Determinants of interregional migration in Italy:A panel data analysis

Ivan Etzo

University of Cagliari

September 2007

Determinants of interregional migration in Italy:

A panel data analysis

Ivan Etzo*

*University of Cagliari

Abstract

After two decades of low internal migration rates, official national statistics report a considerable increase of internal mobility which started in 1996 and still continues to grow at the time of writing. Using panel data analysis on gross migration flows between regions, this study investigates the role of the main economic determinants during the period 1996-2002.. The analysis distinguishes between the role played by the same explanatory variable in the sending region (*push factor*) and in the destination region (*pull factor*). The per capita GDP turns out to be the main economic determinant, showing a strong effect both when it acts as a *push factor* and when it acts as an attractive factor. On the contrary, the effect of the unemployment rate estimates is much stronger in the sending region than in the destination region. Moreover, the standard gravity variables like distance and population size are also significant and with the expected sign.

Key words: Interregional migration, gravity model, panel data.

Index

1. Introduction

Internal migration in Italy was very important during the 1960's, when a considerable number of people were leaving the southern regions in favour of the northern (more developed) regions. This important migration wave started to loose strength during the 1970's and became negligible during the 1980's till all the first half of 1990's. The official national statistics point the 1996 as the year when the new migration wave started to grow again[1].

The last decade of decreasing internal migration rates, however, has been also characterized by the growth in regional disparities between the north and the south of Italy. This is in contrast with the standard economic theory which, in such circumstances, predicts an increase of movements from the poorest to the richest region. That is why the "immobility" of people together with high regional disparities has been called "the empirical puzzle". Its possible explanations will be reviewed in section 3 and will constitute an important foundation for the present study.

The present study, indeed, focuses on the post "empirical puzzle". That is, the aim is to investigate the role played by the standard macro-economic variables as the determinants of the new migration wave. Each explanatory variable affects migration in two different ways: by pushing people to leave the region where they are living, and by attracting (or pulling) them from another region (which becomes the destination when they decide to move). In order to separate this double effect, gross migration flows for each pair of origin-destination region has been used as the dependent variable.

Another important aspect is the double dimension of the data. Using panel data analysis make it possible to take advantage of two types of information, that is, the variance of each observation over time and across individuals (Wooldridge J. M., 2002 and Hsiao C., 2003). The fixed effect vector

[1] According to the last official data, which are available for the 2004, the positive trend has not yet stopped (SVIMEZ 2007).

decomposition estimator (FEVD) improves the efficiency of the estimates for those variables whose within variance is negligible compared with the cross-sectional variance (e.g. population size and population density).

The analysis starts with the estimation of the "gravity model", which studies the effect of the so called "gravity variables", that is population size and distance. The model is then extended in order to include the main economic explanatory variables. In the extended version of the gravity model, population size is replaced by the population density (used as a proxy for social networks).

The study is organized as follows. Section 2 presents a brief literature review on internal migration. Section 3 discuss "the empirical puzzle" and contains some statistics of the main economic variable and the migration flows. Section 4 contains the empirical analysis and its results. Conclusions are presented in Section 5.

2. Internal migration: a brief literature review

The movements of people, both internal and international, has been broadly studied by researchers of social sciences. It is common practice to distinguish the study of people movements within the country borders and the one between different countries. The former is named *internal migration* while the latter is called *international migration*.

This study focuses on migration's determinants for people that leave their region in Italy to move into another Italian region.
During the past two decades, most of the attention shifted to the growing flows of immigrants coming from poor and developing countries and to the outflows of Italian people to other countries.
Growing differences in economic development between poor and developed countries, together with growing population in low developed countries, turning into high poverty levels, thus, forcing people to migrate abroad. In particular Italy, which has always been a country of emigrants, turned into a host country of immigrants coming mostly from Morocco Tunisia and Former Yugoslavia (Strozza, Venturini, 2002). With regards to international outflows from Italy, emigration has been widely studied both to find the main determinants and to assess the emigration of high skilled people which causes the *brain drain* (see Becker,Ichino and Perci, 2003).

The low internal migration rates in Italy throughout the eighties negatively affected the number of studies which analyse the determinants of internal migration. Daveri and Faini (1999) study migration from southern regions of Italy during the period 1970-1989, when internal migration was decreasing, focusing on the choice between internal and international migration. They find that real wages affect internal migration negatively while unemployment rate does not affect migration, although only coefficients for the sending region have been included in their equation.

In a recent study, Fachin (2007) analyses the long-run determinants of internal migration in Italy during the period 1973-1996. His results shed some light on the so called "empirical puzzle", that is the observed low internal migration flows with increasing differentials among regions . The outcomes from panel cointegration tests show that the main determinant of the (low) internal migration has been the income growth in the sending region and confirm the weak effect of the unemployment rate. Like Daveri and Faini (1999), Fachin (2007) considers only the migration from south to northern regions and do not separate between *push* and *pull* factors.

Basile and Causi (2005) study the determinants of net interprovincial migration flows during the period 1991-2001. They differentiated the analysis in two periods: the 1991-1995 period, when internal migration was still low, and the 1996-2001 period characterized by increasing internal migration flows. The results from the two periods show the expected signs for both the unemployment rates and the disposal income, with the latter effect substantially stronger than the former. The effect of income and unemployment rate is higher for the second period than for the first, thus confirming the low response of migration to interprovincial differentials when migration was low.

On the contrary, there is an exhaustive recent literature concerning the study of internal migration determinants in other countries. Low internal migration flows during the eighties characterized not only Italy but all Europe, Nahuis and Parikh (2004) focus on the low labour mobility among European union countries in the presence of large regional disparities during the period 1983-1995. Using panel data on net migration rates they find that, despite low intra Europe migration, both unemployment rate and per capita GDP in the sending region affected migration in the expected way.

In Spain, internal migration flows were very high during the 1960's and first half of 1970's, they were moderate till 1982 when they started to grow again reaching high internal migration rates in 1990's (Maza and Villaverde, 2004a). However, first migration wave responded to regional differentials, whilst the second migration wave showed an "inverse" migration from rich regions with low unemployment rates to poor regions with high unemployment rates. Recent internal migration in Spain has been analysed by Maza (2006) which study the main migration determinants during the period 1995-2002 and find that relative per capita GDP strongly affected migration, while the coefficients for relative unemployment turn out to be low and not significant. Angulo and Mur (2005) carry out a panel data analysis of net interprovincial labour migration during the period 1999-2004. They obtain similar results for different sectors, where both per capita GDP and unemployment rates to affected labour mobility[2].

Parikh and Van Leuvensteijn (2003) examined gross interregional migration flows in Germany after reunification in 1989. They find that internal migration flows during the period 1993-1995 responded to differential in unemployment rates and wages between the sending and the destination region. They also distinguish between wages for white-collar and blue-collar workers, finding a concave relationship between migration for the former and a convex relationship for the latter.

Adrienko and Guriev (2004) study the determinants of interregional migration in Russia. Even though Russia is a country with quite different characteristics compared to Italy, their study is indeed similar to the present analysis. In fact, it is based on the "gravity model" and uses gross migration flows together with the explanatory variables measured separately for the sending and destination region. In this way, their results show different coefficients for the effect that the same explanatory has in the sending region and in the destination region.

[2] They results shows a strong (negative) effect of unemployment rates for migration in the agriculture sector compared to the other sectors (Angulo and Mur,2005).

The above studies are of particular interest for the present analysis because they show how the same socio-economic determinants affect internal migration in different countries. Moreover, all of them are based on panel data models and thus they allow with further comparisons with respect to the different econometric techniques adopted.

3. Internal migration flows in Italy

3.1 Past trends of internal migration flows and the "empirical puzzle"

The intensity of internal migration flows within the borders of Italy has experienced different cycles. They were extremely intense during the fifties and sixties when people from the south were leaving their regions in favour of the industrialised northern regions (Bosco,2003). This big wave of internal migration has been broadly and successfully analysed within the Harris-Todaro (1970) framework. People from rural regions in the south were moving pulled by a great labour demand coming from the big industries in the north. The excess of labour supply in the agricultural sector played also an important role as a *push* factor. The big wave of migrants from South to North continued for twenty years, probably network effects also played a decisive role in the long-run trend.

Internal migration flows started to decrease in the second half of seventies due to the big industries crisis and the consequent fall in labour demand in the northern regions. The descending trend continued throughout the eighties and remained insignificant till the second half of nineties. However, during this period regional disparities in per capita incomes and unemployment rates were still substantially high (Faini et al., 1997). The Italian phenomena of falling internal migration with non decreasing in regional disparities, known as "the empirical puzzle", has been studied by different economists. Different hypothesis have been provided in order to explain the *puzzle*.

A first possible explanation is the decline in wages differentials due to the introduction of the national contract. But if we take into account the unemployment rate, as proxy for the probability to find a job, the growing differentials more than compensate the increase in wages. In fact, when it comes to take the decision of whether to migrate or not, the potential migrants look at the expected future incomes in the origin and destination regions. That is, an even slight difference in real incomes between a pair of regions turns into a substantial difference in presence of big differential in unemployment rates. Moreover, this nominal wage equalization was achieved in the end of sixties, while the internal migration flows started to decrease only ten years later, a lag that can hardly be explained (Faini et al., 1997).

A second possible explanation for the decrease in internal mobility is the increasing costs of housing for emigrants, like transaction's costs and taxes. Empirical results show that differential in house prices discouraged internal migration in Italy (Cannari et al., 2000). However, it's unlikely that this was the main reason of falling internal migration.

A third explanation points out to the increase in disposable income in the southern regions due to strong government and family support (Attanasio and Padoa-Schioppa, 1991). Young potential migrants can rely on family support to finance the cost of waiting while old potential migrants can benefit from different social supports like the increased possibilities to anticipate the retirement. However, one might argue that more disposal income could also help to finance the costs of moving especially in the presence of expectations for growing differential among the southern regions and the rest of Italy. Fachin (2007) supports this explanation using a panel cointegration approach to analyse the long-run determinants of internal migrations during the period 1973-1996.

Faini et al. (1999) show that high household income is associated with great mobility. They argue that the empirical puzzle is the result of the combination of interregional job mismatching and high mobility costs. Job agencies in Itay during that period were only public, they were operating inefficiently under a legal monopoly. Lack of information about the possibility of finding a job in another region means more uncertainty for people who are willing to move but don't know to where apply for it. Moreover, technological progress were changing the labour demand and its main geographical place of origin, shifting from the North-West to the North-East. More qualified and specialised workers were asked instead of generic workers that had been hired in the past decades (Murat and Paba, 2001). Thus, new potential migrants could not rely on the old networks of workers as they did during the 1960-70 period. In fact, it has also been shown that Italy's job searching were based mostly on family and friends networks (Casavola and Sestito, 1993).

3.2 The end of the "empirical puzzle" and the new migration trend

After a long break that lasted for more than two decades, internal migration flows started to grow again in 1996 (SVIMEZ, 2004). In 1998 migration from the South reached 129,000, a level that had not been reached since 1974 (Bonifazi, 2001). This trend is still positive and last official data report a significant flow of migrants from South to Centre-North of Italy, which reached 270,000 units in 2004, a level that has been reached only during the sixties (SVIMEZ, 2007)[3].

Why internal migration flows started to rise again? Despite the literature is still scanty, it is possible to give some general explanations. Firstly, the scarce labour mobility during the same period affected not only Italy but also internal mobility in Europe it is reported to be substantially low (Eurostat,2003, Nahuis and Parikh, 2004). Secondly, following the different explanations to the "empirical puzzle" reviewed above it is worth to point out some structural changes that might have boost internal migration flows. Government support to the southern regions of Italy shrank considerably during the nineties due to the big effort that Italy was asked to make in order to join the EMU. Furthermore, another important structural change is the exponential growth of the ICT sector. The increasing use of internet and

[3] These data refer to gross migration flows and take into account both permanent and temporary migration. Permanent migration refers to the change of address in the population Registry and measured in 120,000 units, while temporary migration refers to people that don't change their official residence, that is commuters, they are estimated in 150,000 units. Commuters are reported to be very young (80% of them is under 45 years old) and to turn into permanent migrants after a certain period.

the World Wide Web, makes it possible for a job seeker to search for a company, to visit its web site, to apply for a job, without moving from his/her house. Job agencies today are private and operate through their web sites. The level and the quality of information grew considerably during the nineties with the support of new technologies. This two structural changes together led to a decrease in family disposable income and a better matching between labour demand and supply. The higher information level is fundamental for potential migrants to assess the real differences (in income, in unemployment rates, cost of life, etc..) between their region and the potential destination. In fact, contrary to information about the region of origin, gathering information about the possible destination regions is much more easier today than it was during the period when internal migration flows were substantially low. Attractive factors of a potential destination region need to be known for them to act as (*pull factor*) determinants of migration.

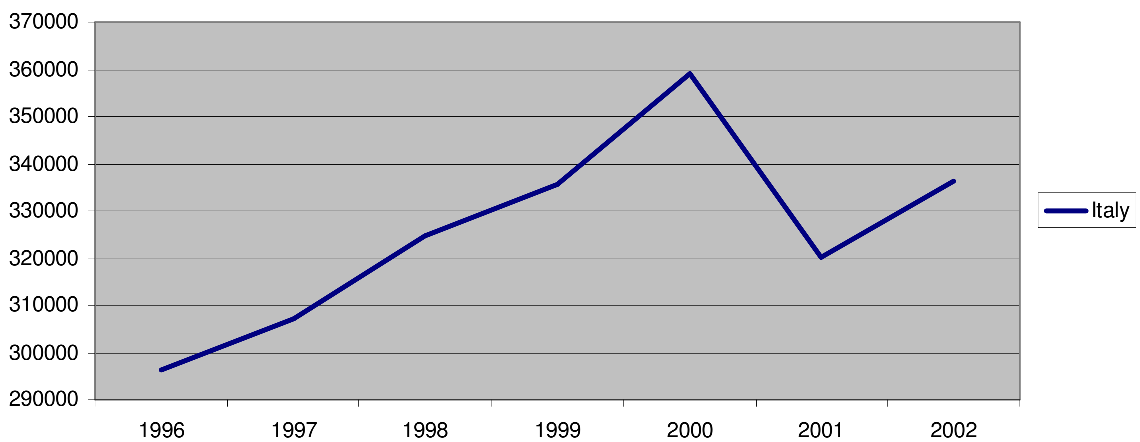

Fig. 3.1 Gross interregional migration flows (ISTAT).

The Fig.3.1 shows the positive trend of interregional migration flows. These data (source: ISTAT) are taken from the municipality population registries, they provide the number of people that, during a year, moved from one region to another one[4].

4 The data measure, for each year, the cancellations from a municipality registry in one region (origin). That is, the number of people who left a region to move in another Italian region.

It is worth to discuss the differences between the actual and the past internal migration flows. In fact, while the main direction of the flows has not changed (from South to Centre-North) its composition appears quite different. During the period 1950-1970 migrants leaving the southern regions of Italy were very young and with a low education level. The young migrants today are five years older in average, between 24 and 29 years old. It is interesting to note their high education level, in 2004 almost half (49.4%) of migrants from southern regions had an high education level (SVIMEZ, 2007). Piras (2005) measured the human capital content of migrants and shows that the southern regions of Italy have been losing human capital during the period 1980-2002[5]. The high level of human capital seems to characterize not only internal migration but also international migration, with Italian northern regions that have been increasingly lost talents during the Nineties (Becker et al., 2003)[6].

3.3 Regional disparities and interregional migration flows

The aim of this study is to focus on the role of the main economic variables as determinants of the new interregional migration trend. Therefore, it is worth to show and discuss the trend of the main macroeconomic variables, that is the unemployment rate and the GDP per capita, during the period 1996-2002. Other studies have already shown that regional disparities were not decreasing during the period 1970-1995, that is when internal migration flows were indeed decreasing (see Fachin, 2007).

In Fig.3.2 it is shown the GDP per capita with the regions aggregated in four repartitions which are very similar to the NUTS I classification[7]. The figure shows that the gap between the South and the Centre-North regions is

5 This aspect bears important implications: a net loss in human capital seriously affects regional growth rates and the convergence process.
6 Compared with the other EU countries Italy is experiencing a “brain drain” instead of a “brain exchange”.
7 The Nomenclature of Territorial Units for Statistics (NUTS) established by EUROSTAT. The classification NUTS I for Italy does not include Sicily and Sardinia in the South.

persistent throughout the period. It is also interesting to show the difference in GDP per capita between the South and the Centre-North.

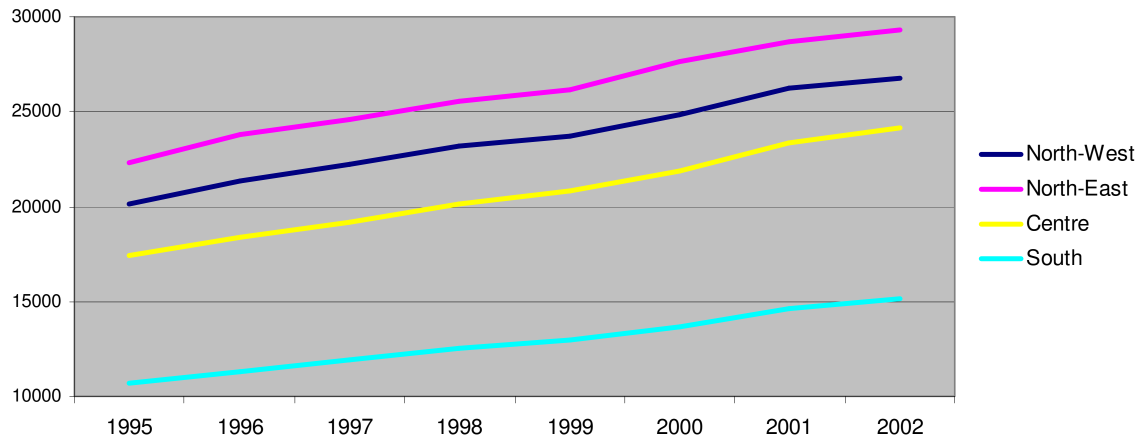

Fig. 3.2 GDP per head at market prices (ISTAT)

The Fig.3.3 shows that during the increase in internal flows the gap between the Centre-North and the South of Italy was also increasing . This fact is not surprising if we consider that internal movements are mostly from South to Centre-North with an average net lost of roughly 60,000 units between 1996 and 2002 (SVIMEZ, 2007) and that this net lost is also a net loss in human capital. The demand for high skilled workers in the South is not enough to cover the increasing supply coming from young high skilled unemployed.

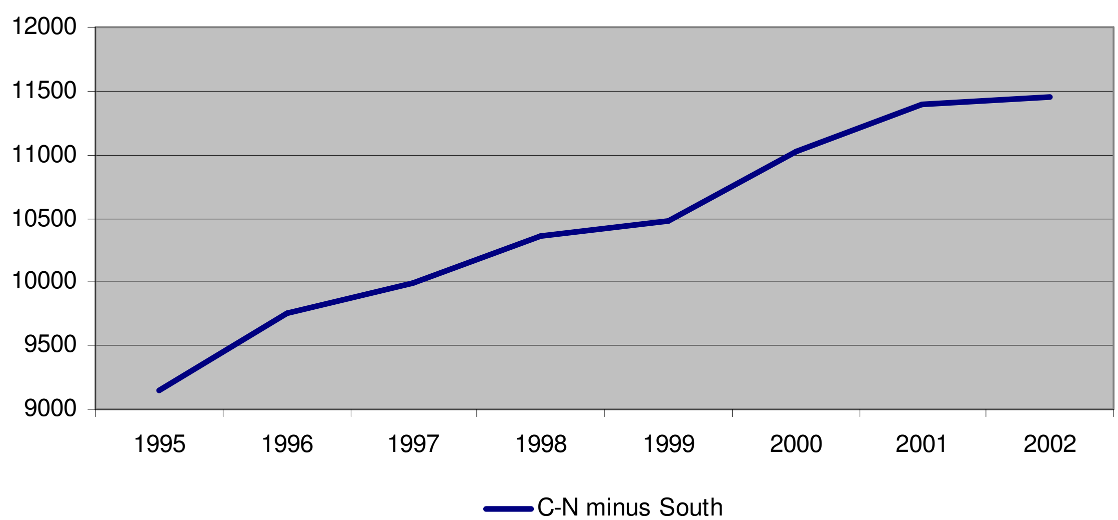

Fig.3.3 GDP per capita: Centre-North minus South (x 1000)

The other variable that is expected to affect the internal migration is the unemployment rate. Fig.3.4 shows that the northern-east regions of Italy have the lowest unemployment rate, followed by the northern-west regions and the centre. It is possible to notice the convergence process that started in 1995 among the region in the Centre-North. The difference between the North-East and the Centre slowed down from 3.6 points in 1995 to 2.9 points in 2002. On the contrary, the southern regions did not experienced the same decreasing trend, the unemployment rate increased from 1995 (18.1%) till 1999 (19.7%) and then started to decrease till 2002 (16.4%).

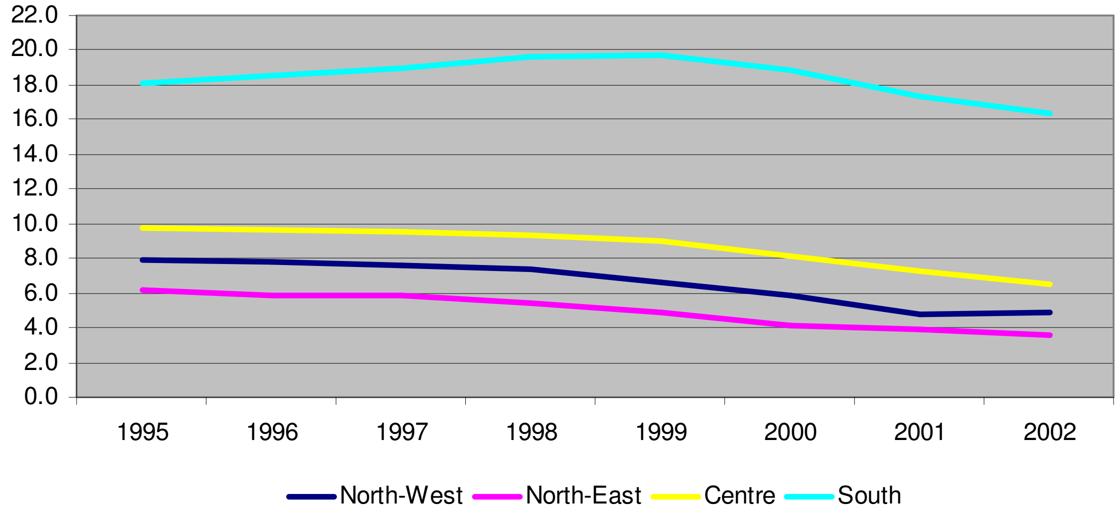

Fig. 3.4 Unemployment rates

The Fig.3.5 shows, in fact, that the gap in the unemployment rates between the South and the Centre-North worsened from 1995 till 1999. In 2000 the difference in the unemployment rates between southern regions and the Centre-North started a decreasing trend[8].

[8] The gap was about 10.2 percentage points in 1995, 12.9 in 2000 when the gap started to decrease till the last available data in 2006 when the gap was 7.8 percentage pints (source: ISTAT).

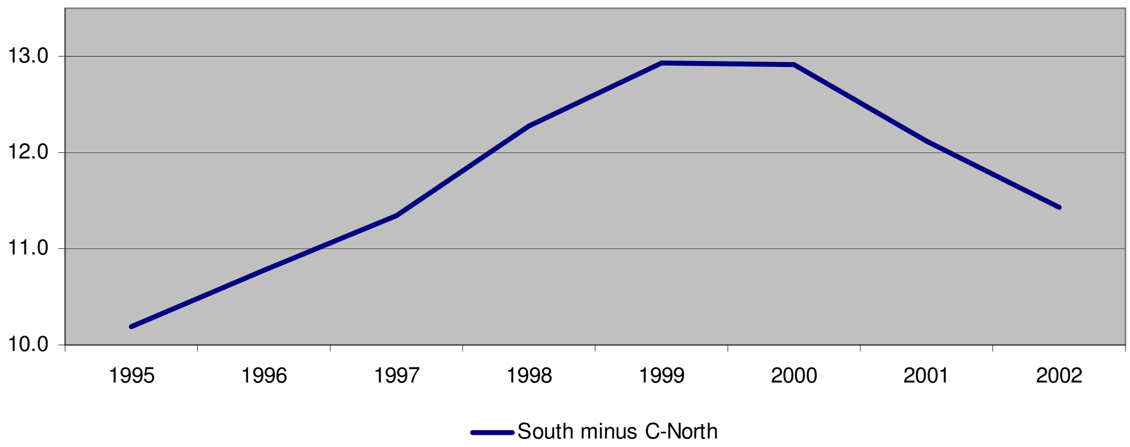

Fig. 3.5 Difference between South and Centre-North unemployment rate

4.The model and results

4.1 The Gravity Model

The gravity model is based on the well known Newton's law about universal gravitation, which states that the attractive force between two bodies is directly related to their size and inversely related to the distance between them (Newton, 1687)[9]. In the basic formulation adapted to migration study (e.g., Lowry 1966), the number of people M_{ij} moving from region *i* to region *j* depends positively to the population size in each region P_i, P_j and negatively to the distance between the two regions D_{ij}. That is:

(4.1) $$M_{ij} = g \frac{P_i^{\alpha} P_j^{\beta}}{D_{ij}^{\gamma}}$$

where *g* is a constant. In the migration contest the parameters are to be estimated, thus they do not have the restricted values as they have in Newton's theory, where $\alpha, \beta = 1$ and $\gamma = 2$.

The Gravity model is characterized by the presence of distance as a key factor. Despite the debate on how to measure distance, previous studies in migration have already shown that it plays an important role (e.g., Greenwood, 1985). The omission of distance or spatial structure in general may seriously affect every empirical study. Distance, in fact, is commonly used as a proxy to measure and capture all the psychic costs that cannot be measured but that surely affect migration flows.

The basic form of the gravitational model can be extended to include economic variables that, together with the population size and distance, may affect migration. Lowry (1966) introduced the wages and unemployment rates to assess for the role played by these economic variables. The basic form of the Lowry model is the following:

---

9 *Philosophiae Naturalis Principia Mathematica.*

(4.2) $M_{ij} = k\left[\frac{u_i}{u_j} \cdot \frac{w_j}{w_i} \cdot \frac{L_i L_j}{D_{ij}}\right] e_{ij}$

Where M_{ij} measure the flows of migrants from region *i* to region *j*, *k* is a gravitational constant, *u* is the unemployment rate, *w* is the hourly wage in manufacturing sector, *L* is the labour force, *D* is the (airline) distance between the two regions and *e* is the error term.

In this formulation the unemployment rate and the wage level play two different roles. An increase (decrease) in the value of one of these variables in the region of origin/destination, relative to the value of the same variable in the destination/origin region, can discourage (encourage) migration. That is, they may act as *push* factors when their value encourage people to leave their country of origin or they may act as *pull* factors in the destination region when they attract people from other regions.

4.2 Estimating the gravity model

The basic version of the gravity model can be generalised to include all the exogenous *push* and *pull* factors, therefore equation (.1) becomes

(4.3) $M_{ij} = k^{\beta_0} \frac{P_i^{\beta_1} P_j^{\beta_2}}{D_{ij}^{\beta_5}} \frac{X_j^{\beta_3}}{X_i^{\beta_4}}$

where X_i may includes all the possible exogenous variables for the origin region that may act as *push* factors for migration, while X_j may includes all the exogenous variables that may attract (*pull*) migrants in the destination region *j*.

The further step is to transform the (4.3) into a linear form in order to obtain an equation that is estimable with the appropriate econometric techniques. Taking the logs of both sides of equation (4.3) yields

(.4) $\ln M_{ij} = \beta_0 \ln k + \beta_1 \ln P_i + \beta_2 \ln P_j + \beta_3 \ln X_j + \beta_4 \ln X_i + \beta_5 \ln D_{ij}$

The equation (.4) is the extended version of the gravity model. This model has been widely applied in migration literature to study the determinants of migration flows in different countries (e.g., Greenwood, 1997).

Population and distance are called *gravity variables* and so characterize equation (4.4) as a gravity model. In gravity models population (both of origin and destination regions) is expected to affects positively migration while distance should discourage migration. The empirical literature widely confirms these expectations but shows also that these two variables can enter in the model in different ways.

The inclusion of population in the model is important to take into account the increase in migration flows which results merely from an increase in population size. That is, the more one region is populate, the higher will be the probability that more people decide to migrate. The population "size effect" can enter in the model in two different ways. The first one is to use it as a weight of migration flows. In this case, the two variables P_i *and* P_j will appear in the left hand side of (.4) and the independent variable will be expressed as the ratio of migration to population. The second one is to leave the population size of both the origin and destination regions as explanatory variables. An advantage of the latter is that there will be less parameter restrictions (J. Fry, T.R.L. Fry and M.W. Peter, 1999). Furthermore, another advantage of including population among the regressors is that the estimated coefficients, β_1 *and* β_2 will provide explicitly information about how differently population affects migration in the origin and in the destination region.

With regards to distance, physical distance can be measured for every pair of regions by the aerial distance between the main cities, by road distance (expressed in km), by the train distance and so on. A considerable number of studies use the physical distance as a proxy to take into account those costs that are (directly or indirectly) related to the distance and might affect migration decisions, like transportation costs, information costs, and psychological costs. However, other studies argue that physical distance does not take into account other important costs (e.g. time of moving or other social costs) that might not be necessary proportional to the physic distance. Particular attention has been given, for example, to the difference between boundaries and not boundaries regions. It has been proved that, all things being equal, boundaries regions have more migration flows between them (citation). Using a dummy variable for regions that share a border can account for differences in migration flows with neighbouring regions. However, the influence of boundaries regions when studying internal migration is expected to be significant for countries with a large territory, while it is likely to be less important for small countries (K. Kumo, 2006)[10]. Itay does not have a wide territory, so in this study distance will be used as an exhaustive proxy variable.

4.3 Panel data approach

The aim of this paper is to study the main determinants of interregional migration in Italy during the period 1996-2002, with a particular focus on the economics and gravitational variables. The study concerns a phenomena that has two dimensions: one involves the different individuals (i.e., regions), and the other one involves the temporal dimension. That is, people move between different regions in the same period and between the same pair of regions in different periods. Within this framework, the natural candidate for an empirical study is therefore a panel data analysis which main characteristic is indeed its double dimension, that is the cross-sectional and time series dimension. For every covariate, the estimated coefficient will result from a variation of data across regions and over the period (i.e., years).

Moreover, although this study focuses on the economic determinants of migration, there are other several aspects that may affect the decision to move from a region and the choice of another specific region among all the

[10] This observation is strengthened when migration data come from municipality population registry. In small countries is more likely that people moving to a neighbour region will not change their official registration. Moreover, people that find a job in a neighbour region are more likely to become commuters in a small country than in a big one.

others. The idea (or assumption) is that all of these *omitted* variables do not significantly affect migration individually but they do it together. Panel data analysis allows to control for the effect of omitted variables, thus reducing the bias of the estimation.

The general form of panel data is the following

(4.4) $y_{it} = \alpha + \beta x_{it} + c_i + \varepsilon_{it}$

Where *i=1,2,…,N* refers to the cross-sectional unit, *t=1,2,…,T* refers to the time period, y_{it} is the dependent variable, α is the constant term, x_{it} is the covariate, β is the coefficient to be estimated and ε_{it} is the error term. The term c_i is the *unobserved effect*, it captures all the unobserved characteristics that vary between individuals but are constant over time. In the interregional migration contest the term c_i may captures different regional propensities to migrate which are related to the region culture, or other characteristics like those related to environmental aspects that might affects migration. The benefit is the possibility to focus on the role played by certain specific variables (in our case economics variables) without losing information from omitted variables.

4.4 Data and descriptive statistics

All the data are obtained from ISTAT.

Lnmig= natural log of gross migration flows from region *i* to region *j*;
Lngdp= natural log of per capita GDP in the origin (lnogdp) and in the destination region (lndgdp);
Lnunr= natural log of regional unemployment rate in region *i* (lnounr) and in region *j* (lndunr);
Lnyunr= natural log of regional unemployment rate for young people (between 15 and 24 years old) in region *i* (lnoyunr) and in region *j* (lndyunr);
Lnpop= natural log of regional population
Lndist= natural log of aerial distance in km between the main city in the sending region and the main city in the destination region.
Lndens= natural log of population density.

Table 4.1. shows the descriptive statistics of the variables. Each region is a sending region and a receiving region at the same time so statistics are reported for only for each variable.

Variable	Obs	Mean	Std. Dev.	Min	Max
lnmig	2660	5.489743	1.62833	0	10.73309
lngdp	2660	9.77504	.2801457	9.182845	10.23085
lnunr	2660	2.272403	.5335669	1.144223	3.197039
lnyunr	2660	3.234327	.4865321	2.158715	3.972365
lnpop	2660	14.44237	1.05541	11.67048	16.01646
lndens	2660	4.972992	.6376476	3.583519	6.042633
lndist	2660	5.943098	.6080883	4.356709	6.979145

Tab. 4.1 Descriptive statistics of the variables

Table 4.2. reports descriptive statistics with decomposition in between and within standard deviation.

Variable		Mean	Std. Dev.	Min	Max	Observations
lnmig	overall	5.489743	1.62833	0	10.73309	N = 2660
	between		1.609892	.7158051	8.835129	n = 380
	within		.2560391	3.799775	9.896946	T = 7
lndpilp	overall	9.77504	.2801457	9.182845	10.23085	N = 2660
	between		.2655184	9.335746	10.11315	n = 380
	within		.0902257	9.602678	9.949805	T = 7
lndunr	overall	2.272403	.5335669	1.144223	3.197039	N = 2660
	between		.5166986	1.411364	3.153704	n = 380
	within		.1353463	1.872488	2.511619	T = 7
lndtdg	overall	3.234327	.4865321	2.158715	3.972365	N = 2660
	between		.4585402	2.507965	3.940361	n = 380
	within		.1640999	2.697181	3.621488	T = 7
lndpop	overall	14.44237	1.05541	11.67048	16.01646	N = 2660
	between		1.056583	11.68212	16.00729	n = 380
	within		.0065698	14.42405	14.46131	T = 7
lnodens	overall	4.972992	.6376476	3.583519	6.042633	N = 2660
	between		.6383259	3.591347	6.041274	n = 380
	within		.0073431	4.955496	4.992725	T = 7
lndist	overall	5.943098	.6080883	4.356709	6.979145	N = 2660
	between		.6087755	4.356709	6.979145	n = 380
	within		0	5.943098	5.943098	T = 7

Tab. 4.2 Between and within descriptive statistics of the variables.

4.5 Empirical Results

4.5.1 The basic gravity model

First step of this study is to estimate the basic gravity model, which is expressed by the following equation

(4.5) $M_{ijt} = \beta_o + \beta_1 Opop_{it} + \beta_2 Dpop_{jt} + dist_{ij} + c_i + u_{ijt}$

Equation (4.5) is in log-log format, that is all variables are expressed in logarithmic. Panel data analysis offers three main techniques to estimate equation (4.5), the fixed effects model, the random effects model and the between effects model. The choice between the random effect model and the fixed effect model depends on the different assumptions requested. Random effects models treat the term c_i as random, not related to the individual *i*, thus included in the (composite) error $v_{it} = c_i + u_{it}$ for which is requested the assumption of independency. For this reason they are called *error component models* and for they estimates to be consistent the regressors have not to be correlated with the error component c_i[11]. Fixed effects estimation, on the contrary, assumes c_i to be deterministic, thus correlated with the regressors. The former uses a GLS (or FGLS) estimation, whereas fixed effects model is based on the *within estimation*, namely each observation is within the individual *i* throughout the period. Despite fixed effect estimation is widely used, one of its drawback is that it fails to give an estimation of those variables that are time-invariant. The third panel data technique is the *between estimation* which focuses only on the cross-section dimension.

The estimation results for the interregional migration gravity model are reported in table 4.1. Estimates are from three different estimators, namely pooled OLS, random effects and fixed effects regression with vector decomposition (FEVD). The latter, is a three stage panel fixed effects vector decomposition model which allows for the estimation of time invariant

[11] The Hausman' specification test is used to test the consistency of the random effect model. If the assumption $Cov(x_{it},c_i)=0$ is satisfied, the random effects estimation is consistent and more efficient than the fixed effects estimation.

variables[12] (see T. Plümper and V. E. Troeger, 2007). The different gravity variables have all the expected signs and coefficients are very similar[13]. It is worth to notice here that long distance between the origin and destination regions seems to play an important role despite distances between Italian regions are much less important compared with internal migration in bigger countries (e.g., Russia, Andrienko and Guriev, 2004). This result emphasises the role of distance as a proxy to control for other aspects concerning regional differences, but that cannot be measured.

	OLS pooled		***Random Effects***		***FEVD***	
lnmig	***Coef.***	***Std. Err.***	***Coef.***	***Std. Err.***	***Coef.***	***Std. Err.***
Lndpop	.9552098	.0135659 ***	.9560363	.0338689 ***	.9281413	.0004482 ***
Lnopop	1.01799	.0135659 ***	1.023108	.0338689 ***	1.025151	.0005085 ***
Lndist	-.3233018	.0235158 ***	-.3231335	.0587596 ***	-.3327224	.0005296 ***
_cons	-21.08651	.320085 ***	-21.17337	.799259 ***	-20.989	.0110687 ***
R-sq (adj)	0.7951		n.c.		0.8153	.9716226
Obs	2660		2660		2660	
θ[14]			.85			
H-test (p-val)			0.1697			

Tab. 4.1 Notes: *=significant at 10%, **=significant at 5%, ***=significant at 1%. All the variables have been treated as (almost) time invariant in the FEVD estimation.

4.5.1 The extended gravity model

The extended version of the gravity model includes the main macroeconomic variables, namely the unemployment rates and the per capita GDP for both

[12] The standard fixed effects (or within effect) estimation cannot be used here because gravity variables are time invariant (e.g., distance) or almost time invariant (e.g., population size).Their within variance is zero or quite small, thus very little information is provided in a within estimation framework.
[13] In the FEVD estimation population has been treated as an almost time invariant explanatory variable in order to take into account its high between variance component.
[14] The parameter θ is the weight of the between variance in the GLS estimation. When $\theta \to 1$ the random effects coefficients estimates approach to OLS , conversely they approach to the fixed effect estimation when $\theta \to 0$.

the origin and the destination region. The following equation is to be estimated

(4.6)

$$M_{ijt} = \beta_o + \beta_1 Odens_{it-1} + \beta_2 Ddens_{jt-1} + \beta_3 Ounr_{it-1} + \beta_4 Dunr_{jt-1} + \beta_5 Ogdp_{it-1} + \beta_6 Dgdp_{jt-1} + c_i + u_{ijt}$$

where all variables are in logarithms and all explanatory variables are lagged one year to avoid simultaneity with the dependent variable. Population has been replaced by population density, which is used as a proxy for social networks[15]. In order to choose the most appropriate technique for the estimate, it is important to take into account the different variability composition of each covariate. In fact, they are of two types: variables with balanced cross-sectional and time series variability (per capita GDP and unemployment rate), variables where cross-section dimension is dominant (distance and population density)[16].

The particular specification of the model is imposed by the nature of the phenomena which is the object of the study. Migration decision is indeed composite, it involves the decision of whether (and when) to move and the choice of a destination. Different types of variables can capture different aspects depending on their different variability over time or over groups. For example, a variable with a relative high cross section variability is likely to affect more the decision of where to migrate, while a variable with adequate within variability is likely to affect more the decision of whether to migrate or not in a certain period.

Random effect estimation captures both types of variances but are often inconsistent due to the failure in rejecting the Hausman test[17]. In the latter case, fixed effect are still consistent but their efficiency can turn out to be

[15] Population density can also serve as proxy for the size of the public services system (Ibarra and Soloaga, 2005).

[16] A third type of variables, which is not present in this study, are those with dominant time series variability, these variables have a zero or almost zero cross-sectional dimension (e.g., the national price index)

[17] The restrictive assumption for random effects estimation, that is $cov(x_{it},c_i)=0$ is not necessary for the fixed effects estimation.

seriously affected by the presence of variables with very little within variables (T. Plümper and V. E. Troeger, 2007)[18].

Table 4.2 presents the results for fixed effect (within estimation), the fixed effect (FEVD) and the random effect estimation (when the Hausman test is not rejected). Results show that the coefficients estimated have the expected sign for all the covariates. Thus, all things being equal, during the period 1996-2002 (and for the same sending region) emigrants increased when the per capita GDP decreased (increased) and when the unemployment rate increased (decreased). From a cross section point of view, migrants preferred to move in regions with low unemployment rates and high per capita GDP. Yet, *ceteris paribus*, they moved to regions that were relatively close to the sending region. Furthermore, the higher coefficients of the per capita GDP with respect to the unemployment rates, both in the sending and in the destination region, show that the former has been the main economic determinant of interregional migration flows. On the contrary, unemployment rate seems to have played a more important role in the sending region, that is "*push*ing" people to migrate, rather than in the sending region, that is as "*pull*" factor.

The other gravity variable, namely population density, seems to have affected interregional migration in the same positively way. Thus, migration has been higher in region with higher population density, which can also be observed as the positive role played by social networks in fostering interregional movements. Moreover, more population density in a region with a high GDP level means also more (and better) public services and amenities.

Another specification of the model has been tested using the unemployment rate of young people, however, results are quite similar with the previous specification.

---

[18] They show that FEVD estimates of almost time invariant covariates are more reliable than fixed effects model' estimates (less RMSE). They also show that FEVD estimates of time varying covariates are more reliable than random effects model' estimates.

lnmig	*FE within (robust se)*		*FEVD (robust se)*		*Random Effects (robust se)*	
	I	**II**	**I**	**II**	**I**	**II**
Lndgdp	.6071	.5565	.6071 ***	.5565 ***	.4278 **	.3280 *
Lnogdp	-.5492	-.6744	-.5492 ***	-.6744 ***	-.4942 ***	-.5035 ***
Lnounr	.1793 ***	-	.1265 ***	-	.1127 **	-
Lndunr	-.0518	-	-.0119 **	-	-.0955 *	-
Lnoyunr	-	.1167 **	-	.0724 ***	-	.1015 **
Lndyunr	-	-.1547 **	-	-.0291 **	-	-.1719 **
Lndist	-	-	-.1405 ***	-.1387 ***	-.1296 **	-.1256 *
Lnddens	2.793 ***	2.798 **	1.388 ***	1.389 ***	1.398 ***	1.403 ***
Lnodens	3.724 ***	2.727 **	1.496 ***	1.493 ***	1.518 ***	1.511 ***
_cons	-27.77 ***	-20.71 **	-8.848 ***	-7.00 ***	-7.63 ***	-6.31 ***
Adj R-sq	-	-	.9716	.9717	-	-
H-Test					$\chi = 0.16$	$\chi = 1.55$
(p-value)					(0.999)	(0.956)
θ[19]					0.89	0.89

Tab. 4.1

Notes: *=significant at 10%, **=significant at 5%, ***=significant at 1%.

Lnunr, lnyunr, lndist and lndens have been treated as (almost) time invariant in the FEVD estimation. Robust standard errors control for cross-sectional heteroschedasticity.

[19] The parameter θ is the weight of the between variance in the GLS estimation. When $\theta \to 1$ the random effects coefficients estimates approach to OLS , conversely they approach to the fixed effect estimation when $\theta \to 0$.

5. Conclusions

Internal migration rates in Italy started to grow again in 1996 after two decades characterized by high regional disparities and negligible interregional movement of people. This study investigated the role played by the main economic variables through the estimation of an "extended gravity model".

The results have been obtained from panel data fixed effects analysis of gross migration flows across the twenty Italian regions. Contrary to what was observed during the 1980's, the main macroeconomic variables, along with population density and distance, seem to be important determinants of the last internal migration wave. In particular, per capita GDP played a strong role both in the sending region and in the destination. This outcome is supported by the statistics on net migration which highlight that the northern (rich) regions are gaining population while the southern (poor) regions are losing population. On the contrary, the effect of unemployment on regional migration appear to be stronger in the sending region (i.e., as a *push* factor) than in the destination region. The last result, however, might reflects two aspects: the disparities among the southern regions and the northern regions along with the within variance component of each observation. That is, if the unemployment rate differentials are high enough, people leaving in regions with very high unemployment rates may be still willing to migrate in another region with a lower but increasing unemployment rate[20].

[20] This is the case of Italy where the unemployment differentials between the South and the Centre-North are high and persistent (see section 3)

References

[01]Adolfo Maza & José Villaverde, 2004. "Migratory flows in Spain: a nonparametric and semiparametric approach", ERSA conference papers ersa04p50, European Regional Science Association.

[02] Andrienko, Yuri; Guriev, Sergei. Determinants of Interregional Mobility in Russia: Evidence from Panel Data. [Journal Article] Economics of Transition. Vol. 12 (1). p 1-27. 2004.

[03] Ana Angulo & Jesus Mur, 2005. "Geographical Labour Mobility In Spain - A Panel Data Approach," ERSA conference papers ersa05p247, European Regional Science Association.

[04] Attanasio, Orazio P; Padoa Schioppa, Fiorella. Regional Inequalities, Migration and Mismatch in Italy, 1960-86. [Collective Volume Article] Mismatch and labour mobility. Padoa Schioppa, Fiorella, ed., Cambridge; New York and Melbourne: Cambridge University Press. p 237-320. 1991.

[05] Becker, S.O., Ichino, A. e Peri, G. (2004), How large is the brain drain from Italy? ,in Giornale degli Economisti e Annali di Economia, vol. 63, pp. 1-32.

[06] Bonifazi, Corrado, International and internal migration in Italy. Context and perspectives of Italian migration flows. Growth in immigration and recover in inter-regional migration, «Demotrends», n. 2, 2001, p. 2.

[07] Cannari, Luigi; Nucci, Francesco; Sestito, Paolo. Geographic Labour Mobility and the Cost of Housing: Evidence from Italy. [Journal Article] Applied Economics. Vol. 32 (14). p 1899-1906. November 2000.

[08] Casavola P, Gavosto A, Sestito P (1995), Salari e Mercato del Lavoro Locale, Banca d.Italia, Servizio Studi Banca d.Italia.

[09] Daveri, Francesco; Faini, Riccardo. Where Do Migrants Go? [Journal Article] Oxford Economic Papers. Vol. 51 (4). p 595-622. October 1999.

[10] Fachin, Stefano. Long-Run Trends in Internal Migrations in Italy: A Study in Panel Cointegration with Dependent Units. [Journal Article] Journal of Applied Econometrics. Vol. 22 (2). p 401-28. March 2007.

[11] Faini, Riccardo & Galli, Giampaolo & Gennari, Pietro & Rossi, Fulvio, 1997."An empirical puzzle: Falling migration and growing unemployment differentials among Italian regions," European Economic Review, Elsevier, vol. 41(3-5), pages 571-579, April.

[12] Greene, W H (2003) Econometric Analysis, 3rd Edition, Prentice Hall International Edition.

[13] Greenwood, M. J. (1985) Human migration: theory, models, and empirical studies, Journal of Regional Science, 25, 521–44.

[14] Harris J.R. and Todaro M.P. (1970), "Migration, Unemployment and Development: A Two Sector Analysis", American Economic Review, 60, 126-142.

[15] Hsiao, Cheng. Analysis of panel data. [Book] Second edition. Econometric Society Monographs, no. 34. Cambridge; New York and Melbourne: Cambridge University Press. p xiv, 366. 2003.

[16] Ibarra G.L. and Soloaga I. (2005), Determinants of migration in Mexico: a spatial econometrics approach, mimeo Universidad de las Amèricas, Puebla.

[17] ISTAT, Movimento migratorio della popolazione residente. Iscrizioni e cancellazioni anagrafiche. Roma, ISTAT, (anni vari, a).

[18] ISTAT, Forze di lavoro, Roma, ISTAT, (anni vari, b).

[19] Luigi Bosco, "Divari Nord-Sud e migrazione interna",Sviluppo Locale, 2003.

[20] Lowry, I.1966. Migration and metropolitan growth: two analytical models. San Francisco: Chandler.

[21] Maza, Adolfo. Migrations and Regional Convergence: The Case of Spain. Jahrbuch fur Regionalwissenschaft/Review of Regional Research. Vol. 26 (2). p 191-202. 2006.

[22] Murat e Paba (2001), "Flussi migratori e modelli di sviluppo industriale. L'esperienza italiana dal dopoguerra agli anni novanta". Mimeo.

[23] Parikh, Ashok; Van Leuvensteijn, Michiel. Internal Migration in Regions of Germany: A Panel Data Analysis. [Journal Article] Applied Economics Quarterly. Vol. 49 (2). p 173-92. 2003.

[24] Piras, Romano, "Il contenuto di capitale umano dei flussi migratori interregionali: 1980-20022", Politica economica,volume: 21, (2005).

[25] Piras, Romano, Un'analisi dei flussi migratori interregionali dei laureati: 1980-1999, «Rivista Economica del Mezzogiorno», vol. XIX, 2005, pp. 129-162.

[26] Plümper, T. and V. Troeger, 2007, Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects, Political Analysis 15, 124-139

[27] Richard Nahuis & Ashok Parikh, 2004."Factor Mobility and Regional Disparities: East, West, Home's Best?". CPB Discussion Papers 4, CPB Netherlands Bureau for Economic Policy Analysis.

[28] Roberto Basile & Marco Causi, 2005."Le determinanti dei flussi migratori nelle province italiane: 1991-2001", Departmental Working Papers of Economics - University 'Roma Tre' 0049, Department of Economics - University Roma Tre.

[29] SVIMEZ, "Report on the economy of southern Italy", SVIMEZ annual reports, 2004.

[30] SVIMEZ, "Rapporto SVIMEZ 2007 sull'economia del Mezzogiorno", SVIMEZ 2007.

[31] Wooldridge, Jeffrey M. "Econometric analysis of cross section and panel data." [Book] Cambridge and London: MIT Press. p xxi, 752. 2002.