View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

ISSN 1825-0211

SPATIAL CLUSTERING AND NONLINEARITIES IN THE LOCATION OF MULTINATIONAL FIRMS

Roberto BASILE — Luigi BENFRATELLO
Davide CASTELLANI

Quaderno n. 90 — Maggio 2011

QUADERNI DEL DIPARTIMENTO
DI ECONOMIA, FINANZA
E STATISTICA

Spatial clustering and nonlinearities in the location of multinational firms

Roberto BASILE[♣]
Second University of Naples

Luigi BENFRATELLO[♠]
University of Turin and Ceris-CNR

Davide CASTELLANI[♦]
University of Perugia and Centro Studi Luca d'Agliano

Abstract

We propose a semiparametric geoadditive negative binomial model of industrial location which allows to simultaneously address some important methodological issues, such as spatial clustering and nonlinearities, which have been only partly addressed in previous studies. We apply this model to analyze location determinants of inward greenfield investments occurred over the 2003-2007 period in 249 European regions. The inclusion of a geoadditive component (a smooth spatial trend surface) allows to control for omitted variables which induce spatial clustering, and suggests that such unobserved factors may be related to regional policies towards foreign investors Allowing for nonlinearities reveals, in line with theoretical predictions, that the positive effect of agglomeration economies fades as the density of economic activities reaches some limit value.

Keywords: Industrial location, Negative binomial models, Geoadditive models, European Union
JEL classification: C14, C21, F14, F23

[♣] Corresponding author: Corso Gran Priorato di Malta, 1, 81043, Capua (CE). E-mail: roberto.basile@unina2.it
[♠] E-mail: benfratello@econ.unito.it
[♦] E-mail: davide.castellani@unipg.it

1. Introduction

New plants location draws considerable attention among economists and policy makers. By attracting new plants regions can foster their economic development, thereby justifying the investments of (local, national and supranational) institutions to attract firms' establishments. This is all the more true when it comes to the location of multinational firms which can bring foreign technology into a local context and eventually generate significant knowledge and pecuniary externalities (Barba Navaretti and Venables, 2005). The alleged positive effect of foreign establishments contributed to motivate a considerable amount of economic research into the determinants of such a location process. A recent survey of empirical works on the topic reports more than fifty econometric studies, mostly from the last decade and in large proportion focussing on foreign-owned firms (Arauzo-Carod et al., 2010). Despite this extensive amount of research, two issues have hitherto received very little attention: spatial effects and nonlinearities. Whereas the former refers to the spatial clustering of foreign firms' investments, the latter implies that not all regions obey a common linear specification of the industrial location model. Empirical models neglecting these issues may suffer from biased estimates and unreliable significance tests.

In this paper, we address these two issues in the case of the location of greenfield foreign investments in the NUTS2 regions of the Enlarged Europe over the 2003-2007 period. We submit that both spatial clustering and nonlinearities may be relevant in this case. On the one hand, it has been widely documented that multinational firms do not locate randomly in space, but rather they tend to concentrate in few neighbouring regions. This clearly violates the assumption of spatial independence. Regional characteristics included as explanatory variables in the model may partially account for the clustering of foreign firms. However, most likely some spatial correlation remains in the errors, affecting efficiency and consistency of the estimates. This may occur because of partially unobservable spatial factors, reflecting the role of regional amenities and regional policies.

On the other hand, nonlinearities are also very likely to occur in industrial location analysis. For example, theory suggests that a variable effect of agglomeration economies on firms' location decisions might exist according to level of agglomeration. In fact, as agglomeration reaches some critical value a congestion effect may eventually kick-in reducing the attractiveness of a given location. To explore this issue some authors have postulated an inverted-U shaped relation modelled by using squared terms for agglomeration measures (Arauzo-Carod, 2005; Viladecans-Marsal, 2004). Admittedly, this is only one of several competing parametric restrictions which may capture a nonlinear relation. Indeed, nonlinearities can be better accommodated in a semiparametric framework, where the actual shape of the partial effect can be assessed using smooth functions.

The two aforementioned methodological issues are not only far from trivial *per se* but, in our case, are complicated by the fact that, in line with the vast literature on plants location and due the lack of FDI data at regional level, we are forced to use as dependent variable the number of new ventures, i.e. a count which takes discrete and non-negative values. A Generalised Linear Model (GLM) framework, assuming a negative binomial distribution for the conditional expectation of the number of new foreign plants in each region, provides a relatively flexible framework for the analysis of such count data. In fact, as it will be discussed in greater details in Section 4, GLMs lend themselves to an extension into the semiparametric framework, by adding smooth functions of covariates in the conditional expectation. This class of models is known as Generalized Additive Models (GAMs) (Wood, 2006a). Also, by including a *smooth spatial trend surface* - i.e. a nonparametric interaction of latitude (easting) and longitude (northing) - among the regressors, a GAM can be turned into the so-called Geoadditive model (see, *i.a.*, Kammann and Wand, 2003; Wood, 2003; Fahrmeir and Echavarrìa, 2006; Augustin et al., 2009), which allows to take unobserved spatial effects into account.

In sum, we use a Geoadditive Negative Binomial Model (Geo-NB-GAM) and apply it to estimate the determinants of multinational firms' greenfield investment location in the NUTS-2 regions of the Enlarged Europe. To the best of our knowledge only another work uses a semiparametric approach to investigate the determinants of new plant creation in Spanish provinces (Arauzo-Carod and Liviano, 2007). However, our work differs from this one since we thoroughly address spatial effects alongside with nonlinearity, we focus on multinational plants and we take a much broader perspective by studying location in the Enlarged Europe.

Results suggest that multinational firms' location choices are spatially clustered: even controlling for several regional characteristics, such as employment density, market size, Jacobs externalities, human capital, labour cost, unemployment, and density of transport infrastructure, the residuals of a semiparametric additive model turn out to be spatially auto-correlated. A Geo-NB-GAM, which incorporates a smooth spatial trend surface, is able to purge the errors from spatial dependence, leading to unbiased estimates and showing how some omitted factors, mainly regional attraction policies, may lead to cluster of foreign multinationals. The flexibility of the semiparametric approach also allows us to appreciate that whereas some regional characteristics have indeed a linear effect on foreign investments counts, others show some important nonlinearities. In particular, in line with theoretical predictions, the effect of agglomeration economies appear to fade as the density of economic activities reaches some limit value. However, this nonlinear relation does not seem to be well captured by an inverted-U shape: it is monotonically positive but the marginal effect decreases as agglomeration rises and, for a significant portion of our sample, the relation is flat. Therefore, no matter how dense the economic activity becomes, our data suggest that congestion (or competition) effects would never overcome positive agglomeration externalities.

The rest of the paper is organized as follows. Section 2 introduces the dataset on foreign greenfield investments in the European regions and reports the results of an exploratory spatial data analysis which provides some important insights for modelling foreign investment counts. Section 3 presents the theoretical framework which motivates the choice of location determinants included in the empirical model. Section 4 introduces the econometric methodology whereas Section 5 reports the econometric results. Section 6 concludes.

2. Spatial distribution of multinational firms' investments within the European Union

The data we use are retrieved from fDi Markets, an online database maintained by *fDi Intelligence* - a specialist division of the Financial Times Ltd - which monitors cross-border investments covering all sectors and countries worldwide. Relying on media sources and company data, *fDi Markets* collects detailed information on cross-border greenfield investments (available since 2003). Data are based on the announcement of the investment and are daily updated[1]. The database is used as the source for foreign investment project

[1] A team of in-house analysts search daily for investment projects from various publicly available information sources, including, Financial Times newswires, nearly 9,000 media, over 1,000 industry organizations and investment agencies, data purchased from market research and publication companies. Each project identified is cross-referenced against multiple sources, and over 90% of projects are validated with company sources. More information at http://www.fdimarkets.com/.

information in UNCTAD's World Investment Report and in publications by the Economist Intelligence Unit.

We selected 1,930 greenfield investments in the creation of manufacturing plants carried out by both European and non-European multinational firms in the European Union over the 2003-2007 period. For each project, detailed information is available on the investor (name, country of origin and sector of activity, including both manufacturing and services) and on the destination area (country, state and city). This allowed us to count the number of projects in each NUTS-2 region. Five countries (Bulgaria, Latvia, Cyprus, Luxemburg and Malta) and three Spanish regions (Comunitad Autónoma de Ceuta, Comunitad Autónoma de Melilla and Canarias) have been excluded due to the lack of data. The dataset included therefore 22 EU member states (249 NUTS-2 regions).

The distribution of the 1,930 greenfield investments in the manufacturing sector is right skewed (Panel A in Figure 1), with a share of zeros of about 14%, suggesting a high degree of over-dispersion in the raw data and reveals a substantial degree of geographical clustering (Panel B in Figure 1). The latter is all the more evident if we regress the number of investment projects[2] on the smooth interaction between latitude and longitude, $h(no, e)$. Figure 2 plots the geographical components of such a model, showing a *saddle-pattern* in the spatial distribution of foreign investments. They are clustered in two peripheral areas: one includes regions belonging to the New Member States (Eastern countries), while the other characterizes some western peripheral areas (mainly in Ireland, but also in Spain, Portugal and France).

- **Insert Figure 1 about here** -

[2] The details on how to estimate this effects are provided in Section 4. Suffice here to mention that we have used a thin plate regression spline as smoother (Wood, 2006a).

In sum, the exploratory spatial data analysis suggests to model foreign investment counts bearing in mind on the one hand the issues of non-normality, skewness and over-dispersion and, on the other hand, the presence of spatial clustering. While accounting for the former issue is rather straightforward using negative binomial models, the latter can be tackled modelling the mean function though a set of covariates intended to capture the spatial clustering. In the next section, we discuss some theoretical hypotheses on the location determinants of foreign ventures which suggest a proper set of explanatory variables needed to model the expected mean function. Should regional observable characteristics not be able to account for the clustering entirely, residual spatial dependence would remain in the error term, causing possible biases in the estimated coefficients. We will tackle this issue by including smooth spatial trends surfaces. This will also allow us to highlight where are the unexplained regional clusters, and assess what would be the factors behind their formation.

3. Determinants of the spatial distribution of inward foreign investments

The spatial distribution of foreign investments can be modelled as the result of the interaction between centripetal (or agglomeration) and centrifugal (or competition) forces Among agglomeration forces, we focus on the role of urbanization externalities, while as for centripetal forces we consider the effect of labour cost along with other labour market characteristics[3].

[3] Admittedly, data availability prevents us to include additional determinants. The role of regional policy measures aimed at attracting foreign firms' investments in order to boost regional development have been investigated by several cross-regional studies (see, *i.a.*, Wheeler and Mody, 1992; Head et al., 1999; Crozet et al., 2004). Other studies have analyzed the effect of national policies and national institutional settings

3.1. Urbanization economies

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 urbanization externalities). According to Marshall, industrial firms tend to localize where other firms of the same industry are already established. The well known benefits of this form of externality are three-fold: *i*) access to a more stable labour market, *ii*) availability of intermediate goods, production services and 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 (mainly traditional) industry (for example, clothing or footwear). Unfortunately, we do not have detailed information on the specific industry where the new plant operates so that we are only able to count the total number of manufacturing foreign investments in the regions. In turn, we are forced to consider only the role of urbanization externalities[4] which we model using four different variables: 1) the size of the regional market, 2) the overall employment density in the manufacturing sector, 3) the degree of sectoral diversification and 4) the level of road infrastructures.

(corporate tax, labour market institutions, bureaucratic efficiency and corruption, legal system and intellectual property right protection, product market regulation and openness to FDI) on regions' performance in attracting foreign investors (Basile et al. 2006; Barrios et al., 2008). Finally, European policies (such as the Structural and Cohesion funds allocated to EU laggard regions) can also be important factors affecting the attractiveness of a location (Basile et al., 2008). Unfortunately, comparable data on institutional variables are not be available for all the regions in our sample.

[4] It is worth mentioning that there exist some empirical evidence that urbanization economies outweigh industry-specific localisation economies (Guimarães et al. 2000, Arauzo-Carod and Liviano, 2007).

The *size of regional market* (measured by the log of total value added) is intended to capture externalities which have to do with the "*home market effect*". As first noted by Krugman (1980), under increasing returns to scale and in presence of transport costs, the appeal of a country (or region) as a production site depends crucially on the size of its domestic market. Firms will locate in the region where they can exploit economies of scale to a greater extent and, eventually, export to neighbouring regions.

Employment density in manufacturing (measured as total manufacturing employment per square km) represents the scale of agglomeration economies. We expect that regions with higher density of economic activity attract more foreign investments due to agglomeration effects. However, the occurrence of congestion costs (including higher land prices, higher crime rates, environmental pollution, traffic jams, excess commuting and so on) may compensate the positive effect of agglomeration economies and, thus, determine a threshold effect in the positive impact of employment density. In other words, regions tend to attract foreign investors if, *ceteris paribus,* agglomeration economies overcome congestion costs. Therefore, a nonlinear effect of employment density on the number of inward foreign investments is expected. Some empirical studies upfront assume an inverted-U shaped relationship between agglomeration and location, thereby inserting the measure of agglomeration economies squared as additional regressor (Viladecans-Marsal, 2004; Arauzo-Carod, 2005). Although it is the easiest way to deal with such a nonlinearity in a parametric framework, this is only one of many possible nonlinear parameterizations. In particular, this specification assumes that at some point congestion costs would be higher than positive agglomeration externalities so that an increase in employment density would discourage new investments.

Sectoral diversification within regions (measured by the median of the sectoral specialization indexes[5]) is meant to capture urbanization externalities deriving from diversity or variety of the regional economy (*Jacobs externalities*). According to Jacobs (1969), a diverse sectoral structure increases the chances 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 extent of *road infrastructures* (kilometres of motorways per squared kilometres) should pick up the component of urbanization economies due to the provision of *public goods*. A higher level of public goods (in particular infrastructures) is likely to increase firm productivity and to reduce transport costs, lowering the cost of inputs sourced and facilitating the access to markets. The ensuing increase in private returns to investments makes locations with better infrastructure provisions more attractive for both domestic and foreign investments.

3.2. Labour market characteristics

The role of labour market characteristics as a determinant of inward foreign investments and new plant creation is well established (Friedman et al., 1992). In this paper we follow previous literature by specifying the regional labour market characteristics using three different variables: the average *wages* (measured by the total compensation to labour divided by the number of employees in the region), *labour availability* (approximated by the

[5] The specialisation index for each sector *s* and region *i* is the following employment location quotient: $S_{si} = (E_{si} / \sum_s E_{si}) / (\sum_i E_{si} / \sum_i \sum_s E_{si})$, where E denotes employment. The median of S_{si} is a measure of the number of sectors in which a region shows a revealed comparative advantage: when the median is high it means that a region has a comparative advantage in a large number of sectors and it is therefore diversified; when the median is low, it means that a region is specialized (see De Benedictis and Tamberi, 2004).

unemployment rate) and *human capital* endowment (approximated by the share of population aged 24 or more holding a tertiary education degree). The impact of wages and unemployment is not univocal, however. Lower wages may in fact attract firms seeking lower labor costs (that is firms pursuing *cost reducing strategies*), but high wages may signal highly skilled workers which in turn attract location of higher value added activities. Furthermore, firms may interpret unemployment both as a measure of a large supply of labor, which would attract firms, and as an indicator of a relatively rigid labor market, which would discourage them. In sum, the effects of basic labor market conditions may be in principle characterized by some nonlinearities which should be properly accounted for when modeling foreign investors location decisions.

4. Modelling regional inward foreign investment counts: a geoadditive negative binomial model

The empirical literature on foreign firms' location choice usually appeals to discrete-choice models (conditional, nested and mixed logit models) based on the Random Utility Maximization (RUM) framework. Decision probabilities are therefore modelled in a partial equilibrium setting where foreign firms maximize profits subject to uncertainty that derives from unobservable characteristics. The use of discrete choice models is often hindered by the large dimension of the choice set (in our case 249 regions) which makes estimation very burdensome, however. An alternative modelling strategy aggregates data at the elementary choice level by counting the number of times a given alternative is chosen (i.e. the number of investments in each region) leading to discrete, non-negative integer valued dependent variables (so-called count data). Under relatively mild conditions, the coefficients of a Poisson regression can be proven to be equivalent to those of a conditional logit model

(Guimaraes et al., 2004). Therefore, also the Poisson regression model can be thought as being directly derived from a RUM process.

The Poisson regression model can be accommodated as a special case of the Generalized Linear Model (GLM) framework (McCullagh and Nelder, 1989). Let *y*, be the dependent variable, X a $k \times 1$ vector of explanatory variables and β a $k \times 1$ vector of regression parameters. The canonical smooth monotonic *link function* in the Poisson GLM is

$$g(\mu) = \log(\mu) = \eta = X^{'}\beta\,; \quad \mu = E(y)\,; \qquad y = Poi(\mu) \quad (1)$$

resulting in a log-linear relationship between mean and linear predictor. A characteristic of the Poisson regression model is the assumption of equidispersion, that is $\mu = E(y) = Var(y)$[6].

In practice, however, the classical Poisson regression model is often of limited use in a regional location analysis since empirical inward foreign investment counts typically exhibit over-dispersion, i.e. $Var(y) > E(y)$. A way of dealing with over-dispersed count data is to assume a negative binomial (NB) distribution for $y \mid X$ which can arise as a gamma mixture of Poisson distributions. One parameterization of its probability density function is

$$P\left(Y = y \mid X^{'}\beta, \theta\right) = \frac{\Gamma(y+\theta)}{\Gamma(\theta)\cdot y!} \cdot \left(\frac{\mu}{\mu+\theta}\right)^{y} \cdot \left(\frac{\theta}{\mu+\theta}\right)^{\theta} \quad (2)$$

[6] It is worth mentioning that for GLM, maximum likelihood estimation (MLE) by a Newton type method can expressed as an Iteratively Re-weighted Least Squares (IRLS) scheme.

with mean μ and shape parameter θ (an index of over-dispersion) following a gamma distribution with parameters *a* an *b*, $\Gamma(a,b)$. The variance function is now $V(\mu)=\mu+\mu^2\theta^{-1}$. Note that, for large θ, the model approaches the Poisson model.

A large number of studies on regional inward foreign investment counts have used the standard NB regression model with cross sectional data (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) or random effects extensions of NB regression for panel data Blonigen (1997), Basile (2004) and Basile et al. (2006) [7]. [8]

A limit of this recent literature on inward foreign investment counts is the assumption that all regions obey a common linear specification of the location model, disregarding likely nonlinearities reflecting spatial heterogeneity in the behaviour of economic agents. In particular, as stated above, we cannot disregard possible threshold effects in the impact of agglomeration externalities on regional attractiveness.

Nonlinearities can be addressed in different ways. Firstly, polynomial expansions up to a cubic can be considered within a GLM approach. Although rather easy to implement, this solution might introduce severe multicollinearity. Secondly, Geographically Weighted Regression (GWR) models represent a standard method to properly handle spatial

---

[7] For a detailed description of these and other works, see Arauzo-Carod et al. (2010).

[8] Although negative binomial models capture overdispersion quite well, they are not always sufficient for modelling excess zeros. To overcome this problem, zero-augmented models that incorporate a second model component capturing zero counts have been proposed. Zero-inflation models (Lambert, 1992) are mixture models that combine a count component and a point mass at zero. Hurdle models (Mullahy, 1986) instead combine a left-truncated count component with a right-censored hurdle component. Applications of these models to FDI location analyses are in Tadesse and Ryan (2004), Basile (2004), Tomlin (2000) and Iannizzotto and Miller (2002).

instability problems. However, as far as we know, there are no extensions of GWR models for overdispersed data. A third solution, that will be considered in this paper, is the Generalized Additive Model (GAM) taking advantage of the recent development of a Negative Binomial Additive Model (NB-GAM) to handle Negative Binomial responses (Thurston et al., 2000).

The GAM framework (Hastie and Tibshirani, 1990) extends the GLM by allowing nonlinearity in the relationship between η and the covariates:

$$\eta = g(\mu) = X^{'*}\beta^* + f_1(x_1) + \ldots + f_2(x_2) + f_3(x_3, x_4) + \varepsilon \qquad (3)$$

$$\varepsilon \sim N(0, \sigma_\varepsilon^2) \qquad \mu = E(Y) \qquad Y \sim negbin(\mu, \theta) \qquad \theta \sim \Gamma(a, b)$$

where $f_j(.)$ are unknown smooth functions of the covariates, $X^{'*}$ is a vector of strictly parametric components and β^* is corresponding parameter vector.

Each univariate smooth term in (3) can be represented as $f_j(x_j) = \sum_{k=1}^{K_j} \beta_{jk} b_{jk}(x_j)$, where the $b_{jk}(x_j)$ are known basis functions, while the β_{jk} are unknown parameters to be estimated. One or more measures of '*wiggliness*' $\beta_j^{'}\mathbf{S}_j\beta_j$, where $\mathbf{S}_j$ are positive semidefinite matrices, is associated with each smooth function. Typically, the wiggliness measure evaluates something like the univariate spline penalty $\int f_j^{''}(x_j)^2 dx$ or its thin-plate spline generalization (Wood, 2006a). The penalized spline base-learners can be extended to two or more dimensions to handle interactions. Specifically, Wood (2006a) recommends to use thin-plate regression splines for smooth interactions of quantities measured in the same units (such as the spatial coordinates) and tensor products for smooth interactions of quantities measured in different units.

Given bases for each smooth term, model (3) can be re-written as a GLM, $g\{\mu\} = X^{'}\beta$ where X includes the columns X^{*} and columns representing the basis functions evaluated at the covariate values, while β contains β^{*} and all the smooth coefficient vectors β_j. The model is estimated by minimizing the penalized deviance:

$$D(\beta) + \sum_j \lambda_j \beta^{'} \mathbf{S}_j \beta \qquad (4)$$

with respect to β, where λ_j are positive smoothing parameters and $D(\beta) = 2\{l_s - l(\beta)\}\phi$ is the model deviance, with l the log-likelihood, l_s the saturated log-likelihood, and ϕ the scale parameter.

Penalized maximum likelihood estimation MLE is performed by Penalized Iteratively Re-weighted Least Squares (PIRLS). Let $\hat{\beta}$ and $\hat{\mu}$ be current best estimates, pseudo value - $z = g^{'}(\hat{\mu})(y - \hat{\mu}) + X\hat{\beta}$ - and weights - $W = \left\{V(\hat{\mu})g^{'}(\hat{\mu})^2\right\}^{-1}$ - are firstly computed. Then, the weighted penalized least square problem of minimizing $\left\|\sqrt{W}(z - X\beta)\right\|^2 + \sum_j \lambda_j \beta^{'} \mathbf{S}_j \beta$ w.r.t. β is solved to obtain the updated $\hat{\beta}$ and $\hat{\mu}$. Iterating these two steps to convergence leads to the penalized likelihood maximizing estimates.

Given λ_j, therefore, computing the estimates $\hat{\beta}_\lambda$ is straightforward. But λ_j are unknown and must be estimated. Woods (2006a, 2008) has proposed different methods for automatic and integrated smoothing parameters selection. The first one is called "*performed iteration*", a grid search provides estimates for λ_j at each PIRL step based on Generalised Cross Validation (GCV). In this case, the parameter θ is chosen in order to ensure that the estimate of the scale parameter is as close as possible to 1, the value that the scale parameter should have. With the second method, termed "*outer iteration*", the scale parameter is set to 1 and a selection criterion for λ_j and for θ is defined to optimize them directly by minimizing the Akaike Information Criterion (AIC). In this case, the PIRL scheme is iterated to convergence for each trial set of smoothing parameters and AIC scores are only evaluated on convergence; optimization is then "outer" to the PIRLS loop. Finally, within the "*outer iteration*" method, it is possible to resort to Restricted Maximum Likelihood (REML), instead of the AIC, due to the possibility of rewriting the penalized GAM as a generalized mixed model (see, Wood, 2006b; Ruppert et al. 2003; Kammann and Wand, 2003). Simulation results suggest that the REML method offers some improvement in mean-square error performance relative to GCV and AIC in most cases (Wood, 2011). Wood has implemented these techniques in the R package *mgcv*.

Both GLM and GAM assume that the errors are independent. In regional location analysis the assumption of spatial independence is often violated, however. As shown in Section 2, indeed, investments from multinational firms tend to cluster over space and this does not necessarily reflect the distribution of observed explanatory variables. Spatial autocorrelation may indeed occur (and should be controlled for when specifying a location model) when unobserved variables are spatially correlated. For example, a number of factors related to culture, policy actions (incentives, corporate taxes and other institutional characteristics) and various forms of amenities can affect the regional attractiveness. Unfortunately, these factors are often either unobservable or cannot be properly measured, especially in samples composed of a lot of small geographical units (as in our case). In so far as these variables are spatially correlated, the residuals will be spatially correlated too.

The issue of spatially correlated unobserved variables can be solved in many ways, among which the most diffused one in the empirical economic literature is the application of spatial econometrics tools (Anselin, 1988; LeSage and Pace, 2009). While the use of spatial lag and/or spatial error models to the analysis of industrial location would be desirable due to their useful interpretation in terms of spatial externalities (Anselin, 2004)[9], there are no extensions of spatial econometrics tools to semiparametric additive models with non-Gaussian response variables.[10] Thus, an alternative solution to account for spatial effects is to extend the GAM framework by incorporating the spatial location as an additional covariate, that is by including the bivariate smooth term of longitude (*northing*) and latitude (*easting*), $h(no, e)$ in model (3), thus generating what is known as the Geostatistical Additive Model (in our case Geo-NB-GAM):[11]

$$\eta = g(\mu) = X^{\prime *}\beta^{*} + f_1(x_1) + \ldots + f_2(x_2) + f_3(x_3, x_4) + h(no, e) + \varepsilon \qquad (5)$$

$$\varepsilon \sim N(0, \sigma_{\varepsilon}^{2}) \qquad \mu = E(Y) \qquad Y \sim negbin(\mu, \theta) \qquad \theta \sim \Gamma(a, b)$$

Geoadditive models are widely used in environmental studies and in epidemiology (see, *i.a.*, Kelvyn and Wrigley, 1995; Kammann and Wand, 2003; Augustin et al. 2009), but are rarely considered for modelling economic data, and, to the best of our knowledge, this is the first application to the location of industrial plants.

5. Evidence from parametric and semi-parametric regressions

In this section we present the results of our econometric analysis. First, the results from a standard GLM results will be illustrated, and then the we highlight to what extent

---

9 Example of spatial econometric methods to FDI location models are in Coughlin and Segev (2000), Blonigen et al. (2007), Baltagi et al. (2007, 2008). All these articles adopt a parametric approach, thereby neglecting possible nonlinearities.

10 A semiparametric extension of the spatial lag model in presence of Gaussian responses has been proposed in Basile (2009) and Basile and Girardi (2010).

11 As emphasised in Wood (2003), the spatial term can be estimated using a low-rank thin plate regression spline.

estimating a GAM and a Geo-GAM, which allow for nonlinearities and controlling for spatial dependence, affects the results.

5.1 GLM estimation results

Table 1 reports coefficients and standard errors (in parenthesis) estimated with parametric GLMs. The dependent variable is the regional number of greenfield investment projects from foreign multinationals directed to each region over the 2003-2007 period. As discussed in Section 3, the explanatory variables are *mkt* (the market size, approximated by the regional total value added), *infra* (a measure of transport infrastructure), *Jacobs* (a proxy for Jacobs externalities), *empdens* (the employment density in manufacturing), *wage* (the average labour cost), *ur* (the regional unemployment rate) and *ter* (the level of tertiary education).[12] All coefficients turn out to be statistically significant and with the expected sign both in the Poisson and in the NB regression models. Since all variables are in logarithms, coefficients can be interpreted as elasticities. First, the positive coefficient associated to *mkt* confirms that foreign firms concentrate where demand is highest and possibly serve smaller markets via exporting. Second, the expected number of foreign greenfields increases with the density of transport infrastructures. Third, Jacobs externalities have a strong positive effect, indicating that a more diversified regional economy is conductive of new foreign firms. Fourth, a higher employment density increases the expected number of greenfield investments into the region. Therefore, on average, it seems that congestion costs are more than counteracted by agglomeration externalities. Finally, high wages seem to discourage foreign investments, while high regional unemployment and tertiary education attract foreign investors.

- **Insert Table 1 about here** -

Table 1 also reports some diagnostics tests and measures of goodness of fit. The value of the Akaike Information Criterion clearly works in favour of the NB model. The NB model seems to perform rather well both against the Poisson model, in that is able to account for over-dispersion[13], and against a Zero-Inflated NB model (ZINB) and a Hurdle model, as the null hypothesis of no excess of zeros with respect to the prediction of the NB distribution cannot be rejected.[14] To assess the extent of spatial autocorrelation, we regress the vector of residuals on a spatial trend surface ($res = f(no, e) + \zeta$) and compute an F-test for the overall significance of this smooth term. As Table 1 shows, the GLM models display significant spatial dependence[15].

5.2 GAM estimation results

Tables 2 and 3 summarize the results of different semiparametric GAMs. As a first step, we estimate a nonparametric model where all covariates are allowed for possibly nonlinear effects (Table 2). Although we have some theoretical priors about the functional form of

---

12 See the appendix for a thorough definition of the variables.

13 The over-dispersion test is based on the estimation of the simple nonparametric model $|e| = f(y) + u$, where $|e|$ is the absolute value of the residuals of the GLM and y is the vector of fitted values. Under the null hypothesis of equi-dispersion, the smooth term $f(y)$ must be estimated with one degree of freedom and, according to a F test, it should have an insignificant effect on $|e|$ (see, Thurston et al., 2000).

14 As the standard ZINB is not-nested in the NB model, a Vuong test is applied. 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 our case, the Vuong test statistic proves to be not significant, so that the existence of zero-inflation can be excluded. The Wald test for the NB against an Hurdle model also points towards the use of the NB model.

the relationship between inward investments and some explanatory variables (see Section 3), we prefer to be agnostic about which covariate should enter the model linearly. Therefore, we specify all terms non-parametrically (using cubic regression splines) and test for which variable a parametric specification could not be rejected. This test is based on the Effective Degrees of Freedom (*edf*, henceforth) estimated for each smooth function. If the *edf* is equal to 1, a linear relationship cannot be rejected. However, in order to assess the robustness of the results of the linearity test, we compare the evidences obtained from the three different methods for automatic smoothing parameter choice recalled in Section 4: *i*) GCV, *ii*) AIC and *iii*) REML. The evidence clearly reveals that the *edf* is equal to 1 for *Jacobs, infra* and *ter* regardless of the method used for estimation, while mixed results emerge for *ur* and *mkt*. Since the AIC values of the three models are almost equal, in statistical terms the three models are equivalent. Considering the relatively small size of our sample (249 observations), we believe that a more parsimonious strategy, which minimizes the number of non-parametric terms, has some advantages. Thus, we decided to rely on the REML results and assume a linear parametric form for the effect of *ur* and *mkt*. Finally, since the *edf* is always higher than 1 in the case of *empdens* and *wage*, these two variables undoubtedly enter the model non-linearly.

- **Insert Table 2 about here** -

Given this evidence, in Table 3 we estimate two semiparametric models with all liner terms but for *wage* and *empdens* using REML. In the first specification, we do not introduce any control for spatial effects, whereas the second model introduces the smooth interaction

---

[15] As a robustness check, we also computed the Moran test and the presence of spatial correlation is confirmed. However, it should be born in mind that no studies have investigated how accurately this test detects spatial autocorrelation in GLM and GAM errors.

between latitude and longitude (specified using a thin plate regression spline) to capture nonlinear spatial trends. As measured by the AIC, the fit of the more parsimonious model is not worse than the fully non-parametric specifications presented in Table 2. Thus, we are reassured that some terms could enter linearly. At the same time, the AIC of the GAM is much lower than the for the GLM (presented in Table 1), suggesting that indeed allowing for non-linearity in *empdens* and *wage* improves the model fit, It also worth noticing, , that the Geo-NB-GAM encompasses all the other models. Moreover, diagnostics on the residuals, reported in Table 3 reveal that spatial dependence is still significant in the GAM, and only in the Geoadditive model, where the spatial trend surface is included, residuals are purged from spatial dependence. In other words, even controlling for a number of key regional characteristics and allowing for some non-linearities, we are not able to explain all of the actual clustering of foreign investments in European regions. This suggests that some unobserved characteristics may still affect the location patterns. We will return on this issue later in this section.

- **Insert Table 3 about here** -

The upper part of Table 3 reports the estimated parameters and standard error for the parametric terms. Despite the fact that the magnitude of most coefficients do change from the GLM to the GAM and Geo-GAM, they keep the same sign and statistical significance. Thus, our results suggest that controlling for non-linearities and spatial dependence do help reducing some bias in standard GLM estimates, but in this case, the bias is not as strong to induce a wrong interpretation of the effects of our location determinants,

The middle part of the Table reports χ^2-tests and p-values for the overall significance of the smooth terms for *wage* and *empdens*, as well as their *edf*. Low values of the

χ^2-test imply a high probability that the estimated smooth term is not different from zero, while as stated above *edf* is a measure of its nonlinearity. Results from both models support the hypothesis that both *wage* and *empdens* are statistically significant determinants in the location of new foreign-owned plants in European regions and that their relationship with the number of new investments is nonlinear. The spatial smoothing term, $h(no, e)$, is also highly significant, suggesting the presence of a geographical pattern in regional location of foreign investments in Europe even after controlling for all the other covariates.

Figure 3 shows the smoothed partial effects of *wage* (Panel A) and *empdens* (Panel B) on the expected number of foreign investments estimated through the Geo-GAM. The shaded areas highlight the 95% confidence intervals. The wage-plot (Panel A) suggests that regions with low average labour costs tend to attract more foreign investments, after controlling for the other variables. However, the effect of a wage drop appears higher for intermediate wages levels and decreases in regions with either very low or very high wages. This latter results is consistent with the idea that in high-end regions the wage rate does not only captures labour cost, but also proxies for its quality. Thus, an increase in wages may not discourage multinationals after all.

As for *empdens,* the graph in Panel B of Figure 4 shows that the expected number of inward foreign investments increases with the employment density in manufacturing, up to a point where the relation becomes basically flat and not significantly different from zero. This is consistent with the hypothesis that a more dense industrial activity can exert a positive externality which promotes the location of foreign firms but, when the level of agglomeration becomes too high, congestion costs kick-in and gradually reduce the magnitude of the positive externality, up to the point where an increase in employment density as no further effect on foreign entries. It should be noted that, in our sample, the relationship between employment density and location of foreign plants is nonlinear but does not appear to be inverted-U shaped, as most studies using parametric specifications had anticipated. In fact, an inverted-U relation would predict that investments would eventually decline for very high values of *empdens*, whereas our smoothing function does not show such a declining pattern.

- **Insert Figure 3 about here** -

Finally, Figure 4 displays the perspective (Panel A) and the map (Panel B) plot of the geographical components of the geoadditive model. The plots show that, after controlling for the most relevant variables and allowing for nonlinearities, there are still some unexplained clusters of foreign investments in the north England and in Ireland and, to a lesser extent, in Spain and in Portugal. This evidence can be interpreted in the light of the quite effective policies to attract foreign investors that some national and regional institutions have adopted. Due to the lack of comparable information across countries and regions on policies towards foreign investors, we cannot account explicitly for the different effectiveness of such policies. However, these unobserved regional characteristics are picked up by the spatial trend surface, so from a statistical point of view, in our Geo-GAM the omission of these variables does not lead to biased estimates.. It is worth mentioning, that if we compare the plots in Figure 2 (where we did not control for any regional characteristic) with those in Figure 4, we notice that the strong clustering in Eastern European countries, which was evident in Figure 2, has almost vanished in Figure 4, suggesting that it can be explained mainly by the explanatory variables (most likely the wage rate).

- **Insert Figure 4 about here** -

6. Conclusions

This paper contributes to the extensive literature on the determinants of industrial location by addressing two largely unexplored issues: spatial clustering and nonlinearities. Using data on greenfield projects in the NUTS-2 European regions, we have estimated first a semiparametric count data models and then a geoadditive one. The use of these models presents some advantage over standard parametric models which neglect spatial correlation. First, the estimated parameters of the variables entering linearly the models can have some remarkable change in magnitude with respect their parametric counterparts. In other terms, failing to take into account non linearities and spatial correlation might lead to biased estimates. Beyond this, the use of these models allows us to highlight the role of spatial correlation and nonlinearities. As for the former, it cannot be fully accounted for in our sample spatial by several spatially varying covariates, such as urbanization externalities and labour market characteristics. The use of a geoadditive model for overdispersed count data incorporating a spatial trend surface, not only is able to purge spatial correlation in the residuals but also allows us to show that, after controlling for the most relevant variables, some relevant clusters of foreign investors in the UK, Ireland, Spain and Portugal still emerge. Not surprisingly, these countries are also those that have introduced many policy actions to attract foreign firms' investments within their regional development strategies. However, the introduction of the additive geographical component prevents these unobserved variables to bias the results. Finally, we have also been able to identify some important nonlinearities. In particular, we have provided evidence that, in line with theoretical predictions, the effect of agglomeration economies fades as the density of economic activities reaches some limit value.

To sum up, our result do suggest that the use of these flexible and general models has some clear advantages over those traditionally employed in the analysis of industrial location. Given the availability of ready-to-use routines, a widespread use of these models is to be welcomed.

Appendix: definition of explanatory variables

- ***Market size***: log of total value added in the region (source: Cambridge Econometrics).
- ***Jacobs externalities***: median specialisation index for each sector *i* and region *j*. Each sectoral index is calculated as the following employment location quotient: $S_{ij} = (E_{ij} / \sum_i E_{ij}) \big/ (\sum_j E_{ij} / \sum_j \sum_i E_{ij})$, where E denotes employment (source: Cambridge Econometrics)
- ***Employment density:*** number of people employed in the manufacturing industry per km^2.
- ***Public infrastructure:*** Length of highways and other roads network (in kilometres) divided by total population in the region.
- ***Tertiary education:*** share of adults (population aged 25-64) with tertiary education (ISCE97 codes 5 and 6) averaged over the 1999-2002 period (source: Eurostat). For the regions DE41 and DE42 data on tertiary education were available only for the years 2004 and 2005.
- ***Labour cost***. Source: Cambridge Econometrics. For German and UK NUTS2 regions, for which data are available only at the NUTS1 level, we have attributed the value of the NUTS1 they belong to.
- ***Unemployment rate***. Source: Cambridge Econometrics.

Since we estimated the effect of all these variables over the 2003-2007 period, we used – if possible – all explanatory variables averaged over the 2000-2002.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models,* Dordrecht: Kluwer Academic Publishers.

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

Arauzo-Carod, J.M. (2005). "Determinants of Industrial Location: An Application for Catalan municipalities", *Papers in Regional Science* 84 (1), 105-120.

Arauzo-Carod, J.M., and D. Liviano. (2007). "Agglomeration and Location: a Nonparametric Approach", Working Paper 5-2007, Universitat Roviri I Virgili.

Arauzo-Carod, J.M., Liviano-Solis, D. and Manjon-Antolin, M. (2010). "Empirical Studies in Industrial Location: An Assessment of their Methods and Results", *Journal of Regional Science*, 50 (3): 685-711.

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.

Augustin, N., Musio, M., von Wilpert, K., Kublin, E., Wood, S. and Schumacher, M. (2009). "Modelling spatio-temporal forest health monitoring data", *Journal of the American Statistical Association*, 104 (487), 899-911.

Baltagi, B. H., Egger, P. and Pfaffermayr, M. (2007). "Estimating models of complex FDI: Are there third-country effects?," *Journal of Econometrics*, 140(1), 260-281.

Baltagi, B. H., Egger, P. and Pfaffermayr, M. (2008). "Estimating regional trade agreement effects on FDI in an interdependent world," *Journal of Econometrics*, 145(1-2), 194-208.

Barba Navaretti, G. and Venables, A. (2005). *Multinational Firms in the World Economy*, Princeton University Press.

Barrios, S., Huizinga, H., Laeven, L., and Nicodème, N. (2008). "Tax and multinational firm location decisions", CEPR DP 7047.

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. (2009). "Productivity polarization across regions in Europe: The Role of Nonlinearities and Spatial". *International Regional Science Review*, 31, 92-115.

Basile, R., Benfratello, L. and D. Castellani (2006). "Attracting Foreign Direct Investments in Europe: Are Italian Regions Doomed?", In Marco Malgarini 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.

Basile, R. and Girardi, A. (2010). "Specialization and Risk Sharing in European Regions", *Journal of Economic Geography*, 5, 645-659.

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

Blonigen, B., Davies, R., Waddell, G. and Naughton, H. (2007). "FDI in space: Spatial autoregressive relationships in foreign direct investment", *European Economic Review*, 51 1303–1325.

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

Crozet, M., Mayer, T. and J. Mucchielli (2004). "How do firms agglomerate? A study of FDI in France", *Regional Science and Urban Economics*, 34, 27– 54.

De Benedictis, L., Tamberi, M. (2004). "Overall specialization empirics: techniques and applications". *Open Economies Review*, 15: 323–346.

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

Diggle, P.J., Ribeiro, P.J. (2007). "Model-based Geostatistics". New York: Springer.

Fahrmeir, L. and Echavarrìa, L.O. (2006). "Structured additive regression for overdispersed and zero-inflated count data". *Applied Stochastic Models in Business and Industry* 22, 351–369.

Friedman, J., Gerlowski, D., Silberman, J. (1992). "What attract foreign multinational corporations? Evidence from branch plant location in the United States". *Journal of Regional Science* 32, 403-418.

Guimaraes, P., Figueiredo, O. and D. Woodward (2000). "Agglomeration and the Location of Foreign Direct Investment in Portugal", *Journal of Urban Economics*, 47, 115-135.

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

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

Head, K., Ries, J. and D. Swenson (1999). "Attracting foreign manufacturing: Investment promotion and agglomeration", *Regional Science and Urban Economics*, 29 (1999) 197–218.

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.

Jacobs, J. (1969). *The Economy of Cities*, Random House.

Kammann, E.E., and M. P. Wand. (2003), "Geoadditive Models". *Applied StatIstics*, 52, 1–18.

Kelvyn, J., and N. Wrigley. (1995), Generalized Additive Models, Graphical Diagnostics, and Logistic Regression, Geographical Analysis, Geographical Analysis, 27, 1.

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

Krugman, P. (1980). "Scale economies, product differentiation and the pattern of trade", *American Economic Review* 70, 950-959.

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

LeSage, J., and R. K. Pace. (2009). *Introduction to spatial econometrics*, CRC Press Inc.

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

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

Ruppert D., M. P. Wand, and R. J. Carroll. (2003). *Semiparametric Regression*, Cambridge University Press.

Schabenberger, O., and C. A. Gotway. (2005). *Statistical Methods for Spatial Data Analysis*, Chapman & Hall/CRC, Boca Raton.

Tadesse, B. and M. Ryan. (2004). 'Host market characteristics, FDI, and the FDI – trade relationship, *The Journal of International Trade & Economic Development*, 13: 2, 199—229.

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

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

Viladecans-Marsal E. (2004). "Agglomeration economies and industrial location: city-level evidence", *Journal of Economic Geography* 4/5: 565-582.

Wheeler, D., and A. Mody. (1992). "International investment location decisions. The case of US firms", *Journal of International Economics*, 33, 57-76.

Wood, S.N. (2003). "Thin plate regression splines". *Journal of the Royal Statistical Society*, Series B, 65, 95–114.

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

Wood, S. N. (2006b). "Low-Rank Scale-Invariant Tensor Product Smooths for Generalized Additive Mixed Models", *Biometrics*, 62, 1025–1036.

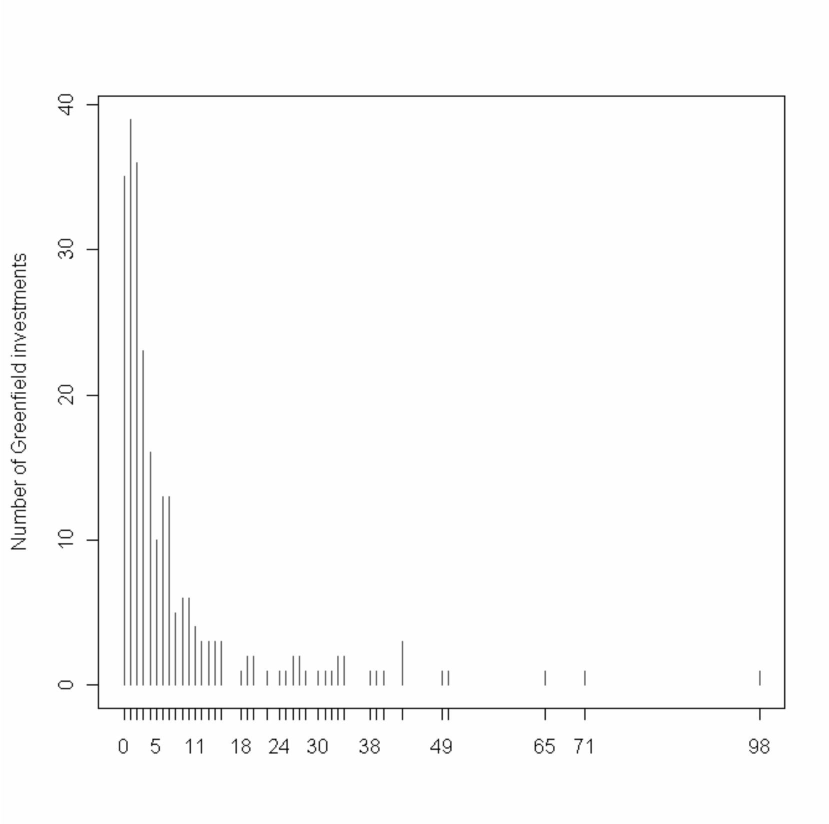
Wood, S. N. (2008). "Fast stable direct fitting and smoothness selection for generalized additive models" *Journal of the Royal Statistical Society, Series B* (Statistical Methodology), 70, 495-518.

Wood, S. N. (2011). "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models", *Journal of the Royal Statistical Society, Series B* (Statistical Methodology), forthcoming.

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

Figure 1 – Regional distribution of foreign investments in the European Union

A – Hystogram



B – Circle map

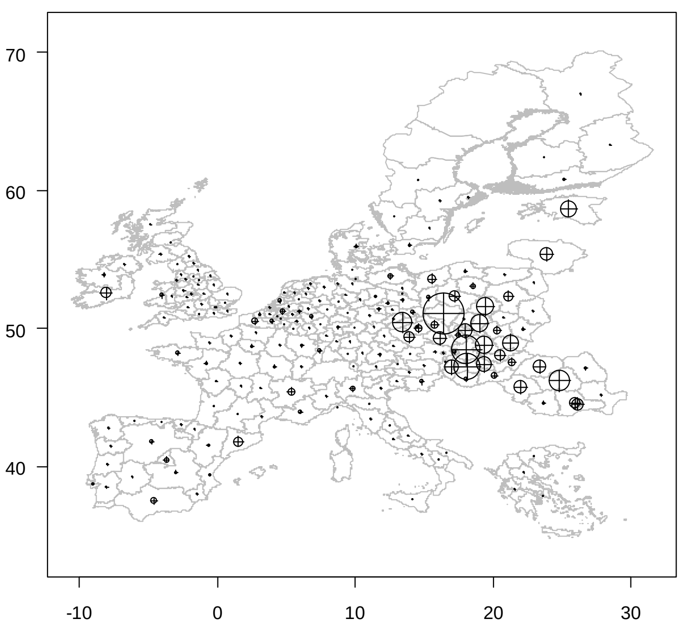
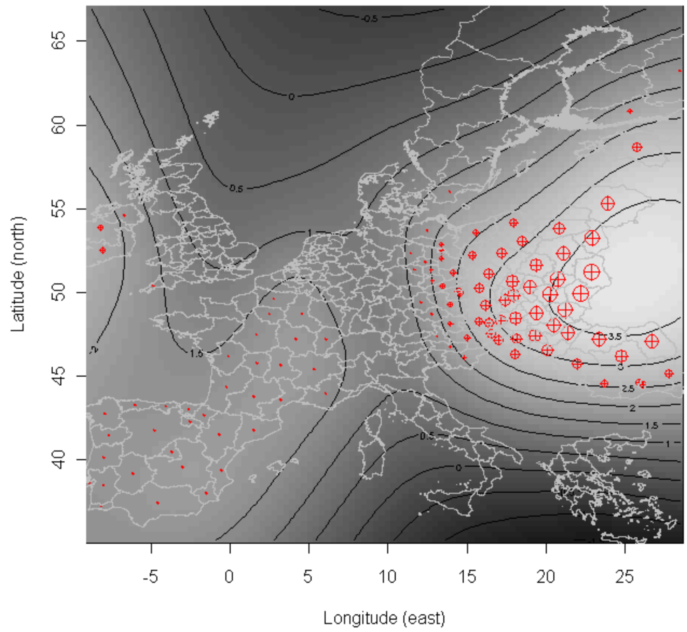


Figure 2 – Smooth trend surface of foreign investments in the European Union

A - Map plot



B – Fit perspective plot

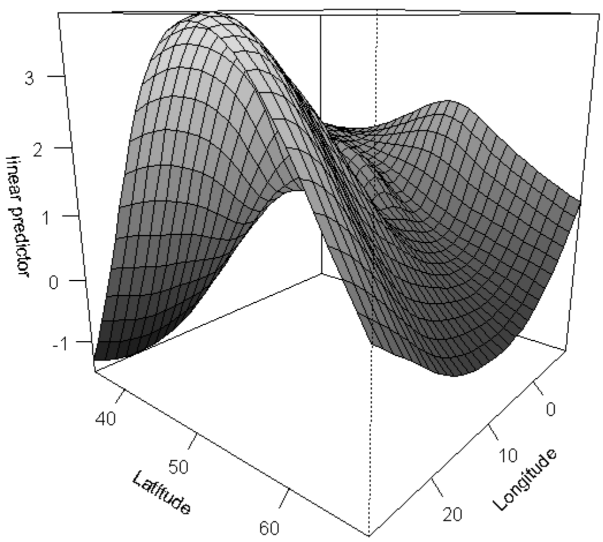
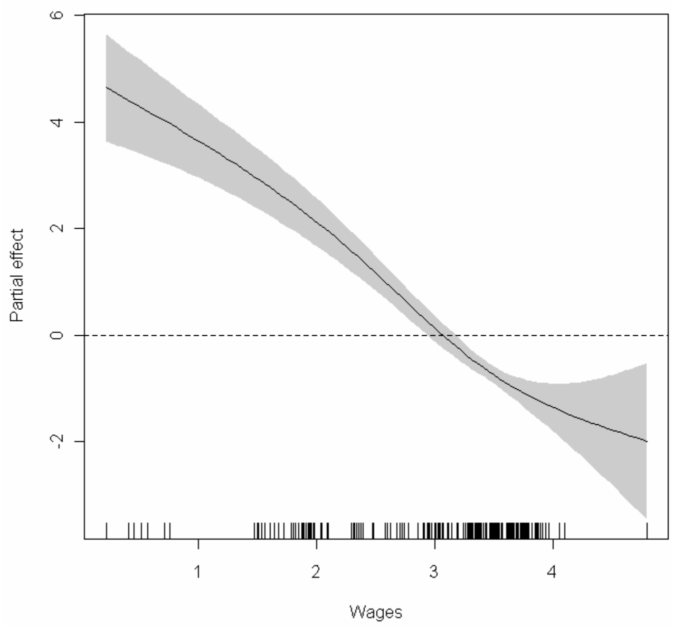


Figure 3 – Smooth effects in the semiparametric Geo-GAM

A – Wages



B – Employment density

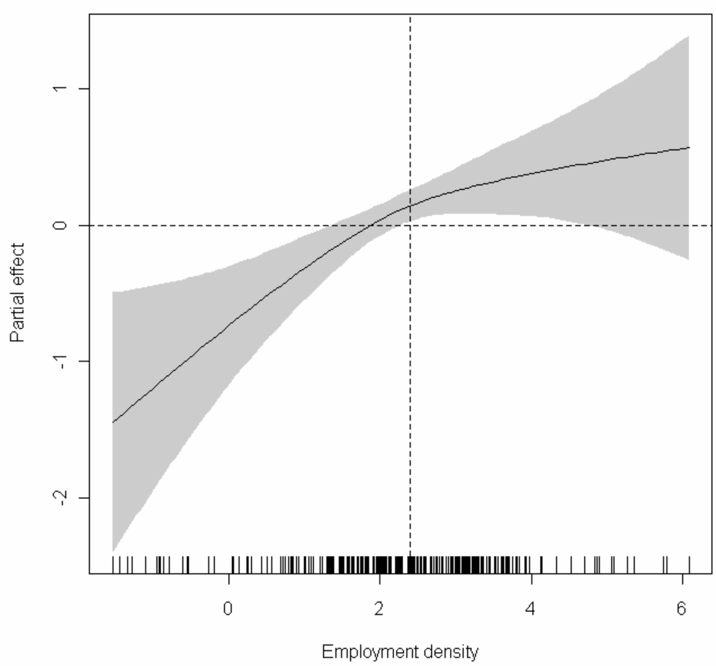
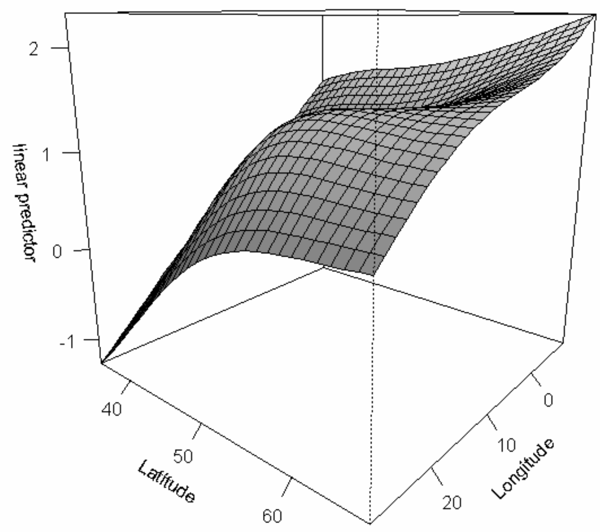


Figure 4 – Geographical components of geoadditive model

A – Perspective plot



B – Map plot

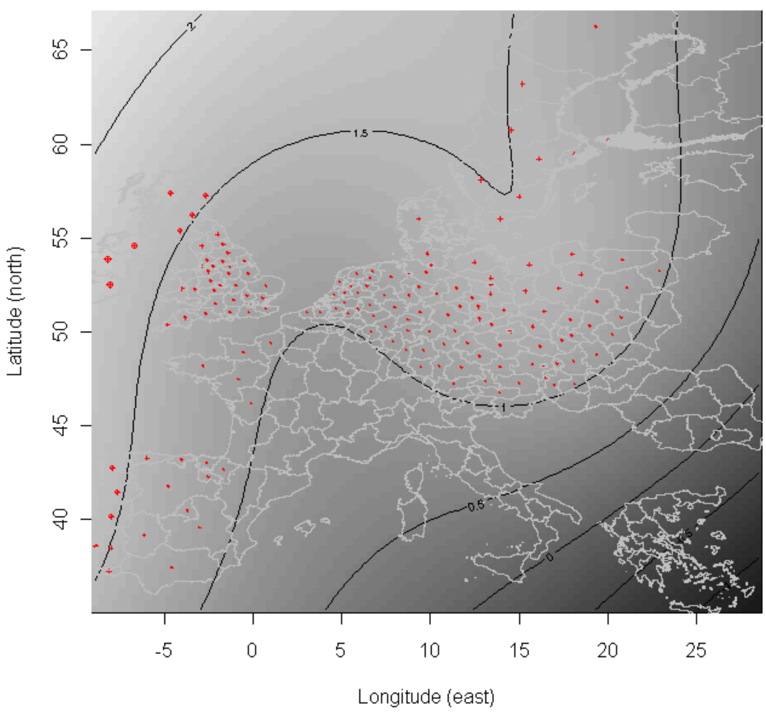


Table 1 – GLM results

	Poisson	**Negative Binomial**
mkt	0.349***	0.522***
	(0.037)	(0.098)
infra	0.329***	0.305***
	(0.033)	(0.087)
Jacobs	1.263***	1.295***
	(0.151)	(0.380)
empdens	0.422***	0.351***
	(0.030)	(0.075)
wage	-1.232***	-1.576***
	(0.046)	(0.146)
ur	0.520***	0.309**
	(0.042)	(0.130)
ter	0.452***	0.669***
	(0.067)	(0.203)
AIC	2,078	1,335
Over-dispersion	10.390	0.031
	[0.000]	[0.860]
Spatial dependence	2.449	2.215
	[0.000]	[0.002]
$\hat{\theta}$		1.183
Vuong		0.309
		[0.378]
Wald-Hurdle		0.281
		[0.998]

Notes: The over-dispersion test is based on the F test of the overall ("approximate") significance of smooth term $f(y)$ in the nonparametric estimation of absolute residuals, $|e|$, on fitted values, y. Spatial dependence is assessed by means of an F-test on the overall significance of a bivariate smooth term of longitude and latitude in explaining the residuals ($res = f(no, e) + \zeta$). $\hat{\theta}$ is the estimated NB shape parameter. Vuong is a non-nested hypothesis test statistic asymptotically distributed $N(0,1)$ under the null that the models ZINB and NB are indistinguishable. 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. Standard errors are in parentheses and p-values are in square brackets.

Table 2 - Nonparametric NB-GAM results

		Performance. iteration (GCV)	Outer iteration (AIC)	Outer iteration (REML)
Nonparametric terms		F *test and p-values*	χ^2 *test and p-values*	
$f(mkt)$		13.113***	47.432***	42.234***
	Edf	2.658	2.810	1.000
$f(infra)$		8.996***	10.852***	9.859***
	Edf	1.000	1.000	1.000
$f(Jacobs)$		7.164***	8.385***	7.375***
	Edf	1.000	1.000	1.000
$f(empdens)$		7.309***	25.344***	27.383***
	Edf	2.688	2.668	2.855
$f(wage)$		33.298***	132.983***	149.513***
	Edf	3.080	2.999	3.089
$f(ur)$		4.735**	8.087**	5.178**
	Edf	1.000	2.547	1.000
$f(ter)$		9.235***	17.350***	18.121***
	Edf	1.476	1.426	1.390
AIC		1,308	1,306	1,307
Over-dispersion		0.161	0.277	0.123
		[0.689]	[0.599]	[0.726]
Spatial dependence		2.370	2.264	2.289
		[0.000]	[0.001]	[0.000]
$\hat{\theta}$		1.672	2.070	1.867

Notes: F and χ^2 tests are used to investigate the overall ("approximate") significance of smooth terms. *Edf* (effective degrees of freedom) reflect the flexibility of the model. An *Edf* equals to 1 suggests that the smooth term can be approximated by a linear term. In such cases, parametric terms have been used. Standard errors are in round parentheses and p-values are in square brackets. The over-dispersion test is based on the F test of the overall ("approximate") significance of smooth term $f(y)$ in the nonparametric estimation of absolute residuals, $|e|$, on fitted values, y. Spatial dependence is assessed by means of an F-test on the overall significance of a bivariate smooth term of longitude and latitude in explaining the residuals $(res = f(no,e) + \zeta)$. $\hat{\theta}$ is the estimated NB shape parameter.

Table 3 - Semiparametric NB-GAM results

		Semiparametric GAM	***Semiparametric Geo-GAM***
Parametric Terms		*Coefficients and standard errors*	
mkt		0.586***	0.622***
		(0.090)	(0.092)
infra		0.256***	0.249***
		(0.080)	(0.086)
Jacobs		0.938***	1.056***
		(0.342)	(0.327)
ur		0.267**	0.260**
		(0.118)	(0.113)
ter		0.737***	0.672***
		(0.180)	(0.190)
Nonparametric terms		χ^2 *test and p-values*	
$f(empdens)$		27.290***	11.330**
	Edf	2.851	2.213
$f(wage)$		149.410***	114.690***
	Edf	3.085	2.719
$f(lat,long)$			38.560***
	Edf		4.592
REML		656.6	646.4
AIC		1,307	1,281
Over-dispersion		0.155	1.608
		[0.695]	[0.206]
Spatial dependence		2.294	0.931
		[0.000]	[0.550]
$\hat{\theta}$		1.862	2.113

Notes: F and χ^2 tests are used to investigate the overall ("approximate") significance of smooth terms. *Edf* (effective degrees of freedom) reflect the flexibility of the model. An *Edf* equals to 1 suggests that the smooth term can be approximated by a linear term. In such cases, parametric terms have been used. Standard errors are in round parentheses and p-values are in square brackets. The over-dispersion test is based on the F test of the overall ("approximate") significance of smooth term $f(y)$ in the nonparametric estimation of absolute residuals, $|e|$, on fitted values, y. Spatial dependence is assessed by means of an F-test on the overall significance of a bivariate smooth term of longitude and latitude in explaining the residuals $(res = f(no,e)+\zeta)$. $\hat{\theta}$ is the estimated NB shape parameter.

ISSN 1825-0211

QUADERNI DEL DIPARTIMENTO DI ECONOMIA, FINANZA E STATISTICA
Università degli Studi di Perugia

1	Gennaio 2005	Giuseppe CALZONI Valentina BACCHETTINI	Il concetto di competitività tra approccio classico e teorie evolutive. Caratteristiche e aspetti della sua determinazione
2	Marzo 2005	Fabrizio LUCIANI Marilena MIRONIUC	Ambiental policies in Romania. Tendencies and perspectives
3	Aprile 2005	Mirella DAMIANI	Costi di agenzia e diritti di proprietà: una premessa al problema del governo societario
4	Aprile 2005	Mirella DAMIANI	Proprietà, accesso e controllo: nuovi sviluppi nella teoria dell'impresa ed implicazioni di corporate governance
5	Aprile 2005	Marcello SIGNORELLI	Employment and policies in Europe: a regional perspective
6	Maggio 2005	Cristiano PERUGINI Paolo POLINORI Marcello SIGNORELLI	An empirical analysis of employment and growth dynamics in the italian and polish regions
7	Maggio 2005	Cristiano PERUGINI Marcello SIGNORELLI	Employment differences, convergences and similarities in italian provinces
8	Maggio 2005	Marcello SIGNORELLI	Growth and employment: comparative performance, convergences and co-movements
9	Maggio 2005	Flavio ANGELINI Stefano HERZEL	Implied volatilities of caps: a gaussian approach
10	Giugno 2005	Slawomir BUKOWSKI	EMU – Fiscal challenges: conclusions for the new EU members
11	Giugno 2005	Luca PIERONI Matteo RICCIARELLI	Modelling dynamic storage function in commodity markets: theory and evidence
12	Giugno 2005	Luca PIERONI Fabrizio POMPEI	Innovations and labour market institutions: an empirical analysis of the Italian case in the middle 90's
13	Giugno 2005	David ARISTEI Luca PIERONI	Estimating the role of government expenditure in long-run consumption
14	Giugno 2005	Luca PIERONI Fabrizio POMPEI	Investimenti diretti esteri e innovazione in Umbria
15	Giugno 2005	Carlo Andrea BOLLINO Paolo POLINORI	Il valore aggiunto su scala comunale: la Regione Umbria 2001-2003
16	Giugno 2005	Carlo Andrea BOLLINO Paolo POLINORI	Gli incentivi agli investimenti: un'analisi dell'efficienza industriale su scala geografica regionale e sub regionale

17	Giugno 2005	Antonella FINIZIA Riccardo MAGNANI Federico PERALI Paolo POLINORI Cristina SALVIONI	Construction and simulation of the general economic equilibrium model Meg-Ismea for the italian economy
18	Agosto 2005	Elżbieta KOMOSA	Problems of financing small and medium-sized enterprises. Selected methods of financing innovative ventures
19	Settembre 2005	Barbara MROCZKOWSKA	Regional policy of supporting small and medium-sized businesses
20	Ottobre 2005	Luca SCRUCCA	Clustering multivariate spatial data based on local measures of spatial autocorrelation
21	Febbraio 2006	Marco BOCCACCIO	Crisi del welfare e nuove proposte: il caso dell'unconditional basic income
22	Settembre 2006	Mirko ABBRITTI Andrea BOITANI Mirella DAMIANI	Unemployment, inflation and monetary policy in a dynamic New Keynesian model with hiring costs
23	Settembre 2006	Luca SCRUCCA	Subset selection in dimension reduction methods
24	Ottobre 2006	Sławomir I. BUKOWSKI	The Maastricht convergence criteria and economic growth in the EMU
25	Ottobre 2006	Jan L. BEDNARCZYK	The concept of neutral inflation and its application to the EU economic growth analyses
26	Dicembre 2006	Fabrizio LUCIANI	Sinossi dell'approccio teorico alle problematiche ambientali in campo agricolo e naturalistico; il progetto di ricerca nazionale F.I.S.R. – M.I.C.E.N.A.
27	Dicembre 2006	Elvira LUSSANA	Mediterraneo: una storia incompleta
28	Marzo 2007	Luca PIERONI Fabrizio POMPEI	Evaluating innovation and labour market relationships: the case of Italy
29	Marzo 2007	David ARISTEI Luca PIERONI	A double-hurdle approach to modelling tobacco consumption in Italy
30	Aprile 2007	David ARISTEI Federico PERALI Luca PIERONI	Cohort, age and time effects in alcohol consumption by Italian households: a double-hurdle approach
31	Luglio 2007	Roberto BASILE	Productivity polarization across regions in Europe
32	Luglio 2007	Roberto BASILE Davide CASTELLANI Antonello ZANFEI	Location choices of multinational firms in Europe: the role of EU cohesion policy
33	Agosto 2007	Flavio ANGELINI Stefano HERZEL	Measuring the error of dynamic hedging: a Laplace transform approach

34	Agosto 2007	Stefano HERZEL Cătălin STĂRICĂ Thomas NORD	The IGARCH effect: consequences on volatility forecasting and option trading
35	Agosto 2007	Flavio ANGELINI Stefano HERZEL	Explicit formulas for the minimal variance hedging strategy in a martingale case
36	Agosto 2007	Giovanni BIGAZZI	The role of agriculture in the development of the people's Republic of China
37	Settembre 2007	Enrico MARELLI Marcello SIGNORELLI	Institutional change, regional features and aggregate performance in eight EU's transition countries
38	Ottobre 2007	Paolo NATICCHIONI Andrea RICCI Emiliano RUSTICHELLI	Wage structure, inequality and skill-biased change: is Italy an outlier?
39	Novembre 2007	The International Study Group on Exports and Productivity	Exports and productivity. Comparable evidence for 14 countries
40	Dicembre 2007	Gaetano MARTINO Paolo POLINORI	Contracting food safety strategies in hybrid governance structures
41	Dicembre 2007	Floro Ernesto CAROLEO Francesco PASTORE	The youth experience gap: explaining differences across EU countries
42	Gennaio 2008	Melisso BOSCHI Luca PIERONI	Aluminium market and the macroeconomy
43	Febbraio 2008	Flavio ANGELINI Marco NICOLOSI	Hedging error in Lévy models with a fast Fourier Transform approach
44	Febbraio 2008	Luca PIERONI Giorgio d'AGOSTINO Marco LORUSSO	Can we declare military Keynesianism dead?
45	Febbraio 2008	Pierluigi GRASSELLI Cristina MONTESI Paola IANNONE	Mediterranean models of Welfare towards families and women
46	Marzo 2008	Mirella DAMIANI Fabrizio POMPEI	Mergers, acquisitions and technological regimes: the European experience over the period 2002-2005
47	Marzo 2008	Bruno BRACALENTE Cristiano PERUGINI	The Components of Regional Disparities in Europe
48	Marzo 2008	Cristiano PERUGINI Fabrizio POMPEI Marcello SIGNORELLI	FDI, R&D and Human Capital in Central and Eastern European Countries
49	Marzo 2008	Cristiano PERUGINI	Employment and Unemployment in the Italian Provinces
50	Marzo 2008	Sławomir I. BUKOWSKI	On the road to the euro zone. Currency rate stabilization: experiences of the selected EU countries
51	Aprile 2008	Bruno BRACALENTE Cristiano PERUGINI Fabrizio POMPEI	Homogeneous, Urban Heterogeneous, or both? External Economies and Regional Manufacturing Productivity in Europe

52	Aprile 2008	Gaetano MARTINO Cristiano PERUGINI	Income inequality within European regions: determinants and effects on growth
53	Aprile 2008	Jan L. BEDNARCZYK	Controversy over the interest rate theory and policy. Classical approach to interest rate and its continuations
54	Aprile 2008	Bruno BRACALENTE Cristiano PERUGINI	Factor decomposition of cross-country income inequality with interaction effects
55	Aprile 2008	Cristiano PERUGINI	Employment Intensity of Growth in Italy. A Note Using Regional Data
56	Aprile 2008	Cristiano PERUGINI Fabrizio POMPEI	Technological Change, Labour Demand and Income Distribution in European Union Countries
57	Aprile 2008	Simona BIGERNA Paolo POLINORI	L'analisi delle determinanti della domanda di trasporto pubblico nella città di Perugia
58	Maggio 2008	Simona BIGERNA Paolo POLINORI	The willingness to pay for Renewable Energy Sources (RES): the case of Italy with different survey approaches and under different EU "climate vision". First results
59	Giugno 2008	Simona BIGERNA Paolo POLINORI	Ambiente operativo ed efficienza nel settore del Trasporto Pubblico Locale in Italia
60	Ottobre 2008	Pierluigi GRASSELLI Cristina MONTESI Roberto VIRDI	L'interpretazione dello spirito del dono
61	Novembre 2008	Antonio BOGGIA Fabrizio LUCIANI Gianluca MASSEI Luisa PAOLOTTI	L'impatto ambientale ed economico del cambiamento climatico sull'agricoltura
62	Novembre 2008	Elena STANGHELLINI Francesco Claudio STINGO Rosa CAPOBIANCO	On the estimation of a binary response model in a selected population
63	Dicembre 2008	Gianna FIGÀ-TALAMANCA	Limit results for discretely observed stochastic volatility models with leverage effect
64	Maggio 2009	Mirella DAMIANI Andrea RICCI	Factors behind performance-related pay: evidence from Italy
65	Giugno 2009	Alessandra RIGHI Dario SCIULLI	The Timing of the School-to-Permanent Work Transition: a Comparison across Ten European Countries
66	Settembre 2009	Fabrizio LUCIANI	Economia agraria e pianificazione economica territoriale nel Parco nazionale del Sagarmatha (Everest, Nepal)
67	Settembre 2009	Valentina TIECCO	I regimi di protezione dell'impiego

68	Ottobre 2009	Gianna FIGÀ-TALAMANCA	Path properties of simulation schemes for the Heston stochastic volatility model
69	Ottobre 2009	Cristina MONTESI	A comparative analysis of different business ethics in the perspective of the Common Good
70	Ottobre 2009	Luisa FRANZINI Margherita GIANNONI	Determinants of Health Disparities in Italian Regions
71	Novembre 2009	Flavio ANGELINI Stefano HERZEL	Evaluating Discrete Dynamic Strategies in Affine Models
72	Novembre 2009	Giuseppe ARBIA Michele BATTISTI Gianfranco DI VAIO	Institutions and geography: Empirical test of spatial growth models for European regions
73	Gennaio 2010	Mirella DAMIANI Andrea RICCI	Performance-Related Pay, Unions and Productivity in Italy: evidence from quantile regressions
74	Febbraio 2010	Davide CASTELLANI Fabio PIERI	The Effect of Foireign Investments on European Regional Productivity
75	Luglio 2010	Guglielmo M. CAPORALE Davide CIFERRI Alessandro GIRARDI	Time-varying spot and futures oil price dynamics
76	Settembre 2010	Mirella DAMIANI	Labour regulation, corporate governance and varieties of capitalism
77	Settembre 2010	Dario SCIULLI Marcello SIGNORELLI	University-to-work transitions: the case of Perugia
78	Ottobre 2010	Olga DEMIDOVA Marcello SIGNORELLI	The Impact of Crises on Youth Unemployment of Russian Regions: An Empirical Analysis
79	Ottobre 2010	Misbah T. CHOUDHRY Enrico MARELLI Marcello SIGNORELLI	The Impact of Financial Crises on Youth Unemployment Rate
80	Novembre 2010	Marco BELLUCCI	Fusioni ed acquisizioni: determinanti ed effetti in un confronto europeo
81	Dicembre 2010	Silvia MICHELI	Learning Curve and Wind Power
82	Dicembre 2010	Leonardo BECCHETTI Stefano CASTRIOTA Elena GIACHIN RICCA	Beyond the Joneses: inter-country income comparisons and happiness
83	Gennaio 2011	Davide CASTELLANI Fabio PIERI	Foreign Investments and Productivity Evidence from European Regions
84	Febbraio 2011	Stefano HERZEL Marco NICOLOSI Cătălin STĂRICĂ	The cost of sustainability on optimal portfolio choices
85	Marzo 2011	Pierluigi GRASSELLI Cristina MONTESI	Politiche orientate al bene comune e politiche attive del lavoro
86	Marzo 2011	Massimo DE FELICE Franco MORICONI	Un'estensione stocastica del modello "Fisher-Lange"
87	Marzo 2011	Mirella DAMIANI Fabrizio POMPEI Andrea RICCI	Temporary job protection and productivity growth in EU economies

88	Aprile 2011	Gaetano MARTINO Paolo POLINORI	Productive process innovation as sequential adjustment of the hybrid governance structure: the case of the poultry sector
89	Aprile 2011	Stefano CASTRIOTA Marco DELMASTRO	Inside the black box of collective reputation
90	Maggio 2011	Roberto BASILE Luigi BENFRATELLO Davide CASTELLANI	Spatial clustering and nonlinearities in the location of multinational firms

ISSN 1722-618X

I QUADERNI DEL DIPARTIMENTO DI ECONOMIA
Università degli Studi di Perugia

1	Dicembre 2002	Luca PIERONI:	Further evidence of dynamic demand systems in three european countries
2	Dicembre 2002	Luca PIERONI Paolo POLINORI:	Il valore economico del paesaggio: un'indagine microeconomica
3	Dicembre 2002	Luca PIERONI Paolo POLINORI:	A note on internal rate of return
4	Marzo 2004	Sara BIAGINI:	A new class of strategies and application to utility maximization for unbounded processes
5	Aprile 2004	Cristiano PERUGINI:	La dipendenza dell'agricoltura italiana dal sostegno pubblico: un'analisi a livello regionale
6	Maggio 2004	Mirella DAMIANI:	Nuova macroeconomia keynesiana e quasi razionalità
7	Maggio 2004	Mauro VISAGGIO:	Dimensione e persistenza degli aggiustamenti fiscali in presenza di debito pubblico elevato
8	Maggio 2004	Mauro VISAGGIO:	Does the growth stability pact provide an adequate and consistent fiscal rule?
9	Giugno 2004	Elisabetta CROCI ANGELINI Francesco FARINA:	Redistribution and labour market institutions in OECD countries
10	Giugno 2004	Marco BOCCACCIO:	Tra regolamentazione settoriale e antitrust: il caso delle telecomunicazioni
11	Giugno 2004	Cristiano PERUGINI Marcello SIGNORELLI:	Labour market performance in central european countries
12	Luglio 2004	Cristiano PERUGINI Marcello SIGNORELLI:	Labour market structure in the italian provinces: a cluster analysis
13	Luglio 2004	Cristiano PERUGINI Marcello SIGNORELLI:	I flussi in entrata nei mercati del lavoro umbri: un'analisi di cluster
14	Ottobre 2004	Cristiano PERUGINI:	Una valutazione a livello microeconomico del sostegno pubblico di breve periodo all'agricoltura. Il caso dell'Umbria attraverso i dati RICA-INEA
15	Novembre 2004	Gaetano MARTINO Cristiano PERUGINI	Economic inequality and rural systems: empirical evidence and interpretative attempts
16	Dicembre 2004	Federico PERALI Paolo POLINORI Cristina SALVIONI Nicola TOMMASI Marcella VERONESI	Bilancio ambientale delle imprese agricole italiane: stima dell'inquinamento effettivo