UNIVERSITAT
ROVIRA I VIRGILI
DEPARTAMENT D'ECONOMIA

(creip)
CENTRE DE RECERCA EN ECONOMIA
INDUSTRIAL I ECONOMIA PÚBLICA

WORKING PAPERS

Col·lecció "DOCUMENTS DE TREBALL DEL DEPARTAMENT D'ECONOMIA - CREIP"

The Geographical Scope of Industrial Location Determinants: An Alternative Approach

Óscar Martínez
Miquel.A. Manjón
Josep M. Arauzo

Document de treball n.28 - 2012

DEPARTAMENT D'ECONOMIA – CREIP
Facultat d'Economia i Empresa

UNIVERSITAT ROVIRA I VIRGILI
DEPARTAMENT D'ECONOMIA

Edita:

Departament d'Economia
www.fcee.urv.es/departaments/economia/public_html/index.html
Universitat Rovira i Virgili
Facultat d'Economia i Empresa
Avgda. de la Universitat, 1
43204 Reus
Tel.: +34 977 759 811
Fax: +34 977 300 661
Email: sde@urv.cat

CREIP
www.urv.cat/creip
Universitat Rovira i Virgili
Departament d'Economia
Avgda. de la Universitat, 1
43204 Reus
Tel.: +34 977 558 936
Email: creip@urv.cat

Adreçar comentaris al Departament d'Economia / CREIP

Dipòsit Legal: T - 1192 - 2012

ISSN edició en paper: 1576 - 3382
ISSN edició electrònica: 1988 - 0820

DEPARTAMENT D'ECONOMIA – CREIP
Facultat d'Economia i Empresa

The Geographical Scope of Industrial Location Determinants: An Alternative Approach*

O. Martínez-Ibañez[a] M. Manjón-Antolín[a]
J.M. Arauzo-Carod[a,b]

[a]QURE-CREIP Department of Economics, Rovira i Virgili University.

[b]Institut d'Economia de Barcelona, University of Barcelona.

November 5, 2012

Abstract

This paper considers the estimation of the geographical scope of industrial location determinants. While previous studies impose strong assumptions on the weighting scheme of the spatial neighbour matrix, we propose a flexible parametrisation that allows for different (distance-based) definitions of neighbourhood and different weights to the neighbours. In particular, we estimate how far can reach indirect marginal effects and discuss how to report them. We also show that the use of smooth transition functions provides tools for policy analysis that are not available in the traditional threshold modelling.

Keywords: count data models, industrial location, smooth transition functions, threshold models

JEL-Codes: C25, C52, R11, R30

*This research was partially funded by SEJ2010-19088/ECON, SEJ2010-16934/ECON, the "Xarxa de Referència d'R+D+I en Economia i Polítiques Públiques", the SGR Program (2009-SGR-322) of the Catalan Government and the PGIR program N-2008PGIR/05 of the Universitat Rovira i Virgili. Usual caveats apply.

1 Introduction

New firms formation is an important policy concern (Hayter 1997, Lee 2008). However, industrial location policies may backfire if different local authorities implement them without an appropriate assessment of what is their geographical scope (Gramlich 1994, Kitson *et al.* 2004). In this context, having an estimate of how far such policies reach may be extremely helpful. For example, the estimated distance may provide the basis for a coordinated policy design that avoids that the efforts of different local authorities offset each other. It may also be useful to know whether the policy effect is constant or decreasing with the distance from the affected local area and/or what is the shape of this effect with respect to such distance. Similarly, how a change in the characteristics of a particular area affects location decisions in other close-by areas is an information that business managers may find of great value. In a metropolitan area, for example, how is the location of new concerns in the peripheral municipalities affected if the metropolis changes some of its infrastructures?

Unfortunately, policy makers and business managers looking for answers to these questions would find little guidance in the literature. Despite the large number of papers dedicated to empirically analyse the determinants of industrial location (see e.g. Arauzo-Carod *et al.* 2010 for a review), the estimation of their geographical scope is an issue that has received little attention.[1] In general, the effects these determinants may have on the location process are assumed to be restricted to the geographical limits of the considered sites, which means that most previous research in this area has been carried out under the assumption that there are no spatial effects.[2] This is not only at odds with the tenets of the New Economic Geography (Combes *et al.* 2008), but may often result in misleading conclusions (Amrhein 1995, Arbia 2001, Olsen 2002).

This paper makes two contributions to this literature: one is methodological and the other is empirical. First, we propose a new estimating approach based on the use of smooth transition functions (Granger and Teräsvirta 1993). The geographical scope of industrial location determinants has previously been estimated using threshold models in which neighbours are defined in terms of an indicator function. This means that two locations are considered neighbours if the distance that separates them is smaller than a certain threshold. Although simple to implement, this approach imposes strong assumptions on

[1] Rosenthal and Strange (2003) is an early attempt to address this issue.

[2] See, however, Autant-Bernard (2006), Lambert *et al.* (2006), Woodward *et al.* (2006), Blonigen *et al.* (2007), Alañón *et al.* (2007), Basile *et al.* (2010), Lambert *et al.* (2010), Melo *et al.* (2010), Alamá-Sabater *et al.* (2011), Arauzo-Carod and Manjón (2011) and Stewart and Lambert (2011).

the (distance-based) weighting scheme of the spatial neighbour matrix. In contrast, we define the weighting matrix using smooth transition functions, which allows us to explore different definitions of neighbourhood and to give different weights to different neighbours (Getis and Aldstat 2004, LeSage 2004, McMillen and McDonald 2004).

Our second contribution to this literature is to provide a number of empirical results with interesting policy implications. First, we replicate results from previous studies using the threshold approach and compare them with those obtained using continuous functions. In particular, we illustrate the validity of our approach in a sample of new manufacturing plants of Catalonia by comparing results from the Uniform and Normal smooth transition functions (as illustrative examples of threshold and continuous approaches, respectively).[3] Second, we report evidence on two important issues that largely motivated this work: how far can reach the "indirect marginal effects"of industrial locations determinants (see e.g. Arauzo-Carod and Manjón 2011) and how can we report these "indirect marginal effects"(see e.g. Woodward *et al.* 2006). While direct marginal effects measure the impact on the dependent variable of a unitary change in a covariate with respect to the same geographical unit, indirect marginal effects measure the impact on the dependent variable of a unitary change in a covariate with respect to a different geographical unit (LeSage and Pace 2009).

The main limitation of this paper is that, in line with previous literature, we do not address the question of what is the origin of the spatial dependence. Our model seems particularly well suited for addressing spatial dependence arising from a poor match between the spatial extent of the phenomenon of interest and the administrative units for which data are available. However, the presence of unobserved (spatially correlated) covariates may also produce spatial dependence. Our model cannot distinguish between these alternative sources of spatial dependence and, in fact, in the latter case would probably be outperformed by other approaches based on e.g. instrumental variables and panel data methods. We differ from previous studies in that we do not need to assume that such spatial dependence exists, for our model provides a direct assessment of the geographical scope of the determinants of industrial location (albeit not a formal test).[4]

The rest of the paper is organised as follows. In section 2 we briefly review the litera-

[3]Although the dataset allows to distinguish between locations and relocations (see e.g. Manjón and Arauzo-Carod 2011), it is important to stress that in this paper we only consider the information referred to new locations (Arauzo-Carod 2008, Arauzo-Carod and Manjón 2011). We leave the comparative analysis of locations and relocations within the framework developed in this paper for future research.

[4]In this respect it is interesting to note that other methodologies that account for spatial correlation, such as for example the Geographically Weighted Regression (see e.g. Fotheringham *et al.* 2002), do not address the estimation of indirect marginal effects (see e.g. Lambert *et al.* 2006).

ture. In section 3 we present the model. In section 4 we discuss the empirical results. In section 5 we summarise the main conclusions of this study.

2 The role of space in industrial location studies

The location decisions of new plants have been extensively analysed from both theoretical and empirical perspectives (Hayter 1997, Arauzo-Carod *et al.* 2010). This interest arises from the enormous implications of such decisions in terms of, for example, employment growth, demand for infrastructures and innovation activities (Gramlich 1994, Kitson *et al.* 2004, Lee 2008). However, a complete review of this literature is beyond the scope of this paper. Rather, we restrict attention to those investigations that account for the role of space when studying the determinants of industrial location.

In particular, recent research in this area has largely concentrated around two related questions (Olsen 2002, Rosenthal and Strange 2003). First, when looking for a potential site, which are the geographical areas considered by the firms? Second, which are the most suitable methods to analyse the spatial range of the variables influencing these decisions?

The first question has been addressed by comparing estimates obtained from different geographical aggregations. Arauzo-Carod and Manjón (2004) and Arauzo-Carod (2008), for example, compare estimates obtained from different administrative levels (municipalities, counties and provinces in the first case, municipalities and counties in the second) and functional areas (travel-to-work areas, in the second case). Both papers find that the differences across spatial aggregations, albeit minimal, do exist. Still, the size of the differences may be due to the similar size of the considered areas. As Pablo-Martí and Muñoz-Yebra (2009) show, when the comparison is made using areas of different size, results differ considerably.

In addition to these empirical contributions, there are some papers that investigate more methodological issues. Amrhein (1995) shows that omitting the geographical scope of the location determinants may entail a severe specification error, while Briant et al. (2010) show how the use of different spatial units in the analysis of agglomeration economies may result in substantially different coefficients. In general, however, these papers are more related to the so-called Modifiable Area Unit Problem (MAUP).[5]

The second question has been addressed by including in the specification of the model

[5]In essence, the main concern behind the MAUP is that, as Arbia (2001) shows, the distribution of a variable that is spatially georeferenced can lead to completely different conclusions depending on how the geographical area of interest is divided: in squares of equal size, in areas of different shape and size, etc.; in different administrative units, such as municipalities, counties, regions, etc..

some measure of spatial dependence, either in the exogenous explanatory variables (as in e.g. Autant-Bernard 2006, Woodward *et al.* 2006, Lambert *et al.* 2006, Alañón *et al.* 2007, Melo *et al.* 2010, Alamá-Sabater *et al.* 2011, Arauzo-Carod and Manjón 2011 and Stewart and Lambert 2011) and/or in the endogenous explanatory variable (as in e.g. Alañón *et al.* 2007, Blonigen *et al.* 2007, Basile *et al.* 2010 and Lambert *et al.* 2010).[6] Thus, this strand of the literature revolves around the use of different neighbourhood matrices, i.e., weighting matrices constructed under a certain definition of neighbourhood.[7]

One way to define a neighbour is on the grounds of contiguity. In this vein, all the areas that share edges or corners with other areas are considered neighbours. This is the first order contiguity employed by for example Autant-Bernard (2006) to analyse the determinants of the location of R&D labs in French regions and by Alañón *et al.* (2007) to analyse how improvements in the accessibility to road infrastructures affect the creation of new manufacturing establishments in Spain.

Another way to define a neighbour is on the grounds of distance. In this vein, neighbours are all the areas that fall within a given Euclidean distance (5 km, 10 km, etc.). This is the criterion used by for example Woodward *et al.* (2006) to estimate the effect that the universities' R&D expenditures in one US county have on the location of high-tech establishments in other counties; by Melo *et al.* (2010) to estimate the impact of railways and motorways of Portugal on new plant openings in manufacturing (see also Holl 2004 and Alañón *et al.* 2007); by Alamá-Sabater *et al.* (2011) to empirically discern whether the location of new manufacturing firms in the Spanish region of Murcia depends on the characteristics of the chosen municipality rather than those of the neighbourhood; and by Arauzo-Carod and Manjón (2011) to estimate how far can a change in the characteristics of a Catalan municipality affect the location of industrial establishments in the surrounding municipalities.

However, it is interesting to note that there is a certain arbitrariness involved in fixing the number of neighbours to be considered and/or the distance that defines a neighbour. Moreover, these contiguity- and distance-based definitions of neighbourhood give the same weight to all the neighbours. This paper proposes a flexible approach for the construction of the weighting matrix that addresses these shortcomings. However, there have been other attempts in the literature to avoid such arbitrariness and/or to account for differences in neighbouring importance.

[6] These can be seen as particular cases of the spatial Durbin model advocated by LeSage and Pace (2009).

[7] See Griffith (1996) and Getis and Aldstadt (2004) for a review of alternative neighbourhood criteria.

Getis and Aldstat (2004), for example, advocate for giving different weights to different types of neighbours, which amounts to give larger weights to closer areas if one uses Euclidean distances (e.g., an area at 5 km receives three times the weight given to an area at 10 km). This is the approach followed by Lambert *et al.* (2006) in their study of the manufacturing investment location in US counties. In particular, they report results from a geographically weighted regression model and a spatial generalized lineal model.

However, one may also use an inverse-distance based criterion that considers that the intensity of the relationship between two neighbours is inversely proportional to the distance that separates them (Fotheringham *et al.* 2002). Blonigen *et al.* (2007) follow this approach to analyse to what extent Foreign Direct Investment (FDI) in neighbour countries helps to explain FDI into the host country and Basile *et al.* (2010) to analyse the determinants of the number of inward Greenfield FDI in European regions.[8]

Lastly, recent work explores the use of hybrid spatial weighting matrices that combine contiguity- and distance-based definitions of neighbourhood. Lambert *et al.* (2010), for example, propose a spatial lag count data model that accommodates global spatial spillovers by using an inverse distance matrix constructed with the eight nearest neighbours. Also, Stewart and Lambert (2011) analyse ethanol production site location in the U.S. counties using a bivariate probit regression with an inverse distance matrix based on the county's nearest neighbours.

3 The Model

3.1 The basic setting

We seek to asses the geographical scope of the determinants of industrial location using regression analysis. In particular, we restrict attention to regression models with an exponential conditional mean function (Cameron and Trivedi 1998):

$$E\left[Y|X\right] = \mu = e^{W\_X\beta}$$

where the dependent variable Y is a column vector containing the number of new manufacturing establishments or firms created during a time period in one of the n locations considered.

[8] More sophisticated inverse-weighting schemes include the bandwidth distance decay (Fotheringham *et al.* 1996), the Gaussian distance decline (LeSage 2004) and the tri-cube distance decline (McMillen and McDonald 2004).

This specification includes, among others, Poisson and Negative Binomial models, which essentially differ in the form of the conditional variance function: while in the Poisson model is μ, in the Negative Binomial model is $\mu+\alpha\mu^2$ (the so-called "NB2 model", with α being a parameter to be estimated). In addition, the "inflated" versions of the Poisson and Negative Binomial models can easily be accommodated by multiplying the conditional mean above by one minus the parameter of the Bernoulli process that governs the two data generation processes involved (the binary and the count), so that $E\left[Y|X\right] = \mu(1-\delta)$.

We concentrate on these count data models for three reasons. First, in the empirical application we seek to replicate a study that uses these count data models (Arauzo-Carod and Manjón 2011). Second, these models are the most commonly used in industrial location studies (Arauzo-Carod *et al.* 2010). Third, although a discrete choice approach is also possible, there is a close relation between count data models and discrete choice models (Guimarães *et al.* 2003).[9]

As for the covariates, we use a set of spatially lagged variables calculated as

$$W\_X = WX$$

where X is a matrix of k explanatory variables (with at least one having spatial variation) and W is an appropriate row-standardised $n \times n$ spatial neighbour matrix. More specifically, W is a symmetric weighting matrix with elements taken 1/0 values depending on whether every two sites are considered as neighbours, where neighbourhood is defined in terms of a predefined distance and the standardisation is performed by dividing the elements by its row sum. Notice that if the diagonal of W is strictly positive, W_X is constructed from (weighted) sums of the explanatory variables and their spatial lags. On the other hand, a zero diagonal in W implies that only the spatially lagged variables are considered to construct W_X.

Some studies use as regressors not only the spatially lagged variables, but also the explanatory variables without applying any spatial lag (see e.g. Autant-Bernard 2006, Alamá-Sabater *et al.* 2011 and Arauzo-Carod and Manjón 2011). In maths, $W\_X = [WX \vdots X]$. However, it is important to notice that in these studies the diagonal of W is typically made of zeros, a constrain that it is not imposed in our model. This means that in our setting W_X contains essentially the same information as X and its spatial lags, so that we can compute both direct and indirect marginal effects without explicitly including X

[9]Having said that, it is important to stress that our approach can also be applied to discrete choice models (in general, to any model that can be represented in conditional expectation form), an extension of this paper that we leave for future research.

among the covariates, much in the same way you can do it in the Spatial Autorregressive Models (where the weighted spatial lag of the dependent variable is included) and the Spatial Durbin Models (where the weighted spatial lag of the dependent variable, the weighted covariates and the covariates are included).[10]

The implicit assumption behind this result is that marginal effects are continuous when the distance is zero. A discontinuity would arise in zero distances, however, had we included X as an additional regressor (to see this, notice that in the computation of the direct marginal effects one would need take into account the coefficients associated with X, whereas now we only take into account those associated with W_X). Leaving aside this detail, which clearly does not alter the essence of our approach, results presented below still hold.

Bearing in mind these issues, a simple specification of the weighting matrix is the following:

$$W\_X = W^U X$$

with

$$W^U[i,j] = \frac{\mathbf{1}\left(d_{ij} \leq h\right)}{\sum_{l=1}^{n} \mathbf{1}\left(d_{il} \leq h\right)},$$

where $W^U[i,.]$ and $W^U[.,j]$ are the $i-th$ row and $j-th$ column of W^U, respectively, d_{ij} is the distance between locations i and j, h is the distance that defines neighbourhood, and $\mathbf{1}(.)$ is an indicator function that takes value one if the condition in brackets (.) is true and zero otherwise. Thus, this weighting matrix amounts to assume that locations i and j are considered neighbours if d_{ij} is less or equal than h.

Woodward *et al.* (2006), Melo *et al.* (2010), Alamá-Sabater *et al.* (2011) and Arauzo-Carod and Manjón (2011), for example, use alternative versions of this weighting matrix while considering different values of h. In particular, Woodward *et al.* (2006) explore an increasing sequence of 5 miles, starting from 0 and up to 250, to define neighbourhood (i.e., $h = 0, 5, 10, \ldots, 250$) and

$$W\_X = \left[X[.,1] \vdots X[.,2] \vdots \ldots \vdots W^U X[.,s] \vdots \ldots \vdots X[.,k] \right],$$

where s is the covariate of interest (university R&D expenditures) and $X[.,j]$ is the $j-th$ column of X.[11] As for Arauzo-Carod and Manjón (2011), they explore an increasing

[10]See e.g. LeSage and Pace (2009) for details.

[11]Actually, the weighting matrix used by Woodward *et al.* (2006) is slightly more complex, for they allow W to vary across industries (i.e., the definition of neighbourhood differs across industries).

sequence of 10 km, starting from 10 km and up to 100 km, to define neighbourhood (i.e., $h = 10, 20, \ldots, 100$) and apply these criteria to all the spatially-varying explanatory variables they have. Thus, if we denote by S this sub-set of covariates, their weighting matrix is

$$W\_X = \left[X[.,1] \vdots X[.,2] \vdots \ldots \vdots W^U X[.,S] \vdots \ldots \vdots X[.,k] \right].$$

Similarly, Melo *et al.* (2010) use a 40 km threshold to weight the two transport variables they consider. Lastly, Alamá-Sabater *et al.* (2011) explore an increasing sequence of 0.1 km, starting from 25 km and up to 125 km, to define neighbourhood (i.e., $h = 25, 25.1, 25.2, \ldots, 125$) and apply these criteria to all the explanatory variables they have, so that

$$W\_X = \left[W^U X[.,1] \vdots W^U X[.,2] \vdots \ldots \vdots W^U X[.,k] \right].$$

3.2 Using smooth transitions functions to construct neighbourhood matrices

In this paper we follow an alternative approach based on smooth transitions functions (Granger and Teräsvirta 1993). This allows to extend this basic setting to the use of continuous rather than discrete (i.e., threshold) neighbourhood criteria. In maths,

$$W^f[i,j] = f(d_{ij}, h_f)$$

where $f()$ is a continuous function on h_f such that $0 \leq W^f[i,j] \leq 1$ and $\sum_{j=1}^n W^f[i,j] = 1$.

There are many functions that may satisfy these simple conditions. However, Kernel functions are natural candidates, since they are likely to satisfy the first condition (see e.g. Pagan and Ullah 1999) and can easily accommodate the second (a commonly used standardisation, as discussed in e.g. LeSage and Pace 2009). For illustrative purposes, here we consider two of the most widely used: the Uniform kernel density (a threshold function hereby denoted with the upper index U) and the Standard Normal density function (a continuous function hereby denoted with the upper index N). Namely,

$$W^U[i,j] = \frac{\mathbf{1}(d_{ij} \leq h_U)}{\sum_{l=1}^n \mathbf{1}(d_{il} \leq h_U)}$$

and

$$W^N[i,j] = \frac{e^{-h_N d_{ij}^2}}{\sum_{l=1}^n e^{-h_N d_{il}^2}},$$

respectively. Notice that the Uniform kernel corresponds to the weighting matrix used by

Woodward *et al.* (2006), Melo *et al.* (2010), Alamá-Sabater *et al.* (2011) and Arauzo-Carod and Manjón (2011). Also, the Normal kernel has been previously used by for example LeSage (2004) and McMillen and McDonald (2004), albeit in a rather different context, for their goal is to estimate a Locally Weighted regression with varying coefficients (Fotheringham *et al.* 2002).

Interestingly, these functions can be seen as two extreme cases of *what is considered a neighbour* and *what is the weight given to each neighbour*. On the one hand, the uniform case only considers as neighbours those locations in the sample that satisfy the "neighbourhood condition" $\mathbf{1}(d_{ij} \leq h_U)$. However, all the values of the explanatory variables that refer to a neighbour location have the same weight. This means that, given a location of interest i, the marginal effect of the variable m with respect to location j ($X_{m,j}$) is the same as the marginal effect of the variable m with respect to location j^* (X_{m,j^*}) as long as $\mathbf{1}(d_{ij} \leq h_U)$ and $\mathbf{1}(d_{ij^*} \leq h_U)$ both hold:

$$\frac{\partial E[Y_i/X]}{\partial X_{m,j}} = \frac{\mathbf{1}(d_{ij} \leq h_U)}{\sum_{l=1}^{n} \mathbf{1}(d_{il} \leq h_U)} \beta_m \mu = \frac{\partial E[Y_i/X]}{\partial X_{m,j^*}} = \frac{\mathbf{1}(d_{ij^*} \leq h_U)}{\sum_{l=1}^{n} \mathbf{1}(d_{il} \leq h_U)} \beta_m \mu$$

(Obviously, this result also holds for the inflated versions of the Poission and Negative Binomial models, except that in that case the corresponding marginal effects are multiplied by $1-\delta$).

On the other hand, the Standard Normal case considers that all the locations are neighbours. In other words, there is no "neighbourhood condition". However, the values of the explanatory variables that refer to a neighbour location in general do not have the same weight, for the weights in this case are decreasing with distance. That is, given a location of interest i, the marginal effect of the variable m with respect to location j is

$$\frac{\partial E[Y_i/X]}{\partial X_{m,j}} = \frac{e^{-h_N d_{ij}^2}}{\sum_{l=1}^{n} e^{-h_N d_{il}^2}} \beta_m \mu,$$

which in general is not the same as the marginal effect of the variable m with respect to location j^* (unless of course the neighbourhood locations are at the same distance from the location of interest):

$$\frac{\partial E[Y_i/X]}{\partial X_{m,j^*}} = \frac{e^{-h_N d_{ij^*}^2}}{\sum_{l=1}^{n} e^{-h_N d_{il}^2}} \beta_m \mu.$$

Rather, it can be seen that, if $h_N \in (0, \infty)$, the larger the distance between two locations the smaller is the associated marginal effect (and this also holds for the inflated versions of the Poisson and Negative Binomial models once the $1-\delta$ correction is applied).

Lastly, it is interesting to analyse the two limiting cases that arise in this setting with respect to what is considered a neighbour. The first limiting case arises in settings in which none of the locations in the sample are considered neighbours. This means that $W\_X = X$, so that the value of the covariates that refer to other locations are not affecting the conditional mean. This limiting case occurs in the Uniform case when $h_U = 0$, while in the Normal case occurs when $h_N \to \infty$. The second limiting case arises in settings in which all the locations in the sample are considered neighbours. This means that all the values of the covariates are affecting the conditional mean, a result that is achieved in the Uniform case when $h_U \geq \max(d_{ij})$ and in the Normal case when $h_N = 0$.[12]

In applications, however, the values of these parameters would typically lie somewhere in between these extremes. Thus, generalizing the previous results we may conclude that: *i)* low estimated values of h_U should be interpreted as evidence that only a few (close-by) locations in the sample affect the conditional mean, whereas high estimated values of h_U should be interpreted as evidence that many locations in the sample affect the conditional mean; *ii)* high estimated values of h_N should be interpreted as evidence that only a few locations in the sample significantly affect the conditional mean, whereas low estimated values of h_N should be interpreted as evidence that many locations in the sample significantly affect the conditional mean.

4 Empirical evidence

In this section we compare the performance of the two weighting functions considered: the Uniform (as an illustrate example of the threshold approach) and the Normal (as an illustrate example of smooth transition functions). We initially analyse direct marginal effects and present results for both the Poisson and the Negative Binomial model as well as their zero-inflated versions. Next we analyse indirect marginal effects. In particular, we address two questions that have previously been investigated using the threshold approach: what is the distance that may reach a change in the covariates (Arauzo-Carod and Manjón 2011) and how to report these indirect marginal effects (Woodward *et al.* 2006). We show the advantages of using smooth transition functions by presenting graphical results from

[12] These results make clear an important difference between the parameters of the Uniform and Normal functions: whereas h_U can be interpreted as a (threshold) distance, h_N is not at all a distance. Another difference between h_U and h_N is that although both parameters can be estimated by grid search (see Section 4.2 for details), h_N may alternatively be estimated as an additional parameter of the corresponding log-likelihood function because the associated weighting matrix makes the function continuous in h_N.

the Zero Inflated Negative Binomial model, which is the specification that fits the data better according to model selection criteria.

4.1 The Data

Results reported below were obtained from a dataset that has been recently used in several papers investigating the determinants of industrial location.[13] Table (1) reports details on the definition of variables and data sources, as well as some descriptive statistics. In essence, the dataset contains information on the location of new manufacturing establishments in the municipalities of Catalonia and on several characteristics of these municipalities. There are 946 municipalities in Catalonia, the autonomous region in the Northeast of Spain that has about 7 million inhabitants (15% of the Spanish population), covers an area of 31,895 km^2, and contributes approximately 19% of Spanish GDP. The city of Barcelona is the capital of Catalonia.

[Insert Table 1 around here]

In particular, the dependent variable is the number of new manufacturing establishments (codes 12–36 of NACE classification) created in each Catalan municipality in 2002. As for the explanatory variables, the dataset contains a number of proxies for agglomeration economies (residential population change between 1991 and 2001, urbanisation economies, dis-urbanisation economies, jobs and population density), industrial mix (a manufacturing concentration index, the percentage of manufacturing jobs and the percentage of jobs in services), education (percentage of population older than 10 years that completed technical secondary school, secondary school, a 3-year degree and a 4-year degree or a Ph.D.), transport infrastructures (a dummy for the existence of a rail station and travel times to the capital of the province, to the closest airport and to the closest port), the knowledge economy (jobs in high-tech industries and jobs in high-tech manufacturing industries), commuting (population working and living at municipality j over jobs at j and population working and living at j over population living at j and working at j or elsewhere), population (population aged between 20 and 44 years), location (dummies for the municipalities of each of the provinces of Catalonia other than Barcelona, a dummy for the capitals of *comarques* or counties, a dummy for shoreline areas, distance in km to the nearest city with at least 100,000 inhabitants and distance in km to the capital of

[13] See, among others, Arauzo-Carod (2008), Arauzo-Carod and Manjón (2011) and Manjón and Arauzo-Carod (2011).

Catalonia) and entrepreneurship (percentage of small firms). All these covariates refer to 2001, except of course the residential population change.

We determine which explanatory variables have spatial variation (and are thus affected by the corresponding weighting matrix) using the same methodology as Arauzo-Carod and Manjón (2011). This was found necessary to replicate their results and turn out to be useful because it reduces the size of the presented evidence (without compromising the main conclusions). In practice, an explanatory variable is considered to have spatial variation if it shows an statistically significant spatial autocorrelation according to Moran's I. Otherwise the variable remains unchanged, i.e., it is included in the model with its original values (and so does the set of dummy location variables). Similarly, we constructed the weighting matrices following their design. Namely, we used a distance based matrix at the municipality level constructed from data provided by the Catalan Cartographical Institute about the latitude and longitude of Catalan municipalities, measured both at the centroid of each municipality.

4.2 Results

Coefficient estimates

We use a grid search to estimate the parameters associated with the distance that indirect marginal effects can reach. In particular, the search is performed using two optimisation criteria: maximisation of the likelihood function (ML) and minimisation of the Chi-Square Goodness of Fit test (GoF).[14] In the Uniform case, we use an increasing sequence of 1 km to find the sample value of h_U, starting from 1 and up to 100. In the Normal case, we use an non-constant increasing sequence to find the sample value of h_N, starting from 7.85×10^{-5} and up to 0.785.[15]

Tables (2), (3), (4) and (5) report maximum likelihood estimates of the Poisson, Negative Binomial, Zero-Inflated Poisson and Zero-Inflated Negative Binomial models, respec-

[14]The GoF test is a m-type specification test; see Cameron and Trivedi (1998) for an excellent introduction to its use in count data models.

[15]The sequence chosen for h_U, $h_U(t) = t$ for $t = 1, 2, \ldots, 100$, is consistent with values found in related studies (Woodward *et al.* 2006, Alamá-Sabater *et al.* 2011, Arauzo-Carod and Manjón 2011). However, using an analogous sequence for the Normal case would make difficult to compare results between functions. This is because whereas using a constant variation in the sequence of the Uniform case results in constant variations of the weighting function, it does not in the Normal case. In fact, in the Normal case a constant variation sequence results in large/small variations of the weighting function when h_N is small/large. Thus, the sequence of values chosen for h_N, $h_N(t) = \frac{\pi}{4t^2}$ for $t = 1, 2, \ldots, 100$, satisfies $\int_0^\infty e^{h_N(t)d} \mathsf{d}(d) = \int_0^\infty \mathbf{1}\left(d < h_N(t)\right) \mathsf{d}(d)$, where $\mathsf{d}\left(\right)$ denotes the derivative. In this way we guarantee that the aggregated "relative effect" (see below) of all possible distances is the same in both the Uniform and Normal weighting functions.

tively, when the parameters h_U and h_N take the optimal value. That is, we report the estimated coefficients and standard errors obtained when the parameters h_U and h_N either maximise the likelihood function or minimise the Chi-Square Goodness of Fit test. For comparative purposes, in the first column of these tables we also report results under the limiting case that none of the locations in the sample are considered neighbours ($h_U = 0$, $h_N \to \infty$ and $W\_X = X$). Thus, these tables replicate some of the results reported by Arauzo-Carod and Manjón (2011).

[**Insert Table 2 around here**]

The first thing to notice is that results obtained for the limiting case given by $h_U = 0$ and $h_N \to \infty$ only allow to compute direct marginal effects. Put differently, these coefficient estimates are obtained imposing that indirect marginal effects are zero. Thus, they are largely comparable with those obtained in previous studies that do not account for spatial dependence. In addition, the sign and significance of these estimates are fairly similar across the four models considered. There are indeed some differences (e.g., the percentage of population older than 10 years that completed technical secondary school is only significant in the inflated versions of the models and the location dummies are less significant in the inflated versions of the models), but some common patterns also emerge.

[**Insert Table 3 around here**]

We find that agglomeration economies tend to have a positive effect on new business creation. However, the attractiveness of concentrated areas shows decreasing returns up to a point in which the new establishments prefer to locate in less densely populated areas. We also find that the characteristics of the industrial mix (concentration, jobs in manufacturing and jobs in services), population (aged between 20 and 44 years) and, to a certain extent, our proxies of the knowledge economy (jobs in high-tech industries) increase the number of new locations. These estimates are largely consistent with those reported in the literature (see Arauzo-Carod *et al.* 2010), both in terms of the sign and significance of the coefficients. In contrast, we find that education and transport infrastructures have no clear impact on the location of new establishments. Lastly, among our set of location dummies the significance of the capitals of *comarques* or counties suggest that the presence of public administrations and/or certain facilities in these municipalities make a difference for the new concerns.

Indirect marginal effects

Tables (2) to (5) also report results when at least one of the locations in the sample are considered neighbours (i.e., when $h_U > 0$ or $h_N < \infty$). These coefficients are thus employed to compute indirect marginal effects. In particular, given the exponential mean specification of the models considered, the change in the expected number of new manufacturing establishments in location i of a unitary variation of covariate m in location j is:

$$\frac{\partial E\left(y_i/\,X\right)}{\partial X_{m,j}} = \beta_m W[i,j] e^{W_i X[i,.]\beta} = \beta_m W_{ij} \mu_i$$

This shows that, in general, indirect marginal effects depend on the coefficients of the model (β and, in the inflated versions of the Poission and Negative Binomial models, δ) and the weighting function ($W[i,.]$). However, if we replace $W[i,.]$ by the two weighting functions considered we have (again omitting δ for the sake of simplicity):

$$\frac{\partial E\left(y_i/\,X\right)}{\partial X_{m,j}} = \begin{cases} \beta_m \frac{\mathbf{1}(d_{ij} \leq h_U)}{\sum_{l=1}^{n} \mathbf{1}(d_{il} \leq h_U)} \mu_i & \text{for the Uniform case.} \\ \\ \beta_m \frac{e^{-h_N d_{ij}^2}}{\sum_{l=1}^{n} e^{-h_N d_{il}^2}} \mu_i & \text{for the Normal case.} \end{cases}$$

Notice that, given the parameters of the model (β, δ if it is the case, and h_U or h_N) and the distance between locations i and j (d_{ij}), indirect marginal effects depend on the sum $\sum_{l=1}^{n} \mathbf{1}\left(d_{il} \leq h_U\right)$ in the Uniform case and on the sum $\sum_{l=1}^{n} e^{-h_N d_{il}^2}$ in the Normal case. Interestingly, these sums that can be interpreted as measures of concentration around location i: the former corresponds to the number of locations that are considered neighbours and the latter to the sum of the (exponentially inverted) distances to all the locations in the sample. This means that indirect marginal effects are larger/smaller the lower/higher is the concentration of locations around location i. Since rural areas tend to have locations more dispersed than metropolitan areas, for example, the change in the expected number of new manufacturing establishments in a metropolitan location due to a unitary variation of a covariate in another metropolitan location will in general be lower than that in a rural location due to a variation in a close-by location (ceteris paribus).[16]

[Insert Table 4 around here]

Notice also that, in applications, it is the estimated parameters (β, δ if it is the case, and h_U or h_N) what drives indirect marginal effects, for the distances between locations do

[16]This result arises directly from the row-standardisation of the weighting matrices.

not vary across the econometric models, optimisation criteria and weighting functions one may consider. However, estimations of the parameters of the model typically vary across the econometric models, optimisation criteria and weighting functions. This is illustrated in Tables (2) to (5); see also Figure (1) below. It is therefore important to determine what should be the preferred specification to make these inferences.

[Insert Table 5 around here]

We find evidence of over-dispersion in the data (a standard result in the literature). In other words, the statistical significance of the parameter α in all the reported specifications indicate that the (Inflated) Negative Binomial model fits the data better than the (Inflated) Poisson model. However, we also find that the Vuong test shows a strong preference for the inflated versions of these specifications (also a commonly found result). In fact, the Akaike criterion indicates that, irrespective of the optimising criteria and weighting function we use, the Inflated Poisson and Inflated Negative Binomial models provide a better fit than the Poisson and Negative Binomial models, respectively. However, only the Zero Inflated Negative Binomial model shows no signs of misspecification according to the GoF test. Moreover, the AIC indicates that this is the specification that fits the data better.

We consequently conclude that the Zero Inflated Negative Binomial model is the best specification for our industrial location data. Thus, we only report results based on this model in our analysis of how far can reach the indirect marginal effects and how can we report them. More specifically, we report results for the two optimising criteria (ML and GoF) and weighting functions (Uniform and Normal) considered because the values of the AIC and the GoF we obtained are not conclusive about what is the optimising criteria and weighting function that perform better for this model.

Compared to the other models, we find that the Zero Inflated Negative Binomial model provides roughly the same results in terms of the direction of the effects of covariates. However, the significance of the coefficients is often lower, especially with regard to the variables measuring transport time to major infrastructures and principal urban areas. It is also worth noting that the sign, significance and, to a large extent, the value of the Zero Inflated Negative Binomial estimates are rather robust across the optimisation criteria and weighting functions considered. Agglomeration economies and the industrial mix thus arise as the main determinants of the entry on new establishments in a municipality, being also relevant the percentage of population that completed secondary school, the number of jobs in high-tech industries and the fact that the municipality is the capital of the *comarca*.

Distance estimates on how far can reach indirect marginal effects

Columns two to four of Table (5) provide estimates of the Zero Inflated Negative Binomial model coefficients when the parameters h_U and h_N take the optimal value. However, what is that optimal value? How far can reach indirect marginal effects? In order to answer these questions Figure (1) graphically shows the relation between the parameters of interest (h_U and h_N) and the optimising criteria (ML and GoF).

[Insert Figure 1 around here]

In the Uniform case, results indicate that both selection criteria (ML and GoF) reach the optimum at small distances. Namely, $h_U = 4$ km for GoF and $h_U = 1$ km for ML. In the Normal case, however, the parameter h_N cannot be interpreted as a distance. In fact, as previously pointed out, the correct interpretation of this parameter in the Normal case is in terms of the number of locations in the sample that significantly affect the conditional mean. This is because h_N reflects the rate at which indirect effects tend to zero: the higher/smaller is h_N, the higher/lower is the rate and the shorter/further is the distance indirect effects reach before becoming insignificant. In particular, as shown in Figure (2) below, a value of h_N around 0.08 (GoF criterion) indicates that those locations within a distance of approximately 8 km are significantly affecting the conditional mean, whereas a value of h_N around 0.78 (ML criterion) indicates that only locations within a distance of approximately 3 km are significantly affecting the conditional mean. These results are slightly bigger but still in line with the optimal distances found in the Uniform case.

These figures are consistent with the institutional setting we study (see also Arauzo-Carod and Manjón 2011). In Catalonia, practically all the municipalities have another municipality within a distance of less than 8 km and around one out of five have at least another municipality within a distance of less than 2 km. Also, indirect marginal effects affect on average between one and eight surrounding municipalities. Therefore, our results indicate that indirect marginal effects typically reach contiguous municipalities (and possibly second-order contiguous municipalities).

In contrast, Woodward *et al.* (2006) find an optimal value of 60 miles for the US counties and Alamá-Sabater *et al.* (2011) find an optimal value of 43.6 km for the Spanish region of Murcia. Notice, however, that these are institutional settings in which the number of locations within the obtained distances is considerably lower. This is apparent in the US counties (which are in general much larger than Catalan municipalities), but is also the case in the municipalities of the region of Murcia, which, according to Alamá-Sabater *et al.* (2011), "are slightly larger on average in the national space". Notice also that Alamá-Sabater *et al.* (2011) use a grid search that starts at 25 km. As Figure (1) shows, had we used the same starting point we would have obtained distances of around 30 (GoF test criterion) and 40 km (maximum likelihood criterion).

Reporting indirect marginal effects

As it is common in non-linear models, indirect marginal effects are not constant but vary in general across locations (i.e., one may compute one effect for each location in the sample). One way to summarise this information is to report descriptive statistics (the mean and the median being the most popular) and/or graphical methods (e.g., histograms). However, these measures are not independent of the distance in models using neighbourhood matrices, which means that they may provide misleading results. We therefore need to compute these statistics conditional on distance. The downside is that, since they then become covariate-specific, the resulting output can be difficult to handle.

We accordingly propose using the ratio between the indirect marginal effect and the direct marginal effect (that is, the "relative" effect) as an alternative measure that combines accuracy and parsimony:[17]

$$\text{RE}_{ij} = 100 \frac{\frac{\partial E(y_i/X)}{\partial X_{m,j}}}{\frac{\partial E(y_i/X)}{\partial X_{m,i}}} = \begin{cases} 100 \times \mathbf{1}\left(d_{ij} \leq h_U\right) & \text{for the Uniform case.} \\ 100 \times e^{-h_N d_{ij}^2} & \text{for the Normal case.} \end{cases}$$

The first thing to notice about the relative effect of municipality j on municipality i is that only depends on the distance between i and j. Thus, the relative effect varies between 100% (zero distance) and 0% (at the distance in which the indirect marginal effect is zero). How the two weighting functions considered incorporate this feature, however, differs. While in the Uniform case the marginal effect produced in neighbouring municipalities (those who satisfy $d_{ij} < h_U$) is the same as that produced in the municipality of interest (i.e., direct and indirect marginal effects are the same, a result that is only plausible for small values of h_U), in the Normal case this relative effect is a continuous decreasing function on d_{ij} (smoother the smaller the parameter h_N). This is apparent in Figure (2), where we report the relative effects computed for the two weighting functions (Uniform and Normal) and optimising criteria (ML and GoF) we have considered.

[Insert Figure 2 around here]

[17] We have also experienced with the quantiles of the indirect marginal effects conditional on the distance and with the indirect marginal effects evaluated at different quantiles. However, these measures are more difficult to compute and less easy to interpret.

Notice also that policies based on the results obtained from the Uniform case would typically leave a number of locations unattended. This is because the threshold structure would lead us to conclude that those municipalities that are not considered neighbours are unaffected (i.e., the relative effect is zero). However, as results from the Normal case show, these locations may simply be less affected than the others. In particular, the difference between the locations considered unaffected by the Uniform estimates and those considered at least affected by the Normal case will (ceteris paribus) be larger the smaller are the parameters h_U and h_N. Obviously, at the end of the day it is up to the governments to decide whether these locations should be take into account. What is interesting to note here is that in some of these potentially unattended locations the relative effects may vary between 20% to 40% according to our estimates, which seem rather high figures to be ignored. In any case, this just shows that indirect marginal effects are an excellent tool to analyse the geographical scope of the covariates.[18]

5 Conclusions

This paper empirically analyses the geographical scope of the determinants of industrial location. We propose a new estimating approach based on the use of smooth transition functions and illustrate the validity of our approach using a sample of manufacturing plants created in the municipalities of Catalonia. In particular, we address two questions that have previously been investigated using the threshold approach and count data models: what is the distance that may reach a change in the covariates and how to report marginal effects with respect to the other locations in the sample. We show that the use of a continuous parametrisation of the neighbourhood function has a number of advantages over the traditional threshold approach. Among others, it can accommodate alternative views on what is considered a neighbour and what is the weight that should be given to each neighbour. These features turn out to provide a more powerful framework for policy analysis than that based on the traditional threshold modelling.

We provide illustrative results using an example of threshold (the Uniform) and smooth transition functions (the Normal). We find that in the Uniform case all locations within a distance of 1 to 4 km (depending on whether we use the maximisation of the likelihood function or the minimisation of the Goodness-of-Fit test as optimising criterion) are signif-

[18]One may easily construct an analogous argument for those locations that are "over-attended", in the sense that the Uniform function gives them a 100% relative effect whereas the Normal function shows that their relative effect varies between 20% and 100%.

icantly affecting the conditional mean, whereas in the Normal case the range of distances extends to 4 to 8 km (for the maximum likelihood and GoF criterion, respectively). Thus, these are the distances that indirect marginal effects can reach in the institutional setting we study. We also find that these effects differ depending on the degree of isolation of the municipality and the level of manufacturing activities, being larger for those municipalities that are isolated and have higher levels of manufacturing activities. Lastly, we show that the use of threshold functions as the basis for policy making may result in a number of locations being unfairly disregarded.

Among the possible extensions of this work, it is interesting to note that we follow previous literature in assuming that the different explanatory variables either have no geographical scope or have the same geographical scope. In other words, we are using a single spatial neighbour matrix. However, the framework presented in this paper allows to use different matrices for different covariates, so that, for a particular covariate m, locations i and j are considered neighbours if $d_{ij} < h_m$. This extension may not only provide a more realistic setting, but additional economic policy tools.

References

Alamá-Sabater, Luisa, Andrés Artal-Tur and José-Miguel Navarro-Azorín (2011), 'Industrial location, spatial discrete choice models and the need to account for neighbourhood effects', *Annals of Regional Science* **forthcoming**.

Alañón, Angel, Josep-Maria Arauzo-Carod and Rafael Myro (2007), Accessibility, agglomeration and location, *in* J.-M.Arauzo-Carod and M.Manjón-Antolín, eds, 'Entrepreneurship, industrial location and economic growth', Edward Elgar, pp. 247–267.

Amrheim, Carl G. (1995), 'Searching for the elusive aggregation effect: Evidence from statistical simulations', *Environment and Planning A* **27**(1), 105–119.

Arauzo-Carod, Josep Maria (2008), 'Industrial location at a local level: Comments on the territorial level of the analysis', *Tijdschrift voor economische en sociale geografie (Journal of Economic and Social Geography)* **99**(2), 193–208.

Arauzo-Carod, Josep Maria, Daniel Liviano-Solís and Miguel Manjón-Antolín (2010), 'Empirical studies in industrial location: An assessment of their methods and results', *Journal of Regional Science* **50**(3), 685–711.

Arauzo-Carod, Josep Maria and Miguel Manjón-Antolín (2004), 'Firm size and geographical aggregation: An empirical appraisal in industrial location', *Small Business Economics* **22**(3/4), 299–312.

Arauzo-Carod, Josep Maria and Miguel Manjón-Antolín (2011), '(Optimal) spatial aggregation in the determinants of industrial location', *Small Business Economics* **forthcoming**.

Arbia, Giuseppe (2001), 'Modeling the geography of economic activities on a continuous space', *Papers in Regional Science* **80**(4), 411–424.

Autant-Bernard, Corinne (2006), 'Where do firms choose to locate their R&D? A spatial conditional logit analysis on French data', *European Planning Studies* **14**(9), 1187–1208.

Basile, Roberto, Luigi Benfratello and Davide Castellani (2010), Location determinants of greenfield foreign investments in the enlarged Europe: Evidence from a spatial autoregressive negative binomial additive model. University of Torino, Working Papers of the Department of Economics and Public Finance.

Blonigen, Bruce A., Ronald B. Davies, Glen R. Waddell and Helen T. Naughton (2007), 'FDI in space: Spatial autoregressive relationships in foreign direct investment', *European Economic Review* **51**(5), 1303–1325.

Briant, Anthony, Pierre-Philippe Combes and Miren Lafourcade (2010), 'Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations?', *Journal of Urban Economics* **67**(3), 287–302.

Cameron, A. Colin and Pravin K. Trivedi (1998), *Regression Analysis for Count Data*, Cambridge University Press.

Combes, Pierre-Philippe, Thierry Mayer and Jacques-François Thisse (2008), *Economic Geography*, Princeton University Press.

Fotheringham, Stewart A., Chris Brunsdon and Martin Charlton (2002), *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, John Wiley and Sons.

Fotheringham, Stewart A., Martin Charlton and Chris Brunsdon (1996), 'The geography of parameter space: An investigation into spatial nonstationarity', *International Journal of Geographical Information Science* **10**(5), 605–627.

Getis, Arthur and Jared Aldstadt (2004), 'Constructing the spatial weights matrix using a local statistic', *Geographical Analysis* **36**(2), 90–104.

Gramlich, Edward M. (1994), 'Infrastructure investment: A review essay', *Journal of Economic Literature* **32**(3), 1176–1196.

Granger, Clive W. and Timo Teräsvirta (1993), *Modelling nonlinear economic relationships*, Oxford University Press.

Griffith, Daniel A. (1996), Some guidelines for specifying the geographic weights matrix contained in spatial statistical models, *in* S.Arlinghaus, ed., 'Practical Handbook of Spatial Statistics', CRC Press, pp. 65–82.

Guimarães, Paulo, Octávio Figueiredo and Douglas Woodward (2003), 'A tractable approach to the firm location decision problem', *Review of Economic and Statistics* **85**(1), 201–204.

Hayter, Roger (1997), *The dynamics of industrial location : The factory, the firm, and the production system*, John Wiley & Sons.

Holl, Adelheid (2004), 'Transport infrastructure, agglomeration economies, and firm birth. Empirical evidence from Portugal', *Journal of Regional Science* **44**(4), 693–712.

Kitson, Michael, Ron Martin and Peter Tyler (2004), 'Regional competitiveness: An elusive yet key concept?', *Regional Studies* **36**, 113–124.

Lambert, Dayton M., James P. Brown and R.J.G.M. Florax (2010), 'A two-step estimator for a spatial lag model of counts: Theory, small sample performance and an application', *Regional Science and Urban Economics* **40**(10), 241–252.

Lambert, Dayton M., Kevin T. Mcnamara and Megan I. Garrett (2006), 'An application of spatial Poisson models to manufacturing investment location analysis', *Journal of Agricultural and Applied Economics* **38**(1), 105–121.

Lee, Yoonsoo (2008), 'Geographic redistribution of US manufacturing and the role of state development policy', *Journal of Urban Economics* **64**, 436–450.

LeSage, James P. (2004), A family of geographically weighted regression models, *in* L.Anselin, R.Florax and S.Rey, eds, 'Advances in Spatial Econometrics: Methodology, Tools and Applications', Springer, pp. 241–262.

LeSage, James P. and Robert Kelley Pace (2009), *Introduction to spatial econometrics*, CRC Press.

Manjón-Antolín, Miguel and Josep Maria Arauzo-Carod (2011), 'Locations and relocations: determinants, modelling, and interrelations', *Annals of Regional Science* **47**(1), 131–146.

McMillen, Daniel P. and John F. McDonald (2004), Locally weighted maximum-likelihood estimation: Monte carlo evidence and an application, *in* L.Anselin, R.Florax and S.Rey, eds, 'Advances in Spatial Econometrics: Methodology, Tools and Applications', Springer, pp. 225–239.

Melo, P.C., D.J. Graham and R.B. Noland (2010), 'Impact of transport infrastructure on firm formation. Evidence from Portuguese municipalities', *Transportation Research Record* **2163**, 133–143.

Olsen, Joshua (2002), 'On the units of geographical economics', *Geoforum* **33**(2), 153–164.

Pablo-Martí, Federico and Carlos Muñoz-Yebra (2009), 'Localización empresarial y economías de aglomeración: el debate en torno a la agregación espacial', *Investigaciones Regionales* **15**, 139–166.

Pagan, Adrian and Aman Ullah (1999), *Nonparametric Econometrics*, Cambridge University Press.

Rosenthal, Stuart S. and William C. Strange (2003), 'Geography, industrial organization and agglomeration', *The Review of Economics and Statistics* **85**(2), 377–393.

Stewart, Lance and Dayton M. Lambert (2011), 'Spatial heterogeneity of factors determining ethanol production site selection in the U.S., 2000-2007', *Biomass and Bioenergy* **35**(3), 1273–1285.

Woodward, Douglas, Octávio Figueiredo and Paulo Guimarães (2006), 'Beyond the Silicon Valley: University R&D and high technology location', *Journal of Urban Economics* **60**(1), 15–32.

Table 1: Variables: definition, sources and descriptive statistics.

Variable	Definition	Source	Mean	Std dev	Min	Max
ENTRY	New manufacturing establishments (2001-2003)	REI, OC	4.093	14.338	0	258
RES VAR	Residential population change between 1991 and 2001	TB2005	0.154	0.344	-0.863	3.043
URB	Jobs per km2	TB2005, IDESCAT, OC	263.737	4.05×10^{6}	0.137	124000
DISURB	URB2	OC	1.65×10^{7}	5.00×10^{8}	0.019	1.54×10^{10}
JOB	Jobs	IDESCAT	2.97×10^{6}	2.22×10^{7}	12	645682
DENS	Residential population per km2	TB2005, OC	380.107	1.52×10^{6}	0.765	21020
CONC	Manufacturing concentration index	TB2005	1.196	1.002	0	3.896
JOB IND	Percentage of manufacturing jobs	IDESCAT	0.222	0.116	0	0.609
JOB SER	Percentage of jobs in services	IDESCAT	0.473	0.259	0	1
TEC SEC	% of population older than 10 with technical secondary school	TB2005	10.043	3.225	0.585	23.585
SEC	% of population older than 10 with secondary school	TB2005	9.593	3.482	1.695	28.226
DEG	% of population older than 10 with 3 years degree	TB2005	5.397	2.196	0	24
DEG PHD	% of population older than 10 with 4 years degree and PhD	TB2005	4.657	2.534	0	21.062
TT CP	Travel time to capital of the province	TB2005	87.010	23.943	0	190
RAIL	Dummy for rail station	TB2005	0.107	0.309	0	1
TT AIR	Travel time to the closest airport	TB2005	48.872	33.086	0	190
TT PORT	Travel time to the closest port	TB2005	62.182	33.187	0	197
JOB HT	Jobs in high-tech industries	TB2005	824.479	1.22×10^{7}	0	371269
JOB HT MA	Manufacturing jobs in high-tech industries	TB2005	16.652	159.379	0	4303
POP JOB	Population working and living at j / Jobs at j	TB2005	43.444	14.681	0	89.401
POP JOB E	Pop. working and living at j / Pop. living at j and working elsewhere	TB2005	178.869	1.73×10^{6}	0	5.21×10^{7}
POP 20 44	Population aged 20-44	OC	29.983	4.596	0	43.050
RES	Residential population (only used in inflated models)	TB2005	6.70×10^{6}	5.17×10^{7}	26	1.50×10^{6}
GIRONA	Province of Girona	IDESCAT	0.234	0.424	0	1
LLEIDA	Province of Lleida	IDESCAT	0.243	0.429	0	1
TARRA	Province of Tarragona	IDESCAT	0.194	0.396	0	1
CAP CO	Dummy for the capitals of the comarques	IDESCAT	0.043	0.204	0	1
COAST	Dummy for shore-line areas	OC	0.074	0.262	0	1
DIST 100	Distance (km) to the nearest city with at least 100,000 inhabitants	CCI	47.073	29.829	0	13.588
DIST CAT	Distance (km) to the capital of Catalonia (Barcelona)	CCI	86.965	39.671	0	199.590
FIRM SMALL	Percentage of small firms (less than 50 workers)	TB2005	83.701	23.671	0	100

Note: In column labelled Sources REI stands for " Register of Industrial Establishments", OC for "Own Calculations", TB2005 for the 2005 report of J. Trullén and R. Boix (*Indicadors 2005*, Diputació de Barcelona and Universitat Autònoma de Barcelona), IDESCAT for "Catalan Statistical Institute" and CCI for "Catalan Cartographical Institute".

Table 2: Estimation Results for the Poisson Model

Variables	$h_U = 0$, $h_N = \infty$		Uniform (GoF Criteria)		Uniform (ML Criteria)		Normal (GoF Criteria)		Normal (ML Criteria)	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
CONST^{abcde}	-3.5256	0.6121	-3.9731	1.2245	-4.7693	1.0101	-4.9179	1.6202	-4.3607	1.0198
(W) RES VAR	-0.0545	0.1073	0.3955	0.3230	-0.0375	0.2115	0.0491	0.5442	-0.1051	0.1843
(W) URB^{abd}	0.0006	0.0001	0.0015	0.0004	0.0007	0.0004	0.0038	0.0011	0.0002	0.0003
(W) DISURB^{abd}	-5.1e-09	9.8e-10	-1.3e-08	3.3e-09	-5.8e-09	3.0e-09	-3.1e-08	8.5e-09	-1.9e-09	2.6e-09
(W) JOB^{abcde}	3.2e-05	3.8e-06	8.5e-05	2.8e-05	5.6e-05	8.3e-06	0.0002	4.6e-05	3.5e-05	1.4e-05
(W) DENS^{abd}	-0.0002	4.3e-05	-0.0003	0.0001	-0.0001	0.0001	-0.0009	0.0003	-8.2e-05	9.9e-05
(W) CONC^{abcde}	0.3705	0.0880	0.4789	0.2059	0.4459	0.1253	1.0567	0.3254	0.4248	0.1067
(W) JOB IND^{abce}	2.5095	0.5049	3.6649	1.1062	2.6677	0.9300	2.3482	1.4862	3.1321	0.8159
(W) JOB SER^{ae}	1.0038	0.3285	0.9428	0.7162	0.6264	0.4492	1.9276	1.1473	1.2137	0.4150
(W) TEC SEC	0.0066	0.0160	-0.0244	0.0458	0.0144	0.0282	-0.0745	0.0578	-0.0020	0.0254
(W) SEC^{e}	0.0345	0.0149	0.0632	0.0507	0.0686	0.0371	0.0786	0.0744	0.0667	0.0293
(W) DEG	-0.0240	0.0278	-0.0208	0.0729	-0.0845	0.0497	0.0857	0.1129	-0.0564	0.0416
(W) DEG PHD	-0.0237	0.0227	0.1200	0.0663	-0.0121	0.0380	0.0952	0.1085	-0.0372	0.0372
TT CP	0.0188	0.0063	0.0189	0.0101	0.0109	0.0086	-0.0017	0.0085	0.0110	0.0100
RAIL^{bcd}	0.2510	0.0864	0.7277	0.1877	0.4491	0.1564	0.4860	0.1653	0.2339	0.1622
TT AIR	-0.0051	0.0042	-0.0124	0.0070	-0.0068	0.0060	-0.0041	0.0066	-0.0040	0.0069
TT PORT^{a}	-0.0234	0.0065	-0.0124	0.0096	-0.0165	0.0086	-0.0076	0.0098	-0.0210	0.0112
(W) JOB HT^{acde}	-6.5e-05	6.7e-06	-8.9e-05	5.0e-05	-8.2e-05	1.5e-05	-0.0002	6.7e-05	-7.6e-05	2.4e-05
(W) JOB HT MA^{abde}	0.0010	0.0002	-0.0060	0.0020	-0.0016	0.0009	-0.0118	0.0034	0.0016	0.0008
(W) POP JOB^{abcde}	0.0123	0.0027	0.0173	0.0077	0.0177	0.0052	0.0358	0.0114	0.0132	0.0053
(W) POP JOB E	1.3e-06	6.9e-06	-3.5e-05	3.2e-05	2.9e-06	1.1e-05	-2.8e-05	4.5e-05	-3.1e-07	9.3e-06
(W) POP 20 44^{ace}	0.0528	0.0110	0.0322	0.0317	0.0770	0.0191	0.0485	0.0410	0.0759	0.0219
GIRONA^{bd}	-0.2883	0.1993	-0.8455	0.3489	-0.4440	0.2605	-0.7831	0.3068	-0.2941	0.2983
LLEIDA	-0.4117	0.2340	-0.2490	0.3405	-0.5495	0.3107	-0.4161	0.3755	-0.4943	0.3008
TARRA^{ace}	-0.8611	0.2111	-0.5030	0.4065	-0.7181	0.3334	-0.2073	0.3947	-0.8180	0.3404
CAP CO^{abcde}	1.1162	0.1019	1.4170	0.1713	1.2437	0.1537	1.2521	0.2072	1.2526	0.2500
COAST	0.3189	0.1012	-0.2573	0.2371	0.1254	0.1795	-0.0895	0.2319	0.2786	0.2422
DIST 100^{abce}	-0.0170	0.0035	-0.0111	0.0054	-0.0129	0.0049	-0.0059	0.0060	-0.0168	0.0057
DIST CAT^{de}	0.0067	0.0032	0.0081	0.0055	0.0087	0.0047	0.0137	0.0049	0.0107	0.0050
(W) FIRM SMALL^{abde}	-0.0052	0.0018	-0.0183	0.0052	-0.0041	0.0033	-0.0310	0.0089	-0.0064	0.0030
LogL	-1168.89		-1196.93		-1147.51		-1177.30		-1159.96	
AIC	2397.79		2453.86		2355.03		2414.61		2379.93	
GoF Test	121.01		96.97		111.37		96.88		115.64	
p-value	0.000		0.000		0.000		0.000		0.000	

Note: (W) denotes a weighted variable, $Coef$ an estimated coefficient, $S.E.$ the Standard Error of the estimated coefficient, ^{a} significance at 5% in the Poisson baseline model, ^{b} significance at 5% in the Poisson-Uniform-GoF model, ^{c} significance at 5% in the Poisson-Normal-GoF model, ^{d} significance at 5% in the Poisson-Uniform-ML model, and ^{e} significance at 5% in the Poisson-Normal-ML model. Also, CONST denotes the constant term of the model and ENTRY is the dependent variables (see Table 1 for definitions of the variables).

Table 3: Estimation Results for the Negative Binomial Model

Variables	$h_U = 0$ $h_N = \infty$		Uniform (GoF Criteria)		Uniform (ML Criteria)		Normal (GoF Criteria)		Normal (ML Criteria)	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
CONST^{abce}	-2.8516	0.8751	48.6501	20.4243	-4.3862	1.0523	16.2892	153.3568	-4.5221	1.1547
(W) RES VAR^{bd}	0.2112	0.1648	-13.3745	4.8890	0.3046	0.2314	-51.9167	19.0007	0.4026	0.2563
(W) URB^{ac}	0.0009	0.0003	-0.0222	0.0192	0.0010	0.0004	0.1355	0.3003	0.0007	0.0005
(W) DISURB^{ac}	-7.6e-09	2.3e-09	1.8e-07	1.6e-07	-8.1e-09	3.4e-09	-1.1e-06	2.5e-06	-5.5e-09	3.7e-09
(W) JOB^{abce}	5.8e-05	1.2e-05	0.0025	0.0010	6.7e-05	1.6e-05	0.0055	0.0137	7.9e-05	1.9e-05
(W) DENS^{a}	-0.0003	9.1e-05	-0.0045	0.0055	-0.0001	0.0001	-0.0469	0.0864	-6.8e-05	0.0001
(W) CONC^{ace}	0.5062	0.1162	-4.4486	3.0164	0.6917	0.1565	12.1594	11.4728	0.8114	0.1763
(W) JOB IND^{ace}	2.4219	0.7427	-28.0002	16.9960	1.9944	0.9162	-113.1803	82.4438	2.1668	0.9667
(W) JOB SER^{a}	1.1612	0.4394	-22.0776	13.2504	0.7205	0.5893	-63.4784	106.7698	1.2512	0.6720
(W) TEC SEC	-0.0005	0.0227	0.0200	0.3892	0.0068	0.0289	0.9627	2.7340	0.0096	0.0320
(W) SEC^{ce}	0.0292	0.0218	-0.6056	0.6177	0.0705	0.0307	7.2262	4.6394	0.0882	0.0347
(W) DEG	-0.0494	0.0393	-0.4213	0.9702	-0.0742	0.0546	-0.9627	4.7314	-0.0849	0.0618
(W) DEG PHD	-0.0300	0.0330	1.6056	1.1011	-0.0161	0.0457	-5.2024	6.0427	-0.0315	0.0513
TT CP	0.0164	0.0100	-0.0155	0.0110	0.0050	0.0095	-0.0272	0.0150	0.0033	0.0099
RAIL^{bcd}	0.2462	0.1616	0.7340	0.1680	0.3827	0.1580	0.6253	0.1686	0.3039	0.1595
TT AIR	-0.0036	0.0066	0.0069	0.0105	-0.0044	0.0067	0.0080	0.0137	-0.0034	0.0068
TT PORT^{a}	-0.0244	0.0102	-0.0123	0.0128	-0.0148	0.0101	0.0015	0.0173	-0.0146	0.0102
(W) JOB HT^{abce}	-0.0001	2.0e-05	-0.0052	0.0016	-0.0001	2.6e-05	-0.0132	0.0177	-0.0001	3.3e-05
(W) JOB HT MA	0.0005	0.0006	0.1116	0.0580	-0.0020	0.0011	-0.0414	0.5737	-0.0005	0.0017
(W) POP JOB^{bce}	0.0045	0.0042	-0.4306	0.1473	0.0146	0.0056	0.1807	0.7987	0.0147	0.0063
(W) POP JOB E^{b}	-6.0e-06	1.9e-05	-0.0027	0.0013	-6.6e-06	3.4e-05	0.0177	0.0110	1.1e-05	3.4e-05
(W) POP 20 44^{ace}	0.0334	0.0159	0.2928	0.3463	0.0588	0.0217	0.3767	2.3351	0.0563	0.0243
GIRONA	-0.3256	0.3199	-0.3088	0.4264	-0.4556	0.3205	-0.2151	0.4203	-0.4103	0.3230
LLEIDA^{c}	-0.5189	0.3282	-0.2336	0.4071	-0.6861	0.3411	-0.2839	0.5254	-0.6416	0.3468
TARRA^{ace}	-0.9123	0.2892	-0.2185	0.4313	-0.7587	0.2932	0.4306	0.5065	-0.7301	0.3002
CAP CO^{abcde}	1.6112	0.2250	2.0667	0.2185	1.5983	0.2070	2.1564	0.2156	1.6146	0.2151
COAST^{a}	0.4368	0.2043	-0.0793	0.2236	0.3676	0.2067	0.2122	0.2432	0.3462	0.2181
DIST 100^{abd}	-0.0123	0.0054	-0.0311	0.0072	-0.0104	0.0053	-0.0262	0.0128	-0.0104	0.0054
DIST CAT^{bce}	0.0076	0.0045	-0.0304	0.0155	0.0116	0.0046	-0.0221	0.0416	0.0133	0.0046
(W) FIRM SMALL^{e}	-0.0048	0.0027	-0.1059	0.1078	-0.0056	0.0037	-0.4882	1.1027	-0.0102	0.0041
α^{abcde}	0.8004	0.0229	1.0863	0.0441	0.7898	0.0222	1.0568	0.0416	0.7975	0.0227
LogL	-1016.52		-1049.96		-1011.51		-1046.41		-1010.14	
AIC	2095.04		2161.93		2085.02		2154.83		2082.29	
GoF Test	48.15		10.90		31.16		12.34		37.53	
p-value	0.000		0.282		0.000		0.194		0.000	

Note: (W) denotes a weighted variable, $Coef$ an estimated coefficient, $S.E.$ the Standard Error of the estimated coefficient, ^{a} significance at 5% in the Negative-Binomial baseline model, ^{b} significance at 5% in the Negative-Binomial-Uniform-GoF model, ^{c} significance at 5% in the Negative-Binomial-Normal-GoF model, ^{d} significance at 5% in the Negative-Binomial-Uniform-ML model, and ^{e} significance at 5% in the Negative-Binomial-Normal-ML model. Also, CONST denotes the constant term of the model and ENTRY is the dependent variables (see Table 1 for definitions of the variables).

Table 4: Estimation Results for the Zero Inflated Poisson Model

Variables	$h_U = 0$ $h_N = \infty$		Uniform (GoF Criteria)		Uniform (ML Criteria)		Normal (GoF Criteria)		Normal (ML Criteria)	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
CONST^{abce}	-3.5256	0.8023	-3.8010	0.8820	-4.2100	0.8038	-2.3716	1.2217	-4.6325	0.8774
(W) RES VAR^{abe}	-0.0545	0.1237	-0.4150	0.1877	-0.2356	0.1240	-0.5193	0.3057	-0.3704	0.1550
(W) URB^{abcd}	0.0006	0.0001	0.0006	0.0002	0.0005	0.0001	0.0022	0.0004	0.0003	0.0001
(W) DISURB^{abcde}	-5.1e-09	9.5e-10	-5.0e-09	1.6e-09	-4.4e-09	9.6e-10	-1.8e-08	3.1e-09	-2.6e-09	1.2e-09
(W) JOB^{abcde}	3.2e-05	3.7e-06	5.0e-05	5.3e-06	3.3e-05	3.7e-06	0.0001	1.3e-05	3.7e-05	4.4e-06
(W) DENS^{abcde}	-0.0002	4.1e-05	-0.0002	5.0e-05	-0.0002	4.2e-05	-0.0006	9.8e-05	-0.0001	4.0e-05
(W) CONC^{abcde}	0.3705	0.1463	0.6396	0.1750	0.5990	0.1480	0.5807	0.2448	0.6753	0.1742
(W) JOB IND^{ace}	2.5095	0.6335	0.9944	0.7572	1.7954	0.6361	1.4473	1.0133	2.2212	0.6902
(W) JOB SER^{abce}	1.0038	0.5621	1.2562	0.6382	2.1541	0.5666	0.1000	0.9324	2.5574	0.6616
(W) TEC SEC	0.0066	0.0196	0.0166	0.0237	-0.0029	0.0197	-0.0504	0.0338	-0.0015	0.0228
(W) SEC^{abcde}	0.0345	0.0192	0.0743	0.0268	0.0436	0.0198	0.1278	0.0459	0.0804	0.0253
(W) DEG^{b}	-0.0240	0.0351	-0.1361	0.0507	-0.0241	0.0354	-0.0675	0.0809	-0.0838	0.0432
(W) DEG PHD^{d}	-0.0237	0.0276	0.0174	0.0385	-0.0344	0.0282	0.1580	0.0642	-0.0320	0.0360
TT CP^{ac}	0.0188	0.0068	0.0106	0.0071	0.0168	0.0068	0.0005	0.0078	0.0072	0.0071
RAIL^{bd}	0.2510	0.0875	0.3305	0.0823	0.0906	0.0879	0.3546	0.0822	0.0901	0.0877
TT AIR	-0.0051	0.0047	-0.0060	0.0048	-0.0050	0.0048	-0.0064	0.0051	-0.0026	0.0049
TT PORT	-0.0234	0.0073	-0.0060	0.0071	-0.0111	0.0073	0.0016	0.0072	-0.0089	0.0074
(W) JOB HT^{abcde}	-6.5e-05	6.5e-06	-7.5e-05	8.4e-06	-6.2e-05	6.5e-06	-0.0002	2.1e-05	-7.6e-05	7.4e-06
(W) JOB HT MA^{abcde}	0.0010	0.0002	-0.0012	0.0004	0.0007	0.0002	-0.0059	0.0013	0.0014	0.0004
(W) POP JOB^{abcde}	0.0123	0.0029	0.0144	0.0035	0.0114	0.0029	0.0296	0.0067	0.0112	0.0033
(W) POP JOB E	1.3e-06	7.1e-06	-9.8e-06	1.4e-05	-9.0e-06	7.1e-06	-4.4e-05	3.2e-05	-1.0e-05	8.2e-06
(W) POP 20 44^{abce}	0.0528	0.0123	0.0561	0.0162	0.0535	0.0122	0.0266	0.0244	0.0642	0.0146
GIRONA^{d}	-0.2883	0.2215	-0.4256	0.2236	-0.3236	0.2223	-0.7427	0.2405	-0.2728	0.2230
LLEIDA	-0.4117	0.2843	-0.4244	0.2781	-0.1008	0.2845	-0.4076	0.3118	-0.2978	0.2877
TARRA	-0.8611	0.2380	-0.0886	0.2388	0.0038	0.2382	0.1167	0.2585	-0.0160	0.2404
CAP CO^{abcde}	1.1162	0.1009	1.0166	0.0907	0.7971	0.1005	0.9541	0.0885	0.9225	0.0986
COAST	0.3189	0.1032	0.0253	0.1017	0.0847	0.1035	-0.1907	0.1241	0.0889	0.1102
DIST 100^{abcde}	-0.0170	0.0039	-0.0141	0.0039	-0.0166	0.0039	-0.0120	0.0041	-0.0165	0.0040
DIST CAT	0.0067	0.0034	0.0021	0.0036	0.0023	0.0034	0.0058	0.0039	0.0057	0.0036
(W) FIRM SMALL^{d}	-0.0052	0.0022	-8.3e-05	0.0028	0.0002	0.0022	-0.0179	0.0045	-0.0003	0.0026
LogL	-1005.39		-1013.99		-1007.54		-1034.76		-1004.54	
AIC	2074.78		2091.98		2079.09		2133.53		2073.08	
GoF Test	57.22		44.18		55.79		43.02		52.89	
p-value	0.000		0.000		0.000		0.000		0.000	
Vuong Test	7.053		6.504		7.018		5.913		6.966	
p-value	0.000		0.000		0.000		0.000		0.000	

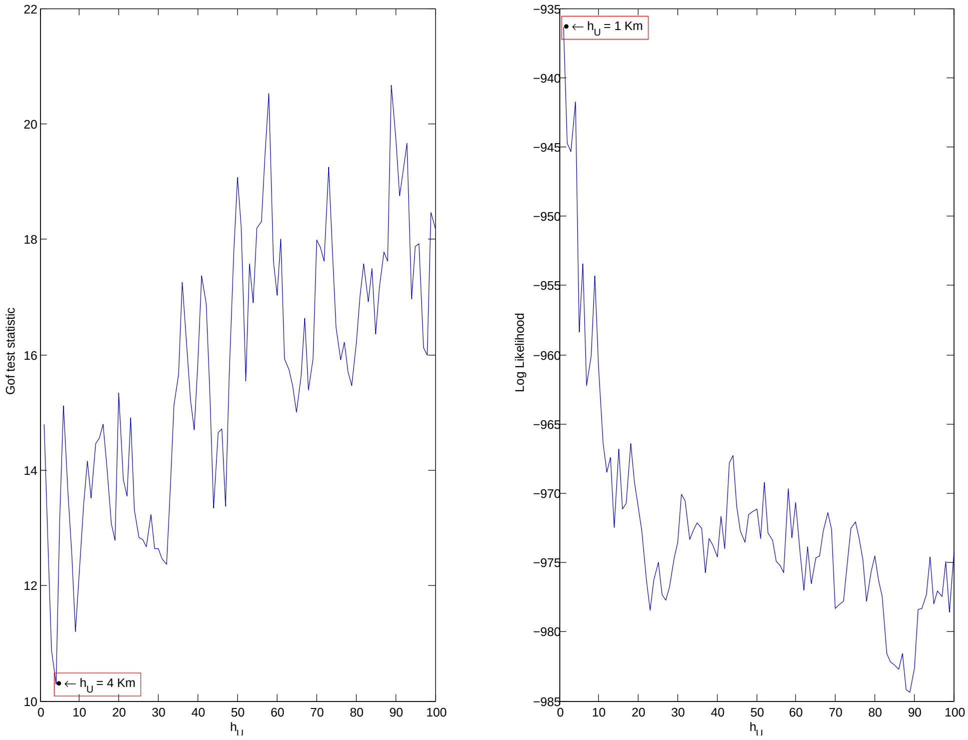
Note: (W) denotes a weighted variable, $Coef$ an estimated coefficient, $S.E.$ the Standard Error of the estimated coefficient, ^{a} significance at 5% in the Inflated-Poisson baseline model, ^{b} significance at 5% in the Inflated-Poisson-Uniform-GoF model, ^{c} significance at 5% in the Inflated-Poisson-Normal-GoF model, ^{d} significance at 5% in the Inflated-Poisson-Uniform-ML model, and ^{e} significance at 5% in the Inflated-Poisson-Normal-ML model. Residential population is the only explanatory variable in the inflated part of the model, being in general the coefficient associated with this variable negative and statistically significant. Also, CONST denotes the constant term of the model and ENTRY is the dependent variables (see Table 1 for definitions of the variables).

Table 5: Estimation Results for the Zero Inflated Negative Binomial Model

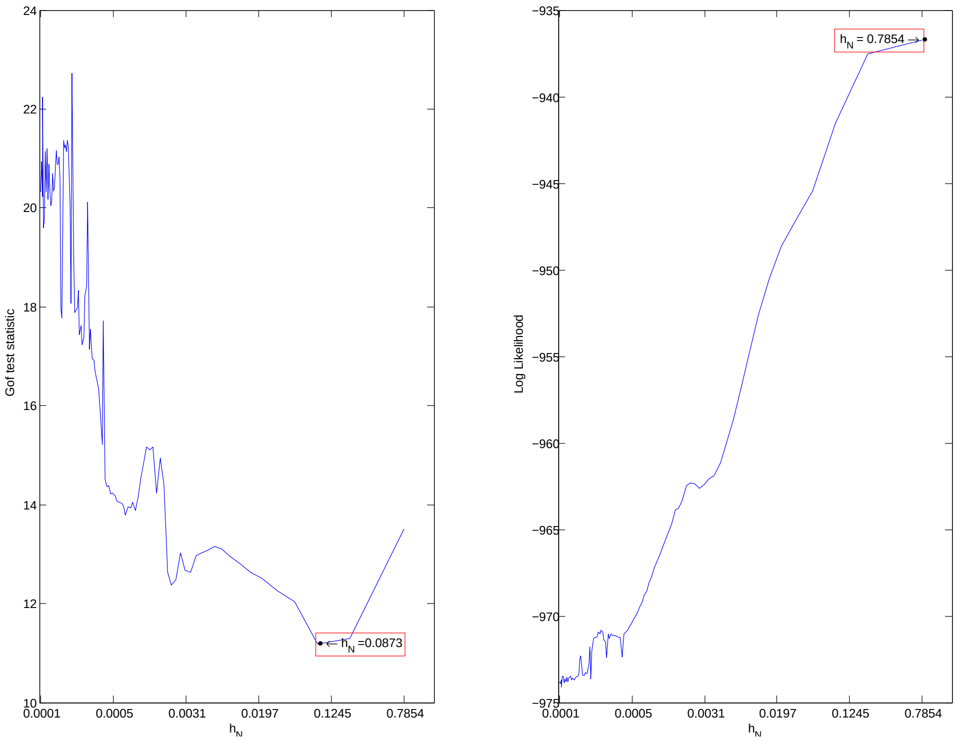
Variables	$h_U = 0$ $h_N = \infty$		Uniform (GoF Criteria)		Uniform (ML Criteria)		Normal (GoF Criteria)		Normal (ML Criteria)	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
CONST^{abcde}	-3.2363	1.0013	-3.3971	1.1428	-3.2726	1.0039	-3.8249	1.2772	-3.2526	1.0164
(W) RES VAR	-0.1450	0.1585	-0.2721	0.2426	-0.1307	0.1591	-0.2772	0.2639	-0.1591	0.1637
(W) URB^{abce}	0.0008	0.0002	0.0008	0.0003	0.0007	0.0002	0.0006	0.0004	0.0008	0.0002
(W) DISURB^{abce}	-6.1e-09	1.7e-09	-6.6e-09	2.8e-09	-6.0e-09	1.7e-09	-5.1e-09	3.0e-09	-6.6e-09	1.8e-09
(W) JOB^{abcde}	4.9e-05	8.7e-06	5.4e-05	1.2e-05	4.9e-05	8.8e-06	6.6e-05	1.5e-05	4.6e-05	8.6e-06
(W) DENS^{ace}	-0.0003	6.9e-05	-0.0002	9.2e-05	-0.0003	6.8e-05	-0.0001	9.6e-05	-0.0003	6.9e-05
(W) CONC^{abcde}	0.6790	0.1658	0.7212	0.2089	0.6647	0.1671	0.8393	0.2384	0.6742	0.1706
(W) JOB IND	1.1304	0.8118	0.5993	0.9891	1.2395	0.8173	0.8038	1.0618	1.1755	0.8258
(W) JOB SER^{acde}	2.1528	0.6290	1.1765	0.7631	2.0423	0.6329	1.7949	0.8858	2.1084	0.6438
(W) TEC SEC	-0.0014	0.0247	0.0125	0.0306	-0.0023	0.0248	0.0097	0.0345	-0.0026	0.0251
(W) SEC^{abcde}	0.0538	0.0259	0.1170	0.0383	0.0626	0.0267	0.1425	0.0431	0.0774	0.0273
(W) DEG	-0.0391	0.0446	-0.1095	0.0636	-0.0350	0.0450	-0.1050	0.0702	-0.0437	0.0456
(W) DEG PHD	-0.0399	0.0359	-0.0139	0.0506	-0.0493	0.0366	-0.0361	0.0578	-0.0523	0.0374
TT CP	0.0142	0.0096	0.0033	0.0095	0.0138	0.0096	-0.0006	0.0101	0.0124	0.0096
RAIL	0.0171	0.1299	0.2215	0.1295	0.0250	0.1309	0.1371	0.1314	0.0379	0.1311
TT AIR	-0.0065	0.0065	-0.0063	0.0067	-0.0065	0.0065	-0.0050	0.0068	-0.0060	0.0065
TT PORT	-0.0170	0.0099	-0.0105	0.0102	-0.0171	0.0099	-0.0107	0.0103	-0.0169	0.0100
(W) JOB HT^{abcde}	-8.6e-05	1.4e-05	-8.4e-05	2.0e-05	-8.7e-05	1.4e-05	-0.0001	2.5e-05	-8.5e-05	1.4e-05
(W) JOB HT MA	0.0003	0.0004	-0.0012	0.0009	0.0004	0.0004	6.3e-05	0.0013	0.0005	0.0005
(W) POP JOB^{bd}	0.0056	0.0039	0.0112	0.0050	0.0050	0.0039	0.0123	0.0057	0.0057	0.0040
(W) POP JOB E	-1.3e-05	1.4e-05	-1.6e-05	2.8e-05	-1.3e-05	1.4e-05	-1.1e-05	2.6e-05	-1.2e-05	1.4e-05
(W) POP 20 44^{ac}	0.0336	0.0160	0.0429	0.0220	0.0333	0.0160	0.0457	0.0245	0.0315	0.0163
GIRONA	-0.4030	0.2993	-0.5013	0.3110	-0.4087	0.3007	-0.4613	0.3144	-0.4108	0.3017
LLEIDA	-0.0250	0.3531	-0.3487	0.3677	-0.0586	0.3538	-0.3227	0.3728	-0.1140	0.3556
TARRA	-0.2222	0.3019	-0.2969	0.3114	-0.2374	0.3020	-0.2471	0.3186	-0.2445	0.3032
CAP CO^{abcde}	1.0387	0.1740	1.2028	0.1643	1.0846	0.1729	1.2032	0.1714	1.0924	0.1712
COAST	-0.0360	0.1686	-0.0032	0.1746	-0.0456	0.1691	-0.0507	0.1843	-0.0741	0.1702
DIST 100	-0.0071	0.0055	-0.0059	0.0056	-0.0071	0.0055	-0.0060	0.0057	-0.0075	0.0055
DIST CAT	0.0022	0.0044	0.0057	0.0046	0.0028	0.0044	0.0078	0.0048	0.0035	0.0045
(W) FIRM SMALL	0.0011	0.0029	-0.0010	0.0039	0.0016	0.0030	-0.0036	0.0044	0.0010	0.0030
α^{abcde}	0.3218	0.0305	0.3765	3.8e-02	0.3273	3.1e-02	0.3762	3.7e-02	0.3327	3.7e-02
LogL	-936.46		-941.71		-936.38		-941.53		-936.73	
AIC	1938.92		1949.43		1938.77		1949.06		1939.46	
GoF Test	14.46		10.30		14.80		11.18		13.50	
p-value	0.106		0.326		0.096		0.263		0.141	
Vuong Test	6.291		6.104		6.262		5.999		6.233	
p-value	0.000		0.000		0.000		0.000		0.000	

Note: (W) denotes a weighted variable, $Coef$ an estimated coefficient, $S.E.$ the Standard Error of the estimated coefficient, ^{a} significance at 5% in the Inflated-Negative-Binomial baseline model, ^{b} significance at 5% in the Inflated-Negative-Binomial-Uniform-GoF model, ^{c} significance at 5% in the Inflated-Negative-Binomial-Normal-GoF model, ^{d} significance at 5% in the Inflated-Negative-Binomial-Uniform-ML model, and ^{e} significance at 5% in the Inflated-Negative-Binomial-Normal-ML model. Residential population is the only explanatory variable in the inflated part of the model, being in general the coefficient associated with this variable negative and statistically significant. Also, CONST denotes the constant term of the model and ENTRY is the dependent variables (see Table 1 for definitions of the variables).

Figure 1: **Optimal h_f Parameters and Distance (Zero Inflated Negative Binomial Model Estimates).**



(a) Uniform case



(b) Normal case

Figure 2: **Relative Effects (Zero Inflated Negative Binomial Model Estimates).**