions.

	CCRE-NB	PQML
Dependent variable: Migrants flows from Origin i to Destination j	Total Flow	Total Flow
Origin characteristics		
Population	0.777*	1.254***
	(0.449)	(0.336)
GDP pc	-0.932**	-1.024***
	(0.468)	(0.219)
Spatial lag - GDP pc	-0.746	-3.090**
	(2.221)	(1.376)
Public transports	-0.067	-0.093**
	(0.059)	(0.040)
Spatial lag - Public transports	-0.588*	0.294
	(0.332)	(0.212)
Crime	0.242***	0.102***
	(0.048)	(0.039)
Spatial lag - Crime	0.448	-0.052
	(0.390)	(0.323)
Tourism attraction	-0.332***	-0.237***
	(0.107)	(0.063)
Spatial lag - Tourism attraction	-1.361**	-0.164
	(0.543)	(0.363)
Human capital	0.115	0.243***
	(0.099)	(0.071)
Spatial lag - Human capital	-1.403**	-0.242
	(0.601)	(0.332)
Social capital	0.021	0.008
	(0.038)	(0.025)
Spatial lag - Social capital	0.053	0.059
	(0.220)	(0.150)
Destination characteristics		
Population	0.185	0.937***
	(0.489)	(0.344)
GDP pc	-0.115	0.564**
	(0.501)	(0.255)
Spatial lag - GDP pc	2.116	1.630
	(1.993)	(1.402)
Public transports	0.224***	0.182***
	(0.061)	(0.043)
Spatial lag - Public transports	0.532	-0.353*
	(0.339)	(0.209)
Crime	0.074	0.075
	(0.049)	(0.052)
Spatial lag - Crime	0.625*	0.920***
	(0.319)	(0.266)
Tourism attraction	0.040	0.113
	(0.097)	(0.073)
Spatial lag - Tourism attraction	-1.393***	-1.239***
	(0.438)	(0.340)
Human capital	0.236**	0.141*
	(0.093)	(0.074)
Spatial lag - Human capital	0.232	-0.721**
	(0.535)	(0.286)
Social capital	-0.025	0.022
	(0.037)	(0.033)
Spatial lag - Social capital	-0.243	0.100
	(0.198)	(0.181)
Origin- Destination characteristics		
Distance (km)	-0.572***	-
	(0.060)	
Observations	4,557	4,557
Log-likelihood	-30410	-34151

Note: All models include a constant time dummies and region-pairs fixed effects
All variables (excluding time dummines, origin dummies and destination dummies) are log-transformed and lagged two years
Cluster-robust standard errors at region-pair level are reported in pharentsis.
Level of significance: *** 1%, ** 5%, * 10%

2.7 Conclusions

In this Chapter we provide new insights on the factors determining internal migration flows across Italian regions for the period 2000-2013. We assess the most relevant push and pull factors of migration flows by applying a gravity model using annual data collection on changes of residence that occurred over the period 2000-2013, distinguishing between Italians and foreign citizens.

Overall, our results seem to be consistent with previous research on the determinants of migration (Piras, 2012, 2015; Mocetti and Porello, 2012; Etzo, 2011) arguing that not only macroeconomic variables matters, but also local amenities (and disamenities) have an important impact on shaping internal migration in Italy. In particular, the empirical analysis shows that the mobility of the Italian and the foreign population presents remarkable differences. In the period 2000-2013, the population size of the regions and their geographical distance appear to be important determinants of migratory flows for both, Italians and the foreigners. However, different patterns for the two groups emerge:

Italians move mainly towards more populated regions, while the foreigners seem to be attracted toward less populated regions. A better endowment of public transports and other tourism related amenities (like bars, restaurants, hotels and so on) seem to be relevant attractive factors thus, they are often find to be negatively related to migration outflows and positively with regional inflows. Similarly, local disamenities, as measured by the number of robberies and thefts per 10,000 inhabitants, which are likely to affect the quality of life, encourage internal movement.

Using different subsamples and different estimators, we find quite mixed results for the role of human capital as driver of migration flows. In general, results seem to confirm the fact that human capital in Southern regions does not restrain regional outflows. On the contrary, among Mezzogiorno regions, are those with the high level of human capital that sustain higher outflows.

The empirical analysis of this study contributes to the literature in two main directions. The first important contribution is related to the proposed empirical approach. The application of a spatially lagged explanatory variable (SLX) model has allowed us to examine the issues related to spatial dependence patterns in a panel migration context, overcoming some limitations of other 'traditional' spatial models. In fact, including the spatial lags of the explanatory variables in the gravity model, the total effect of a given variable can be decomposed into two parts: a *direct* one, due to changes occurred in a region's own variable, and an *indirect* one, caused by changes in the same variable taking place in neighbouring regions, at origin or destination. The econometric results show highly significant evidence of the existence of both origin and destination regional spillovers. In the great majority of empirical results the spatially lagged variables have a statistically different impact according to whether they operate at origin or at destination. These findings offer further empirical

support to the role played by a number of economic mechanisms and by migrant's behaviour, which make regions increasingly interdependent in the repulsion and attractiveness of potential migrants.
Moreover, as far as the econometric methodology is concerned, we carry out the empirical analysis by means of two different regression models: the NB, and the PPML. This choice is the result of two different approaches of modelling the dependent variable (represented by the number of internal movers): the fist considers the purely counts nature of the data, while the second, given the quite large values, favours its linear approximation.
The second contribution is related to the sample disaggregation. Using two different subsamples of population and three different migration patterns (total flows, South to Centre- North flows, Centre-North to South flows) we provide a new and better understanding of the multiple aspects of a complex phenomenon like migration. Despite its novelties, this study leaves ample room for future investigations. On the basis of the evidence found so far, it is our intention to extend the analysis by addressing some limitations. In particular, we aim to investigate the dynamic properties of the series under analysis in order to detect the existence of region to-region network effects. Moreover, conditional on data availability, it would be interesting to investigate how the geographical (re)distribution of foreigners may, in some ways, affects native internal mobility.

Chapter 3

Who moves across Italy? Modelling internal migration using LFS data

3.1 Introduction

As we discussed in the first Chapter, the vast majority of evidence for Italy have focused on the analysis and assessment of pull and push factors for the motivations underlying migrations. Economic considerations (such as unemployment rate, available income, economic structure, population density, living costs), appear to be the main relevant features driving the migration choice.

Only recently, regional analyses have paid more attention to the composition of migration flows, focusing especially on the subsample of skilled people (Piras 2005, Dotti et al. 2013; Marinelli, 2013) which represent the principal category of the new migration flows and that clearly have important implications on the socio-economic development of Italian regions.

Thus, from almost all existing literature, we now know a fair amount regarding the general extent and direction of internal mobility, some of the determinants of these flows, certain implications for economic adjustment, and so on. However, as far as we know, the empirical investigation is partially hindered by a lack of comprehensive individual data: the figures on internal migration are limited and do not offer detailed information on the observable personal characteristics. The purpose of this study is to shed new light on the topic of migration by presenting the results of an empirical analysis of interregional migration in Italy based on micro data. Following seminal contributions by Pissarides and Wadsworth (1989) and Antolin and Bover (1997) and many others, we use data from the Labour Force Survey (LFS) to estimate the probability of migration within a large and particularly interesting stratum of the Italian population, the labour force. More specifically, our research addresses the question: 'who moves?'.

In order to do so, we estimate a weighted logit model where the probability that a person moves from one region to another is taken to be a function of various personal characteristics (age, marital status, child, etc.) and some key economic attributes.

The Chapter is organized as follows. In Section 2 we briefly outline the recent literature on the determinants of migration flows. Section 3 presents the dataset under analysis and some descriptive statistics. The description of the methodology adopted to carry out the empirical analysis follows in Section 4. The econometric results and some robustness analysis are presented in Sections 5 and 6 while Section 7 concludes.

3.2 Literature review

Following the neoclassical theory of migration, the direction and the intensity of interregional migratory flows are driven foremost by economic factors. Regional differences in wages, employment or unemployment rates are the main cause of the phenomenon (Hicks, 1932; Sjastaad, 1962).

However, it is worth emphasizing that the decision of an individual to move from his/her region of residence is the outcome of a more complex choice. Alongside strictly macroeconomic determinants, behind migration choices there is a set of factors ranging from observable (age, gender, education...) and unobserved (ability, culture...) personal characteristics. Some of them may be fixed over time while others are time variant. The role of personal characteristics in migration research has been considerably emphasized by the past and more recent literature (Greenwood, 1975, 1985, 1997; Plane and Bitter, 1997; Cushing and Poot, 2004). 'Personal characteristics not only have an important direct effect on migration but they also alter the effect of some regional economic variables on the migration decision.' (Antolin and Bover, 1997, p.230).

Clearly, the profiles of migrant populations vary considerably partially because of the main features of sending and host countries, and in part because of the variety of sources of migration (return migration, migration for political asylum and so on). However, whatever is its source and the areas involved, there are some personal aspects that, even with different magnitude, broadly play an important role in migration decision.

Age is generally found to have a negative effect on moving, reflecting the fact that older individuals face higher moving costs (psychological and economic) and clearly, lower expected future benefits (due to the shorter time horizon). The general conclusion is that, the best decision is to migrate at the youngest possible age, namely, as soon as the potential migrant enters in the labor market (Hartog and Winkelmann, 2003). However, as suggested by Aisa et al. (2014) this conclusion seems to be inconsistent with Italian and Spanish data where migration decision are always taken between the ages of 25 and 29 years. As a matter of fact, the optimal age for migration, is often related to the education attainments.

In the empirical literature there is clear and consistent evidence of a positive correlation between education attainments and the propensity to move. From the Sjaastad's prospective, differential returns to skills in origin and destination regions are main drivers of migration. The expected return to education is to a large extent, determined by migrants' educational background, and how transferable these skills are to the host region labour market. Dustmann and Glitz (2011) noted that 'the decisions about how much education to obtain and whether to migrate are often sequential,

individuals may in many cases make these choices simultaneously, choosing education at home with a view to migrating later' (p. 6).
Moreover, all the set of variables concerning education attainments (like high marks, have a Master or a PhD degree) seem to further increases the migration probability (Coniglio and Prota, 2008; Nifo et al., 2012).
Similarly, the expected effect of gender on the propensity to migrate seems to be changed over the last decades. Traditionally, male individuals have been found to be more likely to migrate (Pissarides and Wadsworth, 198). However, this figure seems no longer typical in more recent data and especially among high skilled individual. Recent empirical studies (Finnie, 2004; Faggian et al., 2007) indicate that women have considerably increased their propensity to move and that they migration probability is higher than their male counterparts. This gender difference is essentially related the economic structure and geographical context of different regional labor markets. However, other household-related factors have to be considered in conjunction with gender. The importance of family ties has been specifically analyzed since the seminal contribution by Mincer (1977). Being married and/or have children are expected to affect both the costs and benefits of moving, and perhaps differently for men and women. The influences of family responsibilities are indeed, generally higher for women than for man.
The influence of cultural factors and family background may be particularly relevant as well.
Individual coming from wealthier backgrounds and with more educated parents are generally found to be more prone to move while, their less well-off counterpart often chose to live in proximity (or in the same household) to their parents (Faini et al., 1997).
Finally, the effect on individuals' own employment situation on the probability to migrate has been often object of study. In particular, researchers have focused on the role of regional unemployment rate differentials on the decision to migrate for unemployed and employed individuals. The general conclusion is that unemployed workers are more likely to move than employed one (DaVanzo, 1978; Pissarides and Wadsworth, 1989) and that, this propensity to move may be, in different ways, restrained by government support and local subsidies (Attanasio and Schioppa, 1991; Antolin and Bover, 1997).

3.3 Data and descriptive statistics

3.3.1 The Dataset: Italian Labour Force Survey

Periodically, ISTAT makes available the official cancellations and registrations between Italian municipality. However, municipal registers data, while offering the official migration flows among Italian regions in a long time perspective and for the whole population, are quite limited from the point of view of socio-demographic information. In order to address this limitation we resort on the information provided by the European Union Labour Force Survey (EU-LFS) for Italy.

The EU-LFS is conducted in the 28 Member States of the European Union, 2 candidate countries and 3 countries of the European Free Trade Association (EFTA). Comparability of the statistics survey between the European countries is considerably high and it has been ensured through various numbers of regulations guaranteeing harmonization of concepts, definitions and methodologies for all EU Member States. The national statistical institutes are responsible for selecting the sample, preparing the questionnaires, conducting the direct interviews, and forwarding the results to EUROSTAT in accordance with the common coding scheme (concepts, definition and classifications).

For Italy this survey has been conducted by ISTAT each year since 1959 and it has been changed many times. The most recent changes in the definitions and design of the survey occurred in 2004[32]. Those changes were principally dictated by the need to achieve complete harmonization with the requirements of EU regulations and full comparability with the estimates and main labour market indicators (such as employment and unemployment rates) provided by other countries, such as those in the OECD area. The restructuring of the survey has led to changes in the definitions of an employed person and a person seeking employment. The abandoning of the use of self-perceived criteria for defining individuals' economic status and ensures more precise and objective observation of this factor, consistent with the principles set out by the *International Labour Organisation* (ILO). Moreover, other relevant changes have been introduced: the increased frequency of the interviews, the use of CATI (*Computer-Assisted Telephone Interviewing*) technique, the implementation of a sophisticated IT system supporting the carrying out of the survey and wider-ranging information content.

Each quarter, the LFS collects information on almost 70,000 households in 1,246 Italian municipalities for a total of 175,000 individuals (representing 1.2% of the overall Italian population). The reference population of the LFS consists of all household members officially resident in Italy, even if temporarily abroad. Households registered as resident in Italy who but live

[32] For a more detailed discussion of the characteristics of the Italian LFS, see Gazzelloni (2006)

abroad and permanent members of collective facilities (hospices, children's homes, religious institutions, barracks, etc.) are excluded.

The Italian LFS sampling design is a two stage sampling design: (municipalities are prime sampling units, PSUs; while households are finite sample units, FSUs[33]) with stratification of PSUs and rotation of FSUs.

In each NUTS-3 domain, PSUs are stratified according to the demographic size. Large municipalities, with population over a given threshold (also called self-representative municipalities-SR), are always included in the sample; smaller municipalities (not self-representative-NSR) are grouped in strata, and then one municipality in each stratum is selected with probability proportional to its population. At the second stage households are randomly selected from the population registers in all the municipalities drawn at the first stage.

For families, a particular rotation system (2-(2)-2) is applied in order to maintain half the sample unchanged in two consecutive quarters and in quarters one year apart. More specifically, households are interviewed during two consecutive quarters. After a two-quarters break, they are again interviewed twice in the corresponding two quarters of the following year. As a result, each household is included in four waves of the survey in a period of 15 months. Conversely, the first stage units (the municipalities) surveyed does not change over time.

Sampling population weights for the Italian LFS are computed in three steps. In the first step, the initial weights are calculated as the inverse of probability of selection; in the second step, non-response adjustment factors are calculated by household characteristic; in the last step, final weights are calculated using a calibration estimator using the auxiliary demographic information regarding the reference population by sex, five-years age groups, nationality and region (NUTS 2 and NUTS 3 level). In order to have consistency between individual and household statistics the sampling weights are computed at household level, which means that each component of the same household has exactly the same final weights of all the others (household weight). Annual weights are computed simply dividing the quarterly weights by four.

[33] Finite Sample Unit (FSU) or finite population correction (FPC) is the proportion of PSU sampled within each stratum (only for sampling without replacement).

3.3.2 *Descriptive statistics of the sample*

Since it is not possible to follow the individuals over time (the identification code of the individual/household is not released for reasons of confidentiality) we use the yearly dataset from the LFS 2012. Thus, the reference population is the labour force (employed and unemployed[34]) in 2012. An individual is classified as a migrant if the individual's current region of residence is different from the region of residence the year before (2011)[35].

As said before, this definition of mobility (i.e. comparing actual region of residence with the previous ones) using LFS is not new in the empirical literature (Mocetti and Porello, 2012; Antolin and Bover, 1997, Paci et al. 2007). More recently, Parenti and Tealdi (2015) follow a similar approach to study the personal determinants of interregional commuting in Italy[36].

In particular, we focus our analysis on the interregional movement excluding the individuals that were abroad the year preceding the interview[37] (13,447 individuals). As we can see from Table 3.1 in 2012, more than 15 thousand people moved from one Italian region to another, thus, the 0.06% of the total labour force. This very low mobility rate which emerges from the LFS is not particularly surprising; similar results are founded for Eastern EU countries (Paci et. al, 2007) and for Spain (Antolin and Bover, 1997).

Table 3.1 Migrant and non migrants, year 2012.

Type of migration	n.	%
Do no move	25,613,290	99.89
Interregional migration	15,612	0.06
International migration	13,447	0.05
Total labour force	25,642,349	100.00

Source: our elaboration on LFS 2012

Table 3.2 shows the matrix between origin (the macro-region of residence one year before the survey) and destination (the macro-region of residence in the reference year) for a total of 15,612 interregional movers in 2012.

[34] We exclude the economically inactive people: individuals who are not in work, but who do not satisfy all the criteria for ILO unemployment (wanting a job, seeking in the last four weeks and available to start in the next two), such as those in retirement and those who are not actively seeking work. For a detailed description of the definitions 'employed', 'unemployed' and 'inactive', see ISTAT (2012).

[35] We recognize and we are aware that, measuring migration phenomenon through LFS data is not completely trustworthy. In fact, the estimation of migration flows through the LFS presents a high level of discrepancy when compared to information from official registers. In general, the LFS substantially underestimates interregional annual flows (Martí and Ródenas, 2007).

[36] They compare the region of residence with the region of work.

[37] Potentially, combing the latter information with the nationality, we may derive a consistent definition for estimating immigration from outside the country.

Table 3.2 Origin and destination matrix between macro-area, year 2012

Region of residence one year before the survey - ORIGIN	Current region of residence - DESTINATION											
	North-West		*North-East*		*Centre*		*South*		*Islands*		*Total*	
	n.	%	n.	%	n.	%	n.	%	n.	%	n.	%
North- West	732	13.7	771	31.5	0	0.0	832	25.5	128	34.7	2462	15.8
North- East	693	12.9	746	30.5	577	13.8	443	13.6	0	0.0	2458	15.7
Centre	959	17.9	388	15.8	1095	26.2	741	22.7	68	18.5	3251	20.8
South	1805	33.7	380	15.5	1803	43.1	1075	32.9	172	46.8	5234	33.5
Islands	1165	21.8	162	6.6	707	16.9	173	5.3	0	0.0	2207	14.1
Total	5353	100.0	2446	100.0	4181	100.0	3264	100.0	368	100.0	15612	100.0

Source: our elaboration on LFS 2012

A detailed origin and destination matrix between the 21 regions is reported in the Appendix (Table A3.1); however, for a better description of these internal migration movements, the Italian regions are grouped into five macro-regions, namely North-West, North-East, Centre, South and Islands[38]. As we can see more than 7 thousands movers come from the South and the Islands. The North of Italy is the preferred destination, with the North-Western regions leading the group. In particular, it is clearly evident the strong orientation toward those regions for Sardinians and Sicilians.

However, it is worth noting that changes in labour demand for industries located in the industrial triangle has partially reduced the supremacy of the North West and, conversely, the attractiveness of the North East regions (Emilia-Romagna in the lead) has noticeably increased, encouraging the movement of labour force from the Western to Eastern regions. It is also interesting to note that in the short-range mobility (across the same macro-region) is relatively high with the exception of the North West where this movement is rather limited and in the Islands where it is completely absent partly because of a general lack of job opportunities.

It is interesting to note that there is a considerable flow of individuals (2,211) from the Northern and Central regions toward the backward Mezzogiorno. This ambiguous pattern can be seen as a return migration. Think, for example, at seasonal or temporary workers that come back to their regions of origin after having worked in another region, or at students that move to a university located outside their region.

In the migration literature the distance between origin and destination regions, acts as a proxy for transportation and information costs and it is expected to exert an adverse effect on individual

[38] The four areas include the following regions: Piemonte, Val D'Aosta, Lombardia and Liguria (North-West); Trentino Alto Adige (Prov. Autonome di Trento and Bolzano), Veneto, Friuli Venezia Giulia and Emilia Romagna (North-East); Toscana, Umbria, Marche and Lazio (Centre); Abruzzo, Molise, Campania, Aquila, Basilicata, Calabria (South), Sicilia and Sardegna (Islands).

mobility. For the sample, (see Table A3.2 in the Appendix) the migration distance is above 250 km for the majority of the movers; in particular the 37% (more than 5 thousands individuals) move for more than 550 km. Table 3.3 provides the sample frequencies of the individual variables distinguishing between migrants and non migrants. As we can observe, it is clear that the female component represents 55% of migrants, while it represents 42% of the non migrant's sample. Migrant labour force is, on average, slightly younger than non migrants. Nearly 50% of migrants were aged between 32 and 46 years old in 2012, while less 42% of the remaining labour force were in that age group. Only 22% of movers were aged more than 47 years comparing to 42% of non migrants.

It is interesting to note that the percentage of graduates among migrants is more than twice that of the whole sample (40% and 18% respectively). This result clearly supports the existence of a positive self selection, i.e. more educated individuals are more likely to relocate. Conversely, the majority of the labour force has a medium education (47%).

Moreover, the descriptive statistics on the family situation of migrants show a less remarkable selection with respect the marital status: singles have a slightly greater propensity to change region of residence comparing to married people. Conversely, widowed or divorced people represent a small percentage of both samples: 8% for the migrants and 9% for the non migrants. The statistics confirm that the presence of children may represent a relevant tie on mobility: only 33% of the migrants sample is represented by couples with children, while the majority lives with other adults but without children. As widely noted in the literature, employment opportunities play a decisive role in the choice of migration of people. In particular, the need to find a job influences the propensity to migrate. It is indeed noteworthy that nearly 30% of migrants were not in employment in the previous year to the transfer of residence. Specifically, 17% of them were unemployed, 8% were students or employed in an unpaid work experience while, the remaining, were fulfilling domestic task or were totally inactive on the labour market (3% and 2% respectively). Conversely, the percentage of none employed among non migrants was almost the half (15%). In particular, the numbers of unemployed drops to 10% while those of students to 2%. Among people who had a job one year before the survey, the large majority (more than 70%) were employees either in the migrants and non migrants sample. As well as, most of them were employed in the service sector followed by industry and agriculture.

Finally, looking at the actual employment status (using ILO definitions) at the time of the interview[39], we can see that the percentage of migrants in an actual state of unemployment is almost

[39] It should be noted that we are excluding economically inactive people at the time of the interview, hence the sample only includes the actual labour force: employed and unemployed.

twice that of the non migrant's sample (24% and 11% respectively). Furthermore, we can see that for this latter sample, the incidence of long-term unemployed is considerably higher: the 52% of unemployed (corresponding to the 6% of the total labour force) was seeking for a job for 12 months or more. This percentage drops for the migrants, only the 18% of them was in a status of long term unemployment.

Table 3.3 Distribution of migrants and non migrants, by selected socio- demographic characteristics, year 2012.

	Migrants		**Non Migrants**	
	obs	%	obs	%
Gender				
Female	8,514	54.5	10,717,830	41.8
Male	7,097	45.5	14,895,464	58.2
Age				
17-31	4,190	26.8	4,095,448	16.0
32-46	7,932	50.8	10,841,972	42.3
47-62	3,053	19.6	9,357,232	36.5
62-76	437	2.8	1,263,029	4.9
over 77	-	-	55,613	0.2
Education				
High	6,299	40.3	4,589,095	17.9
Low	3,567	22.8	9,079,345	35.4
Medium	5,746	36.8	11,944,853	46.6
Family status				
Married	6,752	43.2	14,620,778	57.1
Single	7,630	48.9	8,683,647	33.9
Widowed, divorced	1,230	7.9	2,308,869	9.0
Household composition				
Singles without children	555	3.6	3,136,759	12.2
Singles with children	196	1.3	534,739	2.1
Two or more adults without children	9,582	61.4	10,047,368	39.2
Two or more adults with children	5,279	33.8	15,184,942	59.3
Prior year Labour Force status				
Emplyed	10,831	69.4	21,759,959	85.0
Professional status				
Self-employed	2,074	19.2	5,017,764	23.1
Emplyee	8,757	80.8	16,446,135	75.6
Family worker			296,060	1.4
Economic sector				0.0
Agricolture	191	1.8	819,266	3.8
Construction	229	2.1	1,709,009	7.9
Industry	1,927	17.8	4,430,466	20.4
Service	8,485	78.3	14,801,217	68.0
Unemployed	2,729	17.5	2,791,671	10.9
Student	1,274	8.2	506,212	2.0
In retirement or early retirement	-	-	73,749	0.3
Permanently disabled	-	-	5,645	0.0
Fulfilling domestic tasks	488	3.1	310,876	1.2
Inactive	289	1.9	165,182	0.6
Actual Labour Force status				
Emplyed	11,846	75.9	22,878,884	89.3
Unemployed	3,766	24.1	2,734,410	10.7
For less than 12 months	3,056	19.6	1,296,636	5.1
For more than 12 months	710	4.5	1,437,774	5.6

Source: our elaboration on LFS 2012

The large set of information on LFS dataset allows us to analyze how the status of migrants changes following the transfer of residence. More precisely, we can compare the labour status situation one year before the survey with the current situation. This labour transition matrix is reported in Table 3.4. However, a consideration has to be mentioned: following the LFS user guide, the ILO definitions cannot be applied here, since not all the necessary questions have been asked. Namely, the situation with regard to activity one year before survey is simply based on a personal perception. As a results, the variable "main status one year before the survey" is compared with the variable "current main status" which has exactly the same structure in order to permit these comparisons.

Table 3.4 Labour market transitions of migrants.

Main status one year before the survey	Current main status											
	Employed		Unemployed		Student		Domestic tasks		Inactive		Total	
	n.	%	n.	%	n.	%	n.	%	n.	%	obs	%
Employed	9,578	88.4	778	7.2	122	1.1	-	-	354	3.3	10,831	100.0
Unemployed	848	31.1	1,881	68.9	0	0.0	-	-	0	0.0	2,729	100.0
Student	486	38.2	767	60.2	21	1.7	-	-	0	0.0	1,274	100.0
Domestic tasks	423	86.6	65	13.4	0	0.0	-	-	0	0.0	488	100.0
Inactive	0	0.0	227	78.5	0	0.0	-	-	62	21.5	289	100.0
Total	11,335	72.6	3,718	23.8	143	0.9	-	-	416	2.7	15,612	100.0

Source: our elaboration on LFS 2012

We can observe from Table 3.4 that after one year, more than 80% of the employed were in the same condition, 7% had moved to unemployment, whilst just over 3% had moved to the status of inactivity. Permanence in the status of unemployment is also significantly high (69%), only the 30% find a job after migration. The group of students showed higher dynamics: the majority (60%) moved straight into the state of unemployment, 38% of them started to carry out a job or profession (including unpaid work for a family business or paid traineeship) and only a negligible percentage remained student. It is interesting to note that more than 86% of people fulfilling domestic task in the year before the survey experienced a positive transition into the labour market, 13% were still seeking for a job at the time of the survey and any of them remained in the same position.

Shortly, we may conclude that the two samples differ in many of the observed characteristics and that a process of self selection seems to be in place. In the next Section we will use this valuable information to asses, by an econometric approach, if they effectively affect the probability to migrate. Table A3.3 in the Appendix presents some basic descriptive of the variables used in the estimation.

3.4 Methodology

In migration context, the choice of econometric methodology depends critically on assumptions regarding how individuals make migration decisions. Some authors (for example Finnie, 2004) consider migration as a sequential decision-making process. In other words the decision to move is treated separately from the destination choice. Thus, as the migration decision is the outcome of a dichotomous choice (whether or not to migrate), the natural starting point is to consider a discrete choice model (DCM) such as binary logit or probit models. Alternatively, others (like Davies et al., 2001) suggest that each individual jointly decide if and where to move; thus the decision to move and choice of destination cannot be separated. From an econometric point of view this latter line of reasoning leads to conditional logit models (see Pellegrini and Fotheringham, 2002 for a review in spatial choice model). One of the principal drawbacks of conditional logit models is that the effect of personal characteristics on migration cannot be directly investigated. Since the aim of our study is to investigate the personal characteristics which explain the propensity to migrate, we do not consider a place to place model (such as a conditional logit model) and we rely on a binary logit model.

In what follows we briefly describe random utility model (RUM) theory which represents the basis of DCMs. Moreover, some basic issues and concepts in survey data analysis will be quickly reviewed.

DCMs encompass a wide array of techniques in which the dependent variable is categorical and represents the choice set. In its simplest formulation a DCM includes two choices, normally indicated with values 0 (in our case, if the individual has not changed region of residence) and 1 (if the individual has change region of residence). DCMs can also accommodate larger choice sets[40]. Irrespective of whether the dependent variable is binary or has more than two choices, all the DCMs can be derived in a RUMs framework, as demonstrated by McFadden (1974). In RUM any decision making unit (in our case, the labour force, i.e. employed and unemployed individual) are assumed to be utility maximizers. Each individual, labelled n , faces a choice among J alternatives[41]. The individual obtains a certain level of utility (U_{nj}) from each of the J alternatives which is clearly known to him\her but partially unknown to the analyst. The latter can in fact, observe some attributes of the two alternatives and some personal attributes of the decision maker but there are also a number of unobservable factors that the researcher cannot control for. As a result, the utility function is decomposed as: $U_{nj} = V_{nj} + \varepsilon_{nj} \quad \forall j$. The term V_{nj} is called representative utility (or

[40] In this case, however, it is necessary that the alternatives are mutually exclusive (an individual can chose only one alternative) and exhaustive (all the relevant possible alternatives are included in the set).

systematic component of the utility) while, the term ε_{nj} represents the part of the utility which is unknown by the researcher and therefore it is treats as random.

The probability that decision maker *n* chooses alternative *i* is:

$$
\begin{aligned}
P_{ni} &= Prob\left(U_{ni} > U_{nj} \forall\, j \neq i\right) \\
&= Prob\left(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \;\forall\, j \neq i\right) \\
&= Prob\left(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \;\forall\, j \neq i\right) \qquad [9]
\end{aligned}
$$

Equation [9] represents the probability that each random term $\varepsilon_{nj} - \varepsilon_{ni}$ is below the observed quantity $V_{ni} - V_{nj}$ and it can be rewritten as a multidimensional integral over the density of the unobserved portion of utility:

$$
= \int_{\varepsilon} I\left(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \;\forall \neq i\right) f(\varepsilon_n)\, d\, \varepsilon_n \qquad [10]
$$

where the term $f(\varepsilon_n)$ represents the joint density of the random vector and *I (·)* is the indicator function, equalling 1 when the expression in parentheses is true and 0 otherwise.

Different discrete choice models are obtained from different specifications of the joint density $f(\varepsilon_n)$, that is, from different assumptions about the distribution of the unobserved portion of utility. More specifically, the logit model is obtained by assuming that each ε_{nj} is independently, identically distributed extreme value[42]. The distribution is also called Gumbel and type I extreme value. Under this assumption, the density for each unobserved component of utility is defined as:

$$
f\left(\varepsilon_{nj}\right) = e^{-\varepsilon_{nj}} \qquad [11]
$$

and the cumulative distribution is defined as:

$$
F\left(\varepsilon_{nj}\right) = e^{-e^{-\varepsilon_{nj}}} \qquad [12]
$$

[42] If ε_{nj} is normally distributed with average 0 and variance 1, i.e. it is N (0, 1), then the model belongs to the probit. Probit models have the disadvantage of not allowing a closed form solution and therefore are computationally more onerous.

If ε_{nj} and ε_{ni} are i.i.d. extreme value, then difference between the two extreme value variables is distributed logistic. Using the extreme value distribution for the errors (and hence the logistic distribution for the error differences) means that the unobserved portion of utility for one alternative is unrelated to the unobserved portion of utility for another alternative. Stated equivalently, the unobserved portion of utility is essentially "white noise."

The main advantage of making this assumption is that, it is possible to define a closed form solution for the probability that an individual n will choose an alternative over another (McFadden, 1974). Some algebraic manipulation leads to the following equation for the logit choice probability.

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}} \qquad [13]$$

For computational convenience, and because any function can be closely approximated by a linear function, representative utility V_{ni} is usually assumed to be linear in the parameters. Thus, the logit choice probability becomes:

$$P_{ni} = \frac{e^{\beta\, x_{ni}}}{\sum_{\mathrm{j}} e^{\,\beta x_{nj}}} \qquad [14]$$

In the binary choice situation, the choice probabilities can be expressed in an even more compact form:

$$P_n = \frac{1}{1 + e^{-\beta x_n}} \qquad [15]$$

When simple random sampling (SRS) design are concerned[43], the coefficients of the logit model can be easily estimated by maximum likelihood (ML). The maximum likelihood estimator (MLE) maximizes the logarithm of the likelihood function which is simply the joint probability distribution, treated as a function of the unknown coefficients.

$$l_p(\beta) = \prod_{i=1}^{n} p_i^{y_i}\,(1 - p_i)^{(1-y_i)} \qquad [16]$$

[43] It requires that each element (observation) has an equal probability of being included in the sample and that the list of all population elements is available.

In other words, the MLE chooses the values of the parameters to maximize the probability of drawing the data that are actually observed. MLE is consistent and normally distributed in large samples, and then statistical inference about the logit coefficients based on the MLE proceeds in the same way as inference about the linear regression function coefficient based on the OLS estimator.

It is important to point out that the vast majority of national surveys (like the EU-LFS or EU-SILC) employ complex sample designs and weighting adjustments thus, observations are not selected using a simple random sample. The principal complex design features which affect the analysis of the data[44] are: the sampling weights, the clustering and stratification design. In what follow we provide a brief description of those aspects.

Sampling weights. In sample surveys different observations may have different probabilities of selection. The sampling weight, represent the number of units that the given sampled observation represents in the total population. It is calculated as the inverse of the product of the conditional inclusion probabilities at each stage of sampling,

Clustering. Individuals may not be sampled independently then, collections of individuals (for example, counties, city blocks, or households) may be sampled as a group (defined cluster). The clusters at the first level of sampling are called primary sampling units (PSUs). Moreover, within the clusters there may also be further sub sampling. For example, counties may be sampled, then households and then finally individual within households. A disadvantage generally associated with cluster sampling is that elements from the same cluster are often more homogeneous than elements from different clusters. This results in a positive covariance between elements within a cluster.

Stratification. Separate sub-groups of clusters are often sampled separately. These groups are called *strata*. For example, the counties of a state might be divided into two strata, say, urban counties and rural counties.

Failing to take these factors into account is likely to result in biased point and variance estimation since the observations are no longer independent and as a result, traditional maximum likelihood methods for estimation cannot be used. When dealing with complex sampling design pseudo-maximum likelihood estimation (PMLE) is used (Skinner et al., 1989). Conceptually, PML estimation is a maximum likelihood estimates for the expanded dataset (Archer et al., 2007).

Differently from the traditional ML, the PML function is constructed as the product of the weighted (the sampling weight is defined by w_{ji}) individual contributions to the likelihood, over the m clusters (or PSUs) sampled and n_m observations within the given clusters. The PML function under complex survey design may be written as:

[44] For an extended discussion on complex survey data analysis see Levy and Lemeshow (2008); Scheaffer et al.; (2012); Thompson (2012).

$$l_p(\beta) = \prod_{j=1}^{m} \prod_{i=1}^{n_j} p_{ji}^{y_{ji}\, w_{ji}} \left(1 - p_{ji}\right)^{(1-y_{ji})\, w_{ji}} \quad [17]$$

given the pseudo-likelihood equation [17], the PMLE (pseudo-maximum likelihood estimator) is that value that maximizes the pseudo log-likelihood function.

The survey sampling design may induce correlation among observations, particularly when cluster samples are drawn. To appropriately estimate standard errors associated with model parameters and estimated odds ratios, it is important to account for the sampling design. Statistical software packages (such as STATA, SPSS and SAS) do not implement these procedures as standard, however, a number of useful command are available in order to correctly identify the survey design characteristics and estimate standard errors, confidence intervals, design effects and effective sample sizes. At the same time, currently, for many countries sample design variables are (partially) lacking, inaccurate and/or not very well documented in the dataset.

In our specific case, to appropriately estimate standard errors associated with model parameters we would need more information on the sampling design beyond that we have access to. More precisely, in the Italian LFS sampling design the municipalities are defined as PSUs and are then stratified according to the demographic size. However, information about municipalities is restricted due to various privacy concerns and to maintain respondent anonymity.

Due to this limitation in our analysis the PSUs are defined by default to be the individuals and for this reason the standard errors are likely to be partially distorted, thus, it is advisable to be cautious in interpreting the empirical results.

In the present study, the empirical specification for the probability of observing a migration is formalized on the basis of the following cumulative weighted logistic distribution:

$$Prob(Y_{n,i} = 1|x_n, \gamma_i) = \frac{1}{1 + e^{-[\beta_1 x_n + \beta_2 \gamma_i]}} \quad [18]$$

where $Y_{n,i}$ takes the value of 1 when the individual n living in region i in 2011 change region of residence from one year to another (namely from 2011 to 2012) and 0 otherwise. x_n indicates a vector of individual characteristics (age, education, marital status, education, previous situation and so on). Origin fixed effects, γ_i, are added in order to capture unobservable economic and non economic factors associated with the migration decision, such as amenities. Equation [18] represents our baseline regression. Unfortunately, one of the principal drawback of this model is that it does not allow us to include region-pair-specific variables (like distance and past migration flows) since we would have positive values only for people which actually move.

However, in addition to the base regression, we estimate an extended version of Equation [18]. In order to take into account the effect on regional push factors we added the term (z_{i-m}) which represents a vector of regional economic differentials between the observed levels of the variable at the origin i, as compared to the national average, m. Moreover, we tested the robustness of our results by including interaction terms, $(z_{i-m})x_n$, in order to take into account the different propensity to migrate according to personal characteristics.

This latter extended version is formalized by the following equation:

$$Prob(Y_n = 1|x_n, z_{i-m}, \gamma_i) = \frac{1}{1 + e^{-[\beta_1 x_n + \beta_2 (z_{i-m}) + \beta_3 (z_{i-m}) x_n + \beta_4 \gamma_i]}} \qquad [19]$$

3.5 Econometric Results

In this Section we present the results of the empirical investigation on the effect of a large set of variables on the probability of internal migration for the Italian labour force.

In our analysis an individual is classified as a migrant if the individual's current region of residence is different from the region of residence the year before. In other words, the dependent variable will take value 1 only if the individual has changed region of residence form one year to another. As already discussed in Section 4, the sample selection design cannot be overlooked in the econometric analysis. Indeed, the population weights will be considered by adopting a pseudo-maximum likelihood approach to estimating the logit model. [45]

The estimated model 1 of Table 3.5 represents the basic specification. These estimates represent the relationship between the independent variables and the dependent variable, where the dependent variable is on the logit scale. In other words, they indicate the amount of increase in the predicted log odds of the dependent variable migration that would be predicted by a 1 unit increase in the predictor, holding all other predictors constant.

As we can see, the explanatory variables have been grouped into five areas: individual characteristics, household composition, education, and situation one year before the survey.

The F-adjusted mean residual goodness-of-fit test is reported at the bottom of the different specifications to verify the adequacy of the model taking the sampling weights into account. As it

[45] We also try to estimate the regression assuming a normal rather than a logistic distribution function, thus using a probit model rather than a logit model. The estimated effects with the two models do not significantly differ. All estimations are carried out by using the survey data commands SVY: with STATA 12.

may be observed, the null hypothesis is not rejected with a significance of 5% in all the estimated models. It may thus be concluded that the differences between the observed and fitted values are small and there is no systematic contribution of the differences to the error structure of the model. In other words, the test suggests no evidence of lack of fit (Archer et al., 2007).

3.5.1 Individual attributes

We begin considering the direct effect of the person's characteristics on the probability of migration. As we can see gender seems to be not a significant predictor of migration behaviour. This evidence is in line with the findings from Mocetti and Porello (2012). Traditionally, Italian males have been found to be more likely to migrate (Faini et al. 1997), but, seem that this is no longer typical in recent data, especially due to the increased participation of women in the labour market.

There is evidence that migration attitude decreases with age given the negative and significative sign of the last two age groups variables. According to the literature (Finnie, 2004; Paci, 2007) age is usually found to be negatively related to migration, mainly because, older people have fewer years to reap the benefits of migration and because psychic costs tend to increase with age.

The dummies describing the family structure are interesting as well. Our results do not support the evidence that the probability of moving is negatively related to marital status. Indeed, there is no significant difference in the probability of migration between single (or widowed\separated) and married individual. However, we obtain a strong negative effect of children on the probability of migration, and a negative effect if the observed individual cohabits with his\her parents. This result confirms Mincer (1977) hypothesis of low migration rate for people living with parents or relatives. As suggested by Antolin and Bover (1997) this aspect may be particularly relevant in countries such as Italy (or Spain in their case) where family bounds are strong and welfare state is lacking for formal care of elderly people.

The literature on migration has shown a strong direct link between educational attainment and migration propensity: individuals with different educational background have a different propensity to migrate. The existence of a positive self selection in migration pattern is confirmed by our estimates: an individual with a bachelor’s degree (or master degree) is more likely to change region than one with only a primary education. Thus, the acquisition of higher education in another region does not seem to the main cause of migration decision. The differential returns to skills in origin and destination Italian regions are the main driver of migration.

Conversely, the dummy variable for *medium level*, seems to be not significant for all specifications.

Furthermore, empirical results suggest that previous employment status is a significant predictor of migration behaviour the following period. In the basic model (column 1) the dummy variable *employed* is negative and significant, suggesting that employed individuals are less likely to move compared to not employed individuals. As a matter of fact, the cost of movement for the former is higher: employed migrants must be compensated for a job that they give up in the region of origin. This finding is in line with one of the three hypotheses empirically tested by Pissarides and Wadsworth (1989) using LFS data for UK.

In column 2 of Table 3.5 we include a finer disaggregation of the situation one year before the survey taking as reference group people employed in the construction sector and disaggregating the variable *not employed* distinguish between students, inactive people, individual fulfilling domestic task or unemployed[46]. As we can observe the magnitude of the different parameter estimates (in log-odds units) change from 2.39 for student to 1.78 for unemployed. In order to get a more easily interpretation of those coefficients we can express converted them into odds ratios by exponentiating the coefficient. Thus, in term of odd ratio we can say that a student is 10.97 times more likely to migrate comparing to an employed individual in the construction sector, while an unemployed only 5.93 times. Turing to sectors, we can also observe that at more disaggregated levels of employment sectors, the white collars (employed in service sectors) are relatively more mobile than workers in the construction sector in strict sense[47]. This last finding probably reflects the increased mobility of young and high skilled individual in the public sector as confirmed by Mocetti and Porello (2012). Even with mixed results for the different countries, Paci et al. (2007) find quite similar results.

In column 3 and 4 we check whether the results of the first two basic models are robust with respect to the inclusion of macro regional dummies for the area of origin[48]. The latter have proved to be positive and significant and most of the coefficients do not change in magnitude and significance, excluding those related to the sector of activity. Macro regional fixed effects capture different regional aspects not explained by the explanatory variables and which clearly affect the probability of migration. As we can see, and as we would expect, the estimated coefficients are positive and statistically significant indicating that individual originating from the Mezzogiorno regions have a higher probability to move comparing to those from the North West of Italy (the reference category).

[46] It is worth recalling that, in this case we are refereeing to the situation one year before the survey and we are not restricting the sample to employed and unemployed.

[47] This disaggregation by sector of activity is based on NACE Rev 1 since not available for NACE Rev 2. The sectors are built using the following grouping at 1 digit level: Agriculture: Section A to B; Industry: Section C to E; Construction: Section F; Services: Section G to U.

[48] In order to reduce the number of parameters and to avoid to lose observations we specify 5 macro area fixed effects. As a preliminary examination we also try to include both origin and destination fixed effects however, the computation failed to converge, perhaps due to the large number of parameters to be estimated and the likely correlation between the variables.

Table 3.5 Weighted logit model for the probability of migration (1)

Variables	(1)	(2)	(3)	(4)
Personal carchteristics				
Male	0.277	-0.209	0.314	-0.260
	(0.214)	(0.210)	(0.215)	(0.210)
Aged 17 to 31	0.340	0.289	0.377	0.320
	(0.364)	(0.368)	(0.362)	(0.368)
Age 47 to 61	-0.922***	-0.947***	-0.957***	-0.978***
	(0.293)	(0.297)	(0.294)	(0.300)
Aged over 62	-1.136**	-1.081**	-1.184**	-1.144**
	(0.525)	(0.523)	(0.525)	(0.524)
Household composition				
Children	-0.818***	-0.834***	-0.843***	-0.859***
	(0.294)	(0.298)	(0.298)	(0.302)
Parents	-0.968**	-0.988**	-1.141***	-1.170***
	(0.394)	(0.406)	(0.398)	(0.411)
Single	0.165	0.160	0.273	0.278
	(0.289)	(0.296)	(0.299)	(0.307)
Widowed\separated	0.0787	0.0742	0.167	0.172
	(0.393)	(0.392)	(0.390)	(0.389)
Education				
Medium	0.118	0.0669	0.159	0.105
	(0.270)	(0.267)	(0.271)	(0.267)
High	1.143***	1.046***	1.180***	1.082***
	(0.289)	(0.290)	(0.293)	(0.294)
Situation one year before survey				
Employed	-0.890***	-	-0.725***	-
	(0.240)		(0.244)	
Student		2.395***		2.284***
		(0.786)		(0.780)
Domestic task		2.369***		2.199**
		(0.900)		(0.906)
Unemployed		1.784***		1.557**
		(0.678)		(0.687)
Inactive		2.295**		2.129**
		(1.029)		(1.031)
Sector of activity one year before survey				
Agriculture		0.690		0.468
		(0.970)		(0.975)
Industry		1.115		1.227*
		(0.722)		(0.727)
Services		1.127*		1.076
		(0.660)		(0.661)
Macro-regional dummy (origin)				
North East			0.360	0.354
			(0.345)	(0.345)
Centre			0.617*	0.628*
			(0.336)	(0.343)
South			1.232***	1.266***
			(0.320)	(0.325)
Islands			1.056***	1.098***
			(0.394)	(0.414)
Constant	-6.528***	-8.247***	-7.435***	-8.857***
	(0.585)	(0.672)	(0.488)	(0.720)
Observations	233,309	233,309	233,309	233,309
Population size	25628905.92	25628905.92	25628905.92	25628905.92
F-adjusted mean residual test	1.022	0.55	1.135	0.624
Prob > F	(0.419)	0.838	0.333	0.777

Standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1

*** p<0.01, ** p<0.05, * p<0.1

Note: The baseline group is defined as an individual with the following characteristics: a female, aged between 32 and 46 years, without children and parents living in the same household, married, with a low education, not emplyed (for column 1) or employed in the construction sector (from column 2 to 4), living in the North West in 2011.

3.5.2 Regional features

Once the importance of individual characteristics and regional aspects in migration choice have been assessed, it seems interesting to examine the possible impact of more general conditions, such as those connected with the economic and institutional context. Then, in a next step we test if regional economic differentials will increase the likelihood of migration, controlling for personal characteristics. To this end, following other migration studies (Antolin and Bover, 1997; and Pissarides and Wadsworth, 1989), we include the difference between some economic variables in the region of origin of each individual and the nation as a whole[49]. As we can see in Table 6 we tested for four different indicators[50]. We use two different measures of regional wellbeing. The first is the real gross domestic product (GDP) per capita, a generally accepted measure for income level. It represents a broad indicator of the economic regional development, and often related to the availability of a more large set of public services (health services, transport infrastructures, and so on).

The second is a cost of living index developed by the Bank of Italy which account also for regional housing prices. Housing cost differentials may in fact, reflect local characteristics and, more generally, quality of life differences across regions. In particular, as underline by Massari et al. (2010) housing prices represent the major element of variation, accounting for almost 70% of cost-of-living differences between Northern and Southern Italy. Furthermore we include unemployment rate differential as a measure of labour market efficiency and job opportunities. Moreover, as suggested by the literature (Nifo and Vecchione, 2014), migration choices (especially when high skilled are involved) may be driven not only by the search for favourable socio-economic contexts but also for better institutions. In order to test this last hypothesis we use the regional quality of government index (QoG) developed by the European Commission. The index covers the four dimensions concerned with the major quality characteristics of a national system, namely: voice and accountability, government effectiveness, rule of law, control and corruption.

Empirical results are reported in models 1 to 4 of Table 3.6. As we can see, the inclusion of those variables is relevant and with the expected sign; that is, negative for the two alternative measures of well-begin (GDP and cost of living index) and for the quality of institutions and positive for unemployment rate. In other words, if a person lives in a region with a relatively higher per capita GDP (higher cost of living, or more efficient institutions), he/she will less likely to migrate to

[49] In a preliminary investigation we estimated models including the complete set of regional economic variables at destination and origin both separately and alternatively. As the former turned out to be not significant at conventional levels, we include them only to characterize origin regions. In this way we achieve a more accurate specification for our model.

[50] As the preliminary analysis of the indexes considered show high correlation we include them alternatively.

another region. Conversely, the relative unemployment rate encourages the decision to migrate. These findings are clearly in line with some of the most recent studies like Mocetti and Porello (2010b) and Nifo and Vecchione (2014) and confirm the fact that recent migrants are still responding to economic disparities between regions. However, once we control for structural differences between the regions, including the macro-regional dummies, these effects are completely absorbed.

3.5.3 Predicted probabilities and marginal effects

The estimated parameters of the binary regression models (Tables 3.5, 6 and 7) provide direct useful information about the relationship between the independent variables and the dependent variable. However, because these coefficients are in log-odds units, do not give us any sense of the absolute (or relative) size of the effects[51] on the probability to migrate.

In order to address this issue we explore the predictive margins of the probability of migration from our basic fitted model (reported in the column 1of Table 3.5).

We compare two common approaches for computing adjusted predictions: Adjusted Predictions at the Means (APMs) and Average Adjusted Predictions (AAPs). Computationally, the first approach produce predicted probabilities evaluated at the mean of all the other the covariates. In other words, is the expected probability of a person with average characteristics. For example, the APM for female (0.034%) is the probability of migration, assuming that everyone in the sample is female and all the other variables are fixed at their mean values; respectively the APMs for male assumes that everyone is male and all the other variables are at their mean. Thus, those are not average probabilities but, probabilities evaluated at the average.

Conversely, the second is the average of the probability among actual persons in the data, thus, are the weighted average of the predicted probabilities for each observation in the estimation sample. In this case the AAP for female (0.052%) is the average probability, assuming that everyone in the sample is female, leaving all other independent variable values as they are, while the prediction for male assumes that everyone is male controlling for the distribution of other covariates.

Results, in percentage values, of the two approaches are reported in columns 1 and 2 of Table 3.7. We can see that as long as the predictions at the average of the covariates are technically different from the predictions averages of the predictions, there are not sizeable differences in the predicted effects. However, we must take into account that the first approach (APMs) is often misleading especially when the independent variables are not continuous as in this case. For these reasons,

[51] More precisely, these estimates tell the amount of increase in the predicted log odds of *migration= 1* that would be predicted by a 1 unit increase in the predictor, holding all other predictors constant.

many researchers consider the AAPs more reliable (Williams, 2012). In general, AAPs are slightly higher than APMs but are surprisingly small in both cases.

As said before those measures give the absolute estimate of different group on the probability to migrate, however, researchers in social sciences are usually interested at relative effects, and for this reason we calculate the average marginal effects (or average partial effect) which are simple computed as the difference in probabilities of migration between the group with designated value 1 and the reference group. Average Marginal Effects (AMEs) of independent variables on the probability of migration are presented in column 3 of Table 3.7.

These results confirm that migrant workers tend to be younger, without family ties, highly skilled, unemployed and from the Italian Mezzogiorno. As we can see, there are not significant effects of gender and marital status on migration choices, conversely the probability of migration decreased with age. This negative effect becomes larger with age groups: compared to the young labour force (32-46 years old), people aged between 47 and 61 years and over 62 have lower and significant probabilities of migration (0.08 and 0.09 percentage points respectively). According to the literature, the main source of the negative age effect is likely to be the shortened time period that older people have to reap the benefits of migration. Moreover, as we would expect from the logit model, formal education has a significant, albeit modest, effect on migration outcomes. Compared to workers with no formal or lower secondary education, graduates had slightly higher probabilities of migration (0.09 percentage points). The effect of a high school degree (medium education level) on migration is not statistically significant.

Finally, we can see that the migration difference between employed and not-employed is about 0.05 percentage points. That is a significant, even if not particularly sizable, difference. The marginal effects of regional dummies are significant as well: the migration difference between Southern and West Northern people is about 0.08%.

Table 3.6 Weighted logit model for the probability of migration (2)

Variables	(1)	(2)	(3)	(4)
Personal carchteristics				
Male	-0.245	-0.257	-0.248	-0.262
	(0.209)	(0.211)	(0.209)	(0.212)
Aged 17 to 31	0.304	0.308	0.308	0.301
	(0.366)	(0.366)	(0.365)	(0.366)
Age 47 to 61	-0.966***	-0.973***	-0.971***	-0.973***
	(0.297)	(0.298)	(0.298)	(0.299)
Aged over 62	-1.121**	-1.133**	-1.124**	-1.136**
	(0.523)	(0.523)	(0.524)	(0.523)
Household composition				
Children	-0.850***	-0.856***	-0.850***	-0.859***
	(0.301)	(0.301)	(0.302)	(0.301)
Parents	-1.096***	-1.118***	-1.126***	-1.123***
	(0.409)	(0.410)	(0.411)	(0.408)
Single	0.233	0.246	0.256	0.254
	(0.300)	(0.301)	(0.305)	(0.301)
Widowed\separated	0.144	0.149	0.161	0.148
	(0.389)	(0.387)	(0.388)	(0.389)
Education				
Medium	0.113	0.108	0.107	0.0929
	(0.267)	(0.269)	(0.268)	(0.268)
High	1.094***	1.086***	1.084***	1.066***
	(0.295)	(0.296)	(0.296)	(0.294)
Situation one year before survey				
Student	2.306***	2.289***	2.302***	2.299***
	(0.779)	(0.780)	(0.783)	(0.783)
Domestic task	2.242**	2.206**	2.227**	2.199**
	(0.905)	(0.909)	(0.907)	(0.910)
Unemployed	1.598**	1.572**	1.574**	1.584**
	(0.684)	(0.687)	(0.687)	(0.688)
Inactive	2.196**	2.147**	2.152**	2.152**
	(1.027)	(1.032)	(1.028)	(1.035)
Sector of activity one year before survey				
Agriculture	0.492	0.519	0.478	0.533
	(0.974)	(0.974)	(0.975)	(0.976)
Industry	1.180	1.195*	1.180	1.197*
	(0.724)	(0.723)	(0.723)	(0.721)
Services	1.078	1.065	1.083	1.070
	(0.661)	(0.662)	(0.662)	(0.662)
Regional Economic differential				
Real Gdp pc (Log)	-1.373***			
	(0.417)			
Unemployment Rate	-	0.0979***		
		(0.0292)		
Cost of living			-0.0422***	
			(0.0114)	
QoG			-	-0.0237***
				(0.00657)
Constant	-8.248***	-8.253***	-8.085***	-8.194***
	(0.673)	(0.673)	(0.674)	(0.673)
Observations	233,309	233,309	233,309	233,309
Population size	25628905.92	25628905.92	25628905.92	25628905.92
Prob > F	0.267	0.161	0.213	0.622

Standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1

*** p<0.01, ** p<0.05, * p<0.1

Table 3.7 Adjusted Predictions at the Means, Average Adjusted Predictions and Average Marginal Effects.

Variables	APMs (%)	AAPs (%)	AMEs (%)
Gender			
Female	0.034	0.052	base level
Male	0.046	0.071	0.019
Age			
17-31	0.080	0.115	0.036
32-46	0.055	0.079	base level
47-61	0.021	0.030	-0.048***
over 62	0.017	0.024	-0.055***
Household compostition - Child			
No child	0.052	0.080	base level
At least one child	0.022	0.034	-0.045***
Household compostition - Parents			
No parents	0.049	0.091	base level
At least one parent	0.016	0.029	-0.062***
Marital Status			
Widowed\divorced\separated	0.041	0.063	0.009
Single	0.045	0.070	0.017
Married	0.035	0.053	base level
Level of education			
Low	0.029	0.041	base level
Medium	0.034	0.048	0.007
High	0.094	0.132	0.091***
Situation one year before survey			
Employed	0.071	0.106	-0.054**
Not employed	0.034	0.051	base level
Macro region of origin			
North West	0.022	0.032	base level
North East	0.032	0.046	0.014
Centre	0.041	0.059	0.027*
South	0.076	0.110	0.078***
Islands	0.064	0.092	0.060**

3.6 Robustness analysis

At this point, having found significant evidence of the economic differential variables, in the next step in our analysis we estimate four different specifications of the logit model including some interaction terms between some personal characteristics and the economic variables. Thus, we tested if people with particular characteristics respond differently to the economic incentives. More specifically, in our analysis we consider two personal characteristics: the status of unemployment in the year before the survey and the level of education[52]. We focus on those two personal characteristics mainly because they play an important role in the migration theory. Previous migration studies that have used micro-data confirms the hypothesis that unemployed are particularly sensitive to local unemployment rates (DaVanzo, 1978; Pissarides and Wadsworth, 1989). While some other empirical evidence (Piras, 2009) indicates that differentials of available per capita income between Italian regions seem to affect significantly the decision of graduates to migrate. Others, such as Nifo and Vecchione (2014) focus on the role of institutional quality as one of the main determinants of internal migration decision of high skilled over the more strictly economic variables.

The results are reported in models 1 to 4 of Table 3.8, but unfortunately, they do not offer further support to the relevance of those last interactions. As we can see from model 3 only a single interaction is significant. Namely, interacting cost of living differential with high level of education dummy give significant result. The high skilled people respond more to cost of living differential than low skilled. Clearly there is a quite high correlation between cost of living and per capita GDP. However, the former may better capture the overall standard of living in a given region.

The F-test reported at the bottom of Table 3.8 suggests, in this case as well, no evidence of lack of fit for the four estimated specifications.

[52] This choice is the results of formerly analysis. Those two personal characteristics have been included turned out to be influential to one way to another.

Table 3.8 Weighted logit model for the probability of migration (3)

Variables	(1)	(2)	(3)	(4)
Personal carchteristics				
Male	-0.237	-0.250	-0.235	-0.260
	(0.207)	(0.209)	(0.208)	(0.210)
Aged 17 to 31	0.323	0.323	0.334	0.309
	(0.366)	(0.366)	(0.363)	(0.365)
Age 47 to 61	-0.978***	-0.991***	-0.989***	-0.983***
	(0.298)	(0.300)	(0.299)	(0.301)
Aged over 62	-1.135**	-1.162**	-1.151**	-1.156**
	(0.524)	(0.525)	(0.526)	(0.525)
Household composition				
Children	-0.846***	-0.853***	-0.847***	-0.856***
	(0.301)	(0.301)	(0.302)	(0.301)
Parents	-1.100***	-1.120***	-1.139***	-1.124***
	(0.407)	(0.410)	(0.408)	(0.408)
Single	0.234	0.255	0.258	0.261
	(0.299)	(0.300)	(0.304)	(0.300)
Widowed or separated	0.138	0.149	0.157	0.148
	(0.389)	(0.387)	(0.389)	(0.390)
Education				
Low secondary	0.0911	0.0852	0.0818	0.0819
	(0.266)	(0.269)	(0.268)	(0.269)
Upper secondary	0.999***	0.966***	1.095***	1.019***
	(0.299)	(0.310)	(0.296)	(0.301)
Situation one year before survey				
Student	2.291***	2.252***	2.267***	2.273***
	(0.777)	(0.778)	(0.782)	(0.783)
Domestic task	2.277**	2.221**	2.275**	2.197**
	(0.907)	(0.911)	(0.912)	(0.913)
Unemployed	1.718**	1.802***	1.619**	1.722**
	(0.687)	(0.685)	(0.683)	(0.681)
Inactive	2.188**	2.111**	2.139**	2.124**
	(1.027)	(1.034)	(1.029)	(1.037)
Sector of activity one year before survey				
Agriculture	0.533	0.529	0.541	0.530
	(0.983)	(0.981)	(0.983)	(0.983)
Industry	1.183	1.209*	1.184	1.211*
	(0.725)	(0.725)	(0.723)	(0.722)
Services	1.089	1.072	1.092*	1.074
	(0.662)	(0.662)	(0.663)	(0.663)
Regional Economic differential				
Real Gdp pc (Log)	-1.052*			
	(0.588)			
Unemployment rate		0.0901**		
		(0.0386)		
Cost of living			-0.0277*	
			(0.0156)	
QoG				-0.0239***
				(0.00897)
Interactions				
Real Gdp pc * high education	-1.129			
	(0.822)			
Real Gdp pc * unemployed	0.820			
	(1.047)			
Unemployment rate* unemployed		-0.116		
		(0.0757)		
Unemployment rate * high education		0.0695		
		(0.0584)		
Cost of living*high education			-0.0497**	
			(0.0214)	
Cost of living*unemployed			0.0322	
			(0.0275)	
QoG* high education				-0.00929
				(0.0135)
QoG*unemplyed				0.0216
				(0.0160)
Constant	-8.229***	-8.237***	-8.112***	-8.190***
	(0.672)	(0.674)	(0.678)	(0.673)
Observations	233,309	233,309	233,309	233,309
Population size	25628906	25628906	25628906	25628906
F-adjusted mean residual test	1.201	1.883	0.673	1.327
Prob > F	0.289	0.049	0.735	0.217

Standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1

3.7 Conclusions

After around three decades of theoretical and empirical research on Italian migration, the role of economic differences among regions as one of its most important determinant is not disputed. However, it is worth emphasizing that the decision of an individual to move from his/her region of residence is the outcome of a more complex choice. Alongside strictly economic determinants, personal and family characteristics are important factors that need to be properly taken into consideration. To the best of our knowledge these latter aspects have not been carefully studied mainly because of a lack of suitable micro data. The recent and limited number of studies (Mocetti and Porello, 2010b; Coniglio and Prota, 2008; Di Pietro, 2005) have focused in a small subsample of the population: young graduates. Indeed, the spatial movements of high skilled have attracted the interests of many researchers because of their crucial role in affecting the dynamics of local development. Thus, if from one side the literature on internal mobility based on micro data is rather restricted; on the other hand, there is a fairly substantial literature based on aggregate data and using various analytical approaches.

In this study we tried to fill this potential gap in the migration literature addressing the topic of migration in terms of the question 'who moves?' within a larger stratum of the Italian population, the labour force. Individual data from the Italian LFS are used. The sample includes men and woman, aged 16 to 70, who are in the labour force at the time of the survey (2012); namely more than 25 million individuals. Following other empirical studies (Pissarides and Wadsworth, 1989; Antolin and Bover, 1997; Paci et al., 2007; Parenti and Tealdi, 2015) in order to identify the migrants we exploit the place of residence with the one a year before.

In the econometric analysis, within limits of available information provided, we have accounted for complex survey design, using pseudo maximum likelihood. Weighted logit models have been estimated in order to assess life-cycle and family factors that influence internal migration of Italian labour force. The evidence provided on appeared to be robust to a different set of robustness checks. As a matter of fact, we account for unobservable area specific characteristics, regional economic differentials and interaction terms.

Consistent with the life-cycle model, our findings suggests that age is negatively related to mobility: older individuals face both increased costs (psychological and economic) and lower expected future benefits of moving. Moreover, our results are in line with the argument put forward by Mincer (1977), according to which family ties tend to discourage migration. This may be particular important in Italy where family bonds are strong. In line with our expectations and with other theoretical and empirical studies, having a degree or post-degree qualification is found to increase the probability of migrating.

The empirical evidence provided by this study shows that migration decision is also strongly influenced by the previous labour status. As suggested by the theory, unemployed, inactive, and students have a higher propensity to move comparing with employed. Unable to get proper jobs in their region of origin people out are often forced to move in another region in the hope of a better life and to find an even temporary or seasonal job.

The results also confirm the relevant influence exerted by economic factor such us the GDP, unemployment rate, cost of living and non economic factors such as the quality of life. These findings provide further empirical support for the claim that economic, social, cultural and institutional factors play an important role to attract labour force (especially highly skilled) and thus enhancing regional growth.

Of course, it is necessary to exercise caution when generalising this findings to other situations (different years or different countries). Our analysis represents a specific case study which aims to depict a general profile of a potential migrant, in one of the most difficult years for the Italian economy and the labour market.

On the basis of the evidence found so far, it is our intention to extend the analysis in the future by addressing some limitations of the present study. In particular, we aim to compare our results with LFS data from previous years. The idea is to asses how (and if) the changes to the labour market and the economic crisis have affected the 'selection' of potential internal movers among the labour force. It would be also interesting to explore the factors which affect the location decision of individuals in others European countries with a similar core-periphery structure.

Additionally, conditional on data availability, it would be interesting the extend further the set of explanatory variables, in particular to account for unobserved worker characteristics (like ability or the informal networks) which have been shown important for understanding the selection of migrants (Fernandez-Huertas, 2011; McKenzie et al. 2010; Bartolucci et al., 2014). Past migration networks is a factor which may have an impact on individual mobility. People who have migrated in the past may provide their relatives and friends with valuable information on the labour opportunities and services available in their region of residence and in the contiguous ones.

References

Anderson, J. E. (2011) The Gravity Model. *Annual Review of Economics*, 3, 133-160.

Aisa, R., Cabeza, J., & Larramona, G. (2014). Education and age at migration. *Optimal Control Applications and Methods*, *35* (4), 412-422.

Allison P.D., and Waterman R.P. (2002). Fixed effects negative binomial models. Sociological Methodology, 32: 247–265.

Anderson, J. E., & van Wincoop, E. (2003). Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review*, *93*(1), 170-192.

Antolin, P., & Bover, O. (1997). Regional migration in Spain: the effect of personal characteristics and of unemployment, wage and house price differentials using pooled cross-sections. Oxford Bulletin of Economics and Statistics, 59(2), 215-235.

Archer, K. J., Lemeshow, S., & Hosmer, D. W. (2007). Goodness-of-fit tests for logistic regression models when data are collected using a complex sampling design. Computational Statistics & Data Analysis, 51(9), 4450-4464.

Attanasio, O. P., & Schioppa, F. P. (1991). Regional inequalities, migration and mismatch in Italy, 1960-86. *Mismatch and labour mobility*, 237-324.

Balia, S., Brau, R., & Marrocu, E. (2014). *Free patient mobility is not a free lunch. Lessons from a decentralized NHS* (No. 201409). Centre for North South Economic Research (CRENoS), University of Cagliari and Sassari, Sardinia.

Bartel, A. P., & Koch, M. J. (1991). Internal migration of US immigrants. In *Immigration, trade, and the labor market* (pp. 121-134). University of Chicago Press.

Bartolucci, C., Villosio, C., & Wagner, M. (2014). *Who Migrates and Why? Evidence from Italian Administrative Data*. Working Paper.

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

Belanger, A., & Rogers, A. (1992). The internal migration and spatial redistribution of the foreign-born population in the United States: 1965-70 and 1975-80. *International Migration Review*, 1342-1369.

Belot, M., & Ederveen, S. (2012). Cultural barriers in migration between OECD countries. *Journal of Population Economics*, *25*(3), 1077-1105.

Beugelsdijk S. and T. van Schaik (2005) Social capital and growth in European regions: an empirical test, *European Journal of Political Economy*,39, 301–324

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*(1), 111-131.

Borjas, G. J. (2006). Native internal migration and the labor market impact of immigration. *Journal of Human resources*, *41*(2), 221-258.

Borjas, G. J., Freeman, R. B., Katz, L. F., DiNardo, J., & Abowd, J. M. (1997). How much do immigration and trade affect labor market outcomes?. *Brookings papers on economic activity*, *1997*(1), 1-90.

Borjas, George J., Stephen G. Bronars, and Stephen J. Trejo. 'Self-selection and internal migration in the United States.' *Journal of Urban Economics* 32, no. 2 (1992): 159-185.

Boschma, R. (2005). Proximity and innovation: a critical assessment. *Regional studies*, *39*(1), 61-74.

Bratti, M., & Conti, C. (2014). *The Effect of (Mostly Unskilled) Immigration on the Innovation of Italian Regions*. European Regional Science Association.

Brücker, H., Fachin, S., & Venturini, A. (2009). Do Foreigners Replace Native Immigrants? Evidence from a Panel Cointegration Analysis. *IZA discussion paper 4438*

Brücker, H., Fachin, S., & Venturini, A. (2011). Do foreigners replace native immigrants? A panel cointegration analysis of internal migration in Italy. *Economic Modelling*, *28*(3), 1078-1089.

Brunello, G., Lupi, C., & Ordine, P. (2001). Widening differences in Italian regional unemployment. *Labour Economics*, *8*(1), 103-129.

Cameron, A. C. & Trivedi, P. K. (2005) *Microeconometrics: Methods and Applications,* Cambridge, Cambridge University Press

Cameron, A. Colin, and Pravin K. Trivedi. *Regression analysis of count data*. Vol. 53. Cambridge university press, 2013.

Cannari, L., Nucci, F., & Sestito, P. (2000). Geographic labour mobility and the cost of housing: evidence from Italy. *Applied Economics*, *32*(14), 1899-1906.

Caragliu, A., Del Bo, C., de Groot, H. L., & Linders, G. J. M. (2013). Cultural determinants of migration. *The Annals of Regional Science*, *51*(1), 7-32.

Card, D & DiNardo, J. (2000). Do Immigrant Inflows Lead to Native Outflows?. *American Economic Review*, *90*(2), 360-367.

Card, D. (2001). Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration. *Journal of Labor Economics*, *19*(1), 22-64

Casacchia, O., Reynaud, C., Strozza, S., & Tucci, E. (2010, April). Italians' and Foreigners' Internal Mobility in Italy: an Application of Gravitational Models. In *45th Scientific Meeting of the Italian Statistical Society*.

Chamberlain G. (1982). Multivariate Regression Models for Panel Data, *Journal of Econometrics*, 18: 5-46

Champion, A. G., Fotheringham, A. S., Rees, P., Boyle, P., & Stillwell, J. (1998). *The Determinants of Migration Flows in England: A Review of Existing Data and Evidence: a Report Prepared for the Department of the Environment, Transport and the Regions*. Newcastle upon Tyne: Department of Geography, University of Newcastle upon Tyne.

Ciriaci, D. (2010). University quality, interregional brain drain and spatial inequality. The case of Italy. *LLEE Working Document*, (87).

Clark, X., Hatton, T. J., & Williamson, J. G. (2007). Explaining US immigration, 1971-1998. *The Review of Economics and Statistics*, *89*(2), 359-373.

Coniglio, N. D., & Prota, F. (2008). Human capital accumulation and migration in a peripheral EU region: the case of Basilicata. *Papers in Regional Science*, 87(1), 77-95.

Corrado, L., & Fingleton, B. (2012). WHERE IS THE ECONOMICS IN SPATIAL ECONOMETRICS?*. *Journal of Regional Science*, *52*(2), 210-239.

Curry, Leslie. "A spatial analysis of gravity flows." *Regional Studies* 6, no. 2 (1972): 131-147.

Cushing B., Poot J.(2004). Crossing boundaries and borders: Regional science advances in migration modelling. *Papers in Regional Science*, 83 (1) , 317–338.

DaVanzo, J. (1978). Does unemployment affect migration? Evidence from micro data. The Review of Economics and Statistics, 504-514

Davies, P. S., Greenwood, M. J., & Li, H. (2001). A conditional logit approach to US state-to-state migration. Journal of Regional Science, 41(2), 337-360.

Del Boca, D., & Venturini, A. (2005). European Migration: What Do We Know?.

Dettori, B., Marrocu, E., & Paci, R. (2012). Total factor productivity, intangible assets and spatial dependence in the European regions. *Regional Studies*, *46*(10), 1401-1416.
Devillanova, C. & Garcı`a Fontes, W. (2004) *Migration across Spanish provinces: evidence from the social security records* (1978 1992), Investigationes Economicas, 23, 461 487

Di Pietro, G. (2005). On migration and unemployment: evidence from Italian graduates. *Economic Issues*, *10*(2), 11-28.

Dotti, N. F., Fratesi, U., Lenzi, C., & Percoco, M. (2013). Local labour markets and the interregional mobility of Italian university students. *Spatial Economic Analysis*, *8*(4), 443-468.

Dustmann, C., & Glitz, A. (2011). Migration and education. *Handbook of the Economics of Education*, *4*, 327-439.

Elhorst, J. P. (2014). *Spatial econometrics: from cross-sectional data to spatial panels*. Springer.

Etzo, I. (2008). Internal migration: a review of the literature. *MPRA paper*, *8783*, 1-28.

Etzo, I. (2011). The determinants of the recent interregional migration flows in Italy: a panel data analysis. *Journal of Regional Science*, *51*(5), 948-966.

Faini, R., Galli, G., Gennari, P., & Rossi, F. (1997). An empirical puzzle: falling migration and growing unemployment differentials among Italian regions. *European Economic Review*, *41*(3), 571-579.

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

Finnie, R. (2004). Who moves? A logit model analysis of inter-provincial migration in Canada. *Applied Economics*, *36*(16), 1759-1779.

Florida, R. (2002). The economic geography of talent. *Annals of the Association of American geographers*, *92*(4), 743-755.

Fujita, M., & Thiesse, J.F. (2002) Economics of Agglomeration. Cambridge University Press, Cambridge, MA.

Furceri, D. (2006). Does labour respond to cyclical fluctuations? The case of Italy. *Applied Economics Letters*, *13*(3), 135-139.

Gazzelloni, S. (2006). La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione. 2006. ISTAT - Metodi e Norme n. 32

Glaeser, E. L., Kolko, J., & Saiz, A. (2001). Consumer city. *Journal of economic geography*, *1*(1), 27-50.

Graves, P. E. (1980). Migration and climate*. *Journal of regional Science*, *20*(2), 227-237.

Greenwood, M. J. (1975). Research on internal migration in the United States: a survey. *Journal of Economic Literature*, 397-433.

Greenwood, M. J. (1997). Internal migration in developed countries. Handbook of population and family economics, 1, 647-720.

Greenwood, M.J. (1985). Human migration: theory, models, and empirical studies. Journal of Regional Science, 521 - 544

Guiso, L., Sapienza, P., & Zingales, L. (2004). The Role of Social Capital in Financial Development. *American Economic Review*, *94*(3), 526-556.

Halleck Vega, S., & Elhorst, J. P. (2015). The SLX model. *Journal of Regional Science*.

Harris, J. R., & Todaro, M. P. (1970). Migration, unemployment and development: a two-sector analysis. *The American economic review*, *60*(1), 126-142.

Hartog, J., & Winkelmann, R. (2003). Comparing migrants to non-migrants: The case of Dutch migration to New Zealand. *Journal of Population Economics*, *16*(4), 683-705.

Hausman, J. A., Hall, B. H., & Griliches, Z. (1984). Econometric models for count data with an application to the patents-R&D relationship. *Econometrica*, 52(4), 909-938.

Hicks, J. R. (1932). Theory of Wages, New York, MacMillan.

Hierro, M., & Maza, A. (2010). Foreign-born internal migrants: Are they playing a different role than natives on income convergence in Spain?. *Applied Geography*, *30*(4), 618-628

ISTAT (2010). La popolazione straniera residente in Italia al 1° gennaio 2010. Roma: ISTAT. (Statistiche in breve).

ISTAT (2012). Rilevazione sulle forze di lavoro – dati trasversali trimestrali, 2012. Press release, available on line: http://www.istat.it/it/archivio/10766.

Krugman, P. (1991) Increasing returns and economic geography. *Journal of Political Economy, 9*9: 483-499.

Kulu, H., & Billari, F. (2004). Multilevel analysis of internal migration in a transitional country: the case of Estonia. *Regional Studies,* 38(6), 679-696.

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

Le Sage J.P. and Pace R.K. (2009). *Introduction to Spatial Econometrics*. Boca Raton: CRC.

LeSage, J. P., & Dominguez, M. (2012). The importance of modeling spatial spillovers in public choice analysis. *Public Choice*, *150*(3-4), 525-545.

LeSage, J. P., & Thomas-Agnan, C. (2015). Interpreting Spatial Econometric Origin-Destination Flow Models. *Journal of Regional Science*, *55*(2), 188-208.

Levy, P. S., & Lemeshow, S. (1999). Variance estimation in complex sample surveys. Sampling of Populations: Methods and Applications, Fourth Edition, 367-394.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of monetary economics*, *22*(1), 3-42.

Marinelli, E. (2013). Sub-national graduate mobility and knowledge flows: An exploratory analysis of onward-and return-migrants in Italy. Regional Studies, 47(10), 1618-1633

Marinelli, E. (2013). Sub-national graduate mobility and knowledge flows: An exploratory analysis of onward-and return-migrants in Italy. *Regional Studies*, *47*(10), 1618-1633.

Marrocu, E., & Paci, R. (2013). Different tourists to different destinations. Evidence from spatial interaction models. *Tourism Management*, *39*, 71-83.

Marrocu, E., Paci, R., & Pontis, M. (2012). Intangible capital and firms' productivity. *Industrial and Corporate Change*, *21*(2), 377-402.
Martí, M., & Ródenas, C. (2007). Migration Estimation Based on the Labour Force Survey: An EU-15 Perspective. International Migration Review, 41(1), 101-126.

Massari, R., Pittau, M. G., & Zelli, R. (2010). Does regional cost-of-living reshuffle Italian income distribution?. Available at SSRN 1567345.

Mayda, A. M. (2010). International migration: A panel data analysis of the determinants of bilateral flows. *Journal of Population Economics*, *23*(4), 1249-1274.

Maza, A., & Villaverde, J. (2004). Interregional migration in Spain: a semiparametric analysis. *The Review of Regional Studies*, *34*(2), 156.

Maza, A., Villaverde, J., & Hierro, M. (2013). Explaining the settlement patterns of foreigners in Spain. *Applied Geography*, *40*, 11-20.

McFadden, D. (1974). The measurement of urban travel demand. Journal of public economics, 3(4), 303-328.
Mincer, J. (1977). Family migration decision. *Journal of political economy*, 86: 749-773

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

Mocetti, S., & Porello, C. (2010). How does immigration affect native internal mobility? new evidence from Italy. *Regional Science and Urban Economics*, *40*(6), 427-439.

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

Molloy, R., Smith, C. L., & Wozniak, A. K. (2011). *Internal migration in the United States* (No. w17307). National Bureau of Economic Research.

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

Mundlak Y. (1978). On the pooling of time series and cross-section data, *Econometrica,46:* 69-85

Murat, M., & Paba, S. (2002). Capitale umano specifico e flussi migratori. *Rivista di politica economica*, *92*(2), 63-108.

Muth, R. F. (1971). Migration: chicken or egg?. *Southern Economic Journal*, 295-306.

Newbold, K. B. (2001). Measuring internal migration among the foreign-born: Insights from Canadian data. *The Review of Regional Studies*, *31*(2), 177-195.

Newbold, K. B. (2012). Migration and regional science: opportunities and challenges in a changing environment. *The Annals of Regional Science*, *48*(2), 451-468

Nifo, A., & Vecchione, G. (2014). Do institutions play a role in skilled migration? The case of Italy. *Regional Studies*, *48*(10), 1628-1649.

Nifo, A., Pagnotta, S., & Scalera, D. (2011). *The best and brightest. Selezione positiva e brain drain nelle migrazioni interne italiane*. University Library of Munich, Germany.

Nogle, J. M. (1994). Internal migration for recent immigrants to Canada. *International Migration Review*, 31-48.

Ortega F, Peri G (2009) The causes and effects of international migrations: evidence from OECD countries 1980–2005. NBER working papers 14833, *National Bureau of Economic Research*

Pace, R. K., & Zhu, S. (2012). Separable spatial modeling of spillovers and disturbances. *Journal of Geographical Systems*, *14*(1), 75-90.

Paci, P., Tiongson E.R., Walewski M.., Liwinski J. (2007, January). Internal Labor Mobility in Central Europe and the Baltic Region . *World Bank Publications*

Paldam M. and G.T. Svendsen (2000) An essay on social capital: looking for the fire behind the smoke, *European Journal of Political Economy*, 16, 339–366

Parenti, A., & Tealdi, C. (2015). *Regional Commuting in Italy: Do Temporary Contracts Affect the Decision?*. Dipartimento di Economia e Management (DEM), University of Pisa, Pisa, Italy.

Partridge, M. D., & Rickman, D. S. (2003). The waxing and waning of regional economies: The chicken–egg question of jobs versus people. *Journal of Urban Economics*, *53*(1), 76-97.

Pellegrini, P. A., & Fotheringham, A. S. (2002). Modelling spatial choice: a review and synthesis in a migration context. Progress in Human Geography, 26(4), 487-510.

Piras R (2012a) Internal migration across Italian regions: macroeconomic determinants and accommodating potential for a dualistic economy. *The Manchester School* 80(4): 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(2) 377-396.

Piras R. (2015) A long-run analysis of push and pull factors of internal migration in Italy. Estimation of a gravity model with human capital using homogeneous and heterogeneous approaches. *Papers in Regional Science,* forthcoming.

Piras, R. (2005). Il contenuto di capitale umano dei flussi migratori interregionali: 1980-2002. *Politica economica-Journal of Economic Policy (PEJEP)*, *21*(3), 461-491.

Piras, R. (2009). How does internal migration by educational attainment react to regional unbalances. In Conference on 'Poverty Traps: An Empirical and Theoretical Assessment', Naples, Italy (pp. 30-31).

Pissarides, C. A., & Wadsworth, J. (1989). Unemployment and the inter-regional mobility of labour. *The Economic Journal*, 739-755.

Plane, D. A., & Bitter, C. (1997). The role of migration research in regional science. *Papers in Regional Science,* 76(2), 133–153

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

Putnam, R. D. (1993). *Making democracy work: Civic traditions in modern Italy.* Princeton, NJ: Princeton University Press.

Putnam, R. D. (1995). Bowling alone: America's declining social capital. *Journal of Democracy, 1, 65-78*

Ravenstein, E. G. (1885). The laws of migration. *Journal of the Statistical Society of London*, *48*(2), 167-235.

RitsilÄa, J., & Haapanen, M. (2003). Where do the highly educated migrate? Micro-level evidence from Finland. *International Review of Applied Economics*, *17*(4), 437-448

Rogers, A., & Henning, S. (1999). The internal migration patterns of the foreign-born and native-born populations in the United States: 1975-80 and 1985-90. *International Migration Review*, 403-429.

Saiz, A. (2007). Immigration and housing rents in American cities. *Journal of Urban Economics*, *61*(2), 345-371.

Salvatore, D. (1977). An Econometric Analysis of Internal Migration in Italy. *Journal of Regional Science*, 17(3), 395-408.

Santos Silva, J. S., & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and statistics*, *88*(4), 641-658.

Santos Silva, J. S., & Tenreyro, S. (2011). Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator. *Economics Letters*, *112*(2), 220-222.

Scheaffer, R., Mendenhall III, W., Ott, R., & Gerow, K. (2012). Elementary survey sampling. 7th ed. Boston: Brooks/Cole.

Schündeln, M. (2014). Are immigrants more mobile than natives? Evidence from Germany. *Journal of Regional Science*, *54*(1), 70-95.

Simpson, L., & Finney, N. (2009). Spatial patterns of internal migration: evidence for ethnic groups in Britain. *Population, Space and Place*, *15*(1), 37-56.

Sjaastad, L.A. **(**1962) The costs and returns of human migration. *Journal of Political Economy* 70, 80–93.

Skinner, C. J., Holt, D., & Smith, T. F. (1989). Analysis of complex surveys. John Wiley & Sons

Thompson, S. K.(2012) Sampling. 3rd ed. Hoboken, NJ: Wiley.

Todaro, M. (1980). Internal migration in developing countries: a survey. In *Population and economic change in developing countries* (pp. 361-402). University of Chicago Press.

Venhorst, V., Van Dijk, J., & Van Wissen, L. (2011). An analysis of trends in spatial mobility of Dutch graduates. *Spatial Economic Analysis*, *6*(1), 57-82.

Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. Stata Journal, 12(2), 308.

Appendix

Table A2.1: List of variables

Variable	Mean	Std. Dev.	Min	Max	Definition	Source
Total inter-regional flows	886.65	1311.71	1.00	11324.00	Total residential changes from region of origin i to region od destination j	ISTAT
Foreigners inter-regional flows	118.66	220.63	0.00	2218.00	Residential changes from region of origin i to region od destination j for non-Italian citizens	ISTAT
Italians inter-regional flows	767.98	1157.51	0.00	10685.00	Residential changes from region of origin i to region od destination j for Italian citizens	ISTAT
Population	2938305.00	2345479.00	118879.00	9917714.00	Total resident population in a region (annual average)	ISTAT
Population aged 15-64	1204.22	1008.51	54.50	4505.30	Total resident population in a region aged betweem 15-64 years (thousand, annual average)	EUROSTAT
Total umeployment rate	8.72	4.83	2.50	24.10	Percentage ratio of perople aged 15-74 without work and seeking work and the population of the same age group	ISTAT
GDP pc	23333.27	5858.62	13815.17	33546.66	Regional per capita GDP (euros), constant values (2005)	Elaborazioni SVIMEZ su dati ISTAT
Distance	519.53	271.41	59.98	1240.60	Distance in Km between the centroids of Origin i and Destination j	Own calculations
Tourism attraction rate	8.63	9.26	1.44	43.38	Ratio of bed nights over resident population	ISTAT
Human capital rate	12.82	2.94	6.70	21.00	Percentage ratio of population aged 25-64 with tertiary educational attainment level and population of the same age group	EUROSTAT
Urban public transport	163.19	108.76	50.70	607.70	Urban networks of local public transport in the provincial capitals for 100 square km	ISTAT
Social capital rate	11.64	4.76	4.40	27.70	Percentage ratio of people aged 14 and over who carried out voluntary work in the total population aged 14 and over	ISTAT
Crime rate	21.41	7.91	6.29	42.43	Robberies and thefts per 10,000 inhabitants	ISTAT

Table A3.1: Origin and destination matrix between Italian regions

Region of origin	Region of destination																					Total
	Piemonte	Val D'Aosta	Liguria	Lombardia	Bolzano	Trento	Veneto	Friuli-Venezia Giulia	Emilia-Romagna	Toscana	Umbria	Marche	Lazio	Abruzzo	Molise	Campania	Aquila	Basilicata	Calabria	Sicilia	Sardegna	
Piemonte	0	21	0	253	0	0	0	0	0	0	0	0	0	0	0	0	0	0	253	0	0	527
Val D'Aosta	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Liguria	0	0	0	0	0	0	0	136	0	0	0	0	0	0	0	0	0	0	0	0	0	136
Lombardia	376	5	77	0	0	35	51	88	460	0	0	0	0	0	0	195	154	22	208	106	22	1,799
Bolzano	0	0	0	0	0	39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	39
Trento	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Veneto	39	0	0	354	35	100	0	44	497	0	0	0	0	0	0	0	0	0	0	0	0	1,068
Friuli-Venezia Giulia	41	0	0	0	32	0	0	0	0	68	0	100	0	0	0	0	0	0	0	0	0	241
Emilia-Romagna	162	0	0	97	0	0	0	0	0	51	0	0	357	0	0	139	186	0	119	0	0	1,111
Toscana	0	0	0	398	92	0	0	0	0	0	112	0	254	0	0	106	0	0	41	68	0	1,071
Umbria	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Marche	107	0	0	403	0	0	0	0	168	0	0	0	0	0	0	0	0	0	0	0	0	678
Lazio	0	0	0	52	0	0	128	0	0	573	155	0	0	0	0	298	149	42	106	0	0	1,502
Abruzzo	0	0	0	356	0	0	0	0	0	0	0	0	239	0	53	155	80	0	0	0	0	884
Molise	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Campania	101	15	0	840	0	0	0	0	0	62	0	0	779	91	0	0	0	0	0	0	0	1,889
Aquila	177	0	0	0	0	0	0	0	218	262	0	0	0	86	0	122	0	0	0	172	0	1,038
Basilicata	0	0	0	0	0	0	0	0	0	309	151	0	0	0	0	215	0	0	0	0	0	675
Calabria	124	12	0	179	0	0	0	162	0	0	0	0	0	0	0	0	271	0	0	0	0	749
Sicilia	69	7	0	809	0	0	0	0	162	0	0	125	411	0	0	0	0	0	173	0	0	1,756
Sardegna	187	0	93	0	0	0	0	0	0	0	0	0	171	0	0	0	0	0	0	0	0	451
Total	1,382	61	170	3,740	158	174	179	430	1,505	1,326	419	225	2,212	177	53	1,230	839	64	900	346	22	15,612

Source: LFS data, 2012

Table A3.2**:** Migration distance,

Migration distance	Number of migrants
up to 100 km	213
from 100 km to 250 km	4,552
from 250 km to 400 km	2,896
from 400 km to 550 km	2,177
more than 550 km	5,774
Total	15,612

Source: LFS data, 2012

Table A3.3: List of variables

Variable	Description	Mean	Std. Dev.
Male	1 = male 0 = female	0.572	0.495
Aged 17 - 31	1 = aged between 17 and 31 years old 0 = other	0.148	0.355
Aged 32 - 46	1 = aged between 17 and 31 years old 0 = other	0.397	0.489
Aged 47 - 61	1 = aged between 17 and 31 years old 0 = other	0.393	0.488
Aged over 62	1 = aged over 62years old 0 = other	0.063	0.243
Children	1 = at least a child 0 = no child	0.356	0.479
Parents	1 = mother or father (or both) in the same household 0 = any parents in the same household	0.215	0.411
Single	1 = single 0 = other	0.325	0.468
Widowed\separated	1 = widowed or separated 0 = single or married	0.092	0.289
Low	1 = basic education level 0 = other	0.366	0.482
Medium	1 = upper secondary education level 0 = other	0.461	0.499
High	1 = third education level 0 = other	0.173	0.378
Emplyed	1 = previously employed (one year before the survey) 0 = other	0.850	0.357
Student	1 = previously student (one year before the survey) 0 = other	0.019	0.138
Unemplyed	1 = previously unemployed (one year before the survey) 0 = other	0.107	0.309
Inactive	1 = previously inactive (one year before the survey) 0 = other	0.006	0.080
Domestic task	1 = previously fulfillung domestic task (one year before the survey) 0 = other	0.012	0.111
Agriculture	1 = previously emplyed in agriculture sector (one year before the survey) 0 = other	0.038	0.192
Service	1 = previously emplyed in service sector (one year before the survey) 0 = other	0.576	0.494
Industry	1 = previously emplyed in industry sector (one year before the survey) 0 = other	0.168	0.374
Construction	1 = previously emplyed in construction sector (one year before the survey) 0 = other	0.068	0.252
North East	1 = region of origin in the North East of Italy (one year before the survey) 0 = other	0.225	0.418
North West	1 = region of origin in North West part of Italy (one year before the survey) 0 = other	0.279	0.449
Centre	1 = region of origin in the Central part Italy (one year before the survey) 0 = other	0.179	0.384
South	1 = region of origin in the South of Italy (one year before the survey) 0 = other	0.206	0.405
Islands	1 = region of origin is Sardina or Sicily (one year before the survey) 0 = other	0.110	0.313

Table A3.4**:** List of variables (2)

Variable	Description	Source	Year	Mean	Std. Dev.	Min	Max
GDP per capita (Log)	Gross domestic product per capita, euro, chain-linked values, reference year 2005	ISTAT	2009	-0.007	0.248	-0.437	0.288
Unemployment rate	Percentage ratio of the population aged 15 and over seeking employment to the labour force.	ISTAT	2009	-0.002	3.677	-5.066	7.114
Cost of living	Estimated regional cost-of-living indices including housing costs.	Cannari and Iuzzolino, 2009, Bank of Italy	2006	3.428	9.797	-14.900	14.100
QoG	Eu Quality of Government Index differential. Survey based index. 4 pillars are covered: Rule of Law, Government Effectiveness, Voice & Accountability, Corruption	EU Commission - EUROSTAT	2009-2010	4.604	17.681	-29.030	39.530