XXIX CONFERENZA ITALIANA DI SCIENZE REGIONALI

LOCATION DETERMINANTS OF FOREIGN FIRMS' BUSINESS FUNCTIONS IN THE ENLARGED EUROPE: EVIDENCE FROM NEGATIVE BINOMIAL ADDITIVE MODELS

Roberto BASILE[1], Davide CASTELLANI[2] e Luigi Benfratello[3]

1 ISAE, CERTeT and University of Macerata, r.basile@isae.it, Piazza Indipendenza 4, 00185, Roma
2 University of Perugia, davide.castellani@unipg.it
3 University of Turin and Ceris-CNR, benfratello@econ.unito.it

SOMMARIO

In questo lavoro si analizzano le determinanti della localizzazione delle imprese multinazionali nelle regioni NUTS2 dell'Europa allargata durante il periodo 2003-2007. L'analisi è effettuata usando un modello additivo generalizzato sotto l'ipotesi di una distribuzione binomiale negativa degli errori. Questo modello consente di testare le ipotesi di autocorrelazione e instabilità spaziale dei parametri. L'analisi non è confinata alla frammentazione internazionale dell'attività produttiva, ma è estesa ad alcune importanti funzioni aziendali (*business function*) che compongono la catena del valore delle imprese multinazionali, ovvero l'attività di R&S e gli altri *business service*. I risultati dell'analisi suggeriscono importanti considerazioni in merito alla sostenibilità della dinamica della divisione interregionale del lavoro in Europa.

1 INTRODUZIONE

Accelerating economic integration in Europe over the past decades has favoured, *inter alia*, a significant flow of international investments from both within and outside the European Union (EU) borders. As a matter of fact, the EU has attracted over 40% of total world flows of Foreign Direct Investments (FDIs, henceforth) in the 1990's, becoming the largest recipient of multinational activity. The recent enlargement of the EU has reinforced further this process, fostering a reorganization (relocation) of the European industry, where multinational firms (MNFs, henceforth) exploited mainly lower labour costs and large market access available in most New Member States. Faster technological progress has also contributed to make cheaper and easier the transfer of production activities abroad. Furthermore, relocation has spread to more sectors (extending progressively from industrial sectors to service sectors) and more activities (from manufacturing to R&D activities and other business functions).

While manufacturing aspects of international fragmentation and location determinants of manufacturing FDIs in Europe have been widely analyzed (see, for example, Mucchielli and Puech, 2002; Head and Mayer, 2004; Disdier and Mayer, 2004; Basile et al., 2006; Basile et al., 2008), the literature on location patterns and on the determinants of other stages composing the MNFs' value-chain (R&D, marketing, telecommunications, logistics and other business services) is only at its beginning (recent studies in such direction are those of Devefer, 2006; Alegria, 2007; Nefussi and Schwellnuss, 2008). We intend to contribute to this new strand of literature by using data on the number of foreign greenfield investments over the 2003-2007 period disaggregated by region of the enlarged Europe and by business function. We perform a detailed analysis of the location determinants of foreign investments, using different econometric models for count data and considering a large set of variables as potential determinants of FDI concentration (market size, Jacobs externalities, labour market conditions, R&D intensity, social capabilities and transport infrastructure). We attempt to ascertain whether and to what extent these variables, usually employed to explain manufacturing FDI location behaviour, also influence the location patterns in business services and R&D.

We are particularly interested on testing whether and to what extent FDIs in business services and R&D activity, which in theory have strong input-output relationships with manufacturing firms (at least more than retail services), are actually more attracted by regions with a larger presence of industrial activity. Such hypothesis is based on the consideration that trade costs are prohibitively high in business services and R&D sectors, so that supplying foreign markets through exports is not a feasible alternative to FDI. Testing this hypothesis is very important to understand and evaluate the sustainability of the present and future interregional division of labour in Europe. As our data clearly show (see section 2), international delocation

(and relocation) of manufacturing activity towards Eastern low-labour-cost regions has indeed contributed to change the European map of interregional manufacturing-vs.-services specialization, with EU-15 regions progressively more specialized in service sectors and newly accessed regions gradually more specialized in manufacturing activity. The question is whether a cumulative causation mechanism where manufacturing investments attract other manufacturing investments as well as business services and R&D (as intermediate inputs) yields to a sustainable equilibrium in the long run.

This paper also attempts to contribute to the literature on FDI location by discussing different methodological issues and by using new sophisticated techniques for estimating count data models. Previous works on regional inward FDI counts has shown that considerable overdispersion exists, that is the conditional variance is significantly higher than the conditional mean (Kogut and Chang, 1991; Bloningen, 1997; Zhou et al., 2002; Coughlin and Segev, 2000; Barry et al., 2003; Basile, 2004; De Propis et al., 2005; Basile et al. 2006; Arauzo-Carod and Viladecans-Marsal, 2007). In these studies, a negative binomial model was used to estimate the number of new foreign firms in a given set of regions as a linear function of some covariates under the implicit assumption that the effect of explanatory variables is homogeneous over space. However, parameter heterogeneity is very likely to occur, so that the relationship between the economic environment and the number of foreign start-ups is nonlinear. Moreover, spatial distribution of inward FDI counts often displays spatial autocorrelation, that is locations close to each other exhibit more similar values than those further apart. If this pattern remains present in the errors of a statistical model based on such data, one of the key assumptions of standard statistical analysis, that errors are independently distributed, is violated. We tackle these issues by adopting a Negative Binomial Additive Model which simultaneously takes into accounts the problems of overdispersion and spatial instability. Moreover, we extend the model by including a spatial autoregressive term and a smooth term of the interaction between spatial coordinates which help addressing the problem of spatial autocorrelation in the residuals. As it is well known, the spatial lag of the dependent variable is by construction correlated with the error term of the model. Thus, we adopt a control function approach as a general solution for the endogeneity biases in nonparametric models (Blundell and Powell, 2003, Ng and Pinkse, 1995).

The rest of the paper is organized as follows. Section 2 introduces the dataset on inward FDI and gives a description of the regional distribution of foreign greenfield investments in Europe. Section 3 presents the theoretical frame which motivates the specification of the empirical model. Section 4 focuses on the econometric methodology. Section 5 reports the econometric results. Section 6 concludes.

2 SPATIAL DISTRIBUTION OF FDI WITHIN THE ENLARGED EUROPE

2.1 Data source and sample

The paper exploits a new database, OCOmonitor, which provides information on almost 11,000 greenfield investment projects carried out by both European and non-European MNFs in the enlarged Europe (including both the 'old' Western European Union countries and the new Eastern accession countries) over the 2003-2007 period. For each project detailed information is available on the investor (name and state/country of origin and sector of activity, including both manufacturing and services), on the destination area (country, state and city) and the main business function (including manufacturing, sales and marketing, R&D, logistic, headquarter and business services) involved in the investment abroad. This allowed us to count the number of projects in each NUTS2 region by business function. The overall sample is composed of 249 NUTS2 regions.

Among the different business functions, we focus on manufacturing processes (manfdi), business services (busfdi) and R&D (rdfdi). Precisely, we use a broad definition of business services, which encompasses logistics, mail services and telecommunications, renting of machinery and equipment, information technology services and other business services (legal, accounting, consulting, personnel, marketing and so on).

Inspection of the distribution of FDI counts in the three different business functions (Figure 1) reveals that it is always right skewed and that the share of zeros is 14% in the case of manfdi, 19% in the case of busfdi and 42% in the case of rdfdi.

This is not surprising considering that R&D is one of the MNFs' activities that are less prone to internationalization (561 greenfield investments in our sample) and that, even if a growing number of MNFs are establishing R&D facilities abroad, not all locations are good recipients for this kind of FDI. It is also important to observe that the emergence of FDIs in R&D is part of the broader phenomenon of offshoring services. The number of busfdi in our sample (2,127) is indeed higher than the number of manfdi.

2.2 Distribution by country and area

Table 1 shows the distribution of FDIs by area (EU-15 regions and New Member States) and country. It clearly emerges that the spatial patterns of MNF's location choices vary according to the type to investment. The geographical distribution of FDIs in manufacturing sectors benefited the New Member States, often reflecting the relocation strategies of EU-15 labour intensive undertakings seeking to exploit relatively low-cost locations and proximity to expanding markets. As a result of this process, regional specialization patterns have changed

substantially. In particular, EU-15 regions have reduced significantly their degree of specialization in manufacturing activity in favor of a higher specialization in service activities. On the opposite side, newly accessed European regions became highly specialized in manufacturing sectors. This process comes out very clear from comparing the 1992's and 2004's cluster maps of local G_i^* statistics (Ord and Getis, 1995) applied to the locaton quotients (LQ) in manufacturing value added.

	Total FDI	%	Manuf.	%	Busin.	%	R&D	%
Europe	**10,798**	**100.0**	**1,930**	**100.0**	**2,127**	**100.0**	**561**	**100.0**
EU-15	*7,385*	*68.4*	*754*	*39.1*	*1,611*	*75.7*	*438*	*78.1*
New Members	*3,413*	*31.6*	*1,176*	*60.9*	*516*	*24.3*	*123*	*21.9*
Austria	263	2.4	31	1.6	56	2.6	15	2.7
Belgium	377	3.5	51	2.6	121	5.7	22	3.9
Czech Republic	487	4.5	201	10.4	69	3.2	32	5.7
Germany	1,005	9.3	150	7.8	243	11.4	64	11.4
Denmark	237	2.2	9	0.5	47	2.2	20	3.6
Estonia	134	1.2	39	2.0	34	1.6	1	0.2
Spain	612	5.7	107	5.5	115	5.4	43	7.7
Finland	104	1.0	10	0.5	18	0.8	8	1.4
France	1,214	11.2	136	7.0	265	12.5	79	14.1
Greece	131	1.2	7	0.4	35	1.6	1	0.2
Hungary	656	6.1	209	10.8	103	4.8	21	3.7
Ireland	485	4.5	37	1.9	59	2.8	37	6.6
Italy	465	4.3	48	2.5	103	4.8	24	4.3
Lithuania	158	1.5	32	1.7	19	0.9	–	–
Netherlands	391	3.6	37	1.9	110	5.2	11	2.0
Poland	816	7.6	329	17.0	116	5.5	31	5.5
Portugal	152	1.4	24	1.2	38	1.8	6	1.1
Romania	768	7.1	190	9.8	118	5.5	27	4.8
Sweden	282	2.6	20	1.0	60	2.8	23	4.1
Slovenia	75	0.7	12	0.6	16	0.8	2	0.4
Slovakia	319	3.0	164	8.5	41	1.9	9	1.6
United Kingdom	1,667	15.4	87	4.5	341	16.0	85	15.2

Table 1 - FDI distribution by area and country

Offshore R&D investments are mainly oriented to EU-15 regions, but the number of rdfdi in newly accessed regions is not negligible. In particular, three metropolitan regions, namely Praga (Czech Republic), Bucarest (Romania) and Budapest (Hungary), are amongst the top-ten rdfdi attracting regions in Europe. Finally, more than 75% of busfdi are sited in EU-15 regions.

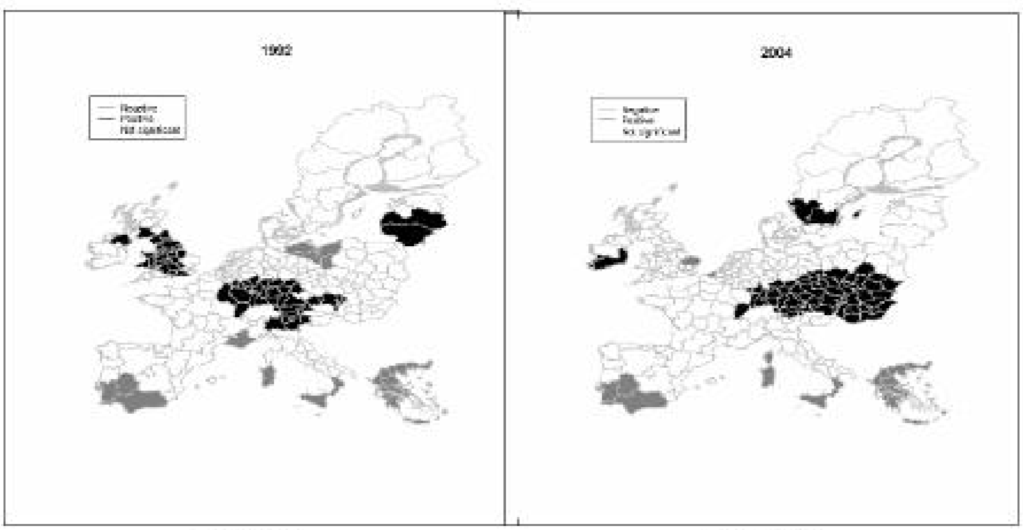

Figure 2a Figure 2b

Figure 1 - Local G* statistics of location quotients in manufacturing activity

a. Regions with a black color are those with a positive G_i^*; those with a grey color are those with a negative G_i^*; while regions with a white color are those with a non-significant value of G_i^*.

2.3 Spatial autocorrelation

The regional distribution of manfdi strongly displays spatial autocorrelation, that is locations close to each other exhibit more similar values than those further apart. Traces of spatial autocorrelation are also present in the case of rdfdi, while busfdi are not correlated in space. We will take this information into account when specifying the econometric model.

3 Theoretical framework

The spatial distribution of economic activities, including FDI investments, can be considered as the result of the interaction between centrifugal (or agglomeration) and centripetal (or competition) forces (including higher land prices, higher factor prices - wages *in primis* - or strong competition). Since Hoover (1948), it is common to distinguish between two sources of agglomeration externalities: a) economies external to the firm but internal to the sector (the so-called Marshallian externalities) and b) economies external both to the firm and to the sector (the so-called location externalities).

According to Marshall, industrial firms tend to localize where other firms of the same industry are already established. The benefits of this form of externality are well known: i) access to a more stable labor market, ii) availability of intermediate goods, production services, skilled manpower and iii) knowledge spillover between adjacent firms. Marshallian externalities are therefore more suitable to explain "small scale" agglomeration phenomena, such as the emergence of Industrial Districts, that is spatial clusters of firms operating in the same sector (for example, clothing or footwear). Yet, in the present context, we do not have detailed information on the specific industrial sector of investment and consider also foreign investments in business services and R&D activity. Thus, we will not consider the role of Marshallian externalities and focus on the effect of location externalities, which occur due to the proximity of various economic activities. This encompasses market size effects, diversity of services, intermediates and final products (Jacobs, 1969) and the presence of social capital (infrastructure, innovation potential and social capability).

The market size effect has been firstly formalized within the international trade theory by Krugman and Venables (1990) and, then, used in the New Economic Geography literature (Krugman, 1991; Venables, 1996) to explain the concentration of industrial activity. Due to this effect, firms of sectors characterized by imperfect competition concentrate their activity in regions with larger markets and export to areas with smaller markets. So, the spatial distribution of the demand influences the geographical distribution of firms. A market size effect may occur both from the consumers' demand and from the demand of other firms, that is input-output linkages between firms, as emphasized by Venables (1996) and Puga (1999).

Moreover, cost externalities can derive from the presence of a large population of intermediate producers.
Jacobs externalities derive from diversity or variety of the regional economy. According to Jacobs (1969), a diverse sectoral structure increases the odds of interaction, generation, replication, modification and recombination of ideas and applications across different industries. Moreover, a diverse industrial structure protects a region from volatile demand and offers the possibility to switch between input substitutes.
The analyses of the relevance of public infrastructures for regional development and for the process of geographical concentration of industrial activities show that regions with a relatively low level of infrastructures have a relatively low level of productivity and return to private investments, which might indeed be smaller than in regions with a higher stock of infrastructures. The relatively low return to private investments within regions that are poor of infrastructure capital reduces the attractiveness for both domestic and foreign investments.
The indigenous innovation potential of each region (usually approximated by traditional R&D intensity measures) represents another important externality that may influence location decisions of foreign firms, especially for those MNF that try to localize an R&D center abroad. It has been observed that these centers are often sited in proximity of other R&D centers in order to exploit important knowledge spillovers. However, it has also been underlined that a certain level of social capability (usually approximated by human capital endowments of the regions) is necessary to convert indigenous and external knowledge into effective innovations and, in general, to exploit R&D spillovers.

4 Spatial modelling of FDI counts

4.1 Overdispersion and zero-inflation

Research on foreign firms' location choice usually appeals to discrete-choice models (conditional logit, nested logit and mixed logit models) that rely on the Random Utility Maximization (RUM) framework. In this framework, decision probabilities are modelled in a partial equilibrium setting where foreign firms maximize profits subject to uncertainty that derives from unobservable characteristics. In the present study, however, we cannot use discrete choice models, since we aggregate the information at regional level. Thus, the dependent variable y_i (i=1,2,...,N) used in the econometric analysis (the number of greenfield investments in each region i) assumes discrete values, that is non-negative integer values (count data). The standard framework for count data is the Poisson regression model. Let $Y_1,Y_2,...,Y_N$ be a set of count data, X_i a k×1 vector of explanatory variables and β a k×1 vector of regression parameters. The Poisson regression model for these data is defined by:

$$P\left(Y_i = y_i | X_i'\beta\right) = \frac{\exp\left(-\mu_i\right)\mu_i^{y_i}}{y_i!} \qquad (1)$$

with mean and variance equal to μ_i (the so-called equidispersion condition). The Poisson regression is a special case of the Generalized Linear Model (GLM) framework. The canonical link is $\eta_i = g(\mu_i) = \log(\mu_i) = X_i'\beta$, resulting in a log-linear relationship between mean and linear predictor.

Fortunately, Guimaraes et al. (2004) have demonstrated that the coefficients of a Poisson regression are equivalent to those of a conditional logit model. Thus, also the Poisson regression model can be thought as derived directly from a RUM process. In practice, however, the classical Poisson regression model for count data is often of limited use in regional location analysis since empirical inward FDI counts typically exhibit overdispersion and/or an excess number of zeros. Overdispersion occurs when the variance is larger than the mean, so that the model generates consistent but inefficient estimates. In the case of location choice analysis, overdispersion is generally observed due to the concentration of foreign firms in a few areas. Zero-inflation may occur when the zero outcome (that is no announcement to invest in a region) can arise from two underlying responses. On the one hand, some regions may never attract a greenfield investment, thus the outcome will be always zero. On the other hand, if the region is an attractive one, the zero outcome may be just the number of investments attracted in a given (sample) period and the response might be some positive number in a different period.

A way of dealing with overdispersed count data is to assume a negative binomial distribution for $y_i | X_i$ which can arise as a gamma mixture of Poisson distributions.[1]

One parameterization of its probability density function is

$$P\left(Y_i = y_i | X_i'\beta, \theta\right) = \frac{\Gamma\left(y_i + \theta\right)}{\Gamma\left(\theta\right) \cdot y_i!} \cdot \left(\frac{\mu_i}{\mu_i + \theta}\right)^{y_i} \left(\frac{\theta}{\mu_i + \theta}\right)^{\theta} \qquad (2)$$

with mean μ and shape parameter θ; $\Gamma(.)$ is the gamma function. The variance function is now $V(\mu) = \mu + \mu^2\theta^{-1}$. Note that, for large θ, the model approaches the Poisson model.

Recently, a large number of studies on regional inward-FDI counts have used the negative binomial regression model to address the overdispersion issue (Kogut and Chang, 1991; Zhou et al., 2002; Coughlin and Segev, 2000; Barry et al., 2003; De Propis et al., 2005; Arauzo-Carod and Viladecans-Marsal, 2007). Random effects extensions of negative binomial regression for panel data have also been considered by Bloningen (1997), Basile (2004) and Basile et al. (2006).

Even though negative binomial regression models capture overdispersion quite well, they are not always sufficient for modelling excess zeros. Mullahy (1986) and Lambert (1992) have

[1] The issue of overdispersion can be also addressed by extending the Poisson regression model in various directions: for example, using sandwich covariances or estimating an additional dispersion parameter (in a so-called quasi-Poisson model).

addressed this problem by introducing zero-augmented models that incorporate a second model component capturing zero counts. Zero-inflation models (Lambert, 1992) are mixture models that combine a count component and a point mass at zero. Hurdle models (Mullahy, 1986) take a somewhat different approach and combine a left-truncated count component with a right-censored hurdle component. Examples of applications of Zero-inflated Poisson (ZIP) and Zero-inflated Negative Binomial (ZINB) models to FDI location analyses are in Tadesse and Ryan (2002), Basile (2004), Tomlin (2000) and Iannizzotto and Miller (2002).

4.2 Nonlinearities: Negative Binomial Additive Models

Poisson and Negative Binomial models, widely used in the recent literature on inward FDI counts, often assume the existence of a log-linear relationship between FDIs and the explanatory variable X_i , that is they implicitly assume that all economies obey a common linear specification of the location model, disregarding the possibility of nonlinearities reflecting spatial instability in the behavior of economic agents. For example, we cannot disregard possible threshold effects in the impact of labor costs or of agglomeration externalities on regional FDI attractiveness.

Nonlinearities (or spatial instability or parameter heterogeneity) can be addressed in different ways. First, polynomial expansions up to a cubic can be considered within a GLM approach, but with the risk of introducing multicollinearity in the model. Second, Geographically Weighted Regression (GWR) models represent a standard method for properly handle spatial instability problems. However, as far as we know, there are no extensions of GWR models for overdispersed data. A third solution, which will be considered in this paper, is the Negative Binomial Additive Model (NB-GAM), recently introduced by Thurston et al. (2000). NB-GAMs are a special extension of Generalized Additive Models (GAM) to handle Negative Binomial responses.

The GAM framework extends the GLM by allowing nonlinearity in the relationship between η_i and the covariates:

$$\eta_i = g(\mu_i) = X_i'^{*}\beta^{*} + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}, x_{4i}) + \ldots + \epsilon_i \qquad (3)$$

where $\mu_i = E(Y_i), Y_i \sim NegativeBinomial, f_j(.)$ are unknown smooth functions of the covariates, X_i^{*} is a vector of strictly parametric components and β^{*} is corresponding parameter vector. A log link function η can be used instead of the canonical link $\eta_i = \ln\{\mu_i/(\mu_i+\theta)\}$.

The most popular approach for estimating GAM models is the back-fitting algorithm (Hastie and Tibshirani, 1990). This method, however, presents some shortcomings with respect to model selection and inference issues. Wood (2000, 2006) and Wood and Augustin (2002) have recently proposed a new methodology to estimate GAM models with spline based penalized regression smoothers which allows for automatic and integrated smoothing

parameters selection via Generalized Cross Validation (GCV). Wood has implemented this approach in the R package mgcv.

4.3 Spatial autocorrelation: a control function approach

If spatial autocorrelation is present in the errors of a statistical model, one of the key assumptions of standard statistical analyses, that errors are independent and identically distributed (*iid*) is violated. The violation may bias parameter estimates and can increase type I error rates (falsely rejecting the null hypothesis of no effect). Spatial autocorrelation in the residuals may occur because of the existence of spatial dependence either in unmodeled effects (when unmodeled variables that are subsumed in the error term jointly follow a spatial random process) and/or in modelled effects (when the X terms affect the left hand side of the model through a 'global multiplier effect', i.e. both x_i as well as a set of x_j throughout the spatial systems affect y) (Anselin, 2004).[2]

Dealing with spatial externalities within a nonparametric framework is a challenging task and at the research frontier in spatial econometrics. In a parametric linear setting, such as $y=X\beta+\varepsilon$, global multiplier effects are modelled by replacing X by $(I-\rho W)^{-1}X$ and ε with $(I-\rho W)^{-1}\varepsilon$, where I is an identity matrix, ρ is the parameter of spatial externality and W is a spatial weights matrix.

In the present context, the inverse spatial transformation of X and ε suggests that the attractiveness of region i is affected not only by its own characteristics and random shocks, but also by the features and random shocks of all other regions. Thus, every location is correlated with every other location in the system. However, given the characteristics of the standardized spatial weights matrix, the strength of spatial dependence between observed regions declines with the distance between them. In other words, neighboring units exhibit a higher degree of spatial dependence than units located far apart ('spatial diffusion with friction'). The introduction of the spatial multiplier effect in the model yields a reduced form as $y=(I-\rho W)^{-1}X'\delta+(I-\rho W)^{-1}\varepsilon$ and the structural form becomes the standard spatial autoregressive model (SAR) $y=\rho Wg+X'\beta+\varepsilon$. These arguments can be extended to the semiparametric GAMs, with the obvious difference that the effect of spatial externalities may not be homogenous over space. So, the NB-GAM framework described above can be extended by including the smooth term $f_4(Wy)$ on the right hand side (SAR-NB-GAM):

$$\eta_i = g(\mu_i) = X_i^{*'}\beta^* + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}, x_{4i}) + f_4(Wy_i) + f_5(Lat_i, Long_i) + \ldots + \epsilon_i \quad (4)$$

Model (4) includes also a smooth term $f_5(lat,long)$, where *lat* and *long* are the spatial coordinates of the region's centroid. This term helps controlling for unobserved heterogeneity in cross-section areal data.

[2] An example would be when inward FDIs are set in function not only of the local market, but also of the market size of the neighbors, and their neighbors' neighbors, etc. (the so-called "market potential").

Because of the feedbacks between y_i and its spatial lag term Wy_i, $f_4(Wy_i)$ enters endogenously in equation (4), that is $f_4(Wy_i)$ and ε_i are correlated.

In linear spatial regression analysis, Kelejian and Prucha (1998) have proposed a 2SLS procedure to estimate the spatial autocorrelation regression model and have suggested to use spatial lags of the strictly exogenous variables as instruments. Following a 2SLS approach, Wy is regressed on a set of exogenous and predetermined variables. In the second stage, the fitted values from the first stage are used in place of the endogenous variable. The motivation for this form of 2SLS is the replacement of the endogenous regressor with that part of Wy (its linear projection on the set of spatial lags of the exogenous variables) that is uncorrelated with the error term. As emphasized by Blundell and Powell (2003), however, this procedure is not suitable for the estimation of nonparametric and semiparametric models. In particular, the replacement of the endogenous term with fitted values of the first stage generally yields inconsistent estimates of $f_4(Wy_i)$. Thus, Ng and Pinkse (1995) and, more recently, Blundell and Powell (2003) have proposed a general solution which is appropriate for the estimation of nonparametric models. This method consists of extending the "control function" method to additive nonparametric models.

The control function approach applied to the linear model $y_i=X_i'\beta+\varepsilon_i$ has its antecedent in the interpretation of the 2SLS estimator β_{2SLS} as the coefficients on X_i in a OLS regression of y_i on X_i and the residuals v_i from a linear regression of X_i on a set of instruments Z_i. Application of the control function approach to nonparametric and semiparametric settings is straightforward. It consists of two steps. In the first one, an auxiliary nonparametric regression of the form $Wy_i=f(X_i)+g(Z_i)+v_i$ is considered, with Z_i being a set of appropriate instruments and v_i a sequence of random variables satisfying $E(v_i \mid Z_i)=0$. Moreover, if Z_i and ε_i are independent, then it yields that $E(\varepsilon_i|v_i,Z_i)=E(\varepsilon_i|v_i)$. It follows from the last assumption that $E(\varepsilon_i|Wy_i)\neq 0$ arises when $E(\varepsilon_i|v_i)\neq 0$. The second step consists of estimating an additive model of the form

$$\eta_i = g(\mu_i) = X_i^{*\prime}\beta^*+f_1(x_{1i})+f_2(x_{2i})+f_3(x_{3i},x_{4i})+f_4(Wy_i)+f_5(Lat_i,Long_i)+f_6(\widehat{v}_i)+\ldots+\epsilon_i \quad (5)$$

5 Empirical evidence

5.1 Model selection

Table 2 reports a series of diagnostics tests and measures of goodness of fit for different specifications of the parametric and nonparametric count data models, namely the parametric Poisson and Negative Binomial GLM models and their nonparametric counterparts with and without spatial controls. The dependent variable, y, is the number of inward FDI in each region. The explanatory variables are Mkt (the market size, approximated by the regional total

value added), Infra (a measure of transport infrastructure), Jacob (a measure of Jacobs externalities), Wage (the average labor cost), Ur (the regional unemployment rate), R&D (the R&D intensity), Ter (the level of tertiary education), Man (the employment density in manufacturing) and Bus (the employment density in business services). A preliminary analysis suggests that Mkt, Jacob and Infra can enter all models linearly (the equivalent degrees of freedom are equal to 1), while all other terms contribute to make up the nonlinear components of the models.[3] So, we estimated the following competing models:

$$\begin{aligned} \eta_i &= g(\mu_i) = \alpha + \beta_1 Mkt_i + \beta_2 Infra_i + \beta_3 Jacob_i + \beta_4 Wage_i + \beta_5 Ur_i + \\ & \beta_6 R\&D_i + \beta_7 Ter_i + \beta_8 Man_i + \beta_9 Bus_i + \epsilon_i \\ & \mu_i = E(Y_i), Y_i \sim Poisson \end{aligned} \quad (6)$$

$$\begin{aligned} \eta_i &= g(\mu_i) = \alpha + \beta_1 Mkt_i + \beta_2 Infra_i + \beta_3 Jacob_i + \beta_4 Wage_i + \beta_5 Ur_i + \\ & \beta_6 R\&D_i + \beta_7 Ter_i + \beta_8 Man_i + \beta_9 Bus_i + \epsilon_i \\ & \mu_i = E(Y_i), Y_i \sim NegativeBinomial \end{aligned} \quad (7)$$

$$\begin{aligned} \eta_i &= g(\mu_i) = \alpha + \beta_1 Mkt_i + \beta_2 Infra_i + \beta_3 Jacob_i + \\ & f_1(Wage_i, Ur_i) + f_2(R\&D_i, Ter_i) + f_3(Man_i, Bus_i) + \epsilon_i \\ & \mu_i = E(Y_i), Y_i \sim NegativeBinomial \end{aligned} \quad (8)$$

$$\begin{aligned} \eta_i &= g(\mu_i) = \alpha + \beta_1 Mkt_i + \beta_2 Infra_i + \beta_3 Jacob_i + \\ & f_1(Wage_i, Ur_i) + f_2(R\&D_i, Ter_i) + f_3(Man_i, Bus_i) + \\ & f_4(Lat_i, Long_i) + f_5(Wy) + f_6(\widehat{v}) + \epsilon_i \\ & \mu_i = E(Y_i), Y_i \sim NegativeBinomial \end{aligned} \quad (9)$$

where *Lat* and *Long* are the spatial coordinates of the regional centroid, *Wy* is the spatial lag of y[4] and v is the residual from the first step.[5]

Nonparametric smooth terms are estimated using thin-plate regression splines and tensor product smoothing splines and applying the method described in Wood (2006) that allows integrated smoothing parameter selection via GCV.

---

[3] See the appendix for a thorough definition of the variables.

[4] W is a standardized spatial weigths matrix. Each element w_{ij} of this matrix summarizes the interaction between regions i and j. The elements w_{ii} on the main diagonal are set to zero whereas elements $w_{ij} = d_{ij}^{-2}$ if $d_{ij} < \overline{d}$ and $w_{ij} = 0$ if $d_{ij} > \overline{d}$, with d_{ij} the great circle distance between the centroids of region i and region j and $\overline{d}$ the cut-off distance (equal to 423 km).

[5] The endogeneity bias has been corrected through the control function method, using spatial lags of the exogenous variables as instruments, i.e. as additional regressors in the first step equation. The results of the first step are available upon request.

The proportion of deviance explained ranges from 59% (the Poiss-GLM model) to 80% (the NSAR-NB-GAM model) in the case of manfdi, from 66% to 77% in the case of busfdi, and from 37% to 80% in the case of rdfdi, clearly suggesting that the SAR-NB-GAM model encompasses all the others. In all cases, the hypothesis of no spatial dependence in the residuals is significantly rejected in all models except for the last one. This suggests that spatial dependence and nonlinearities must be explicitly taken into account in order to avoid misspecification problems.

Table 2a - manfdi

	Poiss-GLM	NB-GLM	NB-GAM	SAR-NB-GAM
Deviance	59.5	56.3	68.3	80.2
Spatial dependence	Yes	Yes	Yes	No
Constant variance	3.77(0.00)	0.45(0.46)	0.50(0.48)	0.01(0.90)
$\widehat{\theta}$		1.81	2.09	4.42
Vuong		0.25(0.40)		
Wald-Hurdle		3.80(0.95)		

Table 2b - busfdi

	Poiss-GLM	NB-GLM	NB-GAM	SAR-NB-GAM
Deviance	66.1	53.3	67.5	76.9
Spatial dependence	Yes	Yes	Yes	No
Constant variance	3.97(0.00)	0.08(0.77)	1.21(0.27)	1.65(0.42)
$\widehat{\theta}$		1.23	1.61	2.43
Vuong		0.39(0.34)		
Wald-Hurdle		40.75(0.00)		

Table 2c - rdfdi

	Poiss-GLM	NB-GLM	NB-GAM	SAR-NB-GAM
Deviance	37.4	36.3	58.9	79.6
Spatial dependence	Yes	Yes	Yes	No
Constant variance	1.81(0.13)	0.62(0.43)	0.03(0.87)	1.30(0.27)
$\widehat{\theta}$		0.97	2.09	40.97
Vuong		0.11(0.46)		
Wald-Hurdle		0.27(0.99)		

Notes: *Deviance* is the proportion of deviance explained. The tests of *spatial dependence* (using different distance neighbors weights matrices ranging from 423 km to 1,423 km) are based on Moran's I statistics. The test of *constant variance* of the residuals is based on the estimation of the simple model $|\widehat{\epsilon}| = \alpha + f(\widehat{y}) + u$, where $|\widehat{\epsilon}|$ is the absolute value of the residuals of the model and $\widehat{y}$ is the vector of fitted values. Under the null hypothesis of constant variance, the smooth term $f(\widehat{y})$ must be estimated with one degree of freedom and, according to a F test, should not have a significant effect on $|\widehat{\epsilon}|$. $\widehat{\theta}$ is the estimated Negative Binomial scale parameter. *Vuong* is a non-nested hypothesis test-statistic, which is asymptotically distributed $N(0,1)$ under the null that the models ZIP-NB and NB-GLM are indistinguishible. *Wald-Hurdle* tests the null hypothesis that no-zero-hurdle is required in hurdle regression models for count data. The same set of regressors is used in the hurdle model for both the count component and the zero hurdle component.

Table 2 - Econometric results: model selection

To evaluate how well the different models fit the data, we also plotted the absolute values of Pearson residuals versus the fitted values on the η scale (McCullagh and Nelder, 1989, p.398). The plots from both GLM Poisson models show that the variance of the residuals increases with the fitted values. The plots from the Negative Binomial models show much less

heteroskedasticity and the F tests for the overall significance of the smoothed term f(y) have p-values higher than 0.1. Overdispersion is also evident from the small values of the estimated θ parameters. Finally, overdispersion can be more formally tested through likelihood ratio tests (comparing the log-likelihood of Poisson and NegBin models).

Zero-inflation is also considered by comparing parametric NB-GLM models with zero-inflated Negative Binomial (ZINB) models. As the standard ZINBs are non-nested, the Vuong test for non-nested model is applied to ascertain which distribution applies. This test calculates the logarithm of the ratio of the conditional probability of the dependent variable, conditional on the independent variables, for two alternative distribution hypotheses. In all cases, the Vuong test statistic comes out to be nonsignificant, so that we can exclude the existence of zero-inflation. Only the Wald test for the hurdle model suggests that the NB-GLM may not be correctly specified.

5.2 Results

In this section, we report the results of the SAR-NB-GAM model. Table 3 reports the F tests for the overall significance of the smoothed terms, the number of effective degrees of freedom (edf) and the β parameters for linear terms.[6] It is immediately important to observe that the endogenous term f(Wy) enters significantly only in the manfdi equation. The lack of spatial externalities in busfdi and rdfdi implies that only the characteristics of region i affect the attractiveness of FDI in business services and R&D activity, while spatial multiplier effects occur in the determination of FDI attractiveness in manufacturing and R&D.

The coefficient of Mkt is 0.7 in manfdi, 1.1 in busfdi and 1.0 in rdfdi, indicating a significant high positive effect of the market size on inward FDI.[7] However, the presence of a significant positive spatial externality in manfdi equations implies that in this case the FDI attractiveness of region i is influenced not only by its market size, but also by the market dimension of all other regions through a 'spatial multiplier effect' which decreases with distance, in line with the concept of 'market potential'. Not surprisingly, FDIs in business services are affected only by the market conditions of the region where they are sited, corroborating the opinion that this kind of services can be rarely exported and, thus, they have to be located in close proximity to final demand. The level of transport infrastructure influences positively and highly significantly only manufacturing FDI attractiveness, for which transport costs are more relevant. The effect of Jacobs externalities (approximated by the median of Balassa indices in industrial sectors) is also positive and significant only in the case of manfdi.

---

[6] All variables are expressed in log scale, so that β parameters are to be interpreted as elasticities.

[7] Expert readers may note that we didn't adopt a measure of 'market potential' which includes information not only on the market size of region i, but also on market size of all other European regions weighted by the geographical distance from i. This was made intentionally, as we wanted to present the results of a specification without spatial lag terms and identify the occurrence of spatial multiplier effects.

		manfdi	*busfdi*	*rdfdi*
Parametric terms				
mkt	β_1	0.752(0.000)	1.092(0.000)	1.039(0.000)
infra	β_2	0.290(0.000)	0.196(0.074)	−0.192(0.125)
Jacob	β_3	0.612(0.041)	0.467(0.196)	0.639(0.082)
Nonparametric terms				
$f_1(wage, ur)$	*F test*	15.907(0.000)	7.298(0.000)	5.651(0.000)
	edf	7.43	6.29	11.33
$f_2(r\&d, ter)$	*F test*	1.135(0.335)	5.508(0.001)	5.602(0.000)
	edf	3.00	3.00	3.66
$f_3(man, bus)$	*F test*	1.141(0.110)	3.801(0.000)	1.639(0.072)
	edf	13.09	13.59	10.45
$f_4(lat, long)$	*F test*	2.995(0.000)	4.248(0.000)	5.459(0.000)
	edf	5.58	12.74	23.52
$f_5(Wy)$	*F test*	6.018(0.000)	0.830(0.507)	1.003(0.317)
	edf	6.77	1.68	1.00
$f_6(\hat{v})$	*F test*	5.619(0.000)	4.792(0.001)	30.403(0.000)
	edf	5.92	2.79	1.00

Notes: *F tests* are used to investigate the overall ('approximate') significance of smooth terms. *e.d.f.* (effective degrees of freedom) reflect the flexibility of the model. An *e.d.f.* equals to 1 suggests that the smooth term can be approximated by a linear term. In such cases, parametric terms have been used and Table 3 reports $\widehat{\beta}$ coefficients. P-values are in parenthesis.

Table 3 - Econometric results of the SAR-NB-GAM

Fig. 2a, 3a and 4a show the smoothed effects of the interaction between employment density in manufacturing and employment density in business sectors on the expected number of manfdi, busfdi and rdfdi, respectively. Table 3 indicates a strong statistical significance of the joint effect of these two variable only in the case of busfdi. One of the main objectives of this study was to test whether the existence of a larger presence of industrial activity had a positive effect on the location choice of foreign firms' business services. The evidence reported in Figure 4a strongly corroborates this hypothesis: the expected number of inward FDIs in business services increases with the employment density in manufacturing. Above a certain threshold in the level of employment density in business services, the two dimensions of economic activity also interact positively in affecting the regional attractiveness of foreign-owned business service firms. Employment density in manufacturing and business services also influences positively the expected number of manfdi and rdfdi, but the statistical significance of the estimated interaction effect is very low.

The effect of the smooth interaction between R&D intensity and the level of tertiary education (shown in Fig. 2b, 3b and 4b) turns out to be significant in the cases of rdfdi and busfdi. Fig. 4b provides a clear evidence in favor of a complementarity between the two variables in the case of rdfdi, confirming the assumption that, while MNFs tend to localize R&D centers abroad in proximity of other R&D centers in order to exploit important knowledge spillovers, the presence of suitable social capabilities (workers with a tertiary education level) represent a necessary factor to exploit this kind of externality. Tertiary education affects positively the

expected number of busfdi, capturing the availability of managers and highly qualified personnel.

Fig. 2c, 3c and 4c display the joint effect of labour cost and unemployment rates, thus synthesizing the role of labour market conditions on FDI attractiveness. The F test reveals a strong statistical significance of this term in all equations. The graphical evidence suggests that regions with low average labor costs tend to attract more FDIs, after controlling for the other variables. However, the relationship between FDI and labor cost is far from being linear in the case of manfdi and busfdi: it declines steeply before rising slightly or remaining constant. This suggests that, in the case of manufacturing and business services, high labor costs are a signal of high labor quality.[8] Moreover, in the case of manfdi, the expected number of inward FDI increases monotonically with the unemployment rate when labor costs are very low, suggesting that in this case high unemployment rates serve as an indicator of labor availability. A positive effect of unemployment rate is also evident in busfdi, while high unemployment rates are a proxy of less-competitive industrial conditions and lower quality of life in the case of rdfdi.

The spatial lag of the dependent variable, Wy, is computed using an inverse distance spatial weights matrix based on a cut-off of 423 km as suggested by the graphical evidence of the spatial autocorrelation tests in the residuals of NB-GAM models. As already mentioned above, the spatial lag term turns out to be significant only in the case of manfdi, along with the control function *f(v)*. Fig. 2e shows the fitted smooth function *f(Wy)* alongside Bayesian confidence intervals (Wood, 2004). The vertical axis reports the scale of the expected values of manfdi; the horizontal axis reports the scale of *Wy*. Spatial externalities are positive but nonlinear, as the effect disappear above a certain threshold of *Wy*. It is important to remark that imposing a linear model on the partial effect of this term would have yield to a downward biased ρ coefficient.

All the interpretations discussed above should be taken with caution since we have used cross-sectional data. Future analyses should be performed by using longitudinal data. One way of dealing with the problem of unobserved heterogeneity consists of including a smooth interaction term for the spatial coordinates, *f(lat,long)*. This term enters significantly in all models and contributes to solve the problem of spatial autocorrelation in the residuals of the NB-GAM.

[8] In fact, we do not control for the level of labor productivity. By including unit labor costs instead of average labor cost measures, indeed, the estimated effect becomes linear.

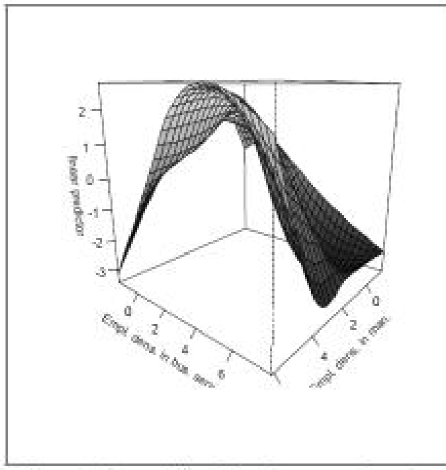

Fig. 4a (*manfdi*) - Employment density

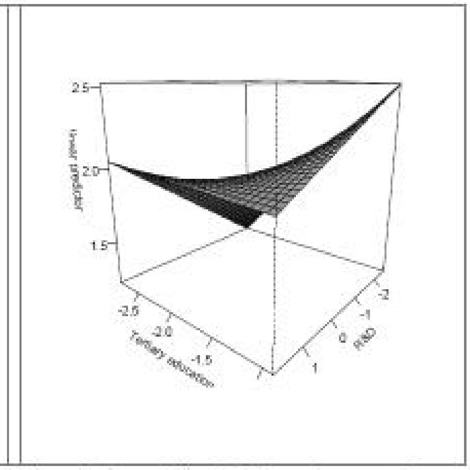

Fig. 4b (*manfdi*) - R&D and tert. educ.

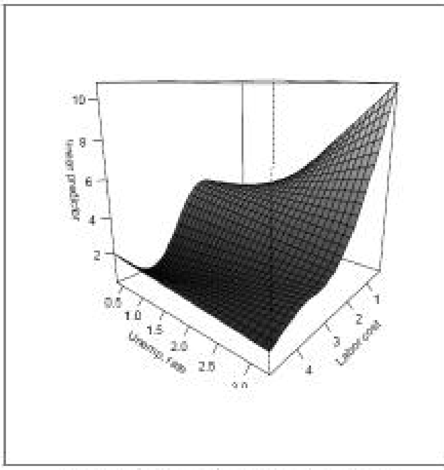

Fig 4c (*manfdi*) - Labor market

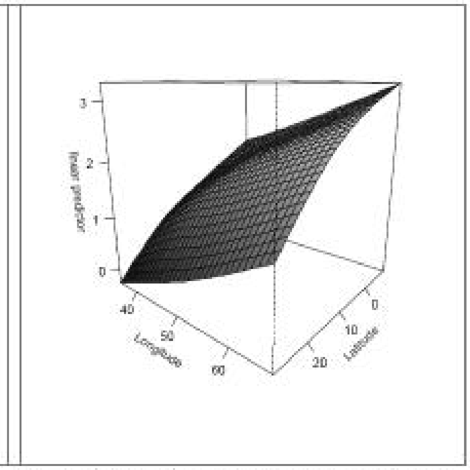

Fig 4d (*manfdi*) - Latitude and longitude

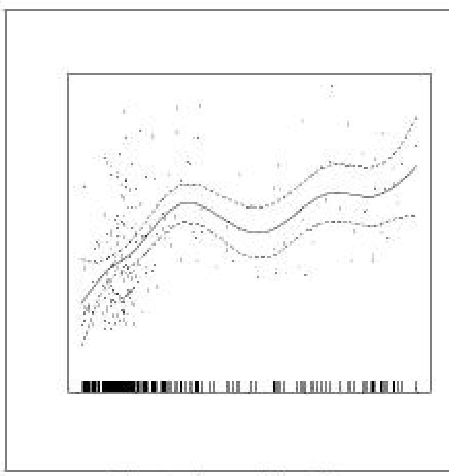

Fig. 4e (*manfdi*) - Wy

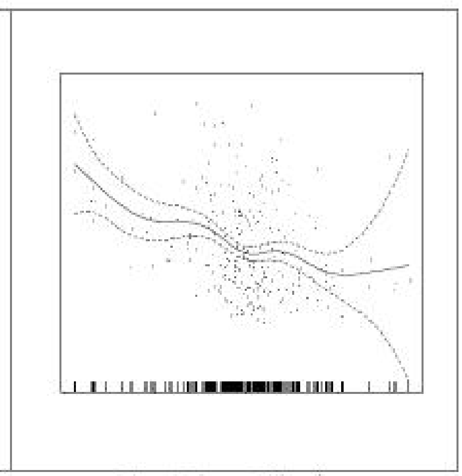

Fig. 4f (*manfdi*) - $\widehat{v}$

Figure 2 - Econometric results (manfdi)

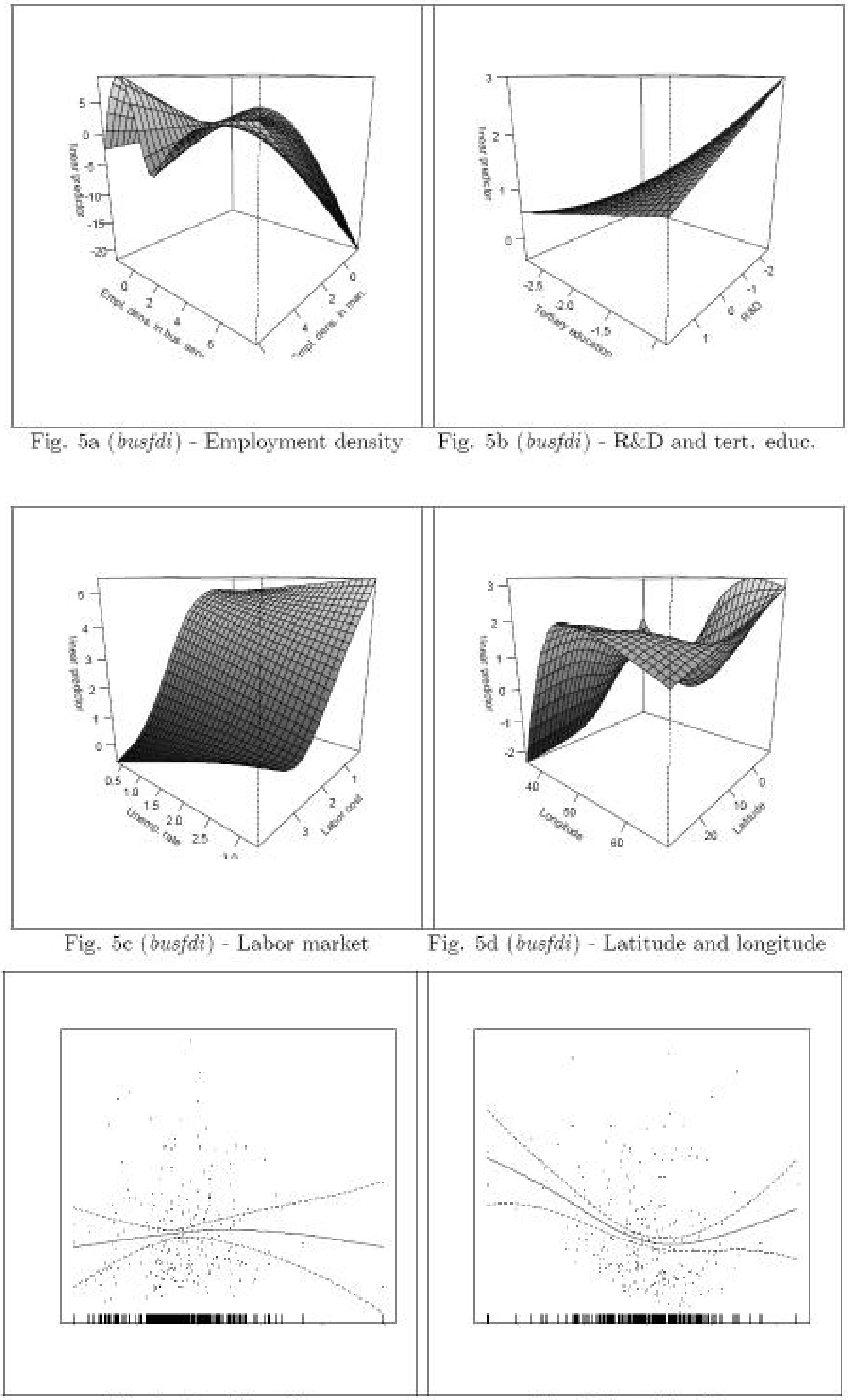

Fig. 5a (*busfdi*) - Employment density

Fig. 5b (*busfdi*) - R&D and tert. educ.

Fig. 5c (*busfdi*) - Labor market

Fig. 5d (*busfdi*) - Latitude and longitude

Fig. 5e (*busfdi*) - Wy

Fig. 5f (*busfdi*) - $\widehat{v}$

Figure 3 - Econometric results (busfdi)

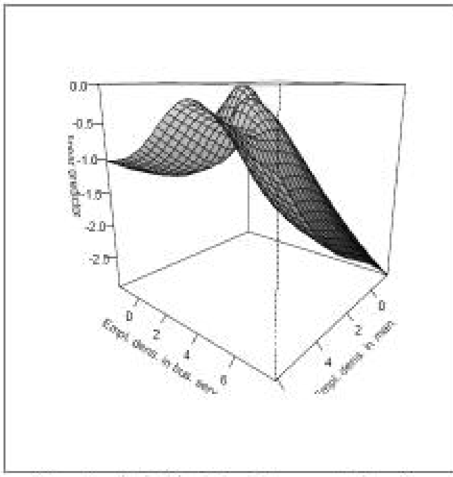

Fig. 6a (*rdfdi*) - Employment density

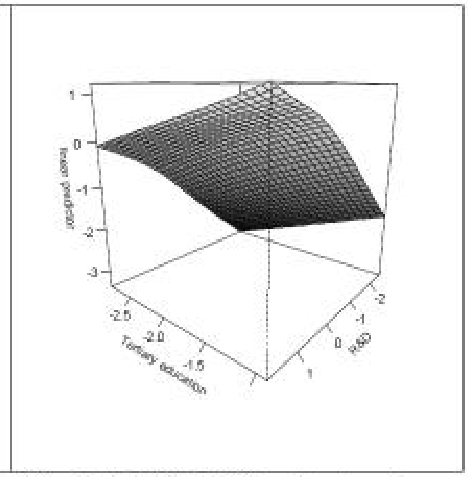

Fig. 6b (*rdfdi*) - R&D and tert. educ.

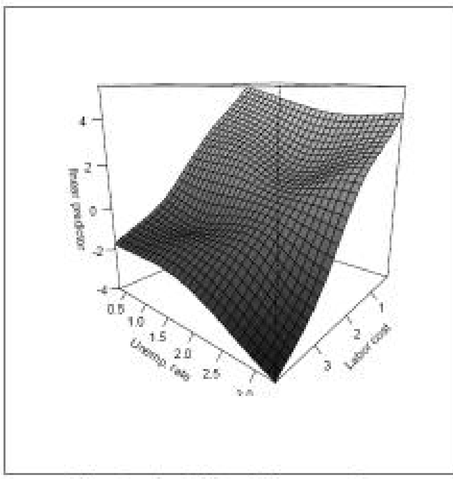

Fig. 6c (*rdfdi*) - Labor market

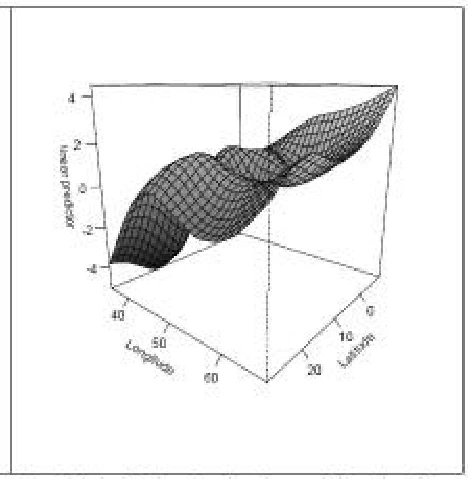

Fig. 6d (rdfdi) - Latitude and longitude

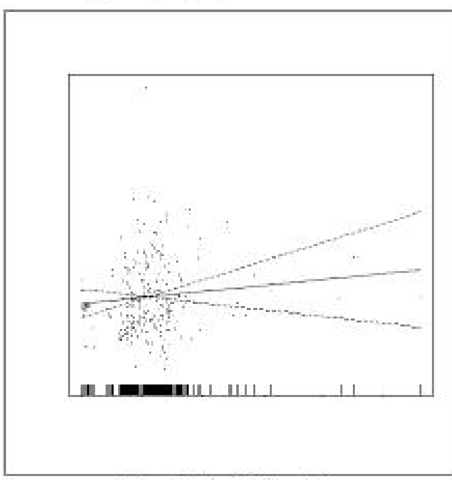

Fig. 6e (*rdfdi*) - Wy

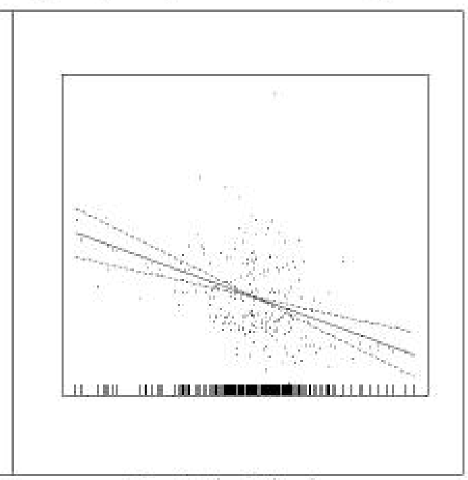

Fig. 6e (*rdfdi*) - $\widehat{v}$

Figure 4 - Econometric results (rdfdi)

6 Conclusions

In this article, we have proposed to employ a spatial autocovariance negative binomial additive count data model (named SAR-NB-GAM) to analyze the regional attractiveness of FDIs. This specification allows to accomodate overdispersion, nonlinearities and spatial dependence simultaneously, thus overcoming many of the methodological issues often raised in the studies on inward FDI based on count data. We have applied the model to European regional data on inward FDI counts. The analysis has been applied not only to manufacturing FDIs, but also to two important business functions that characterize the value-chain of MNFs, namely R&D and business services. The results have provided clear evidence of nonlinearities in the effect of agglomeration economies, labor market conditions, R&D and tertiary education in all three cases. Spatial externalities characterize only the regional distribution of manufacturing FDIs (even after conditioning on observable regional characteristics), thus suggesting that the attractiveness of a region is affected not only by its characteristics, but also by the features of all other regions through the so-called 'spatial multiplier effect'.

In this study we have also investigated whether the existence of a larger presence of industrial activity has a positive effect on the location choice of foreign firms' business services. The SAR-NB-GAM results actually supports this view, as the expected number of inward FDIs in business services increases with the employment density in manufacturing. This evidence may have important consequences in terms of sustainability of the present and future interregional division of labour in Europe. During the ongoing process of European integration and the enlargement to Eastern countries, regional specialization patterns have indeed changed substantially. In particular, regions belonging to the former EU15 have reduced significantly their degree of specialization in manufacturing activity in favor of a higher specialization in service activities. On the opposite side, Newly accessed Eastern European regions became highly specialized in manufacturing sectors. Obviously, FDIs have played a key role on this process. The current off-shoring trend in manufacturing may reflect different patterns of comparative advantage in Western and Eastern European regions. Western regions are expected to specialize increasingly in services while Eastern regions are expected to specialize in manufacturing. However, this kind of division of labour is not necessarily sustainable in the long run. A large share of services, namely the business services, are indeed often used as intermediate input in the manufacturing sector, so that manufacturing location choices may influence the location of FDI in business services. As long as manufacturing activity in absolute terms is mainly located in the West Europe, business services will be attracted towards Western regions. Is this kind of division of labor sustainable in the long run?

7 Appendix: definition of explanatory variables

Market size. We compute market size as the log of total value added (Y_{j}). The source of data is the Cambridge Econometrics database.

Jacobs externalities. Jacobs externalities are measured by means of the median of Balassa indices:

$$MB_i = median\left(\frac{N_{is}/N_i}{N_s/N}\right)$$

where N is the number of employees and s the manufacturing sector. Alternative measure are the negative Herfindahl index and the negative Gini index.

Employment density. We have also included the employment density in manufacturing as a proxy of input-output linkages, based on the consideration that the diversification index (approximated by the median of Balassa indices) does not capture the intensity of the manufacturing presence within the region.

Public infrastructure. We consider only road infrastructure as public infrastructure. The measure is given by the ratio between kilometers of highways plus other roads and total population in the region.

R&D intensity. The measure of R&D intensity at regional level (NUTS2) is taken from the regional statistics provided by Eurostat: percentage of total intramural R&D expenditure on Gross Value Added. Year: average 1999-2002. In some cases, the R&D intensities are imputed to NUTS2 regions according to the differences in terms of EPO applications per million inhabitants. This procedure was chosen after verifying the high correlation between the two variables across the EU NUTS2 regions for which both R&D and patent data were available.

Tertiary education. We approximate social capability with the share of adults (population aged 25-64) with tertiary education (ISCE97 codes 5 and 6). Year: average 1999-2002. Source: Eurostat. For the regions DE41 and DE42 data on tertiary education were available only for the years 2004 and 2005.

Average labour cost. Source: Cambridge Econometrics. For German and UK regions data are available only at the NUTS1 level. We have attributed the same values to NUTS2 regions belonging to the same NUTS1 aggregate.

Unemployment rate. Source: Cambridge Econometrics.

Since we aim at estimating the effect of all these variables over the period 2003-2007, all explanatory variables refer to the previous period 2000-2002.

ABSTRACT

We use a negative binomial additive model which allows for spatial instability and spatial autocorrelation to analyse inward FDI counts in the NUTS2 regions of the enlarged Europe over the 2003-07 period. Following a new strand of literature, we do not consider only manufacturing aspects of international fragmentation, but extend the analysis to the determinants of the location of different functions composing the MNFs' value-chain, namely R&D activity and business services. The results suggest strong implications for the future dynamics of interregional division of labour in the enlarged European system.

8 Bibliografia

Alegria R. (2007) The Location of Multinational Firms in the UK: Sectoral and Functional Agglomeration, WP 400.

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

Arauzo-Carod J.M., Viladecans-Marsal E. (2007) Industrial location at the intra-metropolitan level: the role of agglomeration economies, *IEB Document de Treball* n. 5.

Barry F., Görg H. and Strobl E. (2003) Foreign Direct Investment, Agglomerations and Demonstration Effects: An Empirical Investigation, *Review of World Economics/Weltwirtschaftliches Archiv*, 139, 583-600.

Basile R. (2004) Acquisition Versus Greenfield Investment: the Location of Foreign Manufacturers in Italy. *Regional Science and Urban Economics*, 34, 3-25.

Basile R., Benfratello, L. and D. Castellani (2006) Attracting Foreign Direct Investments in Europe: Are Italian Regions Doomed? In Marco Malagarini and Gustavo Piga (eds.), *Capital Accumulation, Productivity and Growth. Monitoring Italy 2005*, Palgrave Macmillan, 319-354.

Basile R., D. Castellani, and Zanfei A. (2008) Location Choices of Multinational Firms in Europe: the Role of National Boundaries and EU Policy, *Journal of International Economics*, 74, 328-340.

Blonigen B.A. (1997) Firm-Specific Assets and the Link between Exchange Rates and Foreign Direct Investment, *The American Economic Review*, 87, 447-465.

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

Coughlin C. and Segev E. (2000) Location determinants of New foreign-owned manufacturing plants, *Journal of Regional Science*, 40, 323-351.

De Propris L., Driffield N. and Menghinello S. (2005) Local Industrial Systems and the Location of FDI in Italy, *Int. J. of the Economics of Business*, 12, 105--121.

Defever D. (2006) Functional fragmentation and the location of multinationals firms in the enlarged Europe, *Regional Science and Urban Economics*, 36, 658-677.

Disdier A. and Mayer T. (2004) How different is Eastern Europe? Structure and determinants of location choices by French firms in Eastern and Western Europe, *Journal of Comparative Economics*, 32, 280-296.

Guimaraes P., Figueiredo O. and Woodward D. (2004) Industrial location modeling: extending the random utility framework, *Journal of Regional Science*, 44, 1--20.

Hastie T. J. and R. J. Tibshirani (1990) *Generalized Additive Models*, London, Chapman and Hall.

Head, K. and Mayer, T. (2004) Market Potential and the Location of Japanese Investment in the European Union, *The Review of Economics and Statistics*, 86, 959-972

Hoover E.M. (1948) *The Location of Economic Activity*. New York: McGraw Hill.

Iannizzotto M. and N.J. Miller (2002) The effect of exchange rate uncertainty on foreign direct investment in the United Kingdom, Working Paper, *International Economic Association World Congress*, Paris.

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, 99-121.

Kogut B. and Chang S.J. (1991) Technological capabilities and Japanese foreign direct investments in the United States, *The Review of Economics and Statistics*, 73, 401-413.

Krugman P. (1993) First nature, second nature and metropolitan location, *Journal of Regional Science*, 33(2), 129-144.

Krugman P. and A.J. Venables (1990) Integration and the competitiveness of peripheral industry, in C. Bliss and J. Braga de Macedo (eds.) *Unity with Diversity in the European Community*, Cambridge: Cambridge University Press.

Lambert D. (1992) Zero-inflated Poisson regression, with an application to detects in manufacturing, *Technometrics*, 34, 1-14.

McCullagh P. and Nelder J.A. (1989) *Generalized Linear Models*, 2nd edition, London: Chapman and Hall.

Mucchielli J. L. and Puech F. (2003) Internationalisation et localisation des firmes multinationales: l'exemple des entreprises fran¸ caises en Europe, *Economie et Statistiques*, 363-365, 129--144.

Mullahy J. (1986) Specification and testing of some modified count data models, *Journal of Econometrics*, 33, 341-365.

Nefussi B and C. Schwellnus (2008) Does FDI in Manufacturing Cause FDI in Business Services? Evidence from French Firm-Level Data, Working Papers 21, *CEPII research center*.

Ng S. and Pinkse J. (1995) Nonparametric-two-Step Estimation of Unknown Regression Functions when the Regressors and the Regression Error Are not Independent, Cahiers de recherche 9551, Universite de Montreal, Departement de sciences economiques.

Ord J. K. and Getis A. (1995) Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis*, 27, 286--306.

Puga D. (1999) The rise and fall of regional inequalities, *European Economic Review*, 43, 303--334.

Tadesse B. and Ryan M. (2002) The FDI-Trade Relationship: Are Developing Countries Different?, prepared for the 2002 Northeast Universities Development Consortium Conference at Williams College, Williamstown, MA.

Tomlin K. (2000) The Effects of Model Specification on FDI Models: An Application of Count Data Models, *Southern Economic Journal*, 67(2), 460-468.

Thurston S.W., M.P.Wand and J.K. Wiencke (2000) Negative Binomial Additive Models, *Biometrics*, 56, 139-144.

Venables A.J. (1996) Equilibrium locations of vertically linked industries, *International Economic Review*, 37, 341--359.

Wood S. N. (2000) Modelling and smoothing parameter estimation with multiple quadratic penalties, *Journal of the Royal Statistical Society, Series B* (Statistical Methodology), 62, pp. 413-428.

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

Wood S. N. and N. H. Augustin (2002) GAMs with integrated model selection using penalized regression splines and applications to environmental modelling, *Ecological Modelling*, 157, pp.157-177.

Zhou C., Delios A. and Yang J.Y. (2002) Locational Determinants of Japanese Foreign Direct Investment in China, *Asia Pacific Journal of Management*, 19, 63--86