The determinants of the recent interregional migration flows in Italy: A panel data analysis

Etzo, Ivan

University of Cagliari

June 2010

The determinants of the recent interregional migration flows in Italy: A panel data analysis

Ivan Etzo

Abstract

The present study investigates the determinants of interregional migration flows in Italy in the light of the upsurge occurred in 1996, after two decades of decreasing internal migration rates. We apply *the fixed effect vector decomposition estimator* (FEVD) on a gravity model using bilateral migration flows for the period 1996-2005 and show that it improves the estimates with respect to the traditional panel data estimators. We find that omitting distance and in presence of rarely time invariant covariates (e.g., population and income) the standard panel data models significantly bias the estimates. The overall economic level and the probability to find a job (proxied by per capita GDP and unemployment rate) appear to be the key variables whose changes are able to push flows of migrants away from their regions and to direct them to "better off" destinations. We find that migrants leaving the regions in the Centre-North respond differently to the push and pull forces with respect to southern migrants. We then estimate a dynamic model and find evidence for the presence of social networks which in our model take place between each pair of regions.

Keywords: Interregional migration, gravity model, panel data, FEVD.

JEL Classification: R23, J61, O15.

1. Introduction

Over the second half of the last century, internal migration in Italy experienced two main migration trends. The first one, which dominated the fifties and the sixties, was characterized by intense migration flows, mainly from the South to the North of Italy. Large differences in unemployment rates and wages between the North and the South seemed to have been the main migration determinants (Salvatore, 1977). The second one was marked by a dramatic decrease of internal mobility and lasted from the first half of seventies until the first half of nineties. Differentials in income and unemployment, however, did not decrease during this second trend. Falling migration rates despite the presence of strong regional disparities, which is known as "the empirical puzzle" (e.g., Faini *et al.*, 1997, Cannari *et al.*, 2000), persisted until the second half of nineties.

In 1996, official statistics report a considerable upturn in internal migration flows. Is this the end of the empirical puzzle? Are the recent migrants responding to economic disparities between regions? Furceri (2006) finds that, during the period 1985-2001, net immigration responded to regional income differences but not to the unemployment rate differences. Basile and Causi (2007) split their analysis into two periods finding that the macroeconomic determinants have a stronger effect in the second one (1996-2000) rather than in the first one (1991-1995). Our study gives a further contribute that differs from the previous ones in that we use gross bilateral migration flows and estimate a spatial interaction model, as it is strongly suggested by the relevant literature (Greenwood, 1985, Cushing and Poot, 2004). As a result, the analysis is able to assess the impact of distance and to identify the *push* and *pull* factors. Furthermore, we estimate a dynamic specification of the model in order to detect the presence of region-to-region network effects.

We apply panel data analysis in order to take advantage of the double dimension of each observation (i.e. yearly bilateral flows). However, due to the restrictive assumption of the random effect model, only the fixed effect estimator could be used with our data. A common problem that arise when applying the FE estimator to gravity models is that the within transformation does not give the estimates for the time invariant variables (Hsiao 2003, Wooldridge 2002). We overcome this limitation by applying the *fixed effects vector decomposition estimator* (FEVD), a recent panel data technique that allows to estimate the time invariant variables (Plümper and Troeger, 2007). This estimator has also the advantage to improve the estimation efficiency for those variables with a relative low within variance, like the GDP and the population size.

Our aim is, therefore, to provide an empirical investigation on the determinants of the recent migration wave in Italy, using for the first time a spatial interaction model and a recent panel data estimator. The results, in fact, show that distance is an important determinant and that migrants respond differently to the same variable in the region of origin rather than in the destination one. The role of pushing and pulling factors and the impact of distance is emphasized in a further analysis which focuses on the migration between the Centre-North and the South as separate sub-samples.

The paper is organized as follows. Section 2 describes the (past and recent) trends of internal migration in Italy and the interregional disparities. Section 3 introduces the theoretical framework. In section 4 we discuss the econometric technique. In Section 5 we show and discuss the empirical results. Section 6 concludes.

2. Internal migration flows in Italy

2.1 Past trends of internal migration flows and the "empirical puzzle".

After WWII, internal migration in Italy experienced two main cycles. The first one, which dominated the fifties and the sixties, was characterized by extremely intense migration flows. During this period, a considerable stream of people moved mainly from the Southern regions in favour of the more industrialised Northern regions[1]. These intense flows of internal migration can be explained within the rural-to-urban model developed by Harris and Todaro (1970). Accordingly, a considerable increase in labour demand from the big industries, located mostly in the North-West, triggered the migration of people working in the rural Southern regions. The excess of labour supply in the agricultural sector played also an important role as a push factor. This big wave of migrants from Southern to Northern regions continued for twenty years, probably reinforced by social networks, which might have played a decisive role in the long-run trend.

During the first half of seventies internal migration flows started to decrease. The oil crisis, which involved the big industries in the North-West, caused a reduction of the labor demand. The so called "industrial triangle" (i.e., Turin-Genoa-Milan) was, in fact, the main recipient for migrants during the first cycle. The second cycle was characterized by a persistent negative trend of internal migration flows, which continued throughout the eighties until the second half of nineties. During that period the North-West switched from a positive to a negative net migration rate, whilst on the contrary, the Centre and the North-East became the main destinations for internal migrants

[1] In 1961, 240 thousand people moved from the Southern regions to the Centre-North (Bosco, 2003).

(Attanasio and Padoa Schioppa, 1991). In addition, contrary to what had characterized the first cycle, since the mid of Seventies emigration rates seemed not to be strongly correlated with unemployment and income differentials. Although the cost of living and real wages differential changed slightly in favour of Southern regions, during this period regional disparities in per capita incomes and unemployment rates were still substantially high (Faini et al., 1997). The main feature of the second cycle of internal migration flows is, therefore, the mismatch between internal migration and regional disparities. The failure of traditional theory to explain such a phenomenon, which is known as "the empirical puzzle", attracted the interest of many researchers. The main explanations provided by different studies can be summarised as follows.

A possible explanation is the decline in wages differentials between the North and the South due to the introduction of the Labor Contract. However, migrants are supposed to consider the "expected" wage, and for it to decrease also unemployment differentials have to decrease, making the migration a less attractive option to increase the future net expected income. On the contrary the unemployment rates' differentials were not small enough to compensate the increase in real wages.

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

A further explanation points out to the increase in disposable income in the Southern regions favoured by the strong government and family support (Attanasio and Padoa-Schioppa, 1991). Young potential migrants could rely on family support to finance the cost of waiting while old potential migrants could benefit from social supports like the increased possibilities to anticipate the retirement. Fachin (2007) provides empirical supports to this hypothesis using a panel cointegration approach to analyse the long-run determinants of internal migrations during the period 1973-1996. 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.

Faini et al. (1997), in fact, 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 Italy during that period were only public, moreover, they were operating inefficiently under a legal monopoly. Lack of information about the possibility of finding a job in another region brings uncertainty and, consequently, higher migration costs for people willing to move. In addition, technological progress was changing the labour demand and its main geographical place of origin, which shifted from the North-West to the North-East. More qualified and specialised workers were needed in place of generic workers that had been hired in the previous two decades (Murat and Paba, 2001). As a result, new potential migrants could not rely on the old family networks as they did in the past[2].

2.2 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, as reported by the Italian National Institute of Statistics

[2] Casavola and Sestito (1995) point out, in fact, that job searching activities in Italy were based mostly on family and friends networks.

(Fig. 1)[3]. In 1998, the solely out migration from the South to the rest of Italy reached 129 thousand units, a level that had not been reached since 1974 (Bonifazi, 2001).

Fig.1

Fig.2

The Fig.2 shows the positive trend of interregional migration flows during the period chosen for the empirical analysis (i.e. 1996-2005). These data (source: ISTAT), which are collected by each municipality in their official population registries, provide the number of Italian citizens that, during a year, cancelled their residency in their region of origin to move to another region[4].

Another important aspect concerns the differences between the actual migration and the past migration flows. In fact, while the main direction of the flows has not changed (i.e., from South to Centre-North), its composition reveals some relevant changes in terms of age and educational attainment of migrants. During the period 1950-1970 migrants leaving the Southern regions of Italy were very young and with low educational attainment. The young migrants today are more skilled and five years older (on average) than migrants during the sixties.[5] Piras (2005) measured the human capital content of migrants showing that the Southern regions of Italy lost human capital during the period 1980-2002[6].

[3] See also the annual report of SVIMEZ (2004).

[4] The data refer to Italian citizens aged 14 and more and do not include the number of commuters.

[5] They are reported to be between 24 and 29 years old and in 2004 almost half of migrants (49.4%) from Southern regions held a high educational attainment (SVIMEZ, 2007)

[6] This aspect bears important implications: a net loss in human capital might affect regional growth rates and the convergence process.

2.3 Regional disparities and interregional migration flows

Other studies have already shown that regional disparities were not decreasing during the period 1970-1995, when internal migration flows were, indeed, decreasing (see Fachin, 2007). We now show and discuss the trends of the two main macroeconomic indicators, namely, unemployment rate and per capita GDP, during the period 1996-2005.

Since the unification in 1861, Italy has experienced the dualism between the South (also called "Mezzogiorno") and the North. Fig. 3 shows the real per capita GDP trends for the four macro areas[7]. It is clear that the "historical" gap between the South and the Centre-North is still persistent. The distances between the three richer areas are substantially constant, with the North-East regions leading the group.

Focusing on the dualism between the South and the Centre-North, the Fig. 4 shows the real per capita GDP in the South as a percentage of the real per capita GDP in the Centre-North. Despite the evident convergence process, the differential remains remarkable and the real per capita GDP in the South is under the 60% of the real per capita GDP in the Centre-North.

Fig.3

Fig.4

The other macroeconomic indicator, which is expected to affect internal migration, is the unemployment rate. As Fig. 5 shows, the gap between the "Mezzogiorno" and the Centre-North is also relevant with respect to the unemployment rate differential. The North-Eastern regions of Italy have the lowest unemployment rates, followed by the

[7] The four areas include the following regions: Piemonte, Val D'Aosta, Lombardia and Liguria (North-West); Trentino A Adige, Veneto, Friuli V. Giulia and Emilia Romagna (North-East); Toscana, Umbria, Marche and Lazio (Centre); Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria, Sicilia and Sardegna (South or "Mezzogiorno").

North-West and the Centre. It is interesting to note the convergence process that started in 1995 among the regions in the Centre-North, in particular between the Centre and the North-East. The difference between the two areas narrowed from 3.6 points in 1995 to 2.6 points in 2005. On the contrary, the unemployment rate path for the Southern regions experienced an increasing trend from 1995 till 1999 and a decreasing trend from 2000 till 2005.

Fig.5

Accordingly, Fig. 6 shows that the (huge) gap in the unemployment rates between the South and the Centre-North worsened from 1995 until 1999. In 2000 the difference in the unemployment rates between Southern regions and the Centre-North started a decreasing trend[8].

Fig.6

This section showed that interregional differences between the Centre-North and the South of Italy narrowed throughout the period 1995-2005. Is migration playing a role in these convergence process? The answer is not easy but a necessary condition for migration to give a contribution is certainly that migrants move from regions with high unemployment rates and low income to regions where the income is high and the unemployment rate is low. In the next sections we will carry out an empirical analysis aimed at investigating the response of the recent migration flows to the main economic determinants, controlling also for non economic factors.

[8] The gap was about 10.2 percentage points in 1995, 12.9 in 2000 when the gap started to decrease till the last year in 2005 when the gap reached 9.4 percentage points (ISTAT).

3. The model

The empirical analysis is based on the extended version of the gravity model developed by Lowry (1966), which can be generalized to include all the possible push and pull factors, that is:

$$M_{ij} = k^{\gamma_0} \cdot \frac{P_i^{\gamma_1} P_j^{\gamma_2}}{D_{ij}^{\gamma_3}} \cdot \prod_{s=1}^{n} \frac{X_{s,j}^{\alpha_s}}{X_{s,i}^{\beta_s}} \qquad (1)$$

where, the number of people M_{ij} moving from region i to region j depends positively on the population size in each region (P_i, P_j) and negatively on the physical distance between the two regions (D_{ij}). $X_{s,i}$ may includes all the possible exogenous variables for the origin region i that may act as push factors for migration, while $X_{s,j}$ may includes all the exogenous variables that can attract (pull) migrants in the destination region j. According with this model different variables aimed at capturing the economic, labour market, environmental and policy conditions can be used as valid regressors (see Andrienko and Guriev, 2004).

Specifying model (1) in logarithmic form we obtain the following estimating equation:

$$\ln M_{ij} = \gamma_0 \ln k + \gamma_1 \ln P_i + \gamma_2 \ln P_j - \gamma_3 \ln D_{ij} + \sum_{s=1}^{n} (\ln X_{s,j}^{\alpha_s} - \ln X_{s,i}^{\beta_s}) \qquad (2)$$

Population and distance represent the *standard gravity variables* characterizing equation (2) as a gravity model. Population size of both origin and destination regions is expected to affect positively migration while distance should discourage migration. The inclusion of population in the model is important to take into account the increase in migration flows that results merely from an increase in population size. That is, the more one region is populated the higher will be the probability to migrate. The physical distance is commonly used as a proxy to take into account all the costs that are (directly or indirectly) related to the distance and might affect migration decisions, like transportation costs, information costs, and psychological costs. Moreover, including bilateral migration flows together with distance and the push/pull factors qualifies equation (2) as a spatial interaction model. Even though we are aware that spatial interaction models might not assure that the assumption of independence among observations is fully satisfied (see Le Sage and Pace, 2008) it is the first time that spatial interactions are considered to study internal migration in Italy.

4. Panel data approach

4.1 Panel data models

Using yearly data on gross migration flows implies that each observation has a double dimension. A first dimension refers to the spatial distribution of migrants across regions, whilst a second dimension is the temporal one and measures the yearly flows between every pair of regions.

Panel data models take advantage of this double information and can better account for the omitted variables and the individual heterogeneity (Hsiao, 2003; Wooldridge, 2002). These two features turn out to be fundamental in migration where the individuals (i.e., regions) are heterogeneous and many variables are likely to affect the destination choice of migrants.

However, the fixed effects model (FEM) and the random effects model (REM), that is, the most applied estimators in empirical analysis present some shortcomings. The REM has the advantage of using both the cross-sectional and the within variance, but for the estimates to be consistent the unobserved effects must not be correlated with the regressors. On the contrary, the FEM estimator allows the unobserved effects to be correlated with the regressors. However, due to the within transformation, the latter is not fully efficient (Cornwell, C. and Rupert, P.; 1988) and does not give explicit estimate for the time invariant variables. This is a high price in migration studies, particularly when estimating a gravity model where a time invariant variable like distance plays an important role. The Hausman[9] test has systematically rejected the REM in favour of the FEM for all the estimates carried out in this analysis. We overcome the problem using the Fixed Effects Vector Decomposition estimator (FEVD).

4.2 The Fixed Effects Vector Decomposition (FEVD)

Plümper and Troeger (2007) developed a technique that allows the estimation of the time invariant variables using a FEM setup. The authors point out that the FEM does not allow the estimation of time invariant variables and that it provides estimates that are not efficient for variables that have very little within variance[10]. More specifically, while the coefficients for time invariant variables are not computed by the FE estimator, the explanatory power of the quasi-invariant covariates can be seriously harmed by the

[9] The Hausman' specification test (Hausman, 1978) is used to test the consistency of the random effect model. If the unobserved effects are not correlated with the covariates, the random effects estimator is consistent and more efficient than the fixed effects estimator.
[10] They also argue that the inefficiency of the FE model "leads to highly unreliable point estimates and may thus cause wrong inferences in the same way a biased estimator could" (Plümper and Troeger, 2007, page. 3)

within estimator. These two important features make the FEVD a powerful estimator which is able to provide more efficient estimates of the gravity type models. In fact, in our analysis, all the standard gravity variables are time invariant or rarely-time invariant. For instance, population size, as well as the per capita GDP vary considerably across regions but have a very low within variation (see Table 2).

Moreover, variables that differ in their variability structure may also have a different impact on migration. Variables with relative high cross section variability are more likely to affect the decision of where to migrate, therefore, are important in order to identify the pull factors. On the contrary, variables with adequate within variability are likely to affect more the decision of whether to migrate or not. In other words, the variability structure of covariates affects the intensity and the direction of migration flows. For this reason, it is important to use the most appropriate econometric technique which is able to exploit both type of variability.

5. Empirical Results

5.1 FEVD versus standard panel data estimators

The extended gravity model is expressed by the following econometric model:

$$\begin{aligned} M_{i,j,t} = \beta_o + \beta_1 \cdot Opop_{i,t-1} + \beta_2 \cdot Dpop_{j,t-1} + \beta_3 \cdot dist_{i,j} + \\ \beta_4 \cdot Ounr_{i,t-1} + \beta_5 \cdot Dunr_{j,t-1} + \beta_6 \cdot Ogdp_{i,t-1} + \\ \beta_7 \cdot Dgdp_{j,t-1} + \beta_8 \cdot 1997_t + ... + \beta_{16} \cdot 2005_t + u_{i,j,t} \end{aligned} \tag{3}$$

where c_i are the region-to-region fixed effects and u_{ijt} is the error term, all variables are expressed in logarithmic form. The per capita GDP controls for the overall level of prosperity and is also related with the labour market conditions (i.e., it can be seen as a proxy for wages). Changes in per capita GDP are expected to be inversely correlated with migration flows in the sending region. On the contrary, an increase in per capita GDP should pull immigrants in the destination region. The unemployment rate is included to control for the labour market conditions. According with economic theory, it is expected to have a positive effect in the sending region and a negative effect in the destination region, though the empirical literature does not provide clear evidence. As for Italy, Daveri and Faini (1999) find that unemployment in the Southern regions did not affect migration during the period 1970-1989, whilst Fachin (2007) finds a weak impact for the period 1973-1996. In our estimations we need to take account of the possible endogeneity of the unemployment rates, which may be simultaneously determined with migration. Therefore, we carry out the estimates by instrumenting the unemployment rate with the industry-mix employment growth rate from shift-share analysis (Blanchard and Katz; 1992). The model (3) will be estimated using the fixed effect within estimator, the random effect estimator, the two-stage least-squares within estimator and the two stage-least square FEVD estimator. We have also included year dummies in order to model the idiosyncratic temporal effects. The results are reported in Table 3, they show that only the FEVD and the RE give the estimate for distance. The same models seem to perform better than the other two, with all the gravity variables appearing with the expected (and statistically significant) sign. These outcome confirms that the omission of distance affects the estimation results. Moving to the economic determinants, only the positive pull effect for the GDP is statistically significant for the FE within estimator, whilst the RE model and the FEVD reports also the expected negative push effect. As for the unemployment rate, only the FEVD estimates are significant and exhibit the expected sign. The last raw report the Hausman test result for the FE and RE effect models, the latter is rejected forcing us to move on with our empirical estimation using only the FEVD estimator. Summing up, the estimates show that the FEVD estimator appear to be more informative than the within estimator in that it allows the inclusion of distance among the regressors and use also the between information. In addition, the presence of rarely time varying variables (e.g. population and GDP) makes the FEVD estimates more efficient than the REM ones. For these reasons in the next section we will estimate the fully extended gravity model utilising only the FEVD estimator.

Table 3

5.2 The fully extended gravity model

The fully extended gravity model is expressed by the following equation:

$$\begin{aligned} M_{i,j,t} = \beta_o + \beta_1 \cdot Opop_{i,t-1} + \beta_2 \cdot Dpop_{j,t-1} + \beta_3 \cdot dist_{i,j} + \\ \beta_4 \cdot Ounr_{i,t-1} + \beta_5 \cdot Dunr_{j,t-1} + \beta_6 \cdot Ogdp_{i,t-1} + \\ \beta_7 \cdot Dgdp_{j,t-1} + \beta_8 \cdot OTemp_i + \beta_9 \cdot DTemp_j + \\ \beta_{10} \cdot OAirp_{i,t} + \beta_{11} \cdot DAirp_{j,t} + \beta_{12} \cdot 1997_t + ... + \beta_{20} \cdot 2005_t + u_{i,j,t} \end{aligned} \tag{4}$$

We estimated three different specifications of the model presented in (4), including respectively the macroeconomic variables, the amenities and the geographical dummies. The results for the different specifications are shown in Table 4. Being aware of the possibility that dealing with variables in levels some of them might be not stationary we carried out unit root tests to check whether the residuals are I(0). However, the W[t-bar] statistics from the test developed by Im, Pesaran and Shin (2003) for heterogenous panels reject the null hypothesis of nonstationarity. Therefore, we can rely on the estimates and discuss the results. The different gravity variables have all the expected signs and coefficients are statistically significant. The coefficient estimated for physical distance indicates that distance discourages internal migration flows between regions that are located far from each other in favour of migration between closer regions. The population size affects positively migration flows in that the more populated regions experience the highest migration flows. The positive sign for both origin and destination regions suggests that an increase in population size leads more people to emigrate but also attract more immigrants from other regions. The former appear to be stronger than the latter, revealing a negative net effect. These results confirm that for Italy distance is an important determinant of internal migration and cannot be omitted.

Per capita GDP appears to be the main macroeconomic determinant and it is always statistically significant. A high per capita GDP encourages immigration flows, on the contrary, low levels of per capita GDP encourage out migration. The outcomes for the unemployment rate report the results predicted by economic theory, but have less impact on migration than the per capita GDP. The reason probably being that the unemployment rate refers specifically to the labour market, whilst the per capita GDP is an indicator that reflects the overall economic conditions.

Table 4

The results of model one do not change when we control for the presence of airports and for the average temperature, It seems that interregional migration flows are favoured by the presence of airports both at origin and at destination. We tried to test for the highways and railways kilometres as well but the coefficients were never statistically significant. Considering that the coefficient has positive sign both at origin and at destination this results may reveal that for the new migrants it is important to have the possibility to travel back to their region of origin. Climate has relatively low impact but it is statistically significant and show a preference for the warmer regions. .

In the third model we add the geographical dummies in order to control for structural differences between the regions located in the South and those located in the North, the dummies for central regions have been dropped to avoid collinearity. We see that even after controlling for differences in GDP and unemployment rates, still being a region in the south is associated with more outflow migration and less inflow migration. More surprisingly, the same results appear for the northern regions. In the next section we will investigate further the migration flows between these two macro-areas, which have always been characterizing internal migration in Italy.

5.3 South and Centre-North migration determinants

As previously discussed, the only unchanged feature of internal migration flows in Italy during the last sixty years is the leading direction from South to Centre-North. In this section we test the hypothesis that northern migrants and southern migrants might respond differently to the migration determinants. Accordingly, the further analysis is carried out on a two different datasets where intra area flows have not been included[11]. The first model considers the pairs of flows from each of the eight southern regions to each of the twelve regions in the Centre-North. The second model works in the other way around, that is, the dependent variable measures the flows from the regions in the Centre-North to the southern regions. To our knowledge the second model is a novel analysis which has never been carried out before.

Table 5

[11] Only migration flows between the eight Southern regions and the twelve Northern regions are considered. Hence, the panel is: N = 96 (=12 x 8); T = 10.

In this way, the impact of the macroeconomic variables should be emphasized due to the exclusion of short distance migration which is more likely to have been determined by family reasons. In order to control for structural differences between the two areas, we estimate a further specification of model (4) by adding the weight of the industry in a strict sense. The results are shown in Table 5. The gravity variables have the same (expected) signs found for the whole dataset and are all statistically significant. The distance, nevertheless, exhibits a stronger impact on migration flows compared with the previous results. This is, however, not surprising considering that almost all migration flows between neighbour regions are now excluded[12]. It is interesting to notice that, southern migrants appear to be more concerned about long distance than northern ones. There are noticeable differences with regards to the economic determinants. Southern migrants respond strongly to variation in per capita GDP, which act as a strong pull factor. By contrast, the role of per capita GDP for northern migrants is quite different. The positive sign at origin indicates that an increase in per capita GDP allows more people to move from the regions in the Centre-North to the southern regions. A possible explanation to this outcome is that higher income levels allow migrants to finance their move in poorer regions, which might be originated by non economic reasons (e.g., retired people going back to their native regions). Differences between the two macro-areas appear also with regard to the unemployment rate. The latter seems to act as an important push factor for both southern and northern migrants although at destination the former are attracted by low unemployment rates whilst the latter seems to be attracted by high unemployment rate. The explanation to this apparently surprising result is though that northern migrants when moving to a region in the south are always

[12] The only regions that share a border are those located in the Centre-South of Italy. In this way, the physical distance improves its power as a proxy for the psychological and information costs of migration.

going to a region with higher unemployment rate. The outcome for the number of airports confirms the previous results and highlights the rise in the impact of this variable with respect to the whole sample. The different impact reveals also the importance of airports as a transport infrastructure for the long distance migration. Comparing the magnitude of the coefficients for the average temperature we notice that its impact is stronger for the northern migrants. In particular, differently from the southern migrants those living the regions in the north prefer to move to warm regions in the South. The indicator for the economic structure shows that the northern migrants tend to move from regions where the weight of industry in the strict sense is low to regions where it is high. Northern migrants respond at exactly the opposite, that is they move from regions characterized by high value added of the industry sector to the southern regions where industry has a minor weight.

5.4 The dynamic model

In this section, a dynamic version of the extended gravity model is tested to investigate for the presence of networks effects, and to check the robustness of results obtained in the previous sections. The dynamic version of the extended gravity model is expressed by the following equation:

$$\begin{aligned} M_{i,j,t} = {} & \rho_t M_{i,j,t-1} + \beta_o + \beta_1 \cdot Opop_{i,t-1} + \beta_2 \cdot Dpop_{j,t-1} + \beta_3 \cdot dist_{i,j} + \\ & \beta_4 \cdot Ounr_{i,t-1} + \beta_5 \cdot Dunr_{j,t-1} + \beta_6 \cdot Ogdp_{i,t-1} + \\ & \beta_7 \cdot Dgdp_{j,t-1} + \beta_8 \cdot OTemp_i + \beta_9 \cdot DTemp_j + \\ & \beta_{10} \cdot OAirp_{i,t} + \beta_{11} \cdot DAirp_{j,t} + u_{i,j,t} \end{aligned} \tag{5}$$

where $M_{i,j,t-1}$ is the lagged dependent variable, which gives the dynamic specification to the gravity model. We applied two different econometric techniques. The first is the dynamic FEVD estimator. The second is a new econometric technique for dynamic panel data models based on two estimators: the difference GMM developed by Arellano and Bond (1991) and the system GMM developed by Blundell and Bond (1998)[13]. The two GMM estimators are designed for situations with endogenous regressors, fixed effects, heteroskedasticity[14] and autocorrelation within individuals (Roodman, 2006).

The results are presented in Table 6. A first comparison between the FEVD and the GMM shows that the standard gravity variables appear with the same (expected) signs found in the previous estimations and they are all statistically significant. Moreover, the results for the lagged dependent variable show that the past gross migration flows affect positively the current migration. It is worth to note here that this definition is more parsimonious than the one tested by Furceri (2006) in that it is measured for each pair of regions and not with aggregate net migration. Therefore, the results provide empirical evidence for the presence of "social networks" between regions, that is the past immigrants in a region attracts (i.e., pull) future immigrants from the same sending regions. The results for the per capita GDP confirm the previous ones, that is, a positive (pull) effect in the destination region and a negative (push) effect in the sending region. The unemployment rate outcomes confirm the positive (push) effect in the sending region but not in the destination region. Once controlling for the (social) "network effects" unemployment rate in the destination region loses power as a pulling factor, this could be due to the fact that migrants take more advantage from the social network (the presence of past migrants coming from the same region) to find a job. The Arellano and Bond test (AR1) shows that the lag of the dependent variable is endogenous, that is,

[13] Both estimators are designed for panel data where T is small and N is large, which is exactly the structure of the present panel, where T=7 and N=380.

[14] The heteroskedasticity test proposed by Pagan and Hall (1983) for instrumental variables regressions (IV) rejected the hypothesis that the disturbance is homoskedastic. The test is also consistent with the presence of heteroskedasticity in the structural equation (Baum *et al.,* 2002).

there is serial correlation. The test for the validity of the subsequent lags as instruments (AR2) does not reject the null that the lags are exogenous (i.e., valid instruments)[15]. The *J* statistic of Hansen (1982)[16] does not reject the null hypothesis that the other instruments are valid, that is, they satisfy the orthogonality conditions. Finally, following the warnings of Roodman (2007) concerning the weakness of the system GMM when the number of instruments is very high, we report the total number of instruments[17].

Table 6

6. Conclusions

In this paper we investigated the main determinants of interregional migration flows in Italy during the period 1996-2005. The extended gravity model, which is frequently employed in migration studies (Greenwood, 1997; Greenwood and Hunt 2003), has been estimated taking advantage of a recent estimation technique, the fixed effect vector decomposition estimator (FEVD). Our outcomes provide evidence that the FEVD not only it allows the estimation of distance (time invariant) but also it improves the efficiency of the estimates for the rarely time varying variables (Plümper and Troeger, 2007).

The outcomes revealed the significant impact of the main macroeconomic variables, confirming the results found by Basile and Causi (2007) and Furceri (2006). Differently from the previous works, however, our analysis uses gross (instead of net) migration flows and a more advanced econometric techniques. Accordingly, the analysis enabled

[15] That is, there is no serial correlation.

[16] The Hansen J statistic is robust to heteroskedasticity and is thus preferred to the Sargan (1958) test. The Hansen J test is weak to the instruments proliferation (Baum *et al.,* 2002), however, in this case the number of instruments is not large.

[17] According to Roodman (2007) the number of instruments should not be higher than N.

to identify the push and the pull factors, and to consider the spatial interactions between regions. The per capita GDP turned out to be the main economic determinant, representing a strong pull factor. The unemployment rate is, indeed, an important push factor in the sending regions but do not pull long distance migrants. An interesting result is that migrants leaving the regions in the Centre-North seem to respond differently to the push and pull forces with respect to southern migrants. The estimation of the dynamic gravity model provides evidence for the presence of social network effects between pair of regions, which appear to have a strong impact.

Bibliography

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

Arellano M., and Bond S., 1991. "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", *Review of Economic Studies* 58: 277–97.

Attanasio, Orazio P., Padoa Schioppa, Fiorella. 1991. " *Regional Inequalities, Migration and Mismatch in Italy, 1960-86"*, in Padoa Schioppa, Fiorella (ed.), *Mismatch and labour mobility*. Cambridge University Press. Cambridge and New York, 1991, p 237-320.

Basile Roberto & Jaewon Lim. 2006. "Wages Differentials and Interregional Migration in the U. S.: An Empirical Test of the Option Value of Waiting Theory", ERSA conference papers ersa06p263, European Regional Science Association, revised.

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

Baum Christopher F & Mark E. Schaffer & Steven Stillman. 2002. "Instrumental variables and GMM: Estimation and testing", *Boston College Working Papers* in Economics 545, Boston College Department of Economics, revised 14 Feb 2003.

Blanchard O. J., Katz L. F., Hall R. E., Eichengreen B.. 1992. "regional Evolution" Brookings Papers on Economic Activity, Vol. 1992, No. 1, pp. 1-75.

Blundell, R., and S. Bond. 1998. “Initial conditions and moment restrictions in dynamic panel data models.” *Journal of Econometrics* 87: 11–143.

Bonifazi, Corrado. 2001. “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.

Borjas, George J..1994. “The Economics of Immigration,” *Journal of Economic Literature* 32: 1667-1717.

Bosco Luigi. 2003. “Divari Nord-Sud e migrazione interna”, *Sviluppo Locale*.

Cannari, Luigi, Nucci, Francesco and Sestito, Paolo. 2000. “ Geographic Labour Mobility and the Cost of Housing: Evidence from Italy”, *Applied Economics*, Vol. 32 (14). p 1899-1906.

Casavola P., Gavosto A., Sestito P.. 1995. “Salari e Mercato del Lavoro Locale”, *Banca d’Italia*, Servizio Studi Banca d.Italia.

Cornwell C., Rupert P., 1988. “Efficient estimation with panel data: An empirical comparison of instrumental variables estimators”, Journal of Applied Econometrics, Vol. 3 (2), p. 149 – 155.

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

Daveri, Francesco, Faini, Riccardo. 1999. “ Where Do Migrants Go?” *Oxford Economic Papers*, Vol. 51 (4). p 595-622.

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

Faini, R., Galli, G., Gennari, P. and Rossi, F.,.1997. "An empirical puzzle: Falling migration and growing unemployment differentials among Italian regions", *European Economic Review*, Elsevier, vol. 41(3-5), pages 571-579.

Furceri, D., 2006. “Does labour respond to cyclical fluctuations? The case of Italy”, *Applied Economic Letters,* 13, 135-139.

Greenwood MJ, Hunt GL..2003. “The early history of migration research”, *International Regional Science Revie,* 26: 3–37.

Greenwood MJ..1985. “Human migration: Theory, models and empirical studies”, *Journal of Regional Science,* 25: 521–544.

Greenwood, MJ.. 1997. “Internal Migration in Developed Countries”, in Mark R. Rosenzweig and Oded Stark, editors, Handbook of Population and Family Economics, Volume 1B. Amsterdam: Elsevier, pp. 647-720.

Hansen, L.. 1982. “Large sample properties of generalized method of moments estimators”, *Econometrica*, 50(3): 1029-1054.

Harris J.R. and Todaro, M.P.. 1970. “Migration, Unemployment and Development: A Two Sector Analysis”, *American Economic Review*, 60, 126-142.

Hatton, Timothy, J.. 1995. “A Model of U.K. Emigration 1870-1913”, *The Review of Economics and Statistics*, Vol. 77, No. 3, 407-416.

Hausman, Jerry A.. 1978. “Specification Tests in Econometrics” *Econometrica,* 46, 1251-1271.

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

Im K. S., Pesaran M. H., Shin Y., 2003. “Testing for Unit Roots in Heterogeneous

Panels". Journal of Econometrics, Vol. 115, p. 53-74.

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

ISTAT. differen years, b. *Forze di lavoro.* Roma, ISTAT.

Istituto G. Tagliacarne, (2001). *La dotazione di infrastrutture nelle province italiane 1997-2000*. Unioncamere.

Le Sage J. P., Pace R. K.. 2008. "Spatial Econometric Modeling of Origin-Destination Flows", *Journal of Regional Science*, Vol. 48, No. 5, pp. 941-967.

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

Murat, M. and Paba, S.. 2001. "Flussi migratori e modelli di sviluppo industriale. L'esperienza italiana dal dopoguerra agli anni novanta". *Mimeo*.

Pagan, A.R. and D.Hall. 1983. "Diagnostic tests as residual analysis", Econometric Reviews 2(2): 159-218.

Piras, Romano. 2005a. "Il contenuto di capitale umano dei flussi migratori interregionali: 1980-2002", *Politica economica*, volume: 21.

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.

Roodman David. 2007. "A Short Note on the Theme of Too Many Instruments", Working Papers 125, Center for Global Development, revised.

Roodman, David. 2006. "How to Do xtabond2: An Introduction to 'Difference' and

'System' GMM in Stata", Working Paper 103. Center for Global Development. Washington, DC.

Sargan, Jhon, D.. 1958. "The estimation of economic relationships using instrumental variables", *Econometrica*, 26(3): 393–415.

Salvatore Dominick. 1977. "An Econometric Analysis of Internal Migration in Italy", *Journal of Regional Science*, Vol. 17: 395-408.

SVIMEZ. 2004. "Report on the economy of southern Italy", *SVIMEZ* annual reports.

SVIMEZ. 2007. "Rapporto SVIMEZ 2007 sull'economia del Mezzogiorno", *SVIMEZ* annual reports.

Van Dijk J., Folmer H., Herzog H. and Schlottman A..1989. "Labour market institutions and the efficiency of interregional migration: a cross nation comparison", In Van Dijk J., Folmer H., Herzog, H. and Schlottman A. (eds.) Migration and Labour Market Adjustment. Dordrecht: Kluwer.

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

Appendix A. Data and descriptive statistics

The dataset consists of annual observations for the twenty Italian regions over the period 1996-2005. The panel has a matrix structure where each region appears two times, that is as a sending region (denoted with the index *i*) and as a destination region (denoted with the index *j*). Migration flows indicate the number of people that, during each year, cancelled their official residency in one region and registered it in another region. The last year is 2005 since, at the time of writing, this is the last year for which data on gross migration flows are available.

The population size is expressed as the annual average number of people living in region *i*/*j*. The unemployment rate is the ratio between the unemployed males and females (aged 15 years and more) and the total labour force. The real per capita GDP is taken from ISTAT. In table 1 is reported the list of variables names used in this analysis and their definition. Table 2 reports descriptive statistics of between and within standard deviation. The last column of the table shows the ratio between the *between* and the *within* standard deviation. Plümper and Troeger (2007) demonstrate, using Montecarlo simulations, that the FEVD estimator is more efficient than the FEM when the ratio between the between variance and the within variance is large enough. Therefore, the ratio is an important indicator to identify the rarely changing variables for the FEVD estimator. The variables with a star in Table 2 are treated as rarely changing variables in the estimations.

Table 1. Variable Definitions

Variable	Definition
mig	(log) Gross migration flows from region i to region j (source: ISTAT);
pop	(log) Regional population size (source: ISTAT);
dist	(log) Aerial distance (in km) between the main city in the sending region and the main city in the destination region.
gdp	(log) Per capita GDP in the origin (lnogdp) and in the destination region (lndgdp) (source: ISTAT);
unr	(log) Regional unemployment rate in region i (lnounr) and in region j (lndunr) (source: ISTAT);
temp	(log) Regional average yearly temperature (source: ISTAT);
airp	(log) Number of airports (source: ISTAT);

Table 2. Between and within descriptive statistics of the variables.

Variable		Mean	Std. Dev.	Min	Max	Observations	B/W
mig	overall	5.50	1.61	0.00	10.51	N = 3800	
	between		1.60	0.84	8.84	n = 380	
	within		0.22	3.77	9.73	T = 10	
pop*	overall	7.54	1.06	4.76	9.14	N = 3800	
	between		1.06	4.78	9.11	n = 380	91.5
	within		0.01	7.51	7.58	T = 10	
dist*	overall	5.94	0.61	4.36	6.98	N = 4180	
	between		0.61	4.36	6.98	n = 380	TI
	within		0.00	5.94	5.94	T = 11	
gdp*	overall	2.95	0.27	2.43	3.36	N = 3800	
	between		0.27	2.54	3.32	n = 380	5.8
	within		0.05	2.82	3.04	T = 10	
unr	overall	2.19	0.57	0.92	3.20	N = 3800	
	between		0.53	1.32	3.10	n = 380	2.6
	within		0.21	1.11	2.53	T = 10	
airp*	overall	0.71	0.68	0.00	1.61	N = 3800	21.4
	between		0.68	0.00	1.61	n = 380	

	within		0.03	0.54	0.83	T =	10	
temp	overall	6.42	6.89	0.00	18.50	N =	3800	
	between		1.76	1.90	9.07	n =	380	0.3
	within		6.66	-2.65	15.85	T =	10	
vaiss*	overall	-3.47	0.36	-2.41	-1.14	N =	3800	
	between		0.36	-2.36	-1.17	n =	380	8.1
	within		0.04	-1.79	-1.55	T =	10	

Appendix B. Tables and graphs

Fig. 1. Trends in interregional migration flows in Italy

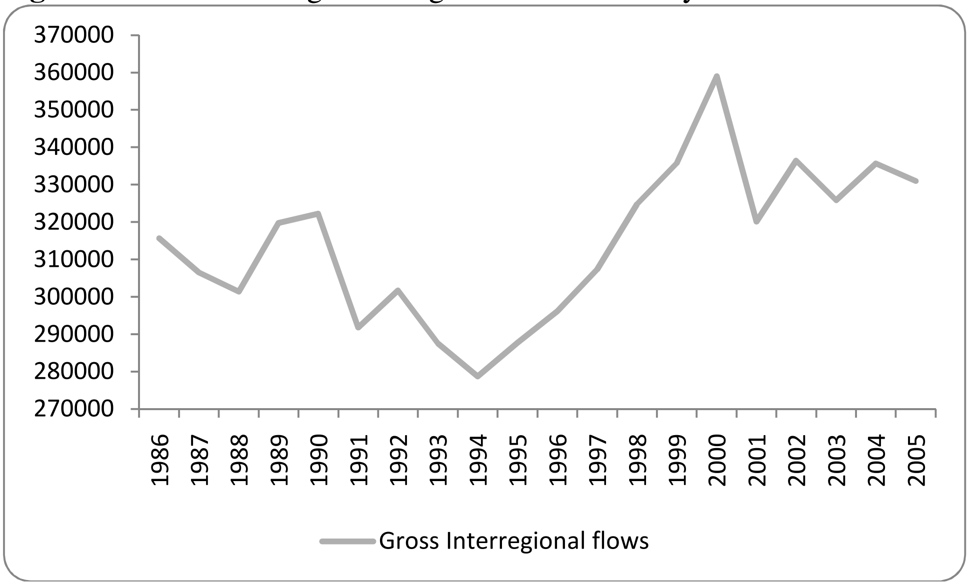


Fig. 2. Interregional migration flows: period 1996-2005 (ISTAT).

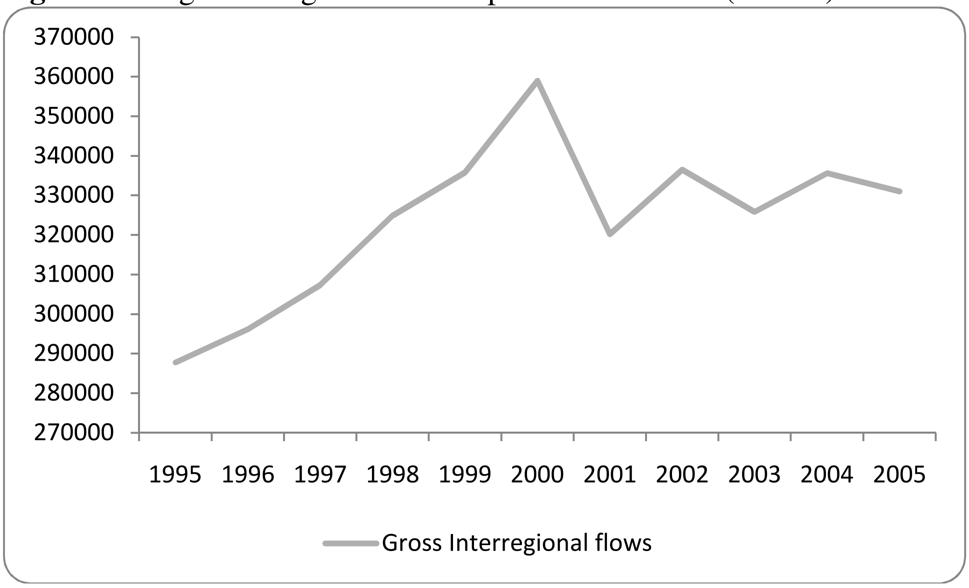


Fig. 3. Real per capita GDP: North-West, North-East, Centre, South

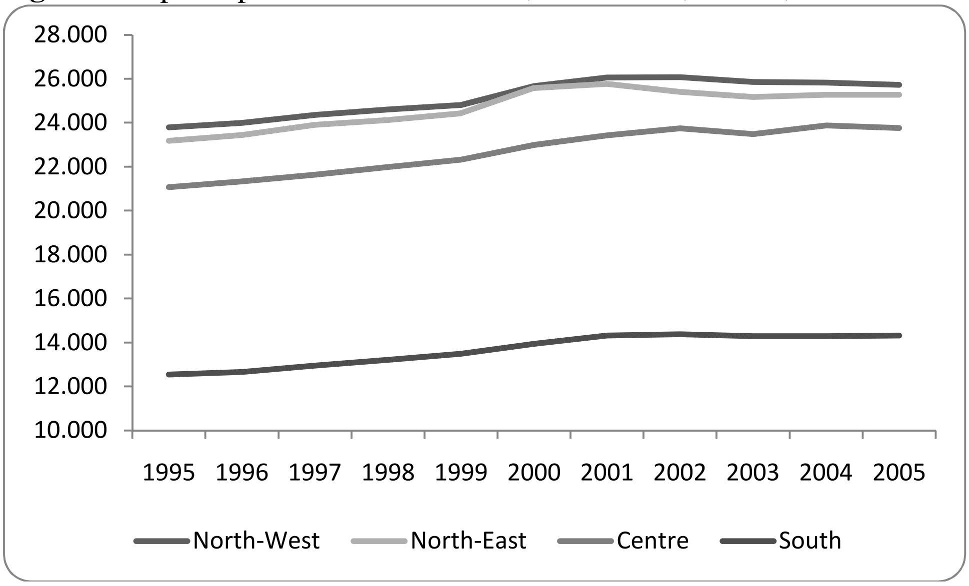


Fig. 4. Gap in real per capita GDP between the South and the Centre-North

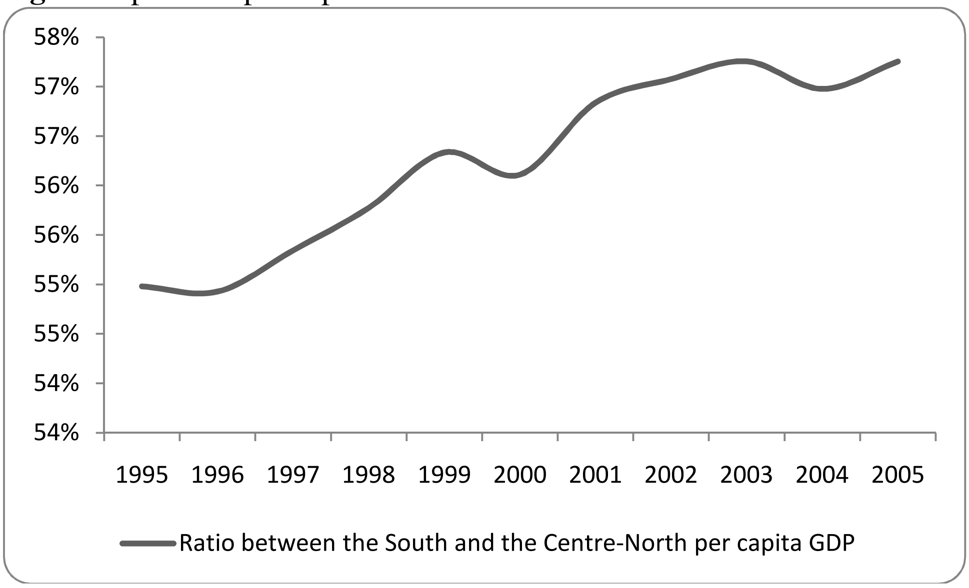


Fig. 5. Unemployment rates differentials

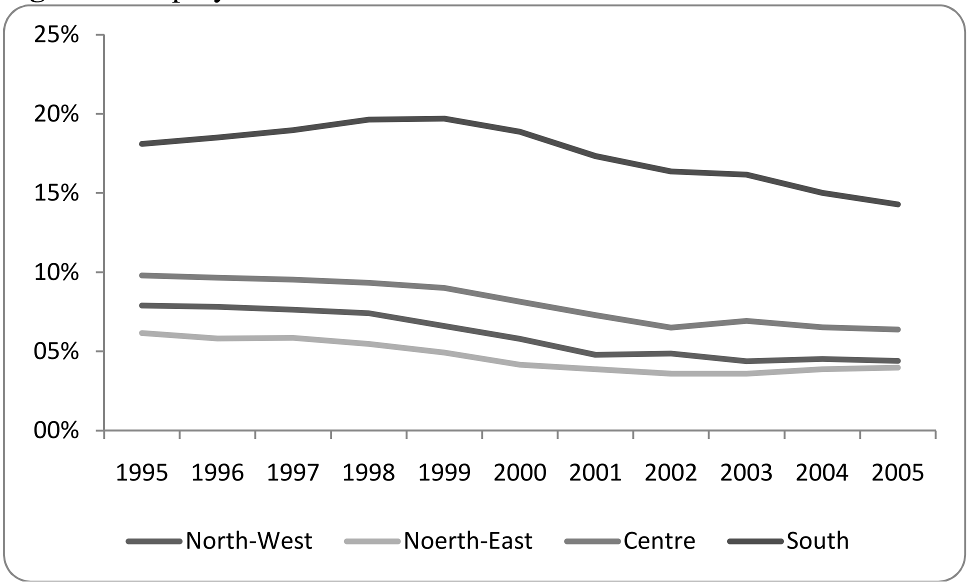
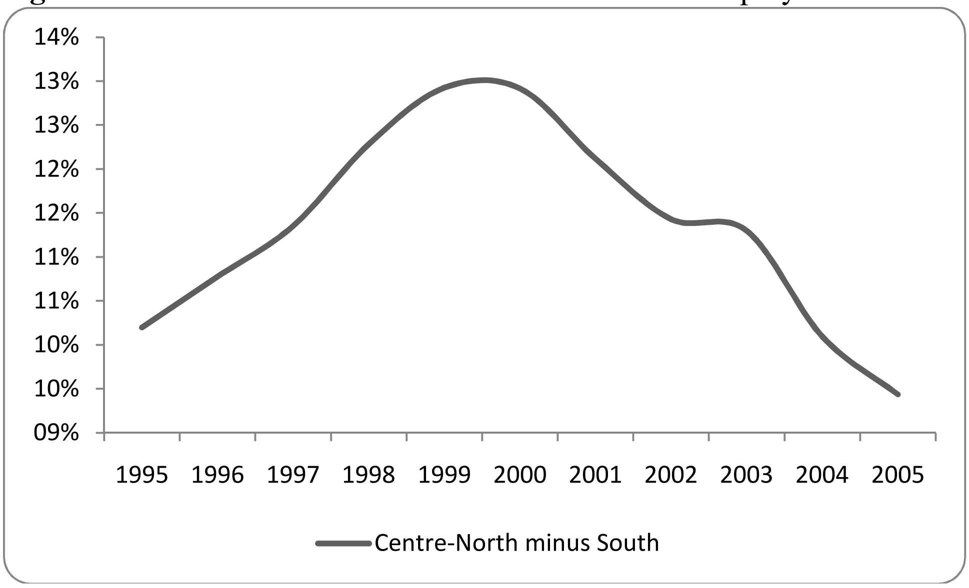


Fig. 6. Difference between South and Centre-North unemployment rate



Tab3. FEVD estimation versus standard panel data estimations

Variable	**FEVD-IV**	**FE-IV**	**FE**	**RE**
opop	1.004***	3.208***	3.897***	1.018***
t-stat (z-stat)	294.69	5.72	8.80	32.77
dpop	0.959***	1.098	1.758***	0.964***
t-stat (z-stat)	281.55	1.96	3.97	31.03
dist	-0.328***	0.00	0.00	-0.344***
t-stat (z-stat)	-52.97	.	.	-6.24
ogdp	-0.328***	-0.191	-0.314	-0.679***
t-stat (z-stat)	-11.47	-0.81	-1.39	-6.39
dgdp	0.368***	0.770**	0.652**	0.535***
t-stat (z-stat)	13.03	3.26	2.90	5.03
eta	1.000***			
t-stat (z-stat)	175.45			
ounr	0.121***	-0.155	0.053	0.022
t-stat (z-stat)	8.42	-1.47	1.9	0.88
dunr	-0.108***	-0.172	0.028	0.019
t-stat (z-stat)	-7.62	-1.63	0.99	0.75
Cons	-7.516***	-27.901***	-38.329***	-7.132***
t-stat (z-stat)	-43.87	-4.20	-7.11	-9.90
Obs	3800	3800	3800	3800
F stat 1st Eq	F(18, 3404) = 3566	F(15,3405)= 415.45		
Hausman chi2(15)	FE vs IV	not rej	7.63	
Hausman chi2(15)	RE vs FE	rej	114.12	

*Note: For the Fixed Effects Regression with Vector Decomposition (FEVD) the covariates opop dpop ogdp dgdp are treated as rarely changing variables and standard errors are robust to heteroschedasticity. Stars denote p-values as follows: * p<0.05; ** p<0.01; *** p<0.001*

Table 4. Extended Gravity model (FEVD)

Variable	**I**	**II**	**III**
opop	1.004***	0.955***	0.961***
t-stat	294.69	193.73	190.93
dpop	0.959***	0.889***	0.887***
t-stat	281.55	180.36	176.40
dist	-0.328***	-0.360***	-0.375***
t-stat	-52.97	-56.87	-57.78
ogdp	-0.328***	-0.382***	-0.394***
t-stat	-11.47	-13.06	-9.28
dgdp	0.368***	0.390***	0.362***
t-stat	13.03	13.46	8.56
ounr	0.121***	0.157***	0.157***
t-stat	8.42	10.62	10.43
dunr	-0.108***	-0.169***	-0.169***
t-stat	-7.62	-11.64	-11.43
oairp		0.131***	0.114***
t-stat		16.9	14.45
dairp		0.104***	0.116***
t-stat		13.40	14.76
otemp		-0.013***	-0.006***
t-stat		-7.46	-3.37
dtemp		0.030***	0.023***
t-stat		16.78	12.61
dnorth			-0.050***
t-stat			-4.31
onorth			0.203***
t-stat			17.59
dsouth			-0.03
t-stat			-1.59
osouth			0.136***
t-stat			7.17
cons	-7.516***	-6.476***	-6.405***
t-stat	-43.87	-34.94	-27.65
Obs	3800	3800	3800
I-P-Shin test W[t-bar]	-2.38	-4.56	-8.55

*Note:Two-stages least-squares Panel Fixed Effects Regression with Vector Decomposition (FEVD). The covariates opop dpop ogdp dgdp otemp dtemp oairp dairp dnorth onorth dsouth osouth are treated as rarely changing variables. Standard errors are robust to heteroschedasticity. Stars denote p-values as follows: * p<0.05; ** p<0.01; *** p<0.001*

Table 5. Extended Gravity Model: South to Centre-North flows

Variable	**South to C-North**	**C-North to South**
opop	0.777***	0.892***
t-stat	1467.62	57.26
dpop	0.750***	0.879***
t-stat	1507.78	61.28
dist	-1.098***	-0.805***
t-stat	-406.45	-26.74
ogdp	-1.102***	2.130***
t-stat	-109.37	18.95
dgdp	2.246***	0.099
t-stat	253.49	0.62
ounr	1.291***	0.424***
t-stat	7.27	10.13
dunr	-0.083***	0.227**
t-stat	-4.71	2.68
oairp	0.238***	0.121***
t-stat	284.57	4.91
dairp	0.210***	0.299***
t-stat	149.12	15.27
otemp	0.004***	-0.026***
t-stat	25.79	-5.93
dtemp	0.011***	0.030***
t-stat	11.27	8.34
ovaiss	-0.572**	0.635***
t-stat	-3.16	15.62
dvaiss	0.657*	-0.284***
t-stat	2.15	-5.22
cons	-6.710***	-11.290***
t-stat	-89.26	-15.41
Obs	960	960

*Note:Two-stages least-squares Panel Fixed Effects Regression with Vector Decomposition (FEVD). The covariates opop dpop ogdp dgdp otemp dtemp oairp dairp dnorth onorth dsouth osouth are treated as rarely changing variables. Standard errors are robust to heteroschedasticity. Stars denote p-values as follows: * p<0.05; ** p<0.01; *** p<0.001*

Table 6. Extended Gravity Model: dynamic analysis

Variable	**FEVD-IV**	**GMM**
mig	0.073***	0.172**
t-stat (z-stat)	4.56	2.95
opop	0.901***	0.898***
t-stat (z-stat)	2012.74	9.01
dpop	0.832***	0.784***
t-stat (z-stat)	1629.29	9.07
dist	-0.365***	-0.213**
t-stat (z-stat)	-850.91	-2.72
ogdp	-0.165***	-1.291**
t-stat (z-stat)	-9.73	-2.59
dgdp	0.543***	1.963**
t-stat (z-stat)	47.62	2.84
dnorth	-0.039***	-0.29
t-stat (z-stat)	-13.84	-0.68
onorth	0.196***	0.805**
t-stat (z-stat)	70.83	3.13
dsouth	-0.055***	0.50
t-stat (z-stat)	-17.79	1.07
osouth	0.096***	0.101
t-stat (z-stat)	19.46	0.35
oairp	0.093***	-0.03
t-stat (z-stat)	43.82	-0.31
dairp	0.095***	0.029
t-stat (z-stat)	58.54	0.26
otemp	-0.011***	-0.019*
t-stat (z-stat)	-6.06	-2.25
dtemp	0.017***	0.021*
t-stat (z-stat)	10.53	2.33
ounr	0.310***	0.284**
t-stat (z-stat)	5.09	3.06
dunr	0.01	-0.01
t-stat (z-stat)	0.11	-0.13
cons	-7.864***	-9.913***
t-stat (z-stat)	-137.61	-5.49
Obs	3800	3800
Nr of Instr.		46
Abond AR(1)		z = -6.55
Abond AR(2)		z = 0.76
Hansen Test		0.405

Note: The GMM model (dynamic panel data) has been estimated using the xtabond2 command in Stata (Roodman, 2006). The number of instruments has been reduced limiting the lags of instruments. The lag *of migration flows and the unemployment rate are treated as endogenous. The industry mix employment growth rates have been added as exogenous instruments.*