Firm Size and Geographical Aggregation: An Empirical Appraisal in Industrial Location*

Josep Maria Arauzo Carod and Miguel C. Manjon Antolin

(*Very preliminary version*)

Comments welcomed.

Facultat de Ciències Econòmiques
(Department of Economics)
Avda. Universitat, 1
43204–Reus (Tarragona)
Spain
E–mail: jmac@fcee.urv.es

* This research was partially funded by the CICYT (SEC2000–0882–C02–02). Usual caveats apply.

Firm Size and Geographical Aggregation: An Empirical Appraisal in Industrial Location

Abstract

This paper asses the relevance of size discrimination and dis–aggregate data in studying the decision to locate a start–up concern. Essentially we compare three econometric specifications using Catalan data: i) a multinomial logit with 4 and 41 alternatives (provinces and *comarques*, respectively) in which firm size is the main covariate; ii) a condicional logit with 4 and 41 alternatives including attributes of the sites as well as interactions size–site; iii) a Poisson with the *comarques* and all the spatial choice set (721 municipalities). The empirical results suggest that ignoring these issues may produce misleading conclusions.

Keywords: Catalonia, cities, industrial location, multinomial response models
JEL: C250, E30, R00, R12

1 Introduction

The choice of the site for a start–up establishment hinges on many factors. In particular, the industrial location literature has focussed on the economies of the territory. Aspects related to the population (e.g., employement and density), the human capital (e.g., qualification) and the incumbents (e.g., sectorial specialization) have been widely used in empirical studies. However, the size of the new comer has comparatively received little attention (Carlton 1979, 1983). Large and small–medium firms decide on the basis of rather different criteria. In the former the election process seems to be built on objective reasons (e.g., external consultants), somehow reflecting the pros and cons of the potential sites. In the latter the arguments tend to be more subjective. As a matter of fact, they are usually associated with personal characteristics of the entrepreneur such as geographical origin, previous experience in the sector and financial status[1].

The sites can be broadly defined as hinterlands outlined by local markets. But in practice this definition is of little use. Rather, researchers resort to administrative units such as states, regions, provinces, counties and the like. This might cause a certain inconsistency with regard to the theoretical framework (Fujita *et al.* 1999), albeit it is very convenient for at least two reasons. First, this is the format in which official statistics are available. Second, the computation time of the likelihood functions increases exponentially with the number of choices. From this point of view, data aggregation is a methodological device that helps to sort out certain econometric problems. However, the "fallacy of composition" observed by McFadden (1974: 134) could bring about drawbacks in the estimation.

This paper aims to test empirically the relevance of these issues in the analysis of industrial location. The discusssion is based on the results obtained from data on start–up establishments in Catalonia[2]. The probability of being located in a particular site is initially addressed in a multinomial logit specification. The main difference with previous studies is that the size of the new establishment is included as an explanatory variable. Choices are given by the Catalan provinces (Barcelona, Tarragona, Lleida and Girona) and the 41

[1] See, for instance, Mueller and Morgan (1962), Johnson and Cathcart (1979), Chapman and Walker (1991), Cotorruelo and Vázquez (1997), Figueiredo and Guimarães (1999) and Meester (2000).

[2] Catalonia is an Spanish administrative region (*comunidad autónoma*) in the northeastern part of Spain. The population is about 6 million people (15% of the Spanish population) and the extension is 31895 km^2. Catalan GDP is approximately 19% of the Spanish GDP. The data employed in this study come from the *Institut d'Estadística de Catalunya* (IDESCAT, the Catalan Statistical Institute) and the *Registro de Establecimientos Industriales* (REI, the Spanish Industrial Establishments Register). The period of analysis is 1987 to 1996.

comarques[3]. In a second stage we estimate a conditional logit model in which the covariates are attributes of the choices as well as interaction terms betwen size variables and dummies for the choices. Both provinces and comarques are employed as alternative sets of choices[4]. Lastly, results are compared with a Poisson model for the *comarques* and the 721 municipalities of the sample. Here the explanatory variables are exclusively attributes of the sites. This approach has recently been proposed by Guimarães *et al.* (2000a), but to our knowledge there are not empirical applications of the kind presented here.

In our view this strategy enable us: i) to examine the robustness of the size coefficient obtained in the first stage of the procedure; ii) to show the effects of geographical aggregation on the estimates. However, this is not a purely technical exercise. There are important economic policy concerns behind these two issues. For example, how effective are the incentives provided to large firms to guide their location decisions? Would not be better to provide these incentives only to small concerns? Should small (large) firms be the main interest of local (regional) governments? Would not be better to cope with entrepreneurship in a broad sense?

The paper is organised in the following way. Next section briefly reviews the empirical literature on industrial location. Later we discus the role of size in industrial location (section 3) and provide insights into the aggregation problem (section 4). Section 5 presents the results of statistical and econometric tests on the Catalonian municipalities, *comarques* and provinces. Last section summarises the main conclusions.

2 The determinants of industrial location: an overview

The conditional logit model proposed originally by McFadden (1974) is the most popular specification in the industrial location literature. Stemming from a profit maximisation program the probability of a establishment being located in a particular site is derived as a funtion of the choice characteristics and a stochastic component. The main limitation of

[3] *Comarques* are territorial units formed by adjacent municipalities belonging to one of the 4 Catalan provinces. There are 11 comarques in Barcelona, 8 in Girona, 12 in Lleida and 10 in Tarragona (41 in total). The average surface and population of Catalan *comarques* are, respectively, 781 km^2 and 145.000 inhabitants (90.000 if we do not consider the city of Barcelona and the *Barcelonès*, Barcelona's *comarca*).

[4] Following Green (2000: cap. 19) it is useful to distinguish between aspects specific to the individual (the characteristics of the establishment) and attributes of the choices (sites) that may also vary accross the individuals. In the multinomial logit model "data are individual specific" whereas in the conditional logit "data consist of choice–specific attributes instead of individual–specific characteristics".

this approach is the independence of irrelevant alternatives axiom[5]. Under mild asymptotic conditions this model provides efficient estimators normally distributed (McFadden 1984).

Early applications for the US can be found in Carlton (1979, 1983) and Bartik (1985). Carlton's seminal work is remarkable in that it brings out the link between location and size[6], while Bartik's paper focuses on taxes and the role of trade unions. In Europe, recent studies include analyses of the *communes* of Brussels (Baudewyns 1999) and the Belgian region of Wallonie (Baudewyns *et al.* 2000). In the former urban transportation networks and agglomeration economies are statistically significant variables and so are in the latter transport infraestructures, agglomeration economies and wage levels. Moreover, a related group of studies has focused on the particular case of Foreign Direct Investments. These include, among others, Coughlin *et al.* (1991), Friedman *et al.* (1992) and Woodward (1992) in the USA; and Guimarães *et al.* (2000b) in Portugal[7].

This paper joins to this literature by analysing the determinants of industrial location in Catalonia. Consequently, the explanatory variables in our models do not differ substantially from those cited above except, of course, for data sources and availability constraints. The main differences with previous studies arise from: i) the emphasis on the size of the new establishment; and ii) the discussion on the implications of data aggregation. Next sections treat both issues in more detail.

[5] "[A]pplications of the model should be limited to situations where the alternatives can plausibly be assumed to be distinct and weighed independently in the eyes of each decision–maker", (McFadden 1974: 113). "This assumption seems implausible for bussiness location decission. (...) Yet the conditional logit approach remains attractive because of its computational feasibility compared with other alternative approaches to the discrete choice problem", (Bartik 1985: 16).

[6] There are other studies that have considered size as an explanatory variable but most of them use survey data. See, for instance, Mueller and Morgan (1962) in the USA, Cotorruelo and Vázquez (1997) in Spain and Meester (2000) in the Netherlands and Germany.

[7] The list of determinants worth noting includes: market demand, agglomeration economies, taxes, wage levels, unemployment rate, transportation infraestructures and promotional expenditures (Coughlin *et al.* 1991); access to markets, promotional expenditures, local labour markets and taxation (Friedman *et al.* 1992); regional markets, taxes, unemployment, education levels, agglomeration economies and population concentration (Woodward 1992); and both agglomeration and urbanisation economies (Guimarães *et al.* 2000b).

3 Size matters

Let assume that the location of a new industrial establishment is guided by the maximisation of the expected profits. Under this decision rule, the firm will thoroughly analyse the costs and revenues of all the potential alternatives. That is, for each site an assesment has to be made of the following non–exhaustive list of elements: labour (skills, wages, etc.); infrastructures (transport, communications, etc.); output markets (prices, competition, etc); etc. Besides, one has to allow for both urbanisation and agglomeration economies (Glaeser *et al.* 1992, Henderson *et al.* 1995). Other factors may be at stake here but, in any case, the residence of the entrepreneur seems to be (at best) a minor determinant of the decision.

This last assertion would be only partially true in the light of certain stylised facts. Namely, big concerns are usually owned by business corporations which probably gather many hints on their potential locations. These are the kind of firms that, for example, would do or buy technical reports on the elements listed above (markets, population, etc.). Thus large firms aiming to open a new establishment appear to hold extensive information on many alternative sites. The majority of the small and medium firms, on the other hand, do not have access to these inputs. Rather, it is unlikely that such firms dedicate too much effort to this goal.

In addition, there exists empirical evidence conforming with this line of reasoning:

- A recent study made by Meester (2000) in the Netherlands and Germany sets the location issue directly to the firms. They were required to evaluate different possible sites for their investments. It turns out that the higher the distance from the original town of the firm, the lower the value assigned to the site. Besides, nearby territories were preferred no matter their characteristics. Variables such as agglomeration economies, availability of specific services and industrial environment were not particularly relevant.

- In Portugal Figueiredo and Guimarães (1999) reach parallel conclusions. The econometric tests reveal that the entrepreneurs' geographical origin is a statistically significant variable. Moreover, conventional factors of location (e.g., external economies and market accessibility) remain unaltered in their explanatory power. However, they are weighted differently depending on whether the entrepreneurs where "movers" or "stayers".

- In Spain external economies are found to be among the main determinants of the industrial location (Callejón y Costa 1996). Moreover, according to Costa *et al.* (2000) the

location decision in Spanish firms appears to be guided by different factors depending on the size of the firm. They conclude that large firms enjoy more discretion when it comes to deciding where to locate their establishments. Small concerns, on the other hand, are randomly spread over different sizes of towns.

All in all, there seems to be reasons to argue that size makes a difference in the decision to locate new industrial establisments. Location theory has traditionally emphasised the role played by the territory, but other factors can be introduced in this neoclassical framework (Fujita *et al.* 1999). Given that entrepreneurship is essentially a matter of an individual (or a reduced group of individuals such as a family), the acquired knowledge by personal experiences could eventually become an asset. Alternatively, the decision to stay near home may be driven by a myopic knowledge. In some cases the range of alternatives could be so limited to familiar sites that locations outside the area in which the entrepreneur lives would not be even considered[8].

This does not mean that searching costs will result in smaller establishments systematically choosing inappropriate locations. What at first blush may look as a drawbak is actually irrelevant as long as the expected payoffs of the investments in information are an increasing function in the establishment size. Therefore, the best strategy for the entrepreneur may well be to exploit a locallly specific knowledge that entails lower start–up costs for the nearby locations (Pred 1967). Similarly, one should not conclude that large establishments are more likely to be optimally located. Expected profits always contain a random component and to discount unknown future events is doubtless subject to error.

3.1 Industrial mix

A natural extension of the previous discussion is to argue that many small start–up concerns stem from the experience and skills of former local employees (Johnson and Cathart 1979, Chapman and Walker 1991). If this is the case, the know–how acquired being employed in other firms of the region becomes a pivotal asset. Thus, one would expect that its activities were akin to those common among the incumbents. By the same token, this behaviour would not show up in footloose firms. Large concerns are unlikely to follow such a pattern of correlation because of the complexity in their decision process. External economies, for

[8]Notice that we are not suggesting a complete determinism in the location of smaller establishments. Actually, the characteristics of the environment do affect (or, at least, are conditioning) the rise of these entrepreneurs.

example, can have an influence in the final decision but so can better conditions in taxes and land prices.

Consequently, the sectorial distribution of the entrants should vary along the range of sizes. In particular, small start–up establishments will tend to reproduce the existing industrial mix. Statistically their sectorial distribution (i.e. the proportion of entrants in each sector) should not differ from the existing one. The opposite would apply to large establishments. A simple test on the equality of proportions can ascertain the validity of this hypothesis. Results for the Catalan municipalities are presented in section 5.

4 The aggregation of individual sites

Most of the work in this field use broad territorial units to define the spatial choice set (and, to a certain extent, this paper is not an exception). Several forceful reasons can explain this: poor quality of the data at the local level, absence of official statistics for non–administrative units, and/or computational constraints imposed by the econometric techniques. Nevertheless, it is important to bear in mind that by employing aggregate data we are at risk of loosing worthwhile information. This is indeed a good solution as long as the territorial heterogeneity is not disguised (Bartik 1985).

But the stretch of some geographical areas is likely to give rise to a problem of representativeness. In some cases one may even end up with the puzzling result that heterogeneity is bigger within–sites than between–sites. To illustrate the importance of this point let consider choice sets like, for instance, the American states or the German Länders. It seems doubtful that "California" or "Baden–Württemberg" are seen as potential sites by American/German agents intending to launch a new business, especially by those of small and medium size. As a consequence, conclusions drawn from the statistical analyses are subject to important caveats.

Actually, inappropriate aggregation procedures can affect the methodological consistency of the study. There is general agreement in the literature that driven factors of the industrial location act at *a local level* (see, e.g., Fujita *et al.* 1999). Therefore, their influence becomes weaker the wider the geographical units. It is also well known that spillovers spread beyond administrative borders. Metropolitan areas like Barcelona, London and Milan are not only big cities but centers of urban *continuums* that benefit from agglomeration economies (Glaeser *et al.* 1992, Henderson *et al.* 1995). In general, the characteristics of the data might undermine the economic foundations of the empirical applications.

Strictly speaking, a valid territorial unit would be a hinterland defined by the local input/output markets. These may be constructed, for example, on the basis of local labour markets and with the help of data on commuting (travel–to–work areas). This is how is done, for example, by the British Department of Employment (Coombes *et al.* 1986). However, this strategy is not free from criticisms either. We can think of at least two difficulties: i) the dynamic nature of these phenomena would produce constant variations in the bounds of the unit; ii) on the practical ground, such official statistics are not available in many countries (e.g., Spain).

In this paper the individual sites are defined at the municipality level. This is far from being optimal, although these units should not be very different from those based on local markets[9]. In any case, results are obviously subject to the pros and cons mentioned above. Interestingly, our data sources enable us to work also with aggregated data. Besides municipalities, in Catalonia there are two broader administrative units: the *comarques* (grouping municipalities) and the provinces (grouping *comarques* and/or municipalities). This provides an excellent chance to test empirically the potential effects of aggregation.

5 An application to Catalan municipalities

5.1 Descriptive analysis

In principle, a firm aiming to open an industrial establishment in Catalonia would have to choose among 942 municipalities. But according to the REI the 17.719 establishments created in Catalonia during the period 1987 to 1996 actually spread over 721 municipalities[10]. Barcelona arises as the principal industrial focus. The spatial distribution of the entrants shows that the province of Barcelona concentrates 77.21% of them, the *comarca* of *Barcelonès* 20.61% (followed by *Vallès Occidental* with 16.52% and *Baix Llobregat* with 11.99%, both in the outskirts of the *Barcelonès*) and the city of Barcelona 12.30% (followed by *Terrassa* with 3.13% and *Mataró* with 2.98%, both nearby Barcelona). The total number of people employed was 138.580, of which 76.9% were hired by the smaller establishments (38.0% by those of less than 10 employees and 38.9% by those of 10 to 49 employees) and 17.0% by the

[9] This is one of the insights provided by Coombes *et al.* (1986). Other studies that employ local data are Carlton (1983), Hansen (1987), Baudewyns (1999), Baudewyns *et al.* (2000) and Guimarães *et al.* (2000b).

[10] The municipalities data set has 17.718 observations. The missing value corresponds to a new municipality created during the period of analysis (*L'Ampolla*). As for the establishments, they are grouped using the old Spanish sectorial classification CNAE-74 on the basis of the following categories:

largest ones (more than 100 employees).

[Insert Table 1 about here]

As for the size of the new comers, they were mostly small and medium concerns: 81.7% of them had less than 10 employees and only 0.4% had more than a hundred. In fact, they are usually smaller than the existing producers (see last column in Table 1). The average size of the start–up establishments during the period 1987 to 1996 is about 40% of those forming the 1986 cohort. This is a commonly found pattern in the industrial dynamics literature (Geroski 1995).

A cross–tabulation with the size of the municipality (population) reveals a certain correlation between both variables (Table 1). This can be seen as an indirect evidence of agglomeration economies. However, statistical tests of the equality of means suggest that this is mostly valid for the smaller cases. Small (rural) sites are more likely to receive small concerns regardless their industrial activity[11]. These results largely agree with those obtained by previous studies in Spain (see, e.g., Callejon and Segarra 1998).

Can we observe differences in the sectorial distribution of the entrants dependig on their size? In other words, is there empirical evidence in Catalonia of the industrial mix hypothesis discussed in section 3.1? A positive answer would be given by a temporal regularity in the figures of small concerns and a rather random behaviour in large (footlose) firms. We employ simple parametric tests of the equality of proportions in the sectorial distribution of the employment. In particular, we compare the means of the 1986 cohort with the entrants in 1987, 1988 and 1989. Entrants are grouped according to their size. Sample sizes and

Code	Description	CNAE–93 code
5	Mining	12, 13, 14 and 26
6	Chemicals industries	23 and 24
7	Metallurgy, electrical machinery and apparatus	27, 28, 29, 30, 31, 32 and 33
8	Transport equipment	34 and 35
9	Food products, beverages and tobacco	15 and 16
10	Textiles, leather clothes and tanning	17, 18 and 19
11	Wood, cork and wood furniture	20 and 36
12	Pulp and paper. Publishing and printing	21 and 22
13	Rubber and plastic products. Other manufacture industries.	25

[11] We have analysed the sensitivity of the results to the use of other clusters – e.g., rural (less than 10.000 people), urban (more than 100.000 people) and mixed (10.000 to 100.000 people) – and the nature of the conclusions remained unaltered.

the Central Limit Theorem enable us to asssume normality in the distribution. Results are displayed in Table 2.

[Insert Table 2 about here]

Differences in the proportions of entrants in each sector are statistical significant except for a few cases. The picture is essentially the same along the range of establishment' sizes. From this viewpoint, there is no empirical evidence in Catalonia of a strong link between small start–ups and the existing industrial mix. Thus, former employees in a sector are able to entry with a rather efficient (i.e. big) size. At the same time, entrepreneurs may be successful in sectors in which they do not have previous experience – for instance, after some industrial reorganising. However, it might also be possible that our results are largely explained by the use of aggregated data. Further research is clearly needed to discern the importance of this caveat.

5.2 Econometric models

Let consider a firm aiming to open a new industrial establishment in Catalonia. We denote the expected benefit derived from this election by π_{ij}, being $i = 1, \ldots, 17719$ and $j = 1, \ldots, J$. Notice that $J = 4, 41$ and 721 depending on the degree of aggregation we are dealing with: provinces, *comarques* and municipalities, respectively[12]. Given the nature of this variable we allow for a deterministic component that takes the form of a lineal combination of variables ($z = Z'_{ij}\beta$) and a stochastic part represented by a random variable ($v = \varepsilon_{ij}$). Thus, $\pi_{ij} = Z'_{ij}\beta + \varepsilon_{ij}$. Assuming that the firm follows a maximaxing principle, the election of a site i would be given by the rule $\pi_{ij} > \pi_{ik}, k \neq j$. Let Y_i a random variable that indicates the election effectively made. Thus, the probability that a firm i locates at site j is $P(Y_i = j) = Prob(\pi_{ij} > \pi_{ik}, k \neq j)$. Assuming that ε_{ij} are *i.i.d.* and Weibull distributed it can be proved that

$$P\left(Y_i = j\right) = \frac{\exp\left(Z'_{ij}\beta\right)}{\sum_{j=1}^{J} \exp\left(Z'_{ij}\beta\right)} \tag{1}$$

[12]Note that in the 221 municipalities actually not being chosen the corresponding (categorical, indicator, count) dependent variable will always take a cero value or its equivalent. In practice this means that these alternatives fall out of the probability and their coefficient cannot be identified in the likelihood function. The underlying selection problem is beyond the scope of this paper (see, e.g., Woodward 1992 and Guimarães *et al.* 2000a).

As it is well known, this result is due to McFadden (1974). In principle, the explanatory variables include both aspects specific to the establishment and the site. In maths, $Z_{ij} = [X_{ij}, W_i]$. This is a purely artificial distinction that helps to see (1) as a general specification embracing two different cases (see Green 2000). When the covariates are characteristics of the individuals (i.e. $Z_{ij} = W_i$) then (1) is known as a multinomial logit model:

$$P(Y_i = j) = \frac{\exp\left(W_i'\beta_j\right)}{\sum_{m=0}^{J} \exp\left(W_i'\beta_m\right)}, j = 0, 1, \ldots, J \tag{2}$$

The main limitations of this model in industrial location studies are computational. In fact, difficulties in the calculation of the likelihood function can make the model unfeasible[13]. Nevertheless, this feature may have contributed to the extensive use of aggregate data. Geographical aggregation reduces the number of alternatives in the choice set and thus simplyfies the estimation procedure.

The expression condicional logit model refers to specifications in which the covariates are attributes of the sites:

$$P(Y_i = j) = \frac{\exp\left(X_{ij}'\beta\right)}{\sum_{j=1}^{J} \exp\left(X_{ij}'\beta\right)} \tag{3}$$

The model is not essentially different from (1) except for the fact that $Z_{ij} = X_{ij}$. Aggregate data has also been employed in many studies, although the reasons argued here are more related to the lack of detailed information at the local level. However, the use of such data relies ultimately on the assumption of homogeneity in the sites. If this does not hold, heterogeneity may produce biases in the estimations (see, e.g., Bartik 1985). Moreover, due to the large number of observations computational difficulties are also important. For instance, in the application presented in this study the number of observations rise from the original 17.719 to 70876, 726.479 and 12.756.960 in dealing respectively with provinces, comarques and municipalities[14].

Guimarães *et al.* (2000a) have recently proposed an alternative approach to sort out these problems. Suppose that Y_{ij} are independent Poisson random variables with means

[13] As an illustration, *Intercooled Stata* allows to use a maximum of 50 alternatives and *Limdep* limits the number of parameters in the models to be 150.

[14] An alternative approach to the direct estimation of the conditional logit model was proposed by McFadden (1978**?**). Consistent (although less efficient) estimates can be obtained by defining smaller choice sets based on sampling alternatives**.** Applications of this technique can be found in Hansen (1987), Woodward (1992) and Guimarães *et al.* (2000b).

$\mu_{ij} = \lambda_i + X'_{ij}\beta$. The joint likelihood function of this model can be divided in a marginal likelihood function based on the marginal totals of the corresponding multiway contingency table and a conditional likelihood function formed by the product of independent multinomial distributions (Birch1963, Palmgren 1981). In particular, the estimates of the parameters of interest obtained from the full and the conditional likelihoods are identical. And so is the covariance matrix. As a result, the Poisson model given by

$$P\left(Y_i = j\right) = \frac{e^{-\lambda_i}\lambda_i^j}{j!} \tag{4}$$

is equivalent to the multinomial response model in (3). However, it is important to bear in mind that the nature of the dependent variable is different from the categorical variable used in (2) and the binary variable of (3). Here were are employing a count dependent variable that does not reflect the maximisation–based election process described above. This caveat aside, from an empirical point of view this a simple and appealing procedure to evaluate the effects of aggregation.

5.3 Covariates, estimation and results

As pointed out above, the multinomial logit specification in (2) is apropriate for individual–specific covariates. Unfortunately, our data set does not contain information on the residence of the entrepreneur. This constraint makes impossible to test directly its relationship with the location patterns of small firms. Nevertheless, we can discuss to which extent vary the location patterns with firm size. We use two variables as proxies of size: number of employees (*NE*) and total investment (*INV*, 1986 pts.). Dummies for the Spanish CNAE–74 industrial classification were also introduced as explanatory variables.

[Insert Table 3 about here]

Results from the estimation of the multinomial logit model for provinces (Table 3) show that size clearly matters, meaning here that the amount of investments made by each entrant firm is an statistically significant variable. The probability of provinces other than Barcelona being chosen depends positively on the size of the establishment. Moreover, dummy variables employed to distinguish among different sizes of establishments reveal that the smaller concers (less than 10 workers) are more likely to be located outside Barcelona. This effect is not apparent for the larger establishments. Sectorial variables were also significant.

These conclusions are not fully robust to the aggregation of the sites. As shown in Table 3, estimates using *comarques* as the choice set largely agree with those obtained for

the provinces. That is, most of the cases look like the *Baix Camp*: size matters and is statistically significant for the small establishments. For the sake of simplicity, Table 3 only includes some selected *comarques*. However, it is clear from its contents that the results are mixed. The significance of the dummies of size varies in each of the examples presented: all are significant (*Baix Llobregat*), only the small and large establishments (*Gironès*), only the medium and large (*Vallès Occidental*), and none of them (*Solonès*). These differences may arise, for example, due to the (non controlled) diversity of the industrial mix in the 41 *comarques* of Catalonia. In any case, an important result remains unaltered in all the sites: the amount of the investment is an statistically significant variable.

The conditional logit specification (3) does allow for differences among sites. We have employed as attributes of the sites the following variables: location economies (*LOC5* to *LOC13* = number of workers per km^2 in each industrial sector, 1986); urbanisation economies (*URB* = total number of workers per km^2, 1986); urbanisation dis–economies ($DIS = URB^2$); density of population per km^2 (*DEN*); industrial diversity, measured by a Hirshmann-Herfindahl index (*DIV*); and human capital (*HC* = number of people with medium and high levels of education per km^2)[15]. We have also included interactions between the characteristics of the individuals (*NE*, *INV*) and dummies representing the choices. This a simple device to explore the individual–site dimension.

[Insert Table 4 about here]

[Insert Table 5 about here]

Table 5 displays results for the Poisson specification based on the Case 1 model of Guimarães *et al.* (2000a). This means that the dependent variable is the number of establishments that have chosen a particular site and the explanatory variables are site–specific. Generally speaking results are similar for *comarques* and municipalities. However, there are some differences worth noting.

The existence of urbanisation economies has a positive effect on the location decision, but the impact is more powerful at the *comarques*' level. As expected, urbanisation dis–economies as well as population density act in the opposite way. This can be interpreted in

[15] Location (urbanisation) economies are cost advantages reached by the concentration of similar (different) activities in a site. Urbanisation dis–economies are cost disadvantages caused by an excessive concentration of firms and population in a site (pollution, input prices, etc.). $DIV_j = \sum_{s=5}^{13} h_{sj}^2$, being s the corresponding industrial sector.

the sense that new firms are prone to locate near the incumbents (i.e. in the same *comarca*), albeit at a distance enough to avoid dis–economies of urbanisation (i.e. not in the same municipality). On the one hand, a dense regional environment (in the *comarca*) is welcomed because of the benefits obtained from the surrounding economic activity – e.g., access to markets, skilled labour, variety of suppliers, etc. On the other, low local density (in the municipality) means lower input prices – e.g., land prices. This structure of preferences is consistent with the economic development of an area like Catalonia where, for instance, mobility among municipalities of the same *comarca* is not costly.

Human capital shows a negative coeffient in both geografical areas. This is probably explained by the specialisation of the Catalan manufacturing sector in products/processes that do not demand highly qualified labour. As a consequence, this is not a major location factor for many Catalan firms.

A comparison of the estimates for the location economies suggests that in most sectors jobs density affects negatively the probability of chosing a particular site. This effect, however, is bigger in the *comarques*' specification. That is, sectorial dis–economies are more powerful at the supra-local level (*comarca*) than at the local level (municipality). By the same token, entrants prefer a specialised environment in the *comarca* and a more diversified one at the local level.

6 Conclusions

This paper intends to maintain that an investigation into the determinants of start–ups locations needs to apply a discriminating criterion based on their size. Territorial aggregation is also relevant and may indeed change the nature of the conclusions. Both issues are worth considering in empirical studies aiming to analyse the reasons behind the rise of new entrepreneurs. Otherwise, conclusions and policy recommendations might be misleading.

The literature on industrial location has traditionally focused on the role of territorial factors. However, personal characteristics of the entrepreneur do influence the location decision of some new firms. There exists empirical evidence showing that bigger firms decide their location according to objective reasons, while smaller ones are mostly oriented by the entrepreneur's preferences. In practice the range of alternatives open to the small concerns is frequently reduced to the nearest geographical area.

Relevant factors for the industrial location do not act uniformly over any territorial aggregation. Rather, their influence varies among them. This requires to use appropriate

territorial dis–aggregation in defining the data set. In particular, unavailability of local or regional data could bring about important biases in the results. This suggests that comparing results from several territorial levels (cities, counties, regions, etc.) might be a good strategy in empirical studies. Here we have employed Catalan data and three levels of administrative aggregation: municipalities, *comarques* and provinces.

The econometric test on the determinants of industrial location aim to be consequent with the previous discussion. As an illustration, results from the Poisson specification show that the observed effects of some variables are more apparent in the *comarques* (rather than in the municipalities). This implies that for the entrants the site of reference is mainly the *comarca*, being less crucial differences among several municipalities in the same *comarca*. Analogous insights can be gained from both the multinomial (individual–specific covariates) and the conditional (site–specific covariates) logit specifications.

References

Bartik, T. (1985), 'Business location decisions in the United States: Estimates of the effects of unionization, taxes and other characteristics of states', *Journal of Business and Economic Statistics* **3**, 14–22.

Baudewyns, D. (1999), 'La localisation intra-urbaine des firmes: Une estimation logit multinomiale', *Révue d'Économie Régionale et Urbaine* **5**, 915–930.

Baudewyns, D., Ayad, M. B. & Sekkat, K. (2000), Infrastructure publique et localisation des entreprises à bruxelles et en wallonie, *in* M. Beine & F.Docquier, eds, 'La Politique de Développement Local et L'infrastructure Publique: Bruxelles et Wallonie', Bruxelles, pp. na–na.

Birch, M. (1963), 'Maximum likelihood in three-way contingency tables', *Journal of the Royal Statistical Society, Series B (Methodological)* **25**, 220–233.

Callejón, M. & Costa, M. (1996), 'Geografía de la producción. incidencia de las externalidades en la localización de las actividades en españa', *Información Comercial Española* **754**, 39–49.

Callejón, M. & Segarra, A. (1998), 'Dinámica empresarial, eficiencia y crecimiento industrial en las regiones españolas (1980-1992)', *Revista Asturiana de Economía* **11**, 137–157.

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

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.

Chapman, K. & Warker, D. (1991), *Industrial Location. Principles and Policies*, Basil Blackwell, Oxford.

Costa, M., A.Segarra & Viladecans, E. (2000), 'Business dynamics and territorial flexibility'. Mimeo.

Cotorruelo, R. & Vázquez, A. (1997), Nuevas pautas de localización de las empresas industriales en España, *in* A. Vázquez, G. Garofoli & J. P. Gilly, eds, 'Gran Empresa y Desarrollo Económico.', Ed. Síntesis, Madrid, pp. na–na.

Coughlin, C., Terza, J. & 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.

Figuereido, O. & Guimaes, P. (1999), 'Start-ups domestic location and the entrepreneur's geographical origin'. Mimeo.

Friedman, J., Gerlowski, D. & 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., Krugman, P. & Venables, A. (1999), *The Spatial Economy*, MIT Press, Cambridge.

Geroski, P. (1995), 'What do we know about entry?', *International Journal of Industrial Organization* **13**, 421–440.

Glaeser, E., Kallal, H., Scheinkman, J. & Shleifer, A. (1992), 'Growth in cities', *Journal of Political Economy* **100**, 1126–1152.

Green, W. (2000), *Econometric Analysis*, Prentice Hall Inc.

Guimaraes, P., Figueiredo, O. & Woodward, D. (2000a), 'A tractable approach to the firm location decision problem'. Mimeo.

Guimaraes, P., Figuereido, O. & Woodward, D. (2000b), 'Agglomeration and the location of foreign direct investment in portugal', *Journal of Urban Economics* **47**, 115–135.

Hansen, E. (1987), 'Industrial location choice in sao paulo, brazil: A nested logic model', *Regional Science and Urban Economics* **17**, 89–108.

Henderson, V., Kuncoro, A. & Turner, M. (1995), 'Industrial development in cities', *Journal of Political Economy* **103**, 1067–1090.

Johnson, P. & Cathcart, D. (1979), 'New manufacturing firms and regional development: Some evidence from the northern region', *Regional Studies* **13**, 269–280.

M. Coombes, A. G. & Openshaw, S. (1986), 'An efficient algorithm to generate official statistical reporting areas: The case of the 1984 travel-to-work areas revision in britain', *Journal of the Operational Research Society* pp. 943–953.

McFadden, D. (1974), Analysis of qualitative choice behavior, *in* P. Zarembka, ed., 'Frontiers in Econometrics', Academic Press, pp. 105–142.

McFadden, D. (1984), Econometric analysis of qualitative response models, *in* Z. Griliches & M. Intriligator, eds, 'Handbook of Econometrics (Vol. II)', Elsevier Science Publishers, pp. 1395–1457.

Meester, W. (2000), 'Locational self-preference of firms'. Mimeo.

Mueller, E. & Morgan, J. (1962), 'Location decisions of manufacturing', *American Economic Review* **502**, 204–217.

Palmgren, J. (1981), 'The fisher information matrix for log linear models arguing conditionally on observed expected explanatory variables', *Biometrika* **68**, 563–566.

Pred, A. (1967), *Behavior and Location: Foundations for a Geographic and Dynamic Location Theory (Part I)*, Lund Studies in Geography B 27 (University of Lund).

Woodward, D. (1992), 'Locational determinants of japanese manufacturing start-ups in the united states', *Southern Economic Journal* **58**, 690–708.

Descriptive statistics

Variables	Territ.	Mean	Std. Dev.	Min	Max
NE	mun	7.821.425	6.453.427	0	8190
	com	7.821.152	6.453.246	0	8190
	pro	7.821.152	6.453.246	0	8190
INV	mun	14510.37	141544	0	1.02×10^{-7}
	com	14509.79	141540.1	0	1.02×10^{-7}
	pro	14509.79	141540.1	0	1.02×10^{-7}
URB	mun	1.358.042	2.053.923	0	6.429.297
	com	1.170.136	2.010.973	1.338.583	5.112.324
	pro	139.807	6.373.291	9.312.061	1.743.736
DIS	mun	6062639	1.34×10^{-7}	0	4.13×10^{-7}
	com	5413001	1.06×10^{-7}	1.791.804	2.61×10^{-7}
	pro	23607.66	12514.83	8.671.448	30406.14
DEN	mun	4.737.991	6.601.493	1.279.461	20571.99
	com	3.840.636	6.482.257	4.020.899	16538.62
	pro	470.184	218.146	2.960.036	5.885.512
DIV	mun	.2761981	.1394086	0	1
	com	.2161676	.0773592	.1477097	.6911405
	pro	.1712078	.0068882	.158111	.1938402
CH	mun	2.960.593	5.832.608	0	1.825.794
	com	2.930.235	5.359.409	.2318051	1.344.635
	pro	277.729	1.340.675	1.501.798	3.505.247
LOC5	mun	2.299.596	2.771.662	0	1.376.366
	com	1.583.665	2.180.004	.0007359	5.784.645
	pro	2.938.377	1.388.428	.1355321	3.691.758
LOC6	mun	554.671	8.205.203	0	2.495.256
	com	4.680.637	7.853.865	.0032817	2.005.523
	pro	5.483.033	2.677.193	.0772331	6.933.587
LOC7	mun	1.467.539	1.906.999	0	7.505.859
	com	1.078.268	1.714.461	.0248713	4.424.751
	pro	1.491.604	7.227.981	.5034255	1.883.916
LOC8	mun	4.392.833	8.933.177	0	276.036
	com	4.339.444	7.909.833	0	1.985.015
	pro	4.059.477	2.107.764	.0395216	5.204.261
LOC9	mun	3.544.785	4.397.566	0	1.510.439
	com	2.722.439	427.609	.0240253	1.108.511
	pro	4.444.536	1.764.159	.6021655	5.397.059
LOC10	mun	1.044.991	1.459.979	0	7.837.347
	com	6.650.411	9.201.555	.0081847	243.525
	pro	1.405.351	6.945.787	.3264495	1.781.989
LOC11	mun	2.080.059	2.716.085	0	1.056.229
	com	1.560.481	2.492.108	.0327765	6.439.213
	pro	2.437.446	.9764477	.185428	2.962.174
LOC12	mun	4.902.131	8.164.275	0	2.445.111
	com	4.292.552	7.615.255	0	1.922.186
	pro	4.471.294	2.248.567	.1117952	569.242
LOC13	mun	3.051.371	4.124.075	0	117.184
	com	224.042	3.633.011	.0007386	9.343.567
	pro	3.144.978	1.552.561	.0469144	3.987.582

Table 1. New Establishments in Catalonia (1987–1996):Average Number of Employees.

Sector	Size1	$t-test$	Size2	$t-test$	Size3	$t-test$	Size4	$t-test$	Size5	$t-test$	Size6	Total	Total / 1986 cohort
5	4.72	-1.31^{***}	5.59	0.52	5.26	-1.79^{**}	7.67	1.11	6.08	2.80	3.43	5.28	35.5%
6	6.52	-1.51^{**}	9.93	0.03	9.84	23.27	-1.03	1.17	8.18	-1.00	179.10	25.53	38.9%
7	5.46	-3.26^{*}	8.38	0.66	7.79	-0.93	8.46	0.90	7.80	-0.87	8.44	7.84	49.1%
8	6.70	-2.84^{*}	17.10	0.82	12.10	-0.73	25.37	0.88	4.52	0.17	4.29	10.71	6.7%
9	5.97	1.24	4.18	-1.40^{**}	5.04	-1.21	8.17	1.65	3.96	-2.51^{*}	5.13	5.04	51.6%
10	7.68	-2.40^{*}	10.08	0.46	9.61	-1.18	12.06	1.99	8.07	1.60	7.21	8.84	34.8%
11	3.92	-1.74^{*}	4.58	-0.60	4.81	-0.52	5.44	-0.24	5.75	2.67	3.88	4.68	99.0%
12	6.11	-1.26^{***}	9.35	0.09	9.10	0.15	8.83	0.86	7.47	-1.52^{***}	8.76	8.59	51.0%
13	8.68	-0.81	10.00	1.49	8.08	-1.31^{***}	10.14	2.55	6.22	-0.38	6.70	8.29	30.1%
Total	5.66	-4.98^{*}	7.79	0.38	7.61	-1.96^{**}	9.7	2.08	6.99	-0.93	10.52	7.82	–

$t-test$ is the t statistic for the alternative hypothesis $H_1 : \mu_m^s < \mu_{m'}^s$ when variances are unequal and unknown, being $s = sector\,(5, \ldots 13)\,; m =$ Size1, . . . ,Size5; $m' = m + 1$. *, ** and * denote 1%, 5% and 10% significance. Size1 denotes municipalies with less than 2.001 inhabitants; Size2 between 2.001 and 10.000; Size3 between 10.001 and 50.000; Size4 between 50.001 and 100.000; Size5 between 100.001 and 1.000.000; Size6 more 1.000.000 (Barcelona city). Source: Own calculations with data from REI, IDESCAT and Encuesta Industrial (1986 cohort).

Table 2. Sectorial distribution of the new establishments in Catalonia (1987-1996)

Sector	TOTAL WORKERS	ENT < 10	ENT 10-49	ENT 50-99	ENT > 99	ENT < 10	ENT 10-49	ENT 50-99	ENT > 99	ENT < 10	ENT 10-49	ENT 50-99
	1986	**1987**	**1987**	**1987**	**1987**	**1988**	**1988**	**1988**	**1988**	**1989**	**1989**	**1989**
5	26821	302	168	0	0	367	183	50	0	350	196	0
6	46183	231	67	75	0	141	102	121	0	179	25	65
7	112833	2213	2313	209	793	2227	2347	358	170	2102	1779	380
8	43614	286	131	82	932	225	166	0	205	333	161	60
9	57614	517	164	0	0	502	215	121	0	452	75	0
10	99449	1204	1409	273	0	1190	1652	234	323	1181	1123	151
11	18140	731	387	123	0	702	373	112	0	715	482	124
12	27423	514	335	205	0	444	463	99	112	510	574	0
13	24979	434	401	0	0	354	363	51	130	409	495	0
TOTAL	**457056**	**6432**	**5375**	**967**	**1725**	**6152**	**5864**	**1146**	**940**	**6231**	**4910**	**780**
		z	*z*	***Z***	*z*	*z*	*z*	*z*	*z*	*z*	*z*	*z*
		3,980*	8,527*	7,764*	10,369*	-0,323	8,920*	2,166*	7,655*	0,838	5,573*	6,973*
		17,279*	21,520*	2,421*	13,922*	20,289*	21,218*	-0,509	10,278*	18,894*	22,290*	1,640
		-17,919*	-30,922*	2,215*	-20,440*	-20,759*	-26,993*	-5,137*	4,690*	-16,424*	-18,631*	-15,543*
		13,860*	17,696*	1,124	-62,285*	15,664*	17,452*	10,994*	-12,772*	11,233*	14,905*	1,757
		10,983*	21,061*	11,808*	15,769*	10,456*	20,572*	2,086*	11,642*	12,672*	23,356*	10,605*
		5,871*	-7,864*	-4,872*	21,891*	4,563*	-11,811*	1,098	-9,351*	5,333*	-1,880	1,623
		-29,805*	-12,009*	-13,893*	8,443*	-29,352*	-9,289*	-10,034*	6,233*	-29,784*	-20,722*	-17,008*
		-6,664*	-0,714	-19,832*	10,492*	-3,988*	-6,063*	-3,755*	-7,621*	-7,197*	-16,622*	7,056*
		-4,486*	-6,385*	7,476*	9,985*	-0,990	-2,425*	1,510	-11,254*	-3,785*	-14,096*	6,715*

z is the statistic for the null hypothesis of the equality of proportions under the assumption of Normality in the distributions * denotes 1% significance. Where ENT<10 is establishments with less than 10 employees; ENT 10-49, between 10 and 49 employees; ENT 50-99, between 50 and 99 employees; ENT>99, more than 99 employees.

Source: own calculations from REI (entrants) and Encuesta Industrial (1986 cohort*).*

Table 3. Multinomial logit models (choices = provinces and *comarques*).

	Girona	**Lleida**	**Tarragona**	**Baix Camp**	**Baix Llobregat**	**Gironès**	**Vallès Occidental**	**Solsonès**
ENT < 10	.7545* (.0947)	.4343* (.1018)	.4611* (.0835)	.9732* (.1848)	-.1725* (.0713)	.6156* (.1883)	-.0909 (.0660)	-.2368 (.3374)
ENT 50-99	.4490 (.3823)	-.3644 (.5965)	.9517* (.2731)	.2426 (.6536)	-.5786** (.3412)	-1.0250 (1.0536)	-.6098** (.3182)	.2198 (1.0779)
ENT > 99	-.2001 (.5918)	-.4365 (.7743)	.4243 (.4303)	.0411 (.7183)	-1.2246* (.4774)	-2.7136** (1.4239)	-1.3837* (.4553)	-.7246 (1.3349)
INV	4.08×10^{-7}* (1.68×10^{-7})	2.21×10^{-7} (2.86×10^{-7})	3.02×10^{-7}* (1.70×10^{-7})	$.1420\times10^{-4}$* (1.60×10^{-6})	$.1400\times10^{-4}$* (1.57×10^{-6})	$.1490\times10^{-4}$* (1.57×10^{-6})	$.1420\times10^{-4}$* (1.56×10^{-6})	$.1470\times10^{-4}$* (1.60×10^{-6})
χ^2		919.44*				3890.40*		
Log likelihood		-13408.99				-47440.67		

** and ** significant at the 5% level and, respectively, at the 10% level. Barcelona (province) and Barcelonès (comarca) are the comparison group in the multinomial logit. Standard errors in brackets.*

Table 5. Poisson models (choices = municipalities and *comarques*).

	Municipalities	***Comarques***
URB	.0061* (.0001)	.1990* (.0070)
DIS	-4.64×10^{-7}* (9.79×10^{-9})	-2.39×10^{-6}* (5.51×10^{-7})
DEN	-.0002* (7.50×10^{-6})	-.0233* (.0010)
DIV	-2.0740* (.0521)	.6369* (.1509)
HC	-.00218* (.0002)	-.3813* (.0214)
LOC5	-.0195* (.0011)	-.3537* (.0296)
LOC6	-.0008 (.0005)	-.3790* (.0163)
LOC7	-.0033* (.0002)	-.0076 (.0056)
LOC8	-.0003 (.0004)	.7792* (.0536)
LOC9	-.0116* (.0005)	-.0079 (.0164)
LOC10	-.0023* (.0001)	-.0280* (.0048)
LOC11	-4.89×10^{-6} (.0013)	-1.0449* (.0603)
LOC12	-.0144* (.0006)	-.1650* (.0280)
LOC13	.0088* (.0009)	-.0045 (.0208)
CONS	2.9190* (.0194)	4.1355* (.0538)
χ^2		
Log likelihood		

** significant at the 5% level. Standard errors in brackets.*