Interregional migration and unemployment dynamics: evidence from Italian provinces

Roberto Basile, r.basile@isae.it

Alessandro Girardi, a.girardi@isae.it

Marianna Mantuano, m.mantuano@isae.it

ISAE (Institute for Studies and Economic Analyses)

Abstract

Does interregional migration equilibrate regional labour market performances? We answer this question focusing on regional unemployment dynamics in Italy over the 1995-2007 period, when a strong flow of out-migration from the South to the North occurred. Building on econometric models for longitudinal data allowing for nonlinearities and spatial dependence, the empirical analysis document that migration flows exert a strong negative effect on regional unemployment growth rates. By falsifying the common wisdom, our results are consistent with recent theoretical contributions pointing out that, in the presence of agglomeration forces, migration flows are likely to magnify spatial disparities in unemployment rates.

Keywords: Unemployment, migration, spatial dependence, nonparametrics

JEL codes: R23, C14, C23

1. Introduction

Lowering unemployment is a policy mission typically challenged at a national level. Only in the textbook case of full efficient markets, however, where equilibrating forces of capital and labour mobility and changes in relative prices are fully at work, no significant spatial unemployment disparities *within* country would exist. In the real world, instead, national averages are likely to hide large regional differences in unemployment rates (Pissarides and McMaster, 1990; Blanchard and Katz, 1992; Decressin and Fatas, 1995; Elhorst, 1995; Taylor and Bradley, 1997; Kostoris-Padoa-Schioppa and Basile, 2002; Overman and Puga, 2002; Bande and Karanassou, 2007). Aside from academic disputes, divergence in unemployment patterns within national boundaries entails welfare losses due to a downward spiral effect for backward regions, which tend to suffer typically from a net loss of population, reduced demand for locally produced goods and services and regional brain drain (selective out-migration of high-skilled workers) (Elhorst, 2003).

Persistent spatial disparities in unemployment rates have been often ascribed to rigidities in labour markets, which have discouraged workforces to move across regions (Blanchard and Kats, 1992; Decressin and Fatàs, 1995; Leonardi, 2004). Recent theoretical contributions (Epifani and Gancia, 2005, among others) have emphasized instead that, in the presence of agglomeration force, migration flows are likely to magnify spatial disparities in unemployment rates rather than mitigate them as predicated by the neoclassical approach.

This paper aims at assessing the ultimate effect of migration on regional unemployment dynamics by using Italian data at a very fine territorial level (103 provinces or NUTS-3 regions) over the 1995-2007 years. The case of Italy is peculiar since the ongoing restructuring of the domestic labor market (Cipollone and Guelfi, 2006; ISAE, 2007) has been leading to a reduction of the nation-wide unemployment rate in the presence of remarkable (and persistent) regional disparities (Faini et al., 1997; Prasad and Utili, 1998; Alesina et al., 1999; Cannari et al., 2000; Brunello et al., 2001, Kostoris-Padoa-Schioppa and Basile, 2002). During the same period, a strong flow of out-migration from Southern towards Northern regions has been started. However, simple descriptive statistics reveal a negative association between unemployment dynamics and out-migration for the backward (Southern) regions, while the opposite relationship holds for Northern regions.

The stylized fact for the South is at odds with the neoclassical view of migration acting as an equilibrating force for unemployment differentials. In an effort to better analyze the effects of migration on regional unemployment dynamics, we propose a methodological framework which innovates along several dimensions with respect to the existent literature. *First*, we employ models for panel data in the presence of spatial dependence (Elhorst, 2009) in place of simpler cross-section methods as in Overman and Puga (2002), among others. *Second*, we allow for possible nonlinearities by specifying semiparametric formulations of the regression models along the lines suggested by Ullah and Mundra (2001) and Mundra (2005) among others. Controlling for a number of determinants suggested by the literature, both parametric and nonparametric estimation results corroborate the hypothesis of a negative effect of migration on the regional unemployment dynamics.

The layout of the paper is the following. Section 2 discusses the role of migration on regional unemployment. Section 3 presents some stylized facts on the regional labour market dynamics and on interregional migration flows in Italy. Section 4 illustrates the set of candidate causative determinants of regional unemployment growth along with the methodological framework. Section 5 discusses the estimation results. Concluding remarks follow.

2. The effect of labor migration on regional unemployment disparities: theoretical controversies and empirical disputes

2.1 Positive or negative effect?

The question whether interregional migration equilibrates regional economic performances has received considerable attention in the traditional as well as in the most recent literature. The question is an issue fraught with controversy, since interregional migration produces both labor supply and labor demand effects (Chalmers and Greenwood, 1985).

Relying on the neo-classical view of homogenous labour, workers moving towards prosperous regions may help reduce regional differences in unemployment rates through a reduction of the pool of job seekers in initially high-unemployment regions and an increase of it in the host regions. On the labour demand side, immigrants are expected to cause an increase in total expenditure due to their requirement of goods and services produced in the host regions. Such an increase in demand for goods and services should lead to higher investment levels and, thus, into new labour demand. However, according to the neo-classical paradigm, supply-side effects (which reduce the unemployment rate differentials) are likely to dominate the demand-side effects (which exacerbate spatial disparities).

By contrast, demand side effects of net migration on unemployment rate differentials may dominate the supply side ones when the neoclassical assumption of perfect competition is relaxed. In keeping with the New Economic Geography (NEG) paradigm, the dynamic core-periphery model with frictions in the job matching process developed by Epifani and Gancia (2005), for example, proves that the same drivers generating agglomeration also determine persistent spatial disparities in unemployment rates. In their setup, higher regional integration (epitomized by a fall in transport costs) activates migration flows which, in turn, stimulate agglomeration economies (through the so-called home-market effect), leading to a core-periphery equilibrium, with strong and persistent inequalities both in terms of productivity and unemployment.[1] This is because agglomeration economies increase profits and, thus, labour demand in the core, which translates into a reduction of the unemployment rate in that area. The opposite occurs in the backward region, where reduced profits increase the unemployment rate, implying the emergence of a core-periphery unemployment gap.

2.2 The empirical content of migration: previous evidence

The discussion above suggests that the effect of net labor migration on interregional unemployment differentials is mostly an empirical question. A group of empirical studies focused on the effectiveness of migration as mechanism of adjustment of negative shocks hitting local labour markets. Sufficiently large labour mobility coupled by massive wage differentials may help absorb negative shocks. In fact, if wages reflect adequately local unemployment rates, then depressed high unemployment regions may be favored if the unemployed move towards low unemployment but high wage regions and if capital moves to high unemployment regions, attracted by the low cost of labour. This type of adjustment mechanism seems to work in a different way in the US and in the EU, producing different outcomes in terms of employment and inactivity rates. For the case of the US, Blanchard and Katz (1992) find that labor mobility has been crucial in achieving regional convergence in unemployment rates. For the case of the EU, Decressin and Fatàs (1995) find that interregional unemployment convergence was achieved though a reduction in the activity rate in high unemployment regions rather than by labor migration. Similarly, Leonardi (2004) finds that the persistence of regional unemployment rates relative to the national mean in Italy may be attributed to the slow response of migration.

Another body of empirical research has produced sizable evidence from regression models designed to analyze the effect of regional migration on spatial unemployment differentials. Groenewold (1997) finds that inter-regional equilibrating forces are slow and do not help equalize regional unemployment rates in Australia. For the case of Canada, Wrage (1981) documents a small but significant symmetric effect of migration on regional unemployment rates (i.e., out-migration has an equal but opposite impact to in-migration). Consequently, the ultimate effect of migration on regional unemployment depends on whether or not a region has a net gain or loss of migrants. Many other empirical studies on Eastern European countries (see, for example, Rutkowski and Przybila, 2002, for Poland; Kertesi, 2000, for Hungary) also provide evidence that net migration flows are positive in low unemployment regions and negative in high unemployment regions, as the neoclassical paradigm would posit, but they are insufficient to compensate large unemployment differentials.[2]

3 Regional labour market dynamics and interregional migration in Italy

Starting from mid-nineties a resurgence of interregional migration movements from the South to the North of Italy has taken place, mainly due to changes in regional policies and some macroeconomic factors (Basile and Causi, 2007). In the light of the revamped flows of workforce across Italian regions, it turns out to be informative to analyze the effect of net migration on the regional unemployment dynamics.

Using the most recent data, we focus on the years 1995-2007, during which the national-wide unemployment rate has dropped from 11.2 percent in 1995 to 6.1 in 2007, although the dichotomy between Northern and Southern regions has increased.[3] The South/Centre-North unemployment rate ratio has indeed moved from 2.3 in 1995 to 3.2 in 2000 as the result of substantially invariant unemployment rates in the South (roughly 18 percent) coupled by a declining pattern in the Centre-North (from 8 to 6 percent). Over the current decade, instead, we observe a slight reduction in the North-South divide, which has led to a ratio of 2.7 in 2007.

In order to understand the regional unemployment dynamics and its relationship with migration, we use data at the NUTS-3 level (provinces). Figure 1 shows the snapshot densities of provincial relative unemployment rates in 1995 (solid line) and in 2007 (dashed line) computed by using a local linear estimator with variable bandwidth selected by generalized cross-validation (Loader, 1996). It emerges an unimodal right-skewed distribution of provincial unemployment rates in 1995, with a higher density for values lower than the national average. The distribution of provincial unemployment in 2007 appears markedly different. We observe a vanishing of the mass around the national average and a corresponding tendency towards polarization, with the main peak much more pronounced than in 1995 and a second lower peak at 1.5 times the national average. Specifically, only one third of Southern provinces shows a reduction of unemployment rates like the one observed for the North, with the remaining Southern provinces entrapped in a condition of high unemployment.

Figure 1

Figure 2 shows the quartile distribution of regional unemployment rates in 1995 and 2007 along with the one of regional migration rates. We observe that unemployment rates at the NUTS-3 level in Italy are closely mirrored by huge migration flows, with high unemployment areas characterized by huge out-migration and *viceversa*. Despite the clear picture emerging from the maps in Figure 2, we also document a strong heterogeneity across spatial units in terms of unemployment rate dynamics, although the lowest growth rates are all in the Centre-North (Figure 3).

Figure 2

Figure 3

During the period 1995-2007, we also document a sort of convergence for spatial units with initial unemployment rates less than 1.8 times the national average, while a slightly diverging pattern characterizes high unemployment areas (Figure 4). Furthermore, the estimated smooth effect of migration on regional unemployment dynamics (Figure 5) shows a U-shaped relationship, which may hide the existence of overlapping forces: when positive and higher than a certain threshold, migration rates seem to increase unemployment dynamics in a way consistent with the neoclassical predictions; in contrast, when negative or lower than the threshold, they do not appear to be able to reduce unemployment dynamics. Specifically, we observe that Southern provinces with higher out-migration are also those with higher unemployment growth rates. Such a pattern is at odds with the idea of migration acting as a re-equilibrating force and supports the predictions of the NEG-styled model by Epifani and Gancia (2005). Obviously, the evidence in Figure 5 can be affected by an identification problem (endogeneity and/or omitted variables). The rest of the paper is devoted to illustrate the econometric model we use in order to analyze the issue at stake.

Figure 4

Figure 5

4. Empirical framework

4.1. Unemployment dynamics and its determinants

In order to assess the effect of migration on regional unemployment disparities in local labour markets, we use longitudinal data for 103 NUTS3 Italian regions and four periods (1995-1998, 1998-2001, 2001-2004 and 2004-2007) to construct our dependent variable, $y = \Delta \ln u$, which measures the three-year dynamics of the provincial unemployment rate, u, in terms of log-difference. All data are taken from the Italian National Institute for Statistics (ISTAT).[4]

Migration. The migration rate (measured as the average net migration balance/total population ratio over each time period, $migr$) constitutes the key causative determinant of regional unemployment rates dynamics in our study. As discussed in Section 2.1 above, no clear-cut predictions can be made on the effects of the migration rate. Relying on the neo-classical view of homogenous labour, workers moving towards prosperous regions may help reduce regional differences in unemployment, leading to a positive effect of $migr$. In the presence of agglomeration forces, we can expect a negative effect of migration on the dynamics of regional unemployment. Thus, assessing the ultimate effect of $migr$ crucially depends on whether supply effects dominate over demand ones, or *vice-versa*.

Other determinants. In keeping with the existent empirical literature, the dynamics of regional unemployment rates is likely to depend on additional factors such as a) local labor market dynamics and b) local economic structures.[5] As for the first class of determinants, we include in the set of regressors the supply-demand mismatch, measured as the average employment growth rate less the labor participation growth rate over each time period, Δeld. Its expected effect is negative almost by definition.

The second set of regressors includes initial conditions (the logarithm of the unemployment rate at the beginning of each period, $\ln u$), the industry mix (the logarithm of the share of agriculture, manufacturing, construction and services employment on total employment at the beginning of each period, $\ln agr$, $\ln man$, $\ln con$ and $\ln ser$, respectively),[6] human capital defined as skill-composition of regional labor forces (the logarithm of the share of adults with upper secondary education at the beginning of each period, $\ln hc$), unit labour cost (the logarithm of the ratio between real wages and labour productivity at the beginning of each period, $\ln ulc$). The expected sign for $\ln u$ and $\ln hc$ is negative: higher initial conditions should imply lower growth rates; highly skilled workers are likely to be more efficient in job search and are less likely to be laid off. Higher unit labor costs are expected to exert a negative effect on labour market performances, so that we expect a positive impact of $\ln ulc$ on the response variable. As for the industry mix, economic intuition suggests that regions specialized in declining economic sectors (such as agriculture) are expected to exhibit larger structural unemployment rates than provinces with production based on manufacturing or services (Elhorst, 2003).

4.2 A panel data approach with spatial interaction effects and nonlinearities

Our starting point is the following parametric specification:

$$y_{it} = \beta X_{it} + \alpha_i + \varepsilon_{it} \qquad i = 1,...,N = 103; \qquad t = 1,...,T = 4 \qquad (1)$$

where i denotes the cross-sectional dimension and t indexes time; X_{it} is a vector of covariates; β is a vector of fixed unknown parameters associated to the covariates, ε_{it} is an independently and identically distributed (*iid*) error term for i and t with zero mean and variance σ^2, while α_i denotes province fixed effects so as to control for all space-specific time-invariant variables whose omission could bias the estimates.

Spatial interaction effects. Stylized facts discussed in Section 3 give pervasive evidence of spatial clustering in provincial unemployment rates. Empirical literature on regional economics has recently shown a growing interest in the possibility to test for spatial interaction (or spatial dependence) effects in standard static linear panel data models (Elhorst, 2009; Kapoor et al., 2007). Within a panel data framework, spatial autoregressive models can be motivated from two different points of view (LeSage and Pace, 2009). First, spatial dependence may help capture the role of externalities arising from neighborhood characteristics. Second, spatial autocorrelation may act as a proxy for omitted variables clustered in space (Niebuhr, 2002).

Two customary specifications are the spatial lag and the spatial error models. The spatial lag or spatial autoregressive (SAR) model includes the dependent variable observed in neighbouring units as an additional regressor with respect to model (1):

$$y_{it} = \delta \sum_{j=1}^{N} w_{ij} y_{jt} + \beta X_{it} + \alpha_i + \varepsilon_{it} \qquad (2)$$

where δ is the spatial autoregressive coefficient and $W = \{w_{ij}\}_{i \neq j}$ is a pre-specified non-negative square matrix of order *N* collecting spatial weights, w_{ij}, which describe the spatial arrangement of the units in the sample. In such a specification, the unemployment rate dynamics in a given location will be affected not only by its exogenous characteristics (e.g. the migration rate) and by its idiosyncratic shocks (ε_{it}), but also by those in all other regions through the inverse spatial transformation $(I - \delta W)^{-1}$.

The spatial error model (SEM) relaxes the assumption of *iid* errors by allowing for their correlation across space. Using the same notation as above, the SEM can be written as:

$$y_{it} = \beta X_{it} + \alpha_i + \phi_{it} \tag{3}$$

$$\phi_{it} = \rho \sum_{j=1}^{N} w_{ij} \phi_{jt} + \varepsilon_{it}$$

where ϕ_{it} denotes the spatially autocorrelated error term and ρ the spatial autocorrelation coefficient.

The choice between fixed effects (FE) and random effects (RE) for models (2) and (3) must be conducted by means of standard Hausman's specification tests. The choice between SAR and SEM could rely on robust Lagrange multiplier (LM) tests. Notice, however, that models (2) and (3) are nested in a more general specification known as the unconstrained spatial Durbin model (SDM) whose reduced form implies the existence of substantive spatial externalities:

$$y_{it} = \delta \sum_{j=1}^{N} w_{ij} y_{jt} + \beta X_{it} + \gamma \sum_{j=1}^{N} w_{ij} X_{jt} + \alpha_i + \varepsilon_{it} \tag{4}$$

The hypothesis $H_0 : \gamma = 0$ can be tested to assess whether SDM collapses to the SAR, while the 'common factor' hypothesis $H_0 : \gamma + \delta\beta = 0$ can be used to verify whether model (4) reduces to the SEM.

Estimation of models (2)-(4) can be carried out through maximum likelihood (ML) or two-stage least square/generalized method of moments (2SLS/GMM) techniques. Both methods assume that ε_{it} are *iid* for all i and t, but only ML estimators rely on the assumption of normality of the errors. Moreover, while δ and ρ are bounded from below and above using ML, they are unbounded using 2SLS/GMM.

An advantage of using 2SLS/GMM consists of the possibility to properly model endogeneity issues (Kelejian and Prucha, 1998): in particular, the first-difference (FD) 2SLS estimator allows using weakly exogenous instruments, while the 2SLS estimation of the FE model leads to inconsistent estimation of β's if the instruments are not strongly exogenous.

Nonlinearities. In models (1)-(4), we treat all terms as globally linear. Such a restriction may lead to biased estimates of the parameters if the data generating process obeys a more articulated specification. For instance, Overman and Puga (2002) introduce a quadratic term in the initial conditions in order to capture some nonlinearities. Unlike these authors, we adopt a semiparametric approach so as to avoid imposing arbitrary functional forms. Semiparametric models for panel data have recently been developed by Li and Ullah (1998), Ullah and Roy (1998), Ullah and Mundra (2001) and Mundra (2005). A semiparametric version of model (4) is:

$$y_{it} = \alpha_i + X_{it}^{*}\beta^{*} + \delta\sum_{j=1}^{N} w_{ij} y_{jt} + f_1\left(x_{1it}\right) + f_2\left(x_{2it}\right) + \ldots + \varepsilon_{it} \qquad (5)$$

where $f_j\left(..\right)$ are unknown smooth functions of the covariates, X_{it}^{*} is a vector of strictly parametric components and β^{*} the corresponding parameter vector. For each k-th smooth term, the estimated function $\hat{f}_k\left(..\right)$ reveals possible nonlinearities in the effect of x_k.[7]

In order to deal with endogeneity problems in the estimation of nonparametric models, we use the procedure proposed by Blundell and Powell (2003), which consists of extending the "control function" method to semiparametric models through a two-step procedure. Considering, for the ease of exposition, only the endogeneity of the spatial lag of the dependent variable, the first step consists of an auxiliary nonparametric regression as:

$$\sum_{j=1}^{N} w_{ij} y_{jt} = \alpha_i + X_{it}^{*}\beta^{*} + f_1\left(x_{1it}\right) + f_2\left(x_{2it}\right) + h(Z_{it}) + \ldots + \upsilon_{1it} \qquad (6)$$

where Z_{it} is a set of conformable instruments and υ_{1it} a random variable satisfying $E(\upsilon_{1it} \mid Z_{it}) = 0$. Moreover, if Z_{it} and ε_{it} are independent, then it yields that $E(\varepsilon_{it} \mid \upsilon_{1it}, Z_{it}) = E(\varepsilon_{it} \mid \upsilon_{it})$ and, thus, $E(\varepsilon_{it} \mid \sum_{j=1}^{N} w_{ij} y_{it}) \neq 0$ when $E(\varepsilon_{it} \mid \upsilon_{1it}) \neq 0$. The second step consists of estimating an additive model of the form:

$$y_{it} = \alpha_i + X_{it}^{*}\beta^{*} + \delta\sum_{j=1}^{N} w_{ij} y_{jt} + f_1\left(x_{1it}\right) + f_2\left(x_{2it}\right) + \ldots + \hat{\upsilon}_{1it} + \varepsilon_{it} \qquad (7)$$

Obviously, in the presence of a number of candidate endogenous terms (for instance, $migr$ and Δeld), different first steps like in (6) - *mutatis mutandis* - are estimated and the corresponding residuals $\hat{\upsilon}$'s are introduced as additional regressors in the second step (7).

5. Estimation results

Table 1 reports the econometric results of a number of alternative parametric specifications. Albeit not reported in the table, all models include fixed time effects so as to capture possible effects of un-modeled factors such as business cycles developments.[8] In Model (1), the migration rate ($migr$) has a negative impact on regional unemployment growth, suggesting that demand side effects dominate over supply side effects, in line with Epifani and Gancia (2005) argumentations. This result gives empirical support to the idea that workforce outflows worsen local labour market performances, exacerbating the divide between backward areas and the rest of the country. As expected, a higher excess labour demand growth rate (Δeld) lowers regional unemployment dynamics. Moreover, WG-OLS estimates advise that Italian provinces with a higher initial unemployment ($\ln u$) and human capital levels ($\ln hc$) are more likely to reduce the unemployment rate than the other provinces, *ceteris* *paribus*. Finally, the remaining covariates ($\ln ulc$ and sector shares of employment over total employment) do exert a negligible role in explaining unemployment dynamics. As for diagnostics, LM tests clearly indicate the existence of sizable spatial dependence. Moreover, the Moran *I* plot (Figure 6.a) shows a positive relationship between residuals (horizontal axis) and their spatial lag (vertical axis).

Table 1

Figure 6

Including spatial interaction effects. On the ground of these findings, resorting to spatial econometric tools appears to be a proper modelling approach in order to better investigate on the dynamics of provincial labour market performances. ML estimates of the SDM specification (Model 2) are consistent with WG-OLS FE results, except for $\ln agr$ which turns out to be clearly significant.[9] Furthermore, there is strong evidence of spatial dependence: the Wy term is statistically significant and signals the presence of global spatial *spillover* in the labour market. This implies that the exogenous characteristics of province i (for example, its level of out-migration) or an idiosyncratic *shock* in that province do not only influence the unemployment dynamics in that location, but affect also the outcome of all other regions with an intensity that decreases with distance (Anselin, 2004). Testing for the SDM against more restricted specifications (namely SAR or SEM) leads to strongly reject the null: the LR statistics for the joint exclusion of spatial lagged terms (except for Wy) as well as for the common factor test indicate that the fully unconstrained SDM appears to be the most reliable specification.

Model (3) collects the estimation results for a more restricted specification of the fully unconstrained SDM where only statistically significant coefficients at the 10 percent level, or better, are retained. Moving from Model (2) to its parsimonious version is corroborated by the LR test: the χ^2-distributed test statistics with 12 degrees of freedom turns out to be equal to 9.4 with an associated p-value of 0.668. The restricted model leads to qualitatively similar conclusions and documents significant effects for just a spatially lagged term ($W\Delta eld$). The positive sign of its coefficient reflects a sort of spatial competition among provinces: regions surrounded by other areas with high excess demand growth rates exhibit higher unemployment growth rates, *ceteris paribus*. Finally, both SDM specifications appear to be able to remove spatial dependence as Moran I plots in Figure 6.b and Figure 6.c show.

Controlling for endogeneity. The ML procedure can take the bias generated by the presence of the endogenous term Wy into account, under the assumption of exogeneity of the other regressors. In our case, however, such an assumption for $migr$ and Δeld (and, thus, for their spatial lags) are clearly violated. The decision to migrate depends indeed on the observed unemployment rate, generating a simultaneity problem. Furthermore, as the employment rate and the participation rate have common components with the dependent variable by construction, a second endogeneity problem is likely to emerge. In order to correct such biases, a FD-2SLS estimation approach is employed. Model (4) is the FD-2SLS counterpart of our preferred ML estimation (Model 3), where irrelevant regressors ($\ln man$, $\ln ser$, $\ln cos$, $\ln ulc$, $W \ln u$, $Wmigr$, $W \ln agr$, $W \ln man$, $W \ln ser$, $W \ln cos$, $W \ln ulc$ and $W \ln hc$) have been excluded from the model for the sake of parsimony. Although the estimation results are qualitatively similar with respect to those for Model (3), the coefficients of the four endogenous variables (Wy, $migr$, Δeld, $W\Delta eld$) get larger in absolute value.

For the application of the FD-2SLS the choice of instruments represents a key issue. As for the endogeneity of the spatially lagged term (Wy), the econometric literature suggests to include within the set of instruments linearly independent spatial lags of the exogenous variables (Kelejian and Prucha, 1998); the literature on internal migration recommends to consider as determinants of regional migration rates ($migr$) the disposable income, the unemployment rate, the age structure of the population, the average house price (and/or the average rental price), environmental variables as well as the time and spatial lags of migration rates themselves (Basile and Causi, 2005; Etzo, 2008); finally, the panel data literature advises to use as internal instruments in a FD-2SLS estimation approach the time lags of exogenous and endogenous variables. Taking into account these different pieces of literature, we construct a large set of instruments. Firstly, it includes the time lag of the: four endogenous variables ($L.migr$, $L.\Delta eld$, $L.W\Delta eld$, $L.\Delta \ln u$), age structure of the population ($L.pop15-29$, $L.pop30-64$), disposable income ($L.inc$), average provincial house price ($L.p-house$), industrial structure variables ($L.\ln agr$, $L.\ln man$, $L.\ln ser$, $L.\ln cos$), other predetermined variables of the model ($L.\ln u$, $L.\ln hc$). Secondly, the set of instruments includes the time lag of the spatial lag of the: migration rate ($L.Wmigr$), disposable income ($L.Winc$),age structure of the population ($L.Wpop15-29$, $L.Wpop30-64$), average provincial house price ($L.Wp-house$), industrial structure variables ($L.W\ln agr$, $L.W\ln man$, $L.W\ln ser$, $L.W\ln cos$), other predetermined variables of the model ($L.W\ln u$, $L.W\ln hc$). Diagnostics tests indicate that the null of excludability of these instruments from the first step is strongly rejected. Moreover, the Sargan test of over-identifying restrictions does not indicate correlation between the instruments and the error term. Finally, residual spatial dependence seems to be properly removed as shown by the Moran I plot in Figure 6.d.

A semiparametric specification. In order to relax unnecessarily restrictive assumptions on the functional form, we estimate a semiparametric version of the SDM which also takes the endogeneity for $migr$, Wy, Δeld and $W\Delta eld$ into account using the same set of instruments employed to estimate Model (4) to apply the control function approach. After considerable experimentation, we present a semi-parametric specification admitting the linearity constraint for $migr$ and $\ln agr$ (Model (5) in Table 2). Both parametric and nonparametric terms are statistically significant. The joint significance of the first-step residuals from the auxiliary regressions for $migr$, Wy, Δeld and $W\Delta eld$ (F test – control functions) corroborates our choice of using the control function approach to control for endogeneity of those terms. The adjusted R^2 confirms some improvements with respect to the fully-linear parametric counterpart. This is also confirmed by the F-test for the equivalence of the linear specification model and its semiparametric counterpart which indicates strong rejection of the null. Finally, there is no evidence of residual spatial dependence (Figure 6.e). Linear terms are very close in magnitude with respect to their counterparts in Model (4). In the rest of the Section we discuss the effects of nonlinear terms focusing on Model (6).

Table 2

Graphs in Figure 7 show the fitted univariate smooth functions (solid line) for Model (5), alongside Bayesian confidence intervals (shaded gray areas) at the 95 percent level of significance, computed as suggested by Wood (2004). In each plot, the vertical axis displays the scale of the expected values of unemployment dynamics, while the horizontal one reports the scale of initial conditions (Figure 7.a), Wy (Figure 7.b), excess demand growth rates (Figure 7.c), human capital (Figure 7.d) and the spatial lag of excess demand growth rates (Figure 7.e). We observe that nonlinearity in all terms but $\ln hc$ and Wy are mainly due to an inflection point in correspondence of the origin of the horizontal axis. In all terms, we also observe a monotonic pattern consistently with the signs of the coefficients estimated in the parametric counterparts. Estimation results document the existence of strong nonlinear spatial dependence: the *edf* of $f(Wy)$ is higher than 2. It is worth mentioning from Figure 7.b that local elasticities are always lower than one (the slope of the curve is lower than that of the 45-degree diagonal) and, thus, explosive spatial multiplier effects are excluded in line with the literature on spatial autoregressive models (Anselin, 1988; LeSage and Pace, 2009). As for human capital, the plot indicates a clear downward pattern for $\ln hc$ only after a threshold. Before such a minimum level of human capital accumulation, $\ln hc$ has no effect on unemployment dynamics, since the confidence intervals include the horizontal axis. This result may suggest that small investments in human capital do not exert any significant effect on regional unemployment and only a large effort guarantees benefits in terms of labour market performances.

Figure 7

6. Concluding remarks

This paper aims at assessing whether interregional migration flows equilibrate local labour market performances in Italy. We focus on regional unemployment dynamics at a very fine territorial level (103 provinces or NUTS-3 regions) over the 1995-2007 period, during which a strong flow of out-migration from the South towards the North occurred. We propose an empirical framework for panel data models in the presence of spatial dependence allowing for possible nonlinearities which innovates along several dimensions with respect to the existent literature. Our results are at odds with the neoclassical view of migration acting as an equilibrating force for unemployment differentials. Empirical estimates from a number of alternative specifications document that migration has in fact a negative effect on unemployment growth rates. This evidence can be rationalized within a theoretical framework where agglomeration forces are at work so that migration flows magnify spatial disparities in unemployment rates rather than mitigate them as pointed out by Epifani and Gancia (2005).

Possible improvements of the research agenda may include a closer look at migration flows disaggregated by levels of schooling. To the extent that labour is not homogenous and that migration propensity increases sharply with education (Greenwood, 2009), new comers will enhance productivity (Chalmers and Greenwood, 1985; Ghatak et al., 1996) and will improve the attractiveness of new (domestic and foreign) firms in the host regions. Under a NEG-type framework, the balance between the supply-side effects and the demand-side effects is expected to be even more negative, since under these circumstances it is harder for the regions of origin to attract investment flows. This is likely to be the case for Italy: as recently pointed out by Mocetti and Porello (2010), the migration out-flows from Southern regions towards the rest of Italy have been particularly relevant for high-skilled workers. As a result, Southern regions appear to be unable to preserve their own human capital with unavoidable detrimental effects not only for local labour market performances but also for long-run local growth. Testing for brain drain effects on interregional are left for future research.

References

Alesina A., Danninger S. and Rostagno M. (1999), "Redistribution through Public Employment", *IMF Working Paper*, n. 177

Anselin L. (2004), "Spatial Externalities, Spatial Multipliers and Spatial Econometrics", *International Regional Science Review*, 26, pp. 153-166

Attanasio O. and Padoa-Schioppa F. (1991), "Regional inequalities, migration and mismatch in Italy, 1960-86", in F. Padoa-Schioppa (ed.), *Mismatch and labor mobility.* Cambridge, Cambridge University Press.

Bande R. and Karanassou M. (2007), "Labour Market Flexibility and Regional Unemployment Rate Dynamics: Spain 1980-1995", *IZA Discussion Papers*, 2593

Basile R. and Causi M. (2007), "Le determinanti dei flussi migratori nelle province italiane: 1991-2000", *Economia e Lavoro*, 2, pp. 139-159

Blanchard O.J. and Katz L.F. (1992), "Regional evolutions", *Brooking Papers on Economic Activities*, 1, pp. 1-75

Blundell R. and Powell J. (2003), "Endogeneity in Nonparametric and Semiparametric Regression Models", in M. Dewatripont, L. Hansen and Turnsovsky S.J. (eds.), *Advances in Economics and Econometrics,* Cambridge: Cambridge University Press

Brunello G., Lupi C. and Ordine P. (2001), "Widening Differences in Italian Regional Employment", *Labour Economics*, 8, pp. 103-129

Cannari L., Lucci F. and Sestito P. (2000), "Geographic Labour Mobility and the Cost of Housing: Evidence from Italy", *Applied Economics*, 132, pp. 1899-1906

Chalmers J.A. and Greenwood M.J. (1985), "The Regional Labor Market Adjustment Process: Determinants of Changes in Rates of Labor Force Participation, Unemployment, and Migration", *The Annals of Regional Science,* 19, pp. 1-17

Cipollone P. and Guelfi A. (2006), "The Value of Flexible Contracts: Evidence from an Italian Panel of Industrial Firms", *Banca d'Italia, Temi di discussione*, n. 583

Decressin J. and Fatàs A. (1995), "Regional Labour Market Dynamics in Europe", *European Economic Review*, 39 (9), pp. 1627-1655

Elhorst J.P. (1995), "Unemployment Disparities between Regions in the European Union", in H. W. Armstrong and R. W. Vickerman (eds.), *Convergence and Divergence among European Unions*, London: Pion

Elhorst J.P. (2003), "The Mystery of Regional Unemployment Differentials: Theoretical and Empirical Explanations", *Journal of Economic Surveys*, 17, pp. 709-748

Elhorst J.P. (2009), "Spatial Panel Data Models", in Fischer M.M. and Getis A. (eds.) *Handbook of Applied Spatial Analysis*, Ch. C.2., Springer: Berlin Heidelberg New York

Epifani P. and Gancia G.A. (2005), "Trade, Migration and Regional Unemployment", *Regional Science and Urban Economics*, 35, pp. 625-644

Etzo I. (2008), "Internal migration: a review of the literature", *MPRA Paper* No. 8783

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*, 4, pp. 571-579

Ghatak S., Levine P. and Price S.W. (1996), "Migration Theories and Evidence: An Assessment", *Journal of Economic Surveys,* 10, pp. 159-198

Greenwood M.J. (2009), "Some potential new directions in empirical migration research", *Rivista di Scienze Regionali* (*Italian Regional Science Review*), forthcoming

Groenewold N. (1997), "Does migration equalize regional unemployment rates? Evidence from Australia", *Papers in Regional Science,* 76(1), pp. 1-20

ISAE (2007), “Modifiche istituzionali e trasformazioni strutturali nel mercato del lavoro italiano”, in Rapporto ISAE, *Le previsioni per l’economia italiana. L’Italia nell’integrazione europea*, marzo

Kapoor M., Kelejian H. and Prucha I.R. (2007), “Panel Data Models with Spatially Correlated Error Components”, *Journal of Econometrics*, 140, pp. 97-130

Kelejian H.H. and Prucha I.R. (1998), “A Generalized Spatial Two Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances”, *Journal of Real Estate Finance and Economics*, 17, pp. 99-121

Kertesi, G. (2000), “Migration and Commuting: Two Potential Forces Reducing Regional Inequalities in Economic Opportunities?", *SOCO Project Paper 77a*, Vienna: Institute for Human Sciences

Kostoris-Padoa-Schioppa F. and Basile R. (2002), “Unemployment Dynamics in the ‘Mezzogiornos of Europe’: Lessons for the Mezzogiorno of Italy”, *CEPR Discussion Paper,* 3594

Leonardi M. (2004), “The Persistence of Italian Regional Unemployment: the Role of Migration and Wages”, mimeo

LeSage J.P. e Pace R.K. (2009), *Introduction to Spatial Econometrics*, CRC Press Taylor & Francis Group, Boca Raton, London, New York

Li Q. and Ullah A. (1998), “Estimating Partially Linear panel Data Models with one-way Error Components”, *Econometric Reviews* 17(2), pp. 145-166

Loader C.R. (1996), “Local likelihood density estimation”, *The Annals of Statistics*, 24, pp. 1602–1618

Mocetti S. and Porello C. (2010), “La mobilità del lavoro in Italia: nuove evidenze sulle dinamiche migratorie”, *Questioni di Economia e Finanza (Occasional papers)*, Banca d’Italia, n. 61

Mundra K. (2005), "Nonparametric Slope Estimators for Fixed-Effect Panel Data", *mimeo*

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

Niebuhr A. (2003), "Spatial Dependence of Regional Unemployment in the European Union", *HWWA Discussion Paper*, 186

Overman H.G. and Puga D. (2002), "Unemployment Clusters across Europe's Regions and Countries", *Economic Policy*, 34, pp. 116-147

Pissarides C.A. and McMaster I. (1990), "Regional Migration, Wages and Unemployment: Empirical Evidence and Implications for Policy", *Oxford Economic Papers*, 42, pp. 812-831

Prasad E.S. and Utili F. (1998), "The Italian Labor Market: Stylized Facts, Institutions and Directions for Reform", *IMF Working Paper*

Rutkowski, J. and M. Przybila (2002), "Poland: Regional Dimensions of Unemployment", in Funck, B. and L. Pizzati (eds.), *Labour, Employment and Social Policy in the EU Enlargement Process. Changing Perspectives and Policy Options*, World Bank, Washington, D.C., Chapter XIV, 157-175

Taylor J. and Bradley S. (1997), "Unemployment in Europe: A Comparative Analysis of Regional Disparities in Germany, Italy and UK", *Kyklos*, 50, pp. 221-245

Ullah A. and Mundra K. (2001), "Semiparametric Panel Data Estimation: An Approach to Immigrant Homelink Effect on U.S. Producer Trade Flows", in *Handbook of Applied Econometrics and Statistical Inferences*, Marcel Dekker

Ullah A. and Roy N. (1998), "Parametric and Nonparametric Panel Data Models", in Ullah A. and Giles D.E.A. (eds.), *Handbook of Applied Economics and Statistics*, Marcel Dekker: New York, 1, pp. 579-604

Van Dijk J. and J. Oosterhaven (1986), "Regional Impacts of Migrants' Expenditures: An Input-Output/Vacancy-Chain Approach," in Batey, P.WJ, and M. Madden (eds.), *Integrated Analysis of Regional Systems* (London Papers in Regional Science),. Pion Press: London

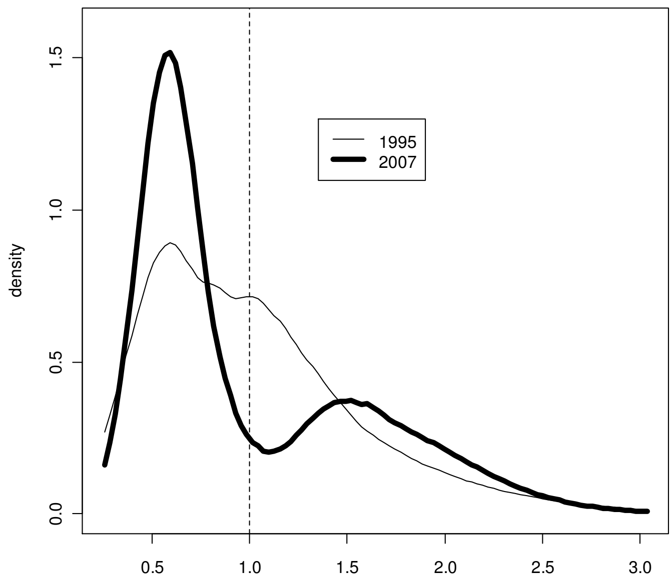
Wood S.N. (2000), "Modelling and Smoothing Parameter Estimation with Multiple Quadratic Penalties", *Journal of the Royal Statistical Society Series B*, 62(2), pp. 413-428

Wood S. N. (2004), "Stable and efficient Multiple Smoothing Parameter estimation for Generalised Additive Models", *Journal of the American Statistical Association*, 99, pp. 673-686

Wood S.N. (2006), *Generalized Additive Models. An Introduction with R*, Boca Ratom: Chapman & Hall/CRC

Wrage P. (1981), "The regional effects of internal migration on regional wage and unemployment disparities in Canada", *Journal of Regional Science*, 21(1), pp. 51-63

Figure 1 – Density estimation of provincial unemployment rates: 1995 and 2007



Notes: Provincial unemployment rates have been normalized with respect to the national average

Figure 2 – Quartile distribution of regional unemployment and migration rates

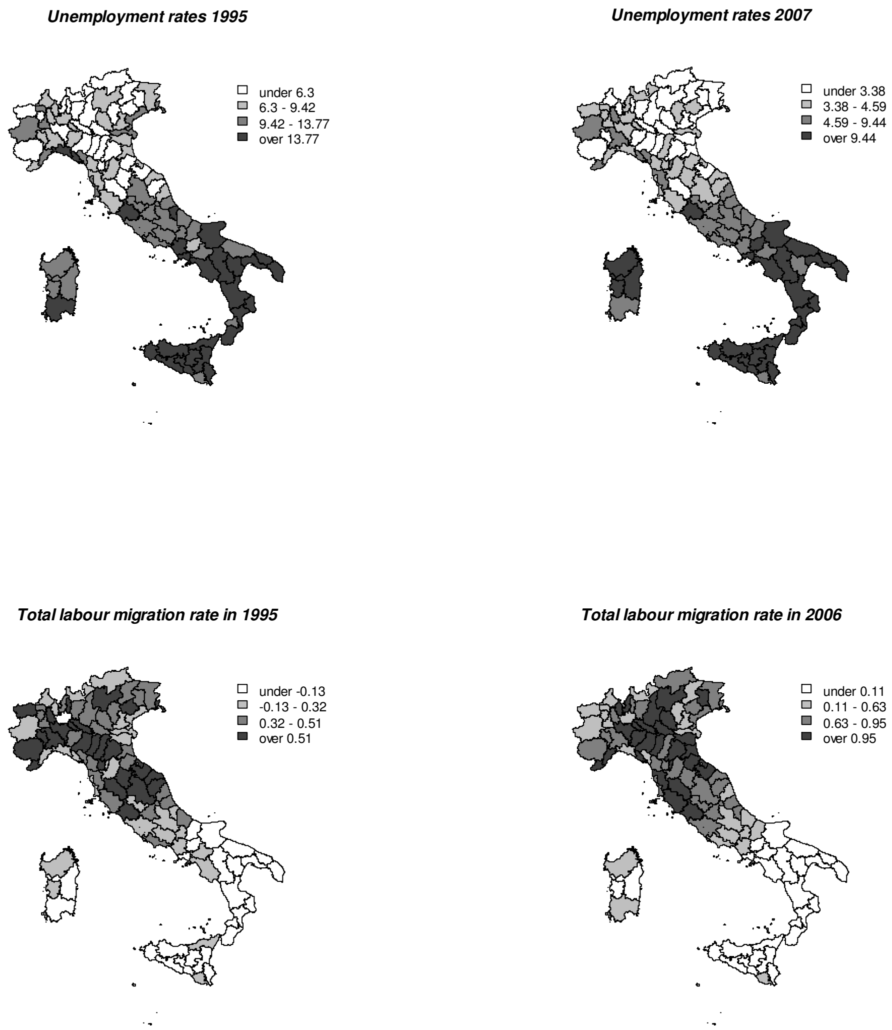


Figure 3 - Unemployment growth rates

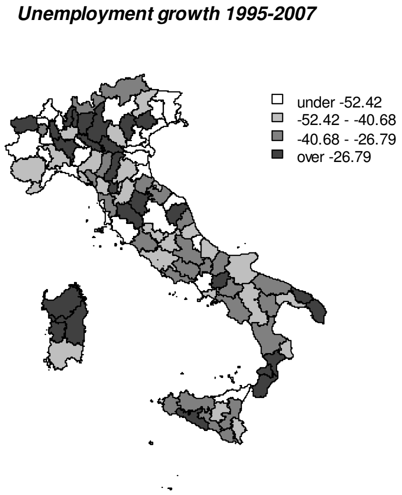


Figure 4 – Initial conditions and unemployment dynamics

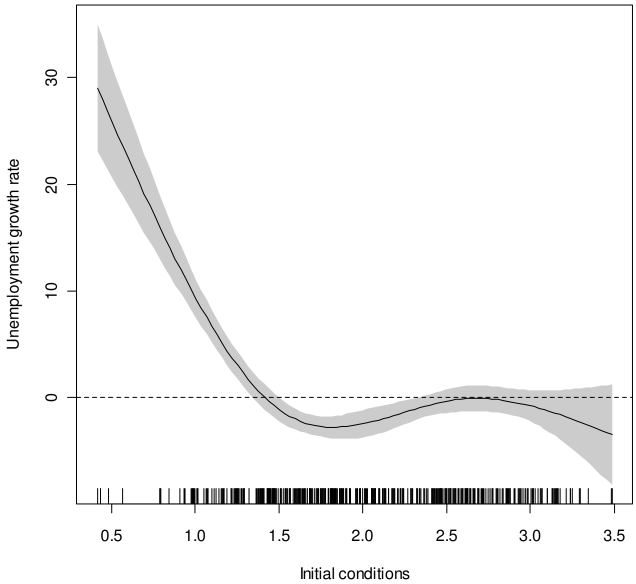


Figure 5 - Migration rates and unemployment dynamics

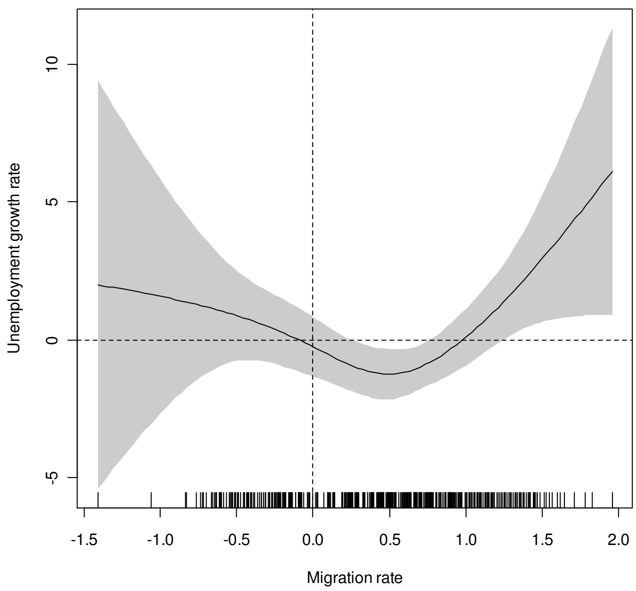
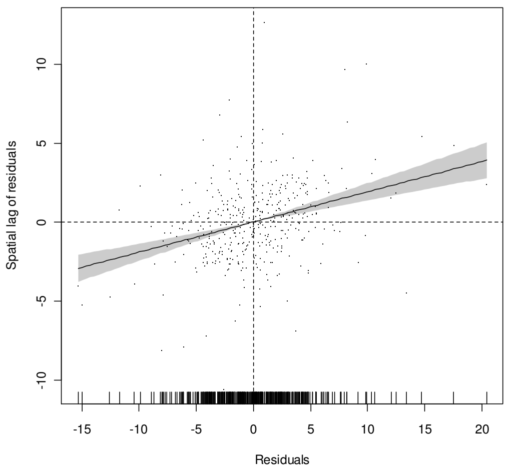
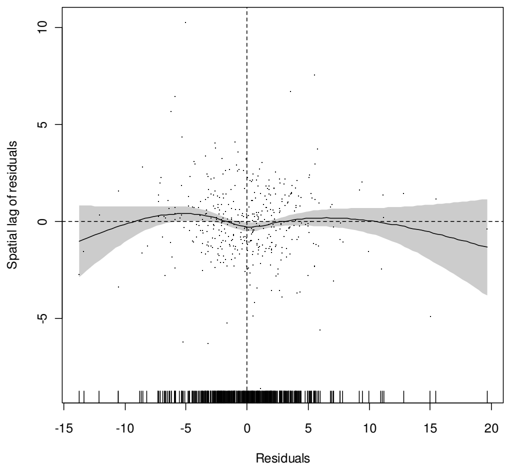


Figure 6 - Moran I Plots

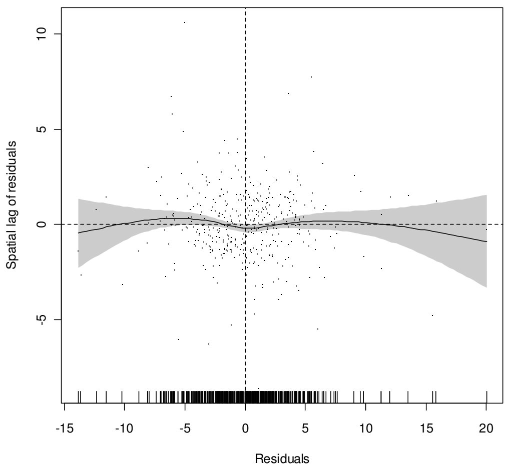
a. Model (1)



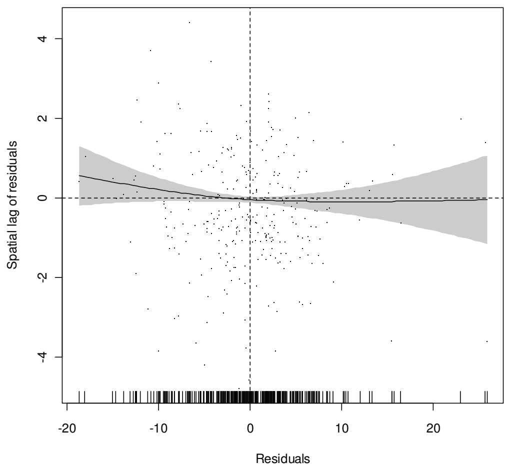
b. Model (2)



c. Model (3)



d. Model (4)



e. Model (5)

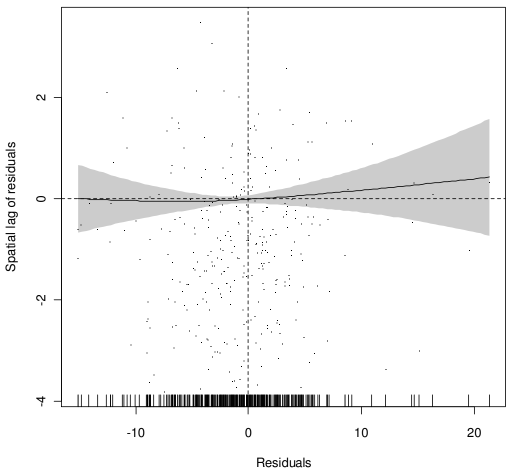
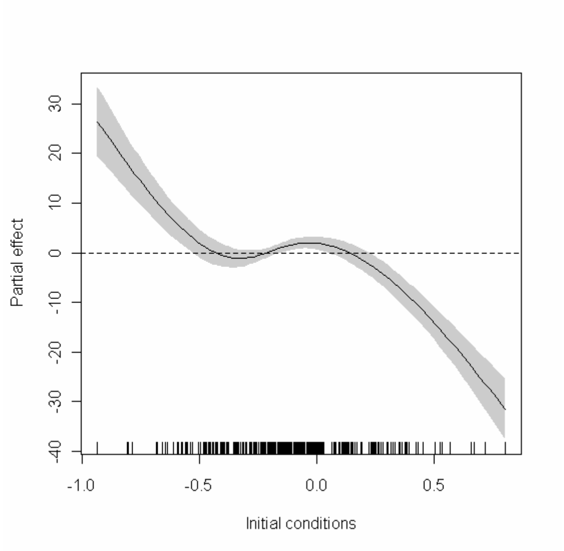
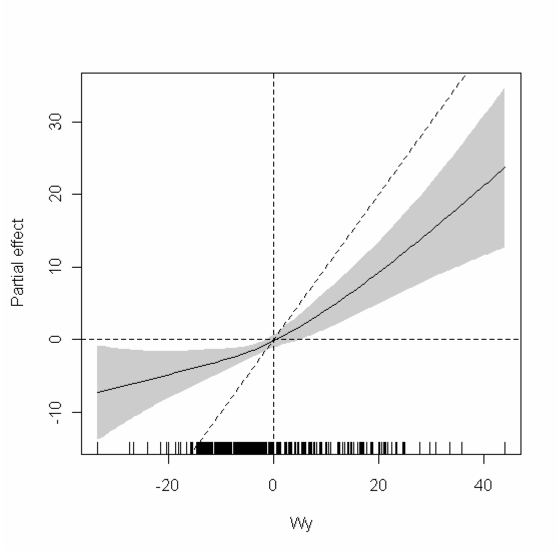


Figure 7 – Partial effects of the univariate smooth term

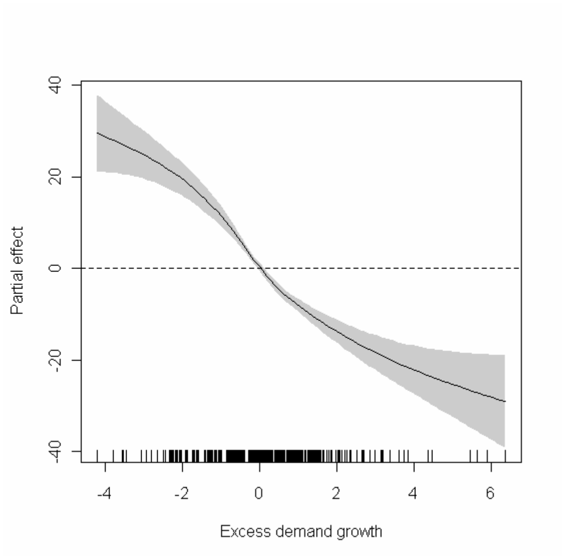
a) $f(\ln u)$



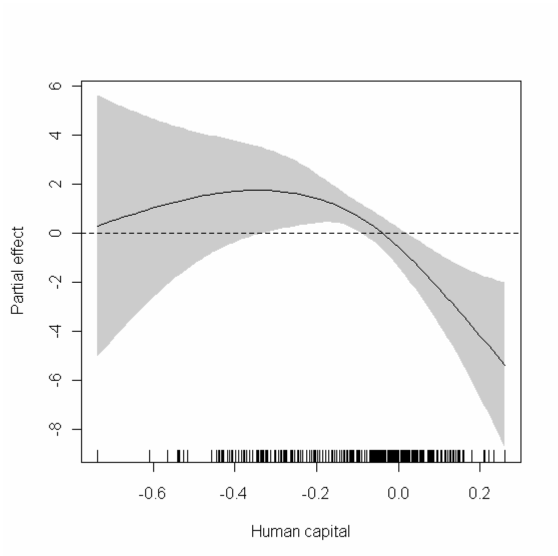
b) $f(Wy)$



c) $f(\Delta eld)$



d) $f(\ln hc)$



e) $f(W\Delta eld)$

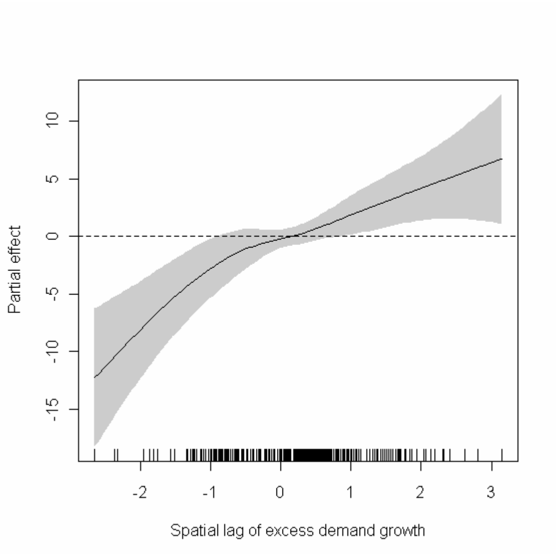


Table 1 – Estimation results

Variables	(1) FE WG-OLS	(2) FE-SDM ML	(3) FE-SDM ML	(4) FD-SDM 2SLS
$\ln u$	-18.178 (0.000)	-16.221 (0.000)	-15.985 (0.000)	-7.276 (0.040)
$migr$	-5.283 (0.001)	-4.366 (0.012)	-4.825 (0.002)	-9.176 (0.010)
Δeld	-4.348 (0.000)	-4.611 (0.000)	-4.517 (0.000)	-9.379 (0.000)
$\ln agr$	6.385 (0.094)	7.524 (0.015)	4.146 (0.085)	6.723 (0.003)
$\ln man$	2.862 (0.484)	0.346 (0.943)	.	.
$\ln ser$	25.570 (0.125)	19.597 (0.177)		
$\ln cos$	2.046 (0.533)	0.418 (0.894)	.	.
$\ln ulc$	4.628 (0.702)	-2.114 (0.834)	.	.
$\ln hc$	-7.332 (0.000)	-9.776 (0.003)	-9.484 (0.003)	-11.360 (0.008)
Wy	.	0.348 (0.000)	0.354 (0.000)	0.564 (0.000)
$W \ln u$	.	0.339 (0.918)	.	.
$Wmigr$	.	-0.618 (0.824)	.	.
$W \Delta eld$	.	1.768 (0.012)	1.937 (0.001)	4.789 (0.000)
$W \ln agr$	.	-5.278 (0.421)	.	.
$W \ln man$	.	18.729 (0.169)	.	.
$W \ln ser$	.	21.251 (0.555)	.	.
$W \ln cos$	.	5.033 (0.462)	.	.
$W \ln ulc$	.	10.613 (0.595)	.	.
$W \ln hc$	.	-1.272 (0.854)	.	.

(*continued*)

Statistics	(1) FE WG-OLS	(2) FE-SDM ML	(3) FE-SDM ML	(4) FD-SDM 2SLS
R-squared adj.	0.640	0.711	0.717	0.862
Hausman's test (RE vs. FE)	(0.007)	(0.000)	(0.000)	
SDM vs SEM: LR test	.	(0.030)	.	.
SDM vs SAR: LR test	.	(0.002)	.	.
LM test no spatial lag	(0.000)	.	.	.
Robust LM test no spatial lag	(0.033)	.	.	.
LM test no spatial error	(0.000)	.	.	.
Robust LM test no spatial error	(0.001)	.	.	.
Model 3 vs Model 2: LR test	.	.	(0.668)	
Hausman's Endogeneity joint test				15.919 (0.000)
Sargan test				14.398 (0.809)

Notes: the dependent variable, $y = \Delta \ln u$, is the average growth rate of the regional unemployment rate. The total number of observations is 412, the number of regions is 103 and the number of periods is 4. Heteroskedasticity-robust *p*-values (White-corrected) are in brackets. A 5NN spatial weights matrix has been used for SDM models.

Table 2 – Econometric results of semiparametric model

Variables	(5) Semiparametric FD	
Parametric terms (β and p values)		
$migr$	-7.795 (0.000)	
$\ln agr$	7.612 (0.001)	
Nonparametric terms		
	F test and p-value	edf
$f(Wy)$	7.493 (0.000)	2.249
$f(\ln u)$	48.267 (0.000)	3.925
$f(\Delta eld)$	35.238 (0.000)	3.893
$f(W\Delta eld)$	4.840 (0.001)	2.991
$f(\ln hc)$	4.080 (0.010)	2.141
R-squared adj.	0.893	
F test First step 1 (*migr*)	(0.000)	
F test First step 2 (*Wy*)	(0.000)	
F test First step 3 (Δ*eld*)	(0.000)	
F test First step 4 (*W*Δ*eld*)	(0.000)	
F test – control functions	(0.000)	
F test – linearities	(0.000)	

Notes: see Notes in Table 1. F tests are used to investigate the overall ("approximate") significance of smooth terms. *edf* (effective degrees of freedom) reflect the flexibility of the model. "F test First step" is the test for the joint exclusion of external instruments from each first step. "F test - control function" is the test for the joint exclusion of the smooth functions of the residuals from the first steps for $migr$, Wy, Δeld and $W\Delta eld$. "F test – linearities" is the test for the null of the equivalence between Model (4) and (5). Bayesian *p*-values are in brackets.

[1] In that setup, the presence of congestion costs guarantees that some workers do not leave the periphery, so that the backward area always exhibits non-zero unemployment levels.

[2] Given the empirical controversies on the role of migration flows in explaining regional unemployment differentials, the academic debate has moved on the issue of the factors hindering internal migration. For the case of Italy, see Attanasio and Padoa-Schioppa (1991) and Murat and Paba (2001), among others.

[3] In the Italian case, it is customary to distinguish between Southern regions, or interchangeably Mezzogiorno (namely, Campania, Abruzzo, Molise, Basilicata, Calabria, Puglia, Sicilia and Sardegna) and Central-Northern regions (namely, Valle d'Aosta, Piemonte, Lombardia, Trentino Alto Adige, Friuli Venezia Giulia, Veneto, Liguria, Emilia Romagna, Marche, Toscana, Lazio and Umbria).

[4] The use of three-year averages and growth rates is related to the construction of the migration variable, which is measured in terms of cancellation from and registration in another province. Since it is customary (at least in Italy) for migrants to formally change the registration only some years after the actual departure from the region of origin, annual data on migration are unlike to properly measure workers mobility. Using three-year averages help reduce such a bias.

[5] Elhorst (2003) gives a comprehensive description of the variables included in recent empirical analyses on regional unemployment differentials.

[6] Notice that a finer sectoral classification would be advisable for this kind of analysis as pointed out by Elhorst (2003). Unfortunately, more articulated sectoral data are only recorded over decades (Census data) and, thus, cannot be used for our purposes.

[7] Wood (2000, 2006) has recently proposed a method to estimate semiparametric additive models with penalized regression smoothers which allows for automatic and integrated smoothing parameters selection. He has also implemented this approach in the R package *mgcv*.

[8] Hausman's test for the consistency of the random effects (RE) estimator provides evidence in favour to the FE estimator in all specifications. The results of a F test confirm the joint significance of fixed spatial effects. Full estimation details are available upon request.

[9] For the estimation of the SDM, we have used the 5-nearest neighbours (5-NN) spatial weights matrix. The results are robust to the choice of the spatial weights matrix, as revealed from estimation results obtained using matrices based on 10- and 20-NN. Full estimation details are available upon request.