



UNIVERSITAT  
ROVIRA I VIRGILI

## WORKING PAPERS

Col·lecció “DOCUMENTS DE TREBALL DEL  
DEPARTAMENT D’ECONOMIA”

**“Empirical Studies in Industrial Location:  
An Assessment of their Methods and Results”**

Josep-Maria Arauzo-Carod  
Daniel Liviano-Solis  
Miguel Manjón-Antolín

Document de treball n° -6- 2008

**DEPARTAMENT D’ECONOMIA  
Facultat de Ciències Econòmiques i Empresariales**



UNIVERSITAT  
ROVIRA I VIRGILI

*Edita:*

Departament d'Economia

[http://www.fcee.urv.es/departaments/economia/public\\_html/index.html](http://www.fcee.urv.es/departaments/economia/public_html/index.html)

Universitat Rovira i Virgili

Facultat de Ciències Econòmiques i Empresariales

Avgda. de la Universitat, 1

432004 Reus

Tel. +34 977 759 811

Fax +34 977 300 661

*Dirigir comentaris al Departament d'Economia.*

Dipòsit Legal: T-1176-2008

ISSN 1988 - 0812

**DEPARTAMENT D'ECONOMIA**  
**Facultat de Ciències Econòmiques i Empresariales**

# **Empirical Studies in Industrial Location: An Assessment of their Methods and Results\***

Josep-Maria Arauzo-Carod\*\*

Daniel Liviano-Solis

and

Miguel Manjón-Antolín

(QURE - Department of Economics, Rovira i Virgili University)

## **Abstract**

This paper surveys recent evidence on the determinants of (national and/or foreign) industrial location. We find that the basic analytical framework has remained essentially unaltered since the early contributions of the early 1980's while, in contrast, there have been significant advances in the quality of the data and, to a lesser extent, the econometric modelling. We also identify certain determinants (neoclassical and institutional factors) that tend to provide largely consistent results across the reviewed studies. In light of this evidence, we finally suggest future lines of research.

---

\* This research was financially supported by grants SEJ2007-64605/ECON and SEJ2007-65086/ECON as well as by the "Xarxa de Referència d'R+D+I en Economia i Polítiques Públiques" of the Catalan Government. We are grateful to P. McCann, I. Mariotti and E. Viladecans for helpful discussions on the topic. Any errors are of course our own.

\*\* Corresponding author ([josepmaria.arauzo@urv.cat](mailto:josepmaria.arauzo@urv.cat)). Address for correspondence: Faculty of Economics and Business, Av. Universitat 1, Reus-43204, Spain.

# 1. Introduction

The location of production units (firms, plants) has been a major topic in Economics ever since the seminal work of Alfred Marshall (1890).<sup>1</sup> However, in recent decades there has been a boost in the number of empirical studies investigating the driving forces behind the location decisions of new industrial concerns. The increasing amount of public programs aiming to attract and promote the creation of new businesses, advances in the analytical foundations and the econometric modelling, as well as wider access to suitable data sets, are some of the reasons that explain the growing interest in the determinants of industrial location (McFadden 2001, McCann and Sheppard 2003, Guimarães et al. 2004). Ultimately, this research has important implications for managers, entrepreneurs and policy makers insofar as, for a new venture, the choice of location can make the difference between failure and success (Strotmann 2007).<sup>2</sup>

Most contributions to this literature consist of new evidence on certain determinants (taxes, wages, agglomeration economies, etc.) and/or new empirical approaches (e.g. Poisson models, etc.), often using new data sets (for smaller geographical areas, with longitudinal structure, etc.). As a result, these investigations differ substantially in terms of econometric specifications, covariates and sampling characteristics (data sources, statistical units, institutional settings, etc.). This heterogeneity has made comparisons difficult and no consensus seems to have emerged on what are the central location factors or what is the best way to estimate their importance. Consequently, a survey paper that provides a retrospective analysis while critically assessing the main findings of this literature would surely be helpful. However, to our knowledge such a review has not yet been done. This is our aim here.<sup>3</sup>

---

<sup>1</sup> See e.g. Isard (1956) for an overview of the early twentieth century contributions.

<sup>2</sup> Although our interest lies in those studies that investigate a firm's/plant's location decisions from a (broadly defined) Regional and Urban Economics perspective, it is interesting to note that there are related studies in other fields such as Marketing (Bradlow et al. 2005) and Industrial Organisation (Seim 2006).

<sup>3</sup> It is worth stressing that we restrict attention to those studies that analyse location decisions of new industrial (national and/or foreign) concerns using appropriate econometric models. This means that we have not considered investigations that merely present descriptive statistics, focus on agriculture or service sectors, and/or analyse the stock (see e.g. Shukla and Waddell 1991) rather than the entry of firms/plants. This selection criterion allows us to keep the number of analysed references within reasonable limits and, more importantly, to make the studies involved in this survey largely comparable. In particular, we have considered those investigations focusing on Foreign Direct Investments (FDI hereafter) as part of the industrial location literature because, nationality of the production units aside,

The rest of the paper is organised as follows. In Section 2 we describe the econometric methods typically used in these investigations and in Section 3 we examine their main results and data features. That is, in Section 2 we focus on the definition of the dependent variables whereas in Section 3 we direct our attention to the explanatory variables and their estimated effects. In Section 4 we extend the analysis to some strands of the literature that deserve particular attention, namely, the idiosyncrasies of FDI and relocations, as well as the problems associated with the spatial aggregation. Section 5 concludes with an overview and suggestions for future research.

## 2. Methods

Since where to locate is a critical decision for a new venture (Seim 2006, Strotmann 2007), it is of great importance to empirically assess what are the factors that shape this decision. One way to proceed is to follow a discrete choice approach that distinguishes between those factors related to the agent taking the decision (i.e. the entrepreneur, the firm) and those related to the set of alternatives from which the choice is made (i.e. the territory, the spatial area). However, rather than examining location decisions from the viewpoint of the agent that makes the choice, one may approach the issue from the viewpoint of the chosen territory. That is, rather than analysing which characteristics make a territory comparatively more attractive than another when it comes to deciding where to locate a new concern (and/or what the relevant decision agent characteristics are when making this choice), one may analyse which characteristics of a territory affect the (per period average) number of new concerns that are created therein.

Needless to say, there are differences in the statistical information and the econometric specifications that are required for each approach (see e.g. Bradlow et al. 2005). On the one hand, if the unit of analysis is the firm/plant and the main concern is how its characteristics (size, sector, etc.) and/or those of the chosen territory

---

there are practically no differences between the econometric specifications, covariates and sampling characteristics of FDI and “national” studies —see, however, Section 4.1 and McCann and Mudambi (2004). Nevertheless, because they constitute a distinguishable group, we have explicitly identified these studies in our analyses.

(population, infrastructures, etc.) affect location decisions, then Discrete Choice Models (DCM hereafter) are used. On the other hand, if the unit of analysis is geographical (municipality, county, province, region, etc.) and the factors that may affect location decisions refer accordingly to the territory, then Count Data Models (CDM hereafter) are used. These differences notwithstanding, both DCM and CDM are consistent with a profit maximization framework in which firms choose the optimal location subject to standard constraints. More specifically, inferences from both DCM and CDM can be interpreted as reduced-form results derived from a structural model of firm/plant location decision (Becker and Henderson 2000, Guimarães et al. 2004).

One may argue that DCM have an advantage over CDM because they may account for both firm/plant and spatial factors. However, firm/plant characteristics cannot be (fully) identified in some DCM. Moreover, computation of the likelihood function in DCM is cumbersome when the number of alternatives, that is, sites, is large. Lastly, the set of alternatives in DCM only includes those locations effectively chosen, since the rest do not contribute to the likelihood function. These issues can turn CDM into our preferred specification (Kim et al. 2008), especially whenever it is possible to recover the parameter estimates of DCM from the estimates of CDM (Guimarães et al. 2003). Also, computational burden is not an issue in CDM, and zero observations not only contribute to the likelihood function but provide interesting insights about the data generation process (Mullahy 1997).

There are thus pros and cons to both DCM and CDM. Yet these are the basic econometric tools in empirical studies on industrial location. Accordingly, next we examine the distinctive statistical features of those studies that have resorted to DCM (summarised in Table 1) and later of those that have resorted to CDM (summarised in Table 2). Also, Section 3 completes this analysis with a description of their main explanatory variables and estimated effects.

## **2.1. Discrete Choice Models (DCM)**

The principal assumptions in the DCM of industrial location are the following (Carlton 1979, 1983; McFadden 2001). First, firm/plant  $n = 1, \dots, N$  chooses its location among a fixed set of  $J$  alternatives or sites. Second, choosing a particular site  $j = 1, \dots, J$  entails a profit of  $\Pi_{nj}$  for the firm/plant. Third, firms/plants choose location  $j$  over

location  $i$  if and only if  $\Pi_{nj} > \Pi_{ni}$ . Fourth, profits are not observable by the researcher but can be additively decomposed into a systematic component ( $\pi_{nj}$ ) - which is a function  $\pi_{nj} = \pi(x_j, w_n)$  that depends on the attributes of the alternatives ( $x_j$ ) and firm/plant characteristics ( $w_n$ ) - and a random component ( $\varepsilon_{nj}$ ) whose joint density is  $f(\varepsilon_n) = f(\varepsilon_{n1}, \varepsilon_{n2}, \dots, \varepsilon_{nJ})$ .

Under these assumptions, the determinants of industrial location decisions can be empirically examined by calculating how *ceteris paribus* changes in the elements of the systematic component of profits,  $\pi_{nj}$ , affect the probability that firm/plant  $n$  chooses site  $j$ ,  $P_{nj}$ . However, to effectively calculate the partial or marginal effects on the choice probabilities we need to specify the functions  $\pi(x_j, w_n)$  and  $f(\varepsilon_n)$ , for it is easy to see that

$$P_{nj} = Pr(\varepsilon_{nj} - \varepsilon_{ni} < \pi_{ni} - \pi_{nj}, \forall j \neq i) = \int_{\varepsilon_n} \mathbf{I}(\varepsilon_{nj} - \varepsilon_{ni} < \pi_{ni} - \pi_{nj}, \forall j \neq i) f(\varepsilon_n) d\varepsilon_n,$$

with  $\mathbf{I}(\cdot)$  being an indicator function (McFadden 1974).

The function  $\pi(x_j, w_n)$  is usually assumed to be linear in parameters, so that without loss of generality we can define  $\pi(x_j, w_n) = \theta'z_{nj}$ , being  $z_{nj} = \{x_j, w_n\}$  and  $\theta$  a suitable parameter vector.<sup>4</sup> As for the joint distribution of  $\varepsilon_n$ , its specification has important implications for both the estimation procedure (choice probabilities may not have a closed form) and the “substitution patterns” between potential locations (how changes in the attributes of a site affect the odds of the other sites being chosen, provided that choice set probabilities sum to one). Ultimately, alternative specifications of  $f(\varepsilon_n)$  lead to different DCM.<sup>5</sup>

---

<sup>4</sup> Since only differences in profits matter in the construction of choice probabilities, firm/plant characteristics are not identified unless “they are specified in ways that create differences in [profits] over alternatives” (Train 2003: 25). One way is to assume that these characteristics affect profits differently for different locations, i.e.  $\pi(x_j, w_n) = \beta'x_j + \delta'w_n = \theta'z_{nj}$  ( $\beta$  and  $\delta$  are suitable parameter vectors). This assumption, however, requires the normalisation of one of the associated coefficients, such as e.g.  $\delta_1 = 0$ . Another way is to interact firm/plant characteristics with the attributes, i.e.  $\pi(x_j, w_n) = \beta'x_j + \delta'(x_j \times w_n) = \theta'z_{nj}$ , or with dummies for the choices,  $d_j$ , so that  $\pi(x_j, w_n) = \beta'x_j + \delta'(d_j \times w_n) = \theta'z_{nj}$ . In this case normalisation is achieved by dropping all the cross-products of one the alternatives (Wooldridge 2002).

<sup>5</sup> The Probit model, for example, assumes that  $\varepsilon_n$  is multivariate Normal. As far as the estimation concerns, this assumption implies that the resulting integral in  $P_{nj}$  does not have, with the exception of the binary case, a closed form and has consequently to be approximated using deterministic or simulation methods (see e.g. Geweke 1996). As for the substitution patterns, the multivariate Normal assumption provides an extremely general setting (see, however, Bunch 1991).

**[Insert Table 1 around here]**

In the empirical studies reviewed here (see Table 1), the joint distribution of  $\varepsilon_n$  is assumed to be a member of a family of multivariate extreme value distributions with cumulative distribution function  $F(\varepsilon_n) = \exp\{-G(e^{-\varepsilon_{n1}}, e^{-\varepsilon_{n2}}, \dots, e^{-\varepsilon_{nj}})\}$ . In particular, which member of this family of distributions is effectively assumed in each study depends on the definition of  $G(\cdot)$ , an homogeneous function of degree one that satisfies certain conditions —see McFadden (1978) for details. This distributional assumption turns out to be extremely convenient because, unlike, for example, the multinomial Probit model, it usually provides a closed-form expression for the choice probabilities:

$$P_{nj} = \frac{e^{\pi_{nj}} \frac{\partial G(e^{\pi_{n1}}, e^{\pi_{n2}}, \dots, e^{\pi_{nj}})}{\partial e^{\pi_{nj}}}}{G(e^{\pi_{n1}}, e^{\pi_{n2}}, \dots, e^{\pi_{nj}})},$$

which means that the estimation of the partial effects can be easily performed using, for example, maximum likelihood. Interestingly, since this result hinges essentially on  $G(\cdot)$ , this applies to a vast family of DCM known as Generalized Extreme Value Models. However, Table 1 shows that in the industrial location literature only two examples of Generalized Extreme Value Models have been used: the Multinomial (Conditional) Logit Model and the Nested Logit Model.<sup>6</sup>

---

<sup>6</sup> We use the term “Multinomial (Conditional) Logit” to refer to the model originally developed by McFadden (1974) as the “Conditional Logit” but which has become known as “Multinomial Logit” in the discrete choice literature (see McFadden 2001 for an account). In fact, econometric textbooks such as Wooldridge (2002: 500-501) use the term Conditional Logit for the specification of the systematic component of profits as a function of variables that “differ across alternatives and possibly across individuals as well”, i.e.  $\pi(x_j, w_n) = \beta'x_j + \delta_j'w_n = \theta'z_{nj}$ , whereas the Multinomial Logit is a particular case in which variables only differ across individuals, so that  $\pi(w_n) = \delta' (d_j \times w_n) = \theta'z_{nj}$ . Among the industrial location studies, Carlton (1979, 1983), McConnell and Schwab (1990), Levinson (1996), Baudewyns et al. (2000) and Figueiredo et al. (2002) use the Conditional Logit Model with only site characteristics as covariates (see also Luger and Shetty (1985), Coughlin et al. (1991), Friedman et al. (1992), Woodward (1992), Head et al. (1999), Guimarães et al. (2000) and Cheng and Stough (2006) in the FDI area), Arauzo and Manjón (2004) likewise use the Conditional Logit Model but include cross-products of firm/plant characteristics (see also Schmenner et al. 1987) and dummies for the choices (see also Autant-Bernard (2006); Head et al. (1995), Luker (1998) and Crozet et al. (2004) for applications to FDI; and Lee (2004) for an application to relocations), and Baudewyns (1999) and Arauzo and Manjón (2004) use the Multinomial Logit Model with only firm/plant characteristics.



In the Multinomial (Conditional) Logit Model,  $G = \sum_j \exp(\pi_{nj})$ ,  $P_{nj} = e^{\pi_{nj}} / \sum_j e^{\pi_{nj}}$  and each random component of the profit function,  $\varepsilon_{nj}$ , is assumed to be i.i.d. extreme value. This is the model used by Carlton (1979, 1983) in his seminal work and, more recently, by, for example, Schmenner et al. (1987), McConnell and Schwab (1990), Levinson (1996), Baudewyns (1999), Baudewyns et al. (2000), Figueiredo et al. (2002) and Arauzo and Manjón (2004). In the Nested Logit Model, it is assumed that the alternatives can be partitioned into  $D_s$  nonoverlapping subsets or nests ( $s = 1, \dots, S$ ),

$$G = \sum_s \left( \sum_{j \in D_s} \exp\left(\frac{\pi_{nj}}{\lambda_s}\right) \right)^{\lambda_s} \quad \text{and}$$

$$P_{nj} = \exp\left(\frac{\pi_{nj}}{\lambda_s}\right) \left\{ \sum_{j \in D_s} \exp\left(\frac{\pi_{nj}}{\lambda_s}\right) \right\}^{\lambda_s - 1} \left\{ \sum_s \left( \sum_{j \in D_s} \exp\left(\frac{\pi_{nj}}{\lambda_s}\right) \right)^{\lambda_s} \right\}^{-1}, \quad \text{with } 0 \leq \lambda_s \leq 1 \text{ being a scale}$$

parameter (Hensher and Greene 2002). Moreover, the marginal distribution of each  $\varepsilon_{nj}$  is univariate extreme value. This is the model used by, for example, Bartik (1985, 1988), Hansen (1987), Henderson and Kuncoro (1996) and Guimarães et al. (1998) and, for FDI and relocations, by Basile et al. (2003) and Crozet et al. (2004), and Strauss-Kahn and Vives (2005), respectively.

One of the attractive features of the Multinomial (Conditional) Logit Model is that marginal effects are easy to calculate. However, computation becomes an issue when the number of alternatives is large, as typically occurs in industrial location choices (as an illustration, Stata, for example, limits the number of alternatives to fifty). Nevertheless, consistent albeit less efficient estimates can be obtained by constructing smaller choice sets from a (random) sampling of the alternatives.<sup>7</sup> On the other hand, the main drawback of the Multinomial (Conditional) Logit Model lies on the independence assumption, for it implies that the ratio of the probabilities of choosing any two sites  $j$  and  $i$ ,  $P_{nj}/P_{ni} = e^{\theta'z_{nj}}/e^{\theta'z_{ni}}$ , depends only on the attributes of these sites.

<sup>7</sup> Applications of this technique, originally proposed by McFadden (1978), can be found in e.g. Hansen (1987) and Guimarães et al. (1998) using the Nested Logit Model and in e.g. McConnell and Schwab (1990) using the Multinomial (Conditional) Logit Model —see also Shukla and Waddell (1991), who use the share of establishments in zip codes rather than the establishments' actual choices as the dependent variable, and, for FDI, see e.g. Friedman et al. (1992), Woodward (1992) and Guimarães et al. (2000). An alternative approach, followed by Bartik (1985, 1988) using the Nested Logit, is to employ aggregated sites and assume that they are representative of the “relatively homogenous (...) disaggregated alternatives” (McFadden 1978: 549).

Consequently, this specification imposes a uniform pattern of substitution between locations “in that one cannot postulate a pattern of differential substitutability and complementarity between alternatives” (McFadden 1974: 112).

Carlton (1983: 441) contends that “[t]he independence assumption (...) is not an implausible one since the possible locations studied in the empirical work are geographically quite distant” and he, indeed, cannot totally reject its validity. In contrast, Bartik (1985: 16; 1988) argues that “[t]his assumption seems implausible for business location decisions” because, “[p]resumably, there are unmeasured attributes (...) that affect [firm/plant] profits and are correlated within [sites]”. However, he does not provide solid inferences supporting this view. Unfortunately, none of the other location studies using the Multinomial (Conditional) Logit Model have statistically addressed this issue (see Table 1), so that little empirical evidence exists regarding the independence of the irrelevant alternatives assumption —see, however, Cheng and Stough (2006). In any case, as Bartik (1985: 16) points out, “the conditional logit approach remains attractive because of its computational feasibility compared with other alternative approaches to the discrete choice problem”.

One such alternative is the Nested Logit Model, which can be consistently estimated by (full-information) maximum likelihood. However, “[n]umerical maximization is sometimes difficult, since the log-likelihood function is not globally concave and even in concave areas is not close to a quadratic” (Train 2003: 89).<sup>8</sup> In fact, the maximum number of nests is restricted in most statistical packages (as an illustration, LIMDEP-NLOGIT, for example, limits the number of levels to four and the total number of choices to a hundred). There is also the question of how to define the nests (see e.g.

---

<sup>8</sup> When facing these difficulties, a sequential, albeit less efficient, estimation procedure may provide consistent starting values for the maximum likelihood method. This involves first estimating the marginal probabilities of choosing an alternative within nest  $D_s$  and then using these to estimate the conditional probabilities of choosing alternative  $j$  given that an alternative in nest  $D_s$  has been chosen —see, however, Hensher (1986); for applications see e.g. Guimarães et al. (1998) and, in the FDI area, Deveraux and Griffith (1998), Crozet et al. (2004) and Disdier and Mayer (2004). Another less efficient procedure to full maximum likelihood is to estimate the coefficients of the lower nest using the Multinomial (Conditional) Logit Model but including dummies for the choices in the upper nest. “The two procedures are intuitively identical because the set of (...) dummy variables absorbs the [intra-sites] correlation” (Bartik 1985: 16, 1988). However, most studies that include dummy variables for larger geographical areas to somehow control for this correlation do not interpret the results as obtained from a Nested Logit Model but simply from a Multinomial (Conditional) Logit Model —see e.g. Levinson (1996) and McConnell and Schwab (1990) and, for FDI, Woodward (1992), Luker (1998), Head et al. (1999) and Cheng and Stough (2006).

Hensher 1986), although this is usually not a major concern in industrial location studies because data is typically collected for administrative divisions of a geographical area (municipalities, counties, provinces, etc.) and these naturally adopt a nested structure —see, however, Schmenner et al. (1987), Guimarães et al. (1998), Basile et al. (2003) and Strauss-Kahn and Vives (2005).

As for the substitution patterns, it is assumed that there is constant correlation between alternatives of the same nest (measured by  $1-\lambda_s$ ) and no correlation with respect to alternatives in other nests. Thus, the independence of irrelevant alternatives holds within alternatives in the same nest but not across nests. In particular, the Nested Logit Model collapses into the Multinomial Logit Model if there is independence among all alternatives in all the nests, that is, if  $\lambda_s = 1 \forall s$  (McFadden 1978). This is tested and rejected in industrial location studies by, for example, Hansen (1987), Henderson and Kuncoro (1996) and Guimarães et al. (1998) and, in FDI applications, by Basile et al. (2003) and Crozet et al. (2004).

It is nonetheless apparent that, to a large extent, the computational and pattern-substitutability drawbacks that characterise the Multinomial (Conditional) Logit Model apply also to the Nested Logit Model. However, we have found no attempt to address these limitations in industrial location studies. In fact, more flexible DCM, such as, for example, the Probit and the Mixed Logit, are virtually unheard of in this literature. Admittedly, these models are not exempt from difficulties either (Bunch 1991, Geweke 1996). Still, less restrictive settings for dealing with unobservables and/or accommodating general patterns of substitution may well pay off. Future research is needed to verify this tenet.

## **2.2. Count Data Models (CDM)**

The principal assumptions in the CDM of industrial location are the following (Becker and Henderson 2000). First, there is a supply of potential entrepreneurs that, at a given point in time, are considering creating a new firm/plant in site  $j (= 1, \dots, J)$ . This supply function is stochastic, not observable by the researcher, and depends essentially on location characteristics ( $x_j$ ) and the number of new firms/plants that are effectively created in this location over a given time period ( $n_j$ ). Second, there is an unobservable stochastic demand function that depends essentially on the same

factors as the supply function does ( $x_j$  and  $n_j$ ), plus some further location characteristics that do not affect the supply function (i.e.  $x_j \subseteq z_j$ ). Third, the number of new firms/plants (effectively) created in site  $j$  over some time period can be implicitly derived from the intersection of supply and demand functions. Thus, there exists an equilibrium given by the following reduced-form equation:

$$n_j = n(z_j, \dots \mid \theta)$$

where  $n(\cdot)$  is a function whose first derivative with respect to the set of covariates  $z_j$  satisfies certain regularity conditions (see Becker and Henderson 2000) and  $\theta$  is a conformable parameter vector.

Under these assumptions, the determinants of industrial location decisions can be empirically examined by calculating how *ceteris paribus* changes in location characteristics affect the conditional expectation of the number of firms/plants created in site  $j$  over a given time period. However, to effectively calculate the partial or marginal effects we need to specify the (conditional) density or probability mass function of  $n_j$ . Given the nonnegative integer nature of the dependent variable ( $n_j = 0, 1, 2, \dots$ ), the Poisson distribution arises as a natural candidate. In particular, alternative parameterisations of the mean or rate parameter with respect to the set of covariates lead to different CDM (Wooldridge 2002). Table 2 reports those used in the industrial location literature.

**[Insert Table 2 around here]**

The standard Poisson regression model, for example, assumes that  $E(n_j \mid z_j) = \exp(\beta'z_j) = \text{Var}(n_j \mid z_j)$ . This is the specification used by, for example, Gabe and Bell (2004), Arauzo and Manjón (2004), Arauzo (2005, 2008), Autant-Bernard et al. (2006), Alañón et al. (2007) and Arauzo and Viladecans (2008) and, for FDI, by, for example, Smith and Florida (1994), Wu (1999), List (2001) and Barbosa et al. (2004). However, since location data tends to reject the assumption of “equidispersion”, that is, equality of conditional variance and mean, and to show an “excess of zeros” relative to what the Poisson distribution produces, results from this model are generally reported for illustrative purposes only. Overdispersion implies that inferences from maximum

likelihood estimates are no longer valid, whereas underestimating the frequency of zeros may result in inconsistent estimates. Although estimates remain consistent if the conditional mean is correctly specified (Gourieroux et al. 1984), there are alternative specifications better suited to accommodate these distinctive features of the data.<sup>9</sup>

Overdispersion and excess of zeros arise ultimately from the existence of unobserved heterogeneity in the conditional mean parameter (Mullahy 1997). Accordingly, “mixture” Poisson regression models allow for this heterogeneity by assuming that  $E(n_j | z_j, \xi_j) = \exp(\beta'z_j)\xi_j = \mu_j\xi_j$ , with  $\xi_j$  being an i.i.d. variable that is independent of the covariates. A leading case is the Negative Binomial Model (NBM), in which  $\xi_j$  has Gamma distribution with unitary mean and constant variance  $\alpha$ . The resulting continuous mixture has a conditional variance function quadratic in the mean,  $\mu_j + \alpha\mu_j^2$ , although the lineal form,  $\mu_j + \alpha\mu_j$ , is also used —see also Barbosa et al. (2004) for an FDI application that uses an encompassing specification of the conditional variance function. The former corresponds to the so-called NB2 Model used by, for example, Arauzo and Viladecans (2008); by Wu (1999) and Cieřlik (2005b) in FDI applications; and by Manj3n and Arauzo (2007) for relocations. The latter is known as the NB1 Model and has been used by, for example, Kogut and Chang (1991) in an FDI application. Other studies that use the NBM but do not make explicit the form of the conditional variance function include Bade and Nerlinger (2000), Gabe and Bell (2004), Egehn et al. (2004), Audretsch and Lehmann (2005), Autant-Bernard et al. (2006), Ala3n3n et al. (2007) and Arauzo (2008) and, for FDI, Coughlin and Segev (2000) and Cieřlik (2005a).

The heterogeneity term in the Poisson-Gamma mixture that characterises the NBM can be interpreted as a location-specific random effect. However, a discrete representation of the unobserved locational heterogeneity is also possible. This involves using “finite mixtures” models that, rather than assuming a continuous distribution for  $\xi_j$ , allow for the existence of an undetermined number of

---

<sup>9</sup> Nevertheless, the use of the standard Poisson regression model can be justified on the grounds of the “equivalence relation between the likelihood function of the conditional logit and the Poisson regression”, which in practice means “that the coefficients of the conditional logit model can be equivalently estimated using a Poisson regression” (Guimar3es et al. 2003: 202-203). See, for example, Gabe and Bell (2004), Arauzo and Manj3n (2004) and Arauzo (2005) for applications using cross-section data and Guimar3es et al. (2004) for an extension to panel data settings —see also Holl (2004c) for an application in this context.

heterogeneous groups in the population of interest (Cameron and Trivedi 1998). In the simplest case, it is assumed that there are two groups of sites: those in which new firms/plants are not, and will not be created (e.g. because they are banned by environmental regulations), and those in which new firm/plants might or might not be created (i.e. in principle there is nothing that prevents this event from happening although it may not happen in certain circumstances, as, for example, when the location is too small, too remote, etc.).<sup>10</sup>

To construct the finite mixture model, this binary-form of heterogeneity is parameterised using, for example, the logistic or Normal cumulative distribution function. The associated Logit or Probit model for the probability of zero entrants in a particular location is then mapped into a count model for the number of new firms/plants created in that location over a given time period. The resulting specification critically differs from the parent count model in that it does not yield the same probabilities for zero and positive outcomes (Mullahy 1986, Greene 1994). The Zero Inflated Poisson Model (ZIPM) uses the standard Poisson model to this end, whereas using NBM generates the Zero Inflated Negative Binomial Model (ZINBM). In industrial location studies, Gabe (2003) uses the ZIPM and Arauzo (2008) the ZINBM; see also List (2001) and Basile (2004) for FDI applications using ZIPM and Manjón and Arauzo (2007) for an application of both models to relocations.

Lastly, an alternative (or maybe complementary) way to deal with the unobserved heterogeneity is to take advantage of the longitudinal structure of the data — see Papke (1991) for a pioneering study. The NBM and the ZINBM previously described enable us to control for unobserved location-specific heterogeneity when the data

---

<sup>10</sup> This is the motivation of the “with-zeros (WZ)” model of Mullahy (1986: 345-347), who also proposes a “hurdle” model that shares some of the appealing features of the WZ formulation but leads to rather implausible implications for location studies (see, however, List 2001). “The idea underlying the hurdle formulations is that a binomial probability model governs the binary outcome of whether a count variate has a zero or a positive realization. If the realization is positive, the “hurdle” is crossed, and the conditional distribution of the positives is governed by a truncated-at-zero count data model” (typically, the Poisson or the NBM). Therefore, only the binomial process generates the observed zeros, which amounts to assuming that, with a certain probability  $\psi$ , the population of sites consists of locations in which new firm/plants are created and locations in which they are not. “Like the hurdle models, the idea motivating the WZ specifications is that the conditional distribution of the positives is properly characterized by the truncated-at-zero version of the parent distribution. The probabilities of the positives relative to the probability of the zero outcome, however, are no longer as specified by the parent distribution. Instead, the WZ model specifies that the probability of the zero outcome is additively augmented or reduced by” a mixing parameter  $\psi_j$ , which in the “zero inflated” formulation is further parameterised using e.g. a Logit or Probit model (see Greene 1994 for details).

consist of a cross section of geographical units (municipalities, counties, provinces, states, etc.). However, in these models the probability mass function of the dependent variable is not assumed to be Poisson. Panel data estimators may also control for such heterogeneity, but they can do it maintaining the assumption that the data is Poisson distributed. The downside is that they impose equidispersion. As in the cross-section case, Negative Binomial specifications can cope with this.<sup>11</sup>

Hausman et al. (1984) derive Poisson and Negative Binomial estimators for panel data under alternative assumptions on the stochastic relation between individual effects and covariates.<sup>12</sup> On the one hand, the fixed effects estimator is consistent regardless of the correlation between covariates and effects. This probably explains why most industrial (re)location studies have opted for this type of specification, either using the Poisson model (Papke 1991, Becker and Henderson 2000, List and McHome 2000, Holl 2004a, 2004c, Manjón and Arauzo 2007) or the NBM (Holl 2004a, 2004b, Manjón and Arauzo 2007). On the other hand, the random effects estimator is efficient as long as there is zero correlation between covariates and effects (and not consistent otherwise). This is the assumption made in the FDI studies of Blonigen (1997), Blonigen and Feenstra (1997) and Basile (2004), although only Blonigen (1997) reports supportive evidence from a Hausman test (see also Manjón and Arauzo 2007).

All the previous specifications can easily be estimated by maximum likelihood using available statistical packages (LIMDEP, Stata, etc.). However, in applications the zero-inflated and panel data models may have convergence problems (Cameron and Trivedi 1998). Still, this is the standard approach in location studies, which do not seem to have considered alternative (semiparametric, simulation-based, generalized method of moments) estimation procedures that allow for more flexible specifications and/or are less sensible to misspecification errors (in the model and/or the mixture). It is also interesting to note that the relatively common use of panel data models in

---

<sup>11</sup> "It should be kept in mind, however, that a common reason for such extensions in using cross-section data is to control for unobserved heterogeneity. The longitudinal data methods already control for heterogeneity, and Poisson longitudinal models may be sufficient" (Cameron and Trivedi 1998: 280).

<sup>12</sup> The random effects estimator is explicitly derived from a mixture model, whereas the fixed effects estimator is based on a conditional maximum likelihood approach. Notice, however, that in terms of model specification the assumption that the effects are "fixed" can be seen as a degenerate case of a mixture model. Therefore, both specifications can somehow be seen as examples of mixture models.

studies using CDM contrasts with their limited use in studies using DCM —see, however, Guimarães et al. (2004) and Holl (2004c).

### **3. Results**

In this section we present the main findings of those empirical studies that resort to the econometric methods described in the previous section to analyse the determinants of industrial location decisions. However, we will not analyse all the explanatory variables proposed in the literature (see Tables 1 and 2 for details). Rather, we will concentrate on the most representative from a theoretical point of view.

The location of economic activity has been analysed from a wide range of theoretical perspectives. However, these can be grouped into three main categories: neoclassical, institutional and behavioural (Hayter 1997). Neoclassical theories consider that rational and perfectly informed agents choose the optimal locations on the grounds of profit-maximizing or cost-minimizing strategies. Thus, neoclassical determinants are profit- or cost-driving factors such as agglomeration economies (proxied by population, number of workers, etc.), transport infrastructures (spatial distribution, distance, etc.), technology and human capital. As for the institutional theories, they extend the neoclassical framework by considering that agents decide locations given a network of economic relations (with clients, suppliers, competitors, unions, public administrations, etc.). Accordingly, institutional factors somehow measure how these relations affect location decisions. Lastly, behavioural theories emphasise the role of individual preferences. Thus, while neoclassical and institutional theories stand on factors that are “external” to the firm, behavioural factors have an “internal” (size, age, etc.) and “entrepreneurial” (previous experience, residence, etc.) nature.

Next we discuss the sign and statistical significance of the estimated effects that these factors have on the location decisions. However, it is important to stress that in practice the distinction between neoclassical, institutional and behavioural factors is not always clear. In fact, there are factors that may plausibly be attributed to different theories (e.g. wages and pollution can be considered as neoclassical or institutional



factors). Yet this does not invalidate the use of this taxonomy for our descriptive purposes.

### 3.1 Neoclassical factors

*Agglomeration economies* are probably the most studied determinant of industrial location.<sup>13</sup> There is general agreement that the relation between the spatial concentration of the economic activity in a particular location and the degree of attraction that such a location has for new concerns shows an inverted U-shape profile, that is, it is initially positive but, once a certain threshold is reached, it turns into negative, i.e. agglomeration becomes a diseconomy.<sup>14</sup> It is also interesting to note that urbanization economies seem to outweigh industry-specific localisation economies (especially in large cities) and that service agglomeration economies seem to have stronger effects than (industry-level) localisation economies (Head et al. 1995, Guimarães et al. 2000).

*Transport infrastructures* have also been extensively studied. Since a substantial part of firm/plant activities involves moving inputs and outputs, better accessibility to transport infrastructures has been hypothesised to have a positive impact on the location decisions of firms. This has been supported by a number of empirical studies in Belgium (Baudewyns et al. 2000), Spain (Holl 2004c, Arauzo 2005, Alañón et al. 2007), Poland (Cieślik 2005b), Portugal (Holl 2004b) and the U.S. (Coughlin et al. 1991, Friedman et al. 1992, Smith and Florida 1994, Luker 1998, Coughlin and Segev 2000, List 2001). However, the importance of this effect differs across manufacturing

---

<sup>13</sup> Although Marshall (1890) is acknowledged as the first to establish their existence, it was Hoover (1936) who provided the most commonly used empirical implementation of the concept by distinguishing between urbanization economies (derived from the concentration of economic activity as a whole) and localisation economies (derived from the concentration of similar economic activities).

<sup>14</sup> Supportive empirical evidence of the positive effect that the agglomeration economies have on location decisions can be found in e.g. Luger and Shetty (1985), Hansen (1987), Coughlin et al. (1991), Friedman et al. (1992), Woodward (1992), Smith and Florida (1994), Wu (1999), Coughlin and Segev (2000), Guimarães et al. (2000), List (2001), Figueiredo et al. (2002), Gabe (2003), Holl (2004a, 2004b and 2004c) and Arauzo (2005). In contrast, in a recent French FDI study Crozet et al. (2004) find that firms from certain countries (e.g. the Netherlands and Italy) “tend to avoid” firms from the same countries. Moreover, although the identification of the threshold from where diseconomies arise has never been explicitly addressed, the subsequent non-linear effect is usually accounted for by including the square of the agglomeration measure as an additional explanatory variable (see e.g. Arauzo 2005) or by using squared population (see e.g. Viladecans 2004). In any case, it is difficult to compare results across studies because of the use of many different variables to proxy for the agglomeration economies (typically, statistics that capture the spatial density of jobs, people and/or firms).

sectors, which indicates that accessibility requirements may vary with technology and/or demand.

In this respect, Bade and Nerlinger (2000) find that German start-ups in *technology intensive industries* prefer to be located in large agglomerations —see also Egelin et al. (2004) for related evidence from German public research spin-offs. This also seems to be the case for small and medium size biotech firms and large R&D labs in France (Autant-Bernard 2006, Autant-Bernard et al. 2006). More specifically, Arauzo and Viladecans (2008) show that Spanish manufacturing establishments in high-tech industries prefer to be located as close as possible to the centre of the metropolitan area.

Lastly, among the *human capital* characteristics, wages and education have been the most actively explored.<sup>15</sup> First, it has been largely demonstrated that firms/plants tend to avoid areas with higher wages (see e.g. Luger and Shetty 1985, Coughlin et al. 1991, Papke 1991, Friedman et al. 1992, Henderson and Kuncoro 1996, Luker 1998, List 2001, Barbosa et al. 2004 and Basile 2004; see, in contrast, Smith and Florida 1994). Second, most studies tend to conclude that geographical areas that have a higher mean level of education in the (working) population are more attractive (Coughlin et al. 1991, Woodward 1992, Smith and Florida 1994, Coughlin and Segev 2000; see, in contrast, Bartik 1985 and Arauzo 2005). In particular, Audretsch and Lehmann (2005: 1200) find that the number of knowledge-based start-ups clustered around German universities “is positively influenced by the knowledge output of the respective university and the innovative capacity of the region”.

### **3.2 Institutional factors**

Firms/plants are not isolated agents. Rather, they operate within an institutional framework that is likely to exert a certain influence on their location decisions. There is, for example, evidence indicating that the level of unionism negatively affects the likelihood of new business start-up activities in a particular location (Bartik 1985, Woodward 1992). Nevertheless, most of the studies that have considered institutional factors among the set of explanatory variables have concentrated on the actions taken

---

<sup>15</sup> See also Coughlin et al. (1991), Woodward (1992) and Cieřlik (2005b) for evidence on the positive effects of the unemployment level of the geographical area and Friedman et al. (1992) on the negative effects that higher labour productivity seem to have on location decisions.

by the public administrations, in particular, taxes, environmental regulations and incentive programs for new business.

According to the earlier studies on industrial location, the effect of taxation is ambiguous (see e.g. Luger and Shetty 1985): Carlton (1979, 1983) found a non-significant effect of tax levels on location decisions in the U.S., whereas Bartik (1985) found that taxation exerts a moderately negative effect on U.S. states location. In contrast, later FDI studies such as Coughlin et al. (1991), Friedman et al. (1992), Woodward (1992), Devereux and Griffith (1998) and Coughlin and Segev (2000) all reported a negative effect of taxes on the location of foreign firms. More recently, however, Gabe and Bell (2004) have argued that there is a trade-off between taxes and the provision of public goods and services in that high-tax locations remain attractive as long as they spend large sums of money in the provision of public goods and services. They indeed show that high-tax locations are on average more attractive than low-tax locations with a poor provision of public goods and services.

As for the environmental regulations, most investigations have focused on the effects of the U.S. Clean Air Act and its Amendments since 1970 (see Jeppesen et al. 2002 for an overview). However, the evidence is not conclusive. On the one hand, Becker and Henderson (2000) studied Census of Manufacturers data on the location decisions of polluting plants between 1963 and 1992 and concluded that these plants have progressively moved from areas where the air-quality standards had not been attained to areas where these standards had been attained. Similarly, List and McHone (2000: 189) find that the main effect of the laws aimed at limiting pollution levels is to move productive resources from polluted areas to areas that are free of pollution. However, while new firms in pollution-intensive sectors were weakly deterred by more stringent pollution regulation in the early 1980s, they were strongly affected in the late 1980s (“particularly after the EPA and state agencies had an opportunity to experiment and develop a systematic procedure to deal with polluters”). On the other hand, Bartik (1988: 37) “does not find any statistically significant effect of state environmental regulations on the location of new branch plants” and neither do McConnell and Schwab (1990) or Levinson (1996). Also, List (2001) reports analogous findings for the particular case of FDI in California.

Finally, what are the effects that public funds invested in incentive programs aiming to attract new businesses have? The answer is not clear, for the extant evidence on this issue is limited and inconclusive. Lee (2004) shows that these programs have had little effect on relocation decisions in the U.S. (although firms located in states that implemented these programs seem to have benefit in terms of growing employment, capital, and output) and Guimarães et al. (1998) reach analogous conclusions with respect to the regional incentive policies of Puerto Rico in the early 1980s (the aim of these policies was to promote industrial decentralisation and attract new activities to the less developed areas of the island, but they do not seem to have had any statistically significant effect on the location decisions of manufacturing plants). In contrast, in the FDI literature the support of the public administration has been found to be a critical determinant (Friedman et al. 1992, Woodward 1992; see, however, Luger and Shetty 1985).

### **3.3 Behavioural factors**

Behavioural factors have comparatively been less studied than neoclassical and institutional factors. One reason for this is that it is difficult to find appropriate data on entrepreneurs and their personal circumstances. However, the scarce empirical evidence suggests that these factors do matter. Figueiredo et al. (2002) compare location alternatives inside and outside the entrepreneur's area of residence and find that some investors are willing to accept much higher labour costs to take advantage of the potential home-field advantages. In contrast, non-home location choices are strongly driven by neoclassical factors such as agglomeration economies and the proximity to major urban centres. Similarly, Arauzo and Manjón (2004) show that large and small firms follow different location patterns. Whereas large firms seem to be mostly guided by "objective" factors (e.g. markets' characteristics), small firms seem to be mostly guided by the entrepreneur's preferences (see also Carlton 1983).

## **4. Extensions**

### **4.1 Foreign Direct Investments (FDI)**

We have previously argued that, nationality of the production units aside, there are no major differences between FDI and "national" studies of industrial location. In particular, in Section 2 we have shown that both FDI and "national" studies share the

same econometric framework and in Section 3 that they use the same basic set of explanatory variables. Still, there are certain issues that can only arise when one analyses the location decisions of foreign firms/plants, such as, for example, border effects, nationality clustering and the impact of variations in the terms of trade.

Border effects in the location of foreign firms in Poland have been investigated by Cieřlik (2005a). His main result is that Polish regions that shared borders with Eastern EU non-accessing countries in the 1990s (Belarus, Russia and Ukraine) were less attractive to foreign investors than Polish regions that shared borders with EU countries. However, these border effects are often elusive. Basile et al. (2003) analyse multinational firms' location choices in Europe and conclude that these firms make location decisions in terms of multi-country regions rather than countries, i.e. they perceive regions far beyond the actual political boundaries of the European countries as genuine geo-economic entities. Also, Disdier and Mayer (2004) analyse location decisions of French multinational firms in Eastern and Western Europe and find that the distinction between these two geographical areas tends to disappear as the transition process in Eastern countries increases.

As for the existence of clusters of FDI from the same country, most of the evidence comes from Japanese FDI. Smith and Florida (1994) and Head et al. (1995, 1995) in the U.S. and Cheng and Stough (2006) in China, all find that Japanese investors prefer to locate their firms/plants in areas where they find concentrations of previous Japanese investments. The evidence in Europe is more limited, but still consistent with the U.S. findings—see, for example, Crozet et al. (2004) for French evidence.

Lastly, Blonigen (1997) and Blonigen and Feenstra (1997) investigate how variations in the terms of trade affect FDI location decisions in the U.S. These studies find, respectively, that real dollar depreciations relative to the yen and the threat of market protection seem to stimulate FDI investments in the U.S. by Japanese firms. However, these conclusions have not been supported by analogous investigations, so that further research on this topic is clearly needed.

## **4.2 Relocations**

The location of economic activity has been analysed from a wide range of theoretical

perspectives (Hayter 1997). However, none of them has dedicated much effort to investigating the idiosyncrasies of relocations. As Brouwer et al. (2004: 336) point out, “[r]elocation theories are hardly applied and are often treated as a special case of location theories”. In fact, location theories tend to overemphasise/minimise the importance of pull/push factors in relocation decisions, thus concluding that (the forces driving) location and relocation processes are basically the same. Consistent with this view, empirical studies on industrial location do not generally distinguish between new and relocated firms/plants.

However, relocations critically differ from strictly new locations in that they are the outcome of a sequence of decisions taken over the history of the firm/plant. In other words, relocation decisions are taken conditionally upon previous location decisions. One may then argue that the information used to take the decision of relocating an existing firm/plant is not the same as the information used to decide where to locate a new firm/plant. In particular, migrations within the same geographical market are likely to have more and better information about the sites than start-ups. All in all, the opening of new concerns and the relocation of existing concerns are different location processes that should consequently be studied separately (Pellenbarg et al. 2002a, 2002b; Lee 2006).

There is a number of empirical investigations that have considered such a distinction. However, as Mariotti (2005) shows, descriptive statistical methods prevailed until practically the late 1990s. As for the more recent studies providing sounded econometric evidence, there is only a handful that use specifications analogous to those described in Section 2. We shall therefore concentrate on these here.<sup>16</sup>

Among those using DCM, we can further distinguish between those that are interested in the decision of “whether to relocate” and those that are interested in the decision of “where to locate”. Among the former, Lee (2004) aims to assess the impact of U.S. states development incentives on the decision to (re)locate (“Shut Down” and “Stay” being the other choices). Among the latter, Baudewyns et al. (2000) analyse the effect that better public infrastructures have on the (re)location decisions of Belgian firms

---

<sup>16</sup> Relocation studies that do not use specifications analogous to those described in Section 2 include Cooke (1983) and Brouwer et al. (2004), who use probit and logit specifications, respectively, and van Dijk and Pellenbarg (2000), who use ordered logit and probit specifications.

from the city of Brussels and the region of Wallonia, whereas Strauss-Kahn and Vives (2005) use data from U.S. metropolitan areas to discern between the decision of “where to locate” and that of “whether to relocate” by U.S. and non-U.S. companies’ headquarters.

Among those using CDM, Holl (2004a: 665) analyses the determinants of Portuguese plant start-ups and relocations and concludes “that [they] are not attracted by the same set of location characteristics”. In contrast, Manjón and Arauzo (2007) using data on Catalan establishments show that although the determinants of start-ups and relocations are practically the same their partial or marginal effects differ. They also find that locations and relocations are positively, albeit asymmetrically, interrelated.

In light of this evidence it is apparent that relocations by existing firms/plants have comparatively received less attention than strictly new location decisions. However, it seems that the main reason for this is not that the topic does not deserve the effort but that there is a lack of appropriate data. We can consequently foresee an increasing research interest in relocations as long as data collection strategies address this drawback.

### **4.3 Spatial aggregation**

An interesting trend seems to emerge when one carefully examines the territorial unit of analysis used by the studies reviewed in Tables 1 and 2. While the early papers of the 1980s and 1990s generally resorted to large territorial units, such as, for example, metropolitan areas (Carlton 1983, 1979) and U.S. states (Bartik 1985, Schmenner et al. 1987, Coughlin et al. 1991, Papke 1991, Friedman et al. 1992, Head et al. 1995, Levinson 1996), more recent investigations tend to rely on smaller units, such as counties (Smith and Florida 1994, List and McHome 2000, Becker and Henderson 2000, List 2001, Coughlin and Segev 2000, Gabe 2003, Arauzo and Manjón 2004, Guimarães et al. 2004), districts (Bade and Nerlinger 2000) and municipalities (Baudewyns 1999, Baudewyns et al. 2000, Guimarães et al. 2000, Figueiredo et al. 2002, Arauzo and Manjón 2004, Holl 2004a, 2004b, 2004c, Arauzo 2005, Manjón and Arauzo 2007). It seems therefore that the empirical research on industrial location decisions has progressively moved towards the study of geographically disaggregated data.

One may argue that the use of more or less aggregated data is essentially a matter of availability, so that this trend may simply reflect the increasing accessibility to spatially disaggregated data sets. However, there are at least two additional factors worth considering to explain this trend. First, the fact that the New Economic Geography concluded that agglomeration economies arise mainly at local level and decrease as distance increases (Fujita et al. 1999, Fujita and Thisse 2002). Second, advances in the econometric modelling alleviated the computational burden of using such data (see Section 2).

Ultimately, the variety of territorial units reported in Tables 1 and 2 raises the question of which is the correct one. The underlying assumption in most industrial location studies is that the answer to this question is irrelevant for the analysis (or it is given by the unit they use). However, although the effects of aggregation on the statistical inferences of non-linear models like the ones typically used in this literature are unclear, they are unlikely to be harmless. Therefore, the use of an inappropriate territorial unit may result in severely misleading policy implications and, obviously, in biased results and conclusions.

Unfortunately, this is an issue that has received little attention in the literature (see, however, Rosenthal and Strange 2003). Arauzo and Manjón (2004), for example, compare marginal or partial effects for different levels of territorial aggregation in Catalonia (municipalities, “*comarques*” and provinces) and show that location factors do not act uniformly across them. In particular, they conclude that Catalan firms seem to choose between “*comarques*” rather than between municipalities. Along the same lines, Arauzo (2008) compares the results of using data on administrative (municipalities and “*comarques*”) and functional territorial units (travel-to-work areas) of Catalonia. He finds little difference in the determinants of industrial location across different administrative units and only minor difference with respect to functional units.

All in all, this limited evidence suggests that comparing results from several territorial units (cities, counties, regions, etc.) may be a good strategy to empirically assess the impact of aggregation on our conclusions. However, it remains unclear what the level of aggregation effectively used by agents is when it comes to deciding where to locate



a new concern. Given the relevance of the question, this seems a promising area for future research.

## 5. Conclusions

The location of manufacturing firms/plants is a major concern for managers, entrepreneurs and policy makers. Naturally, this is an issue that has also attracted the attention of a number of researchers. However, sound econometric evidence on the determinants of industrial location decisions was not first reported until the early 1980's (Carlton 1979, 1983). This paper provides a critical assessment of the methods and results employed in the myriad of studies that followed (around fifty by our reckoning).

We find two basic specifications in this literature: Discrete Choice Models and Count Data Models. This means that the location decisions of new industrial concerns have been empirically studied from two main perspectives: that of the agent taking the decision and that of the chosen territory. Historically, earlier contributions tend to use DCM whereas CDM are more common in the more recent investigations. However, at the end of the day to use one or the other depends on the data at hand: if the available information refers to firms/plants, then DCM arise as the natural choice while, if the available information refers to the administrative divisions of a geographical area (municipalities, counties, provinces, etc.), then it is CDM which one typically resorts to. What is important to bear in mind is that since the unit of analysis differs between CDM and DCM, so do the inferences that one can extract from them (see, however, Guimarães et al. 2003, 2004).

Explanatory variables in DCM are firm/plant- and/or territory-specific and estimates of their marginal or partial effects offer evidence on how *ceteris paribus* variations in the explanatory variables affect the (conditional) probability of choosing a particular territory. In particular, we find that agglomeration economies, unemployment, education and better transport infrastructures seem to have a positive effect. On the other hand, explanatory variables in CDM are territory-specific and the marginal or partial effects offer evidence of how *ceteris paribus* variations in the explanatory variables affect the (conditional) mean of new locations. In particular, we find that

agglomeration economies and market size tend to provide a significant positive effect, while wages and taxes tend to act in the opposite way.

However, two important caveats apply to this evidence. First, most results have been obtained using rather standard econometric techniques that are already implemented in most commercial statistical packages (LIMDEP-NLOGIT, Stata, etc.). This may have facilitated comparative analyses and the development of certain lines of research, but often at the cost of imposing too strong assumptions on our models and restricting research questions to those addressable within this setting. It is illustrative that little effort has been made to jointly consider firm/plant- and territory-specific factors in DCM. Similarly, mixing strategies for CDM have been limited to those that provide closed-form expressions. Second, some of the reported effects may not be robust to the use of alternative geographical units. In general, it is not clear what effects the spatial aggregation may have on the inferences from such non-linear models. Therefore, it appears that the use of flexible and encompassing specifications using different levels of geographically aggregated (panel) data is a promising line for future research.

Lastly, it is worth noting that, although most investigations relate the definition of the explanatory variables to neoclassical, behavioural and/or institutional factors, the link with the associated location theories is usually weak. This is particularly noticeable in some areas such as, for example, relocation (which lacks specific foundations) and FDI (with respect to international trade theories). Admittedly, this is essentially an empirically-driven literature. Nevertheless, it may be interesting to explore frameworks beyond the basic Random Utility (Profit) Maximisation Model and/or more structural approaches to the location decision problem.

## References

Alañón, Á.; Arauzo, J.M. and Myro, R. (2007): "Accessibility, agglomeration and location". In: J.M. Arauzo and M. Manjón (eds.), *Entrepreneurship, Industrial Location and Economic Growth*, Edward Elgar.

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

Arauzo, J.M. (2008): "Industrial Location at a Local Level: Comments on the Territorial Level of the Analysis", *Tijdschrift voor Economische en Sociale Geografie - Journal of Economic & Social Geography* **99**: 193-208.

Arauzo, J.M. and Manjón, M. (2004): "Firm Size and Geographical Aggregation: An Empirical Appraisal in Industrial Location", *Small Business Economics* **22**: 299-312.

Arauzo, J.M. and Viladecans, E. (2008): "Industrial Location at the Intra-metropolitan Level: The Role of Agglomeration Economies", *Regional Studies*: forthcoming.

Audretsch, D. and Lehmann, E. (2005): "Does the Knowledge Spillover Theory of Entrepreneurship hold for regions?", *Research Policy* **34**: 1191-1202.

Autant-Bernard, C. (2006): "Where Do Firms Choose to Locate their R&D? A Spatial Conditional Logit Analysis on French Data", *European Planning Studies* **14**: 1187-1208.

Autant-Bernard, C.; Mangematin, V. and Massard, N. (2006): "Creation of Biotech SMEs in France", *Small Business Economics* **26**: 173-187.

Bade, F.J. and Nerlinger, E.A. (2000): "The spatial distribution of new technology-based firms: Empirical results for West-Germany", *Papers in Regional Science* **79**: 155-176.

Barbosa, N.; Guimarães, P. and Woodward, D. (2004): "Foreign firm entry in an open

economy: the case of Portugal”, *Applied Economics* **36**: 465-472.

Bartik, T.J. (1985): “Business Location Decisions in the U.S.: Estimates of the Effects of Unionization, Taxes, and Other Characteristics of States”, *Journal of Business and Economic Statistics* **3**: 14-22.

Bartik, T.J. (1988): “The effects of environmental regulation on business location in the United States”, *Growth and Change* **19**: 22-44.

Basile, R. (2004): “Acquisition versus greenfield investment: the location of foreign manufacturers in Italy”, *Regional Science and Urban Economics* **34**: 3-25.

Basile, R.; Castellani, D. and Zanfei, A. (2003): *Location Choices of Multinational Firms in Europe: The Role of National Boundaries and EU Policy*. Centro Studi Luca d’Aglia.

Baudewyns, D. (1999): “La localisation intra-urbaine des firmes: une estimation logit multinomiale”, *Revue d’Économie Régionale et Urbaine* **5**: 915-930.

Baudewyns, D.; Sekkat, K. and M. Ben-Ayad (2000): “Infrastructure publique et localisation des entreprises à Bruxelles et en Wallonie”. In M. Beine and F. Docquier (eds.), *Convergence des régions: cas des régions belges*, De Boeck.

Becker, R. and Henderson, V. (2000): “Effects of Air Quality Regulations on Polluting Industries”, *Journal of Political Economy* **108**: 379- 421.

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. and Feenstra, R. (1997): *Protectionist Threats and Foreign Direct Investment*. In R. Feenstra (Ed.), *Effects of U.S. Trade Protection and Promotion Policies*, University of Chicago Press.

Bradlow, E.T; Bronnenberg, B.; Russell, G.J.; Arora, N.; Bell, D.R.; Duvvuri, S.D.; ter

Hofstede, F.; Sismeiro, C.; Thomadsen, R. and Yang, S. (2005): "Spatial Models in Marketing", *Marketing Letters* **16**: 267-278.

Brouwer, A.E.; Mariotti, I. and van Ommeren, J.N. (2004): "The Firm Relocation Decision: An Empirical Investigation", *The Annals of Regional Science* **38**: 335– 347.

Bunch, D.S. (1991): "Estimability in the Multinomial Probit Model", *Transportation Research B* **25**: 1-12.

Cameron, A.C. and Trivedi, P.K (1998): *Regression analysis of count data*, Cambridge University Press.

Carlton, D. (1979): "Why new firms locate where they do: An econometric model". In: Wheaton, W. (ed.), *Interregional Movements and Regional Growth*, The Urban Institute, Washington.

Carlton, D. (1983): "The location and employment choices of new firms: An econometric model with discrete and continuous endogenous variables", *Review of Economics and Statistics* **65**: 440–449.

Cheng, S. and Stough, R.R. (2006): "Location decisions of Japanese new manufacturing plants in China: a discrete-choice analysis", *Annals of Regional Science* **40**: 369-387.

Cieřlik, A. (2005a): "Location of foreign firms and national border effects: the case of Poland", *Tijdschrift voor Economische en Sociale Geografie - Journal of Economic & Social Geography* **96**: 287-297.

Cieřlik, A. (2005b): "Regional characteristics and the location of foreign firms within Poland", *Applied Economics* **37**: 863-874.

Cooke, T.W. (1983): "Testing a model of intraurban firm relocation", *Journal of Urban Economics* **13**: 257-282.

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

Coughlin, C.C.; Terza, J.V. and Arromdee, V. (1991): "State characteristics and the location of foreign direct investment within the United States", *The Review of Economics and Statistics* **73**: 675-683.

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

Deveraux, M. P. and Griffith, R. (1998): "Taxes and the Location of Production: Evidence from a Panel of US Multinationals", *Journal of Public Economics* **68**: 335–367.

Disdier, A. and Mayer, T. (2004): "How Different is Eastern Europe? Structure and Determinants of Locational Choices by French Firms in Eastern and Western Europe", *Journal of Comparative Economics* **32**: 280-296.

Egelin, J.; Gottschalk, S. and Rammer, C. (2004): "Location Decisions of Spin-offs from Public Research Institutions", *Industry and Innovation* **11**: 207-223.

Figueiredo, O.; Guimarães, P. and Woodward, D. (2002): "Home-field advantage: location decisions of Portuguese entrepreneurs", *Journal of Urban Economics* **52**: 341-361.

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

Fujita, M. and Thisse, J.F. (2002): *Economics of Agglomeration. Cities, Industrial Location and Regional Growth*, Cambridge University Press.

Fujita, M.; Krugman, P. and Venables, A. J. (1999): *The Spatial Economy*, MIT Press.

Gabe, T. (2003): "Local industry agglomeration and new business activity", *Growth and Change* **34 (1)**: 17-39.

Gabe, T. and Bell, K.P. (2004): "Tradeoffs Between Local Taxes and Government Spending as Determinants of Business Location", *Journal of Regional Science* **44 (1)**: 21-41.

Geweke, J. (1996): "Monte Carlo Simulation and Numerical Integration" in H.M. Amman, D.A. Kendrick and J. Rust (eds.), *Handbook of Computational Economics*, Elsevier Science.

Gourieroux, C.; Monfort, A.; Trognon, A. (1984): "Pseudo Maximum Likelihood Methods: Applications to Poisson Models", *Econometrica* **52**: 701-720.

Greene, W.H. (1994): "Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models", *Discussion Paper EC-94-10*, Department of Economics, New York University.

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

Guimarães, P.; Figueiredo, O. and Woodward, D. (2003): "A Tractable Approach to the Firm Location Decision Problem", *Review of Economics and Statistics* **85**: 201-204.

Guimarães, P.; Figueiredo, O. and Woodward, D. (2004): "Industrial Location Modeling: Extending the Random Utility Framework", *Journal of Regional Science* **44**: 1-20.

Guimarães, P.; Rolfe, R.J. and Woodward, D. (1998): "Regional Incentives and Industrial Location in Puerto Rico", *International Regional Science Review* **21**: 119-138.

Hansen, E.R. (1987): "Industrial location choice in São Paulo, Brazil: A nested logit model", *Regional Science and Urban Economics* **17**: 89–108.

Hausman, J.; Hall, B.H.; Griliches, Z. (1984): "Econometric Models for Count Data with an Application to the Patents-R&D Relationship", *Econometrica* **52**: 909-938.

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

Head, K; Ries, J. and Swenson, D. (1995): "Agglomeration benefits and location choice: Evidence from Japanese manufacturing investments in the United States", *Journal of International Economics* **38**: 223-247.

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

Henderson, V. and Kuncoro, A. (1996): "Industrial Centralization in Indonesia", *World Bank Economic Review* **10**: 513-540.

Hensher, D.A. (1986): "Sequential and Full Information Maximum Likelihood Estimation of a Nested Logit Model", *Review of Economics and Statistics* **68**: 657-667.

Hensher, D.A. and Greene, W.H. (2002): "Specification and Estimation of the Nested Logit Model: Alternative Normalisations", *Transportation Research B* **36**: 1-17.

Holl, A. (2004a): "Start-ups and Relocations: Manufacturing Plant Location in Portugal", *Papers in Regional Science* **83**: 649-668.

Holl, A. (2004b): "Transport Infrastructure, Agglomeration Economies, and Firm Birth. Empirical Evidence from Portugal", *Journal of Regional Science* **44**: 693-712.



Holl, A. (2004c): "Manufacturing Location and Impacts of Road Transport Infrastructure: Empirical Evidence from Spain", *Regional Science and Urban Economics* **34**: 341-363.

Hoover, E.M. (1936): "The Measurement of Industrial Location", *The Review of Economics and Statistics* **18**: 162-171.

Isard, W. (1956): *Location and Space Economy*, MIT Press.

Jeppesen, T.; List, J.A. and Folmer, H. (2002): "Environmental Regulations and New Plant Location Decisions: Evidence from a Meta-Analysis", *Journal of Regional Science* **42**: 19-49.

Kim, H.; Waddell, P.; Shankar, V.N. and Ulfarsson, G.F. (2008): "Modeling Micro-Spatial Employment Location Patterns: A Comparison of Count and Choice Approaches", *Geographical Analysis* **40**: 123-151.

Kogut, B. and Chang, S.J. (1991): "Technological Capabilities and Japanese Foreign Direct Investment in the United States", *The Review of Economics and Statistics* **73**: 401-413.

Lee, Y. (2004): "Geographic Redistribution of US Manufacturing and the Role of State Development Policy", *Working Paper* 04-15, Federal Reserve Bank of Cleveland.

Lee, Y. (2006): "Relocation Patterns in US Manufacturing", *Working Paper* 06-24, Federal Reserve Bank of Cleveland.

Levinson, A. (1996): "Environmental Regulations and Manufacturers' Location Choices: Evidence from the Census of Manufacturers", *Journal of Public Economics* **62**: 5-29.

List, J.A. (2001): "US county-level determinants of inbound FDI: evidence from a two-step modified count data model", *International Journal of Industrial Organization* **19**: 953-973.

List, J.A. and McHone, W.W. (2000): "Measuring the effects of air quality regulations on "dirty" firm births: Evidence from the neo and mature-regulatory periods", *Papers in Regional Science* **79**: 177-190.

Luger, M.I. and Shetty, S. (1985): "Determinants of Foreign Plant Start-ups in the United States: Lessons for Policymakers in the Southeast", *Vanderbilt Journal of Transnational Law* **18**: 223-245.

Luker, B. (1998): "Foreign Investment in the Nonmetropolitan U.S. South and Midwest: A Case of Mimetic Location Behavior?", *International Regional Science Review* **21**: 163-184.

Manjón, M. and Arauzo, J.M. (2007): "Locations and Relocations: Modelling, Determinants, and Interrelations", *Working Paper 6-2007 Department of Economics URV*.

Mariotti, I. (2005): *Firm relocation and regional policy*, Utrecht / Groningen: Department of Spatial Sciences (University of Groningen), Netherlands Geographical Studies 331.

Marshall, A. (1890): *Principles of Economics*, MacMillan.

McCann, P. and Mudambi, R. (2004): "The Location Behavior of the Multinational Enterprise: Some Analytical Issues", *Growth and Change* **35**: 491-524.

McCann, P. and Sheppard, S. (2003): "The Rise, Fall and Rise Again of Industrial Location Theory", *Regional Studies* **37**: 649-663.

McConnell, V.D. and Schwab R.M. (1990): "The impact of environmental regulation on industry location decisions: The motor vehicle industry", *Land Economics* **66**: 67-81.

McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behaviour", in P. Zarembka (ed) *Frontiers in econometrics*, Academic Press.

McFadden, D. (1978): "Modelling the Choice of Residential Location" in A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull (eds.), *Spatial Interaction Theory and Planning Models*, North Holland.

McFadden, D. (2001): "Economic Choices", *American Economic Review* **91**: 351-378.

Mullahy, J. (1986): "Specification and Testing of Some Modified Count Data Models", *Journal of Econometrics* **33**: 341-365.

Mullahy, J. (1997): "Heterogeneity, Excess Zeros, and the Structure of Count Data Models", *Journal of Applied Econometrics* **12**: 337-350.

Papke, L. (1991): "Interstate business tax differentials and new firm location", *Journal of Public Economics* **45**: 47-68.

Pellenbarg, P.H.; van Wissen, L.J.G. and van Dijk, J (2002b): "Firm Migration", in P. McCann (ed.), *Industrial Location Economics*. Edward Elgar Publishing.

Pellenbarg, P.H.; van Wissen, L.J.G. and van Dijk, J. (2002a): "Firm Relocation: State of the Art and Research Prospects", *SOM Research Report 02D31*, University of Groningen.

Rosenthal, S.S. and Strange, W.C. (2003): "Geography, industrial organization and agglomeration", *Review of Economics and Statistics* **85**: 377-393.

Schmenner, R.; Huber, J. and Cook, R. (1987): "Geographic differences and the location of new manufacturing facilities", *Journal of Urban Economics* **21**: 83-104.

Seim, K. (2006): "An empirical model of firm entry with endogenous product-type choices", *Rand Journal of Economics* **37**: 619-640.

Shukla, V. and Waddell, P. (1991): "Firm location and land use in discrete urban space. A study of the spatial structure of Dallas-Fort Worth", *Regional Science and Urban Economics* **21**: 225-253.

Smith, D.F. and Florida, R. (1994): "Agglomeration and Industrial Location: An Econometric Analysis of Japanese-Affiliated Manufacturing Establishments in Automotive-Related Industries", *Journal of Urban Economics* **36**: 23-41.

Strauss-Kahn, V. and Vives, X. (2005): "Why and Where Do Headquarters Move?", Discussion Paper No. 5070, Centre for Economic Policy Research.

Strotmann, H. (2007): "Entrepreneurial Survival", *Small Business Economics* **28**: 87-104.

Train, K. (2003): *Discrete Choice Methods with Simulation*. Cambridge University Press.

van Dijk, J. and Pellenbarg, P.H. (2000): "Firm Relocation Decisions in the Netherlands: An Ordered Logit Approach", *Papers in Regional Science* **79**: 191-219.

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

Woodward, D. (1992): "Locational determinants of Japanese manufacturing start-ups in the United States", *Southern Economic Journal* **58**: 690-708.

Wooldridge, J.M. (2002): *Econometric Analysis of Cross Section and Panel Data*, The MIT Press.

Wu, F. (1999): "Intrametropolitan FDI firm location in Guangzhou, China: A Poisson and negative binomial analysis", *Annals of Regional Science* **33**: 535-555.

**Table 1: Location studies using Discrete Choice Models (DCM).**

Studies	Spatial unit (Country)	Period	Industry level (firm/plants)	Main determinants	Specification
Arauzo and Manjón (2004)	Counties and provinces (Catalonia, Spain)	1987-1996	Manufacturing sector (plants)	Establishment size (+). Urbanization economies (+). Urbanization diseconomies (-). Population density (-).	CL and ML
Autant-Bernard (2006)	Regions (France)	1995-2001	R&D labs (firms)	Agglomeration economies (+). Knowledge spillovers (+). Academic research (-).	CL
Bartik (1985, 1988)	States (USA)	1972-1978	Manufacturing sector (plants)	Land area (+). Unionisation level (-). Corporate tax rate (+). Manufacturing activity (+). Environmental regulation (0).	CL and NL
Basile et al. (2003)	Regions (8 EU countries)	1991-1999	All sectors (firms)	Cohesion funds countries (+). Skilled workforce (+). Taxes on labour (-). Corporate tax rates (+). Unemployment rates (+). Agglomeration economies (+). Agglomeration of foreign firms (+)	NL
Baudewyns et al. (2000)	Brussel's municipalities and "arrondissements" of Valonia (Belgium)	1981-1991 and 1990-1994	Manufacturing and services sectors(firms)	Transport infrastructures (+). Agglomeration economies (+).	CL
Baudewyns (1999)	Brussel's municipalities (Belgium)	1981-1991	Manufacturing and services sectors(firms)	Transport infrastructures (+). Agglomeration economies (+).	ML
Carlton (1979, 1983)	SMSA (USA)	1967-1971	3 Manufacturing sectors(plants)	Agglomeration economies (+). Firm size (+). Energy price (-). Taxes (0) . Public policies supporting new firms (0).	CL
Cheng and Stough (2006)	Cities (China)	1997-2002	Manufacturing sectors (Japanese FDI plants)	Japanese plants agglomeration (+). Policy incentives to attract FDI (+). Land costs (-). Labour costs (+)	CL
Coughlin et al. (1991)	States (USA)	1981-1983	Manufacturing sectors (plants)	Per capita income (+) . Agglomeration economies(+). Wages (-). Unemployment (+). Unionisation (+). Transport infrastructures (+). Economic promotion (+).	CL
Crozet et al. (2004)	Departments (France)	1985-1995	206 sectors (FDI firms)	Agglomeration effects (+). Number of competitors (+). Proximity to the home market (+). Wages (-)	CL and NL
Disdier and Mayer (2004)	EU and CEE countries.	1980-1999	All sectors (French FDI firms)	Market size (+). Agglomeration economies (+). Institutional quality (+). Distance between France and the host country (-). Wages (-). GDP (+). GDP per Capita (-). Previous existence of French firms (+). Unemployment (+).	CL and NL
Figueiredo et al. (2002)	Distritos (Portugal)	1995-1997	Overall industry(plants)	Localization economies (+). Urbanization economies (+). Labour costs (-). Land costs (-). Distance to urban areas (-).	CL
Friedman et al. (1992)	States (USA)	1977-1988	Manufacturing sector (plants)	Market size (+). Transport infrastructure (+). Wages (-). Local taxes (-). Unionisation (+). Unemployment (+). Productivity (+). Promotional subsidies (+).	CL
Guimarães et al. (2000)	"Concelhos" (Portugal)	1985-1992	Manufacturing sector (plants)	Agglomeration economies (+).	CL
Guimarães et al. (1998)	Municipalities (Puerto Rico)	1979-1986	Manufacturing sector (plants)	Agglomeration economies (+). Main highway distance (-). Distance to the capital (-). Population density (-).	NL
Hansen (1987)	Cities (São Paulo state, Brasil)	1977-1979	Manufacturing sector (plants)	Urbanization economies (+). Localization economies (+). Distance to the State core (-). Wages (-). Land prices (-).	NL
Head et al. (1995)	States (USA)	1980-1989	Manufacturing sectors (Japanese FDI plants)	Localization economies (+). Agglomeration economies (+).	CL
Head et al. (1999)	States (USA)	1980-1992	Manufacturing sectors (Japanese FDI plants)	Agglomeration economies (+). State income (+). Unionisation (-). Wages (+). Taxes (-). Foreign trade zones (+). Labour and capital subsidies (+).	CL
Henderson and Kuncoro (1996)	"Kabupatens" (Java, Indonesia)	1980-1986	Non-food manufacturing (plants)	Local demand (+). Wages (-). Own-industry employment (+). Industrial diversity (+). Distance to urban areas (-).	CL
Levinson (1996)	States (USA)	1983-1987	Industry sector (plants)	Manufacturing climate (+). Existing plants (+). Unionisation (-). Roads (+).	CL
Luger and Shetty (1985)	States (USA)	1979-1983	Drugs, machinery and motor industry (FDI plants)	Agglomeration economies (+). Wages (-). Skills (+). Public policies supporting FDI (+).	ML

**Table 1 (cont.): Location studies using Discrete Choice Models (DCM).**

Studies	Spatial (Country)	unit	Period	Industry (firm/plants)	level	Main determinants	Specification
Luker (1998)	Counties (USA)		1974-1986	Manufacturing plants	sector (FDI)	South location dummy (+). Investment climate (+). Wages (-). Population (+). Distance to highways (-).	CL
McConnell and Schwab (1990)	Counties (USA)		1973-1982	Motor industry (plants)		Attainment of ozone standards (+). Urbanisation economies (+). Taxes (-).	CL
Schmenner et al. (1987)	States (USA)		1970-1980	Manufacturing (plants)	sectors	Unionism (-). Wages (+,-). Education (-). Energy costs (+). Benefits and expenditures (-). Population density (+,-).	ML
Strauss-Kahn and Vives (2005)	Counties (USA)		1996-2001	All sectors (headquarters)		Airport facilities (+). Corporate taxes (-). Wages (-). Business services (+). Agglomeration of headquarters (+).	NL
Woodward (1992)	Counties and States (USA)		1980-1989	Manufacturing (Japanese plants)	sectors	Manufacturing agglomeration (+). Population density (+). Interstate connection (+). Productivity (+). Educational attainment (+). Poverty (-). Unemployment (-). Land area (+).	CL

**Source: own elaboration.**

Notes: (+) and (-) means that the associated coefficient is positively and negatively statistically significant, respectively. CL, ML and NL stand for Conditional Logit, Multinomial Logit and Nested Logit, respectively.

**Table 2: Location studies using Count Data Models (CDM)**

Studies	Spatial (Country)	unit	Period	Industry level (firm/plants)	Main determinants	Specification
Alañón et al. (2007)	Cities (Spain)		1991-1995	Manufacturing (plants)	Human Capital (+). Agglomeration economies (+). Local value added (+). Distance to main cities (-)	NB
Arauzo (2005)	Municipalities (Catalonia, Spain)		1987-1996	5 manufacturing sectors (plants)	Urbanization economies (+). Urbanization diseconomies (-). Industrial diversity (-). Commuting intensity (-). Industrial employment share (+). Services employment share (+). Distance to county capitals and main cities (-).	PO
Arauzo (2008)	Counties and municipalities (Catalonia, Spain)		2001-2005	Manufactures (plants)	Agglomeration economies (-). Market size (+). Distance to main cities (-). Sectoral specialisation (+)	PO, NB and ZINB
Arauzo and Manjón (2004)	Municipalities and counties (Catalonia, Spain)		1987-1996	Aggregated manufactures (plants)	Urbanization economies (+). Urbanization diseconomies (-). Industrial diversity at the county level (+). Industrial diversity at the municipality level (-). Population density (-). Human capital (-). Localization economies (-).	PO
Arauzo and Viladecans (2008)	Metropolitan areas (Spain)		1992-1996	High-, intermediate-, and low-technology sectors (plants)	Population density (+). Previous entries in the own sector (+). Distance from central city (+,-). Human Capital (+,-).	NB
Autant-Bernard et al. (2006)	Regions (France)		1993-1999	Biotech (SMEs)	Concentration of R&D (-). Concentration of biotech market (+). Overall R&D investment (+). Number of scientific publications (+). Specialization of scientific publications (-).	PO and NB
Bade and Nerlinger (2000)	Districts (West Germany)		1989-1996	New-tech based sectors (firms)	R&D facilities (+). Agglomeration economies (+,-). Number of SMEs (+). District size (+). Population density (+). Wages (+).	NB
Barbosa et al. (2004)	Sectors (Portugal)		1982-1990	25 Sectors (FDI firms)	Sector dimension (+). Capital intensity (-). Foreign penetration (+). Product differentiation (+). Wages (-). Productivity (+).	PO and NB
Basile (2004)	Provinces (Italy)		1998-1999	Aggregated manufactures (FDI firms)	Agglomeration economies (-). Infrastructures (+). Labour costs (-). Unemployment (+).	PO-FE, NB-FE and ZIPO
Becker and Henderson (2000)	Counties (USA)		1963-1992	Manufactures (plants)	Manufacturing employment (+). Wages (-). Attainment of air quality standards (+).	PO-FE
Blonigen (1997)	Sectors (USA)		1975-1992	All sectors (Japanese acquisitions)	Real exchange rate (+). Domestic acquisitions (+). Industry value-added share (+). Japan real GDP growth (+). Japan stock market (+).	NB-RE and NB-RE
Cieřlik (2005a)	Provinces (Poland)		1993-1998	All economic activities(firms)	GDP (+). Wages (-). Human capital (-). Unemployment rates (-). Road infrastructures (+). Rail infrastructures (-). Port infrastructures (+).	NB
Cieřlik (2005b)	Regions (Poland)		1993-1998	All economic activities(firms)	Shared border with EU-members (+). Shared border with non EU-members (-). GDP (+). Wages (-). Schooling (+). Unemployment (-). Road infrastructures (+). Rail infrastructures (-). Port infrastructures (+). Urbanization economies (-). Service agglomeration (+). Industry agglomeration (+).	NB
Coughlin and Segev (2000)	Counties (California, USA)		1989-1994	Aggregated manufactures (plants)	Regional dummies (-). Population size dummies (+). Human capital (+). Wages (-). Taxes (-). Road infrastructures (+). Black population (+).	NB
Egeln et al. (2004)	Regions (Germany)		1996-2000	Knowledge intensive sectors (spin-offs)	Population (+). Travel time to airport (-). Human Capital (+). Purchasing power per inhabitant (+). Unemployment (-). Share of employees in R&D sectors (+).	NB
Gabe (2003)	Counties (Maine, USA)		1996-1999	Manufacturing industries(firms)	Establishments (+). Locational Quotient (+). Taxes (-). Government spending (+). Population (+). Industry growth (+). Entry costs (-).	ZIPO
Gabe and Bell (2004)	Counties (Maine, USA)		1993-1995	Manufacturing industries (plants)	Agglomeration economies (+). Subsidies for public education (-). Other local public investments (+). Existence of a high school (+). Distance to main highways (-).	PO and NB
Guimarães et al. (2004)	Counties (USA)		1989-1997	Manufactures (plants)	Labor costs (-). Land costs (-). Taxes (-). Market size (+). Localization economies (+). Urbanization economies (+)	PO
Holl (2004a)	Municipalities (Portugal)		1986-1997	Aggregated manufactures(plants)	Population (+). Road infrastructures (+). Sectoral diversity (+). Wages (-). Human capital (-)	PO-FE and NB-FE

**Table 2 (cont.): Location studies using Count Data Models (CDM)**

Studies	Spatial unit (Country)	Period	Industry level (firm/plants)	Main determinants	Specification
Holl (2004b)	Municipalities (Portugal)	1986-1997	12 manufacturing sectors, 9 service sectors and building (plants)	Population (+). Road infrastructures (+). Sectoral diversity (+). Sectoral specialisation (-) . Wages (-). Human capital (+).	NB-FE
Holl (2004c)	Municipalities (Spain)	1980-1994	10 manufacturing sectors (plants)	Population (+). Road infrastructures (+). Supplier accessibility (+). Demand accessibility (-).	PO-FE
Kogut and Chang (1991)	Sectors (USA)	1976-1987	Manufacturing sectors (new plants, acquisitions and joint ventures)	R&D investment (+). Firm concentration (-). Export restrictions (+).	NB
List (2001)	Counties (California, USA)	1983-1992	Aggregated manufactures (plants)	Population density (+). Agglomeration economies (+). Market size (+). County surface (+). Wages (-).	PO and ZIPO
List and McHome (2000)	Counties (New York, USA)	1980-1990	7 manufacturing sectors (plants)	Environmental regulation (-). Wages (-). Taxes (-).	PO-FE
Manjón and Arauzo (2007)	Municipalities (Catalonia, Spain)	2001-2004	Manufacturing sectors (locations and relocations)	Density of economic activity (+). Urbanization Economies (-). Sectoral diversity (+). Industrial employment (+). County capital (+). Entrepreneurship (+).	ZIPO, NB and ZINB
Papke (1991)	States (USA)	1975-1982	5 manufacturing sectors (plants)	Land price (-). Land Area (-). Taxes (-). Wages (-). Population (+)	PO-FE
Smith and Florida (1994)	Counties (USA)	-	Industry Sectors related to the car industry(Japanese plants)	Distance to Japanese plants (-). Localization economies (+) . Population size (+). Skilled workforce (+). Transport infrastructure (+). Wages (+). Taxes (-). Share of manufacturing workforce (+).	PO and NB
Wu (1999)	Postal zones (Metropolitan Area of Guangzhou, China)	1981-1991	All economic activities (firms)	Distance to the business central district (+) . Population access (+) . Labour market (+). High-ranking hotels (+).	PO and NB

**Source: own elaboration.**

Notes: (+) and (-) means that the associated coefficient is positively and negatively statistically significant, respectively. (ZI)PO and (ZI)NB stand for (Zero Inflated) Poisson and Negative Binomial, respectively. FE and RE stand for Fixed and Random Effects, respectively