UNIVERSITAT ROVIRA I VIRGILI

WORKING PAPERS

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

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

Josep Maria Arauzo Carod
Miguel C. Manjón Antolín

Document de treball nº -5- 2001

DEPARTAMENT D'ECONOMIA
Facultat de Ciències Econòmiques i Empresarials

Edita:

Departament d'Economia
http://www.fcee.urv.es/departaments/economia/public_html/index.html
Universitat Rovira i Virgili
Facultat de Ciències Econòmiques i Empresarials
Avgda. de la Universitat, 1
432004 Reus
Tel. +34 977 759 811
Fax +34 977 300 661

Dirigir comentaris al Departament d'Economia.

Tots els treballs publicats a la col·lecció *"Documents de Treball del Departament d'Economia"* han superat un procés d'avaluació externa.

Dipòsit Legal: T- 1025 - 2006
ISSN 1576 - 3382

DEPARTAMENT D'ECONOMIA
Facultat de Ciències Econòmiques i Empresarials

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

Josep Maria Arauzo Carod
(jmac@fcee.urv.es)
and
Miguel C. Manjón Antolín
(mma@fcee.urv.es)

Department of Economics, Rovira i Virgili University, Spain
Avgda. Universitat, 1
43204-Reus
Spain

Abstract

This paper assesses empirically the importance of size discrimination and disaggregate data for deciding where to locate a start-up concern. We compare three econometric specifications using Catalan data: a multinomial logit with 4 and 41 alternatives (provinces and *comarques*, respectively) in which firm size is the main covariate; a conditional logit with 4 and 41 alternatives including attributes of the sites as well as size-site interactions; and a Poisson model on the *comarques* and the full spatial choice set (942 municipalities) with site-specific variables. Our results suggest that if these two issues are ignored, conclusions may be misleading. We provide evidence that large and small firms behave differently and conclude that Catalan firms tend to choose between *comarques* rather than between municipalities. Moreover, labour-intensive firms seem more likely to be located in the city of Barcelona.

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

1. Introduction

The choice of site for setting up an establishment hinges on many factors. In particular, the literature on industrial location has focused on the economies of the territory. Aspects related to population (e.g. sectorial employment and density), human capital (e.g. qualification) and incumbents (e.g. sectorial specialization and industrial diversity) have been widely used in empirical studies. However, the size of the newcomer has received comparatively little attention (Carlton 1979, 1983). Large firms base their decisions on rather different criteria from small and medium-sized firms. For large firms, the choice process seems to be built on objective arguments (e.g. external consultants), which in some way reflect the pros and cons of the potential sites. For small and medium-sized firms, the arguments tend to be more subjective. In fact, they are usually linked to some personal characteristic of the entrepreneur - geographical origin, previous experience in the sector or financial status, for example[1].

Broadly speaking, the sites are hinterlands defined by local markets. In practice, however, this definition is of little use. Researchers tend to resort to administrative units such as states, regions, provinces, counties and the like. This may cause a certain inconsistency with the theoretical framework but it is very convenient for at least two reasons. First, this is the format in which official statistics are available. Second, the time needed to calculate the likelihood functions increases exponentially as the number of choices increases. From this point of view, data aggregation helps to sort out certain econometric problems. However, the "fallacy of composition" observed by McFadden (1974: 134) means that the estimation may have some drawbacks.

This paper aims to test empirically how relevant these issues are for analysing industrial location. We have based our discussion on the results from data on start-up establishments in Catalonia[2] (see Table I). The probability of being located in a particular site is initially addressed in a multinomial logit specification. The main difference between this study and previous studies is that the size of the new establishment is included as an explanatory variable. Choices are given by the Catalan provinces (Barcelona, Girona, Lleida and Tarragona) and the 41 *comarques*[3]. In a second stage we have estimated a conditional logit model in which the covariates are

attributes of the choices as well as interaction terms between size variables and dummies for the choices[4]. Both provinces and *comarques* are employed as alternative sets of choices. In our view, with this strategy we can examine the robustness of the size coefficient obtained in the first stage because the multinomial logit model precludes the use of site variables. Finally, we have used a Poisson model for the *comarques* and the 942 municipalities of the sample. This eases the computational burden that implies such a huge number of choices in the conditional logit model. Here the explanatory variables are exclusively attributes of the sites. This approach was recently proposed by Guimarães *et al*. 2003).

However, this is not purely a technical exercise. There are important economic-policy concerns behind these two issues. For example, how effective are the incentives provided to large firms in guiding their location decisions? Would it not be better to provide these incentives primarily to small concerns? Should small (large) firms be the principal target of local (regional) governments? Would it not be better to unite the efforts of the local and regional authorities to cope with entrepreneurship in a broad sense, regardless of the size of the firm?

The paper is organised as follows. In the next section, we briefly review the empirical literature on the location decision. We then discuss the role of size (section 3) and provide insights into the aggregation problem (section 4). In section 5 we present the results of statistical and econometric tests on the Catalonian municipalities, *comarques* and provinces. In the final section we summarise our main conclusions.

[Insert Table I about here]

2. The determinants of industrial location: an overview

The conditional logit model proposed originally by McFadden (1974) is the most popular specification in industrial location literature. Stemming from a profit maximisation program, the probability that an establishment is located in a particular site is derived as a function of the choice characteristics and a stochastic component. Under mild asymptotic conditions this econometric model provides efficient normally distributed estimators (McFadden 1984). The disadvantage of this approach is the

strong assumption contained in the ”independence of irrelevant alternatives” axiom. However, this does not seem to be a major concern in many location studies[5]. Given the aim of this paper, we do not address this issue and assume that the necessary and sufficient characterization of the multinomial logit model holds.

Early applications for the US can be found in Carlton (1979, 1983) and Bartik (1985). Carlton's seminal work reveals the link between location and size[6], while Bartik's paper deals with 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 Baudewyns (1999), urban transportation networks and agglomeration economies are statistically significant variables; in Baudewyns *et al*. (2000), transport infrastructures, agglomeration economies and wage levels are. Related studies have focused on the particular case of Foreign Direct Investments. These include Coughlin *et al*. (1991), Friedman *et al*. (1992) and Woodward (1992) in the USA; and Guimarães *et al*. (2000) in Portugal[7].

This paper adds to this literature by analysing the determinants of industrial location in Catalonia. Consequently, the explanatory variables in our models are not so different from those cited above - except, of course, for data sources and availability constraints. The main differences between this study and previous ones is the emphasis on the size of the new establishment and the discussion regarding the implications of data aggregation. The following sections deal with both issues in greater detail.

3. Size matters

Let us 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 assessment has to be made of the following non-exhaustive list of elements: labour (skills, wages, etc.); infrastructure (transport, communications, etc.) and output markets (prices, competition, etc). Firms may even take into account urbanisation and agglomeration economies (Glaeser *et al*. 1992, Henderson *et al*. 1995). Other factors may be at stake here, but the previous section suggests that those directly related to the

entrepreneur (e.g. residence) have implicitly been considered a minor determinant of the decision.

This position seems unwise in the light of certain stylised facts. Large concerns are usually owned by business corporations that probably gather many tips on potential locations. These are the kinds of firms that, for example, would make or buy technical reports on the elements listed above (markets, population, etc.). Besides, large firms intending to open a new establishment appear to hold casual information on many of the alternative sites. The majority of the small and medium firms, on the other hand, do not have access to these inputs and it is unlikely that they spend too much effort on obtaining them. As Figueiredo *et al.* (2002) pointed out, the opportunity cost of information is not the same for all firms. In fact, it is bigger for small firms than it is for large ones.

In addition, some empirical evidence supports this line of reasoning:

- A recent study by Meester (2000) in the Netherlands and Germany required firms to evaluate several possible sites for their investments. It turns out that the greater the distance from the firm's original town, the lower the value assigned to the site. Also, nearby territories were preferred regardless of their characteristics. Other characteristics of the sites, like distance to market and size of agglomerations, are important as well. Variables such as the availability of specific services and industrial environment were not particularly relevant.

- In Portugal, Figueiredo *et al*. (2002) used a yearly survey from the Ministry of Employment to reach parallel conclusions. Econometric tests showed that the Entrepreneurs' geographical origin is a statistically significant variable. Moreover, the explanatory power of conventional factors of location (e.g. external economies and market accessibility) remained unaltered. However, the weight of the factors depended on whether the entrepreneurs were "movers" or "stayers".

- In Spain external economies are among the main determinants of industrial location (Callejón and Costa 1996). According to Costa *et al*. (2000) the location

decision for 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. Small concerns, on the other hand, are randomly spread over different sizes of towns.

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

This does not mean that searching costs will result in smaller establishments systematically choosing inappropriate locations. What at first sight seems a drawback is actually not a problem 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 locally 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 that many small start-up concerns may 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 while being employed in other firms of the region becomes a key determining factor. Thus, one would expect that the activities of new start-ups tend to follow those which are common among the incumbents. By the same token, this behaviour might not show up in large footloose firms. Large concerns are unlikely to follow such a pattern of

correlation because of the complexity of their decision process. External economies, for example, can influence 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, these theoretical arguments suggest that small start-up establishments will tend to reproduce the existing local industrial mix. This would imply that 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 whether this hypothesis is valid. Results for the Catalan municipalities are presented in section 5.

4 The aggregation of individual sites

Most of the work in this field uses broad territorial units for defining the spatial choice set (and, to a certain extent, this paper is no exception). There are several important reasons for this: the poor quality of the data at the local level, the absence of official statistics for non-administratively defined units and/or computational constraints imposed by the econometric techniques. Nevertheless, by employing aggregate data we risk losing some useful information. Aggregation is indeed a good solution provided it does not disguise the territorial heterogeneity (Bartik 1985).

However, the size of some geographical areas is likely to produce 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 how important this point is, let us consider choice sets like the American states or the German Länders. It is doubtful that an American or German agent intending to launch a new business (especially a small or medium-sized one) sees “California” or “Baden-Württemberg” as a potential site. Conclusions drawn from statistical analyses are therefore subject to these caveats.

Further, inappropriate aggregation may affect the methodological consistency of the empirical studies. For example, there is general agreement in the literature that driven factors of the industrial location act at *a local level*. Therefore, they have less and less

influence as the geographical units get wider. Also, 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. Other methodological inconsistencies may arise with other theoretical insights that shape the boundaries of the sites.

Strictly speaking, a valid territorial unit would be a hinterland defined by the local input/output markets. This may be constructed, for example, on the basis of the labour markets and with the help of data on commuting (travel-to-work areas). This is how it is done, for example, by the British Department of Employment (Coombes *et al*. 1986). However, this strategy is not free from criticism either. We can think of at least two disadvantages: i) the dynamic nature of these phenomena would produce constant variations in the boundaries of the unit; ii) from the practical point of view, 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 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 above pros and cons. Interestingly, our data sources enable us to also work with aggregate 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 framework for testing the potential effects of aggregation empirically.

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 from 942 municipalities. But according to the REI the 17719 establishments created in Catalonia in the period 1987 to 1996 actually spread over 721 municipalities[10]. Barcelona is the leading industrial focus. The spatial distribution of the entrants shows that 77.21% of them were concentrated in the province of Barcelona, 20.61% in the *comarca* of Barcelonès (followed by 16.52% in the Vallès Occidental and

11.99% in the Baix Llobregat, both in the outskirts of the Barcelonès) and 12.30% in the city of Barcelona (followed by 3.13% in Terrassa and 2.98% in Mataró, both near Barcelona). The total number of people employed was 138.580, 76.9% of whom were hired by the smaller establishments 38.0% by those with less than 10 employees and 38.9% by those with 10 to 49 employees); 17.0% were hired by the largest ones (those with over 100 employees).

[Insert Table II about here]

Newcomers were mostly small and medium concerns: 81.7% of these had fewer than 10 employees and only 0.4% had more than a hundred. In fact, they were usually smaller than the existing producers (see the final column in Table II). The average size of the establishments started up from 1987 to 1996 was around 40% of that of the firms in the 1986 cohort. This is a common pattern in the industrial dynamics literature (Geroski 1995). The results largely agree with those of previous studies in Spain (see e.g. Callejón and Segarra 1998).

A cross-tabulation with the population of the municipality shows that there is a certain positive relationship between the size of the firm and the size of the municipality (Table II), i.e. large (small) firms tend to locate in large (small) sites. This can be seen as an indirect evidence of agglomeration economies. However, statistical tests of the equality of means suggest that the interdependence is mostly valid for the smaller cases. Small (rural) sites are more likely to receive small concerns regardless of their industrial activity[11].

But is there any empirical evidence in Catalonia of the industrial mix hypothesis discussed in section 3.1? In other words, are there any differences in the sectorial distribution of the entrants depending on their size? A positive answer would be given by a temporal regularity in the figures of small concerns and a rather random behaviour in large (footloose) firms. We used simple parametric tests of the equality of proportions in the sectorial distribution of the employment and compared the means of the 1986 cohort and the entrants in 1987, 1988 and 1989. Entrants are grouped according to size. Sample sizes and the Central Limit Theorem enabled us to assume normality in the distribution. Results are given in Table III.

[Insert Table III about here]

Differences in the proportions of entrants in each sector are statistically significant except in a few cases. The picture is essentially the same for all the sizes of establishment. There is no empirical evidence in Catalonia of a strong link between small start-ups and the existing industrial mix. At the same time, entrepreneurs may be successful in sectors in which they do not have previous experience: for instance, when after some industrial reorganisation a worker becomes an entrepreneur in another sector. However, our results may be distorted by the use of aggregate data. Further research is clearly needed to discern the importance of this caveat.

5.2 Econometric models

Let us consider a firm aiming to open a new industrial establishment in Catalonia. We denote the expected benefit derived from this choice by π_{ij}, where i = 1,..., 17719 and j = 1,... ,J. Notice that J = *4, 41* or *721* (*942*) depending on the degree of aggregation we are dealing with: provinces, *comarques* or municipalities, respectively. As usual, the expected benefit is a random variable made up of a deterministic component that takes the form of a linear combination of variables $\left(Z_{ij}^{'}\beta\right)$ and a stochastic component (ε_{ij}). Thus, $\pi_{ij} = Z'_{ij}\,\beta + \varepsilon_{ij}$. Assuming that the firm i follows a maximaxing principle, the choice of site j would be given by the following rule of thumb: $\pi_{ij} > \pi_{ik}, \quad k \neq j$.

Let Y_i be a random variable that indicates the choice effectively made. Thus, the probability that a firm i locates at site j is $P(Y_i = j) = P(\pi_{ij} > \pi_{ik}, k \neq j)$. Assuming that ε_{ij} are *i.i.d.* and Weibull distributed, it can be proved that:

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

As we know, this result is due to McFadden (1974). In principle, the explanatory variables include aspects specific to both the establishment and the site. In mathematical

terms, $Z_{ij} = [W_i, X_{ij}]$. This is a purely artificial distinction that helps to describe (1) as a general specification embracing two different cases (see Green 2000). When the covariates are characteristics of the individuals (i.e. $Z_{ij} = W_i$), (1) is known as the multinomial logit model:

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

The main limitations of this model in industrial location studies are computational. Difficulties in calculating the likelihood function can make the model unfeasible when the number of choices is high[12]. Geographical aggregation obviously reduces the set of parameters and simplifies the estimation procedure. This fact may have contributed to the extensive use of aggregate data. Incidentally, this also increases the plausibility of the independence of irrelevant alternatives assumptions.

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

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

The model is not essentially different from (1) except for the fact that $Z_{ij} = X_{ij}$. Aggregate data has also been used 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 that the sites are homogeneous. If this assumption does not hold, heterogeneity within the individual spatial choice areas may produce biases in the estimates - see McFadden (1978) and Bartik (1985). Moreover, the large number of observations makes computation cumbersome. For instance, in the application presented in this study the number of observations would rise from the original 17719 to 70876, 726479 and 12756960 for provinces, *comarques* and municipalities, respectively[13].

Guimarães *et al.* (2003) recently proposed an alternative approach to sort out some of these problems. Let Y_j denote the number of establishments created in site *j*. Suppose these are independent Poisson random variables with means $E(y_j) = \lambda_j$, $ln\ \lambda_j = X'_{ij}\beta$. The joint likelihood function of this model can be divided into 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 (Birch 1963, Palmgren 1981). Interestingly, the estimates of the parameters of interest obtained from the full and conditional likelihoods are identical. And so are the covariance matrices. As a result, the Poisson model given by

$$P(Y_j = n_j) = \frac{e^{-\lambda_j}\lambda_j}{n_j!}, \quad n_j = 0, 1, 2, \ldots \tag{4}$$

is equivalent to the multinomial response model in (3). However, we must stress that the nature of the dependent variable is different from the categorical variable in (2) and the binary variable in (3). Here we are employing a count-dependent variable that does not arise directly from the firm's maximisation-based election process described above. Rather, it is the actual outcome of this process. This caveat aside, from an empirical point of view this a simple and attractive procedure for evaluating the effects of aggregation.

5.3 Covariates, estimation and results

As pointed out above, the multinomial logit specification in (2) is appropriate for individual-specific covariates. Unfortunately, our data set does not contain information on the residence of the entrepreneur. This constraint means that it is impossible to directly test its relationship with the location patterns of small firms but we can discuss the extent to which the firm location patterns vary with firm size. We use two variables as proxies for size: the 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 IV about here]

Results from the estimation of the multinomial logit model for provinces (Table IV) show that size clearly matters, which means here that the amount of investment made by the entrant is statistically significant. The probability that provinces other than Barcelona are chosen depends positively on the investment size of the establishment. Also, dummy variables used to distinguish between sizes of establishments show that the smaller concerns (under 10 workers) are more likely to be located outside Barcelona. This is not apparent for the large establishments. Sectorial dummy variables were also significant.

However, these conclusions are not fully robust if we consider other levels of aggregation. On the one hand, estimates using *comarques* as the choice set largely agree with those for the provinces. That is, most cases look like the Baix Camp: size matters and is statistically significant for the small establishments (for the sake of simplicity, Table IV only includes some selected *comarques*). On the other hand, Table IV shows that there are some *comarques* in which the results are contradictory. The significance of the size dummies varies in each case: all are significant in the Baix Llobregat, only the small and large establishments are\ significant in the Gironès, only the medium and large establishments are significant in the Vallès Occidental, and none of them are significant in the Solsonès. These differences may be due, among other reasons, to the (non-controlled) diversity of the industrial mix in the 41 *comarques* of Catalonia. However, an important result remains unaltered in all sites: the amount of investment is a statistically significant variable. The odds that a new establishment is located outside Barcelona increases with the size of the investment.

The conditional logit specification (3) does allow for differences between sites, a potential misspecification error of the multinomial logit model presented above. As attributes of the sites we have used urbanisation economies (*URB* = total number of workers per km^2, 1986), urbanisation dis-economies ($DIS = URB^2$) and density of population per km^2 (*DEN*). Urbanisation economies are cost advantages reached by the concentration of similar activities in a site, while urbanisation dis-economies are cost disadvantages caused by an excessive concentration of firms and population in a site (pollution, input prices, etc.). Density can indirectly measure the size of the markets. We have also included interactions between the characteristics of the individuals (*NE,*

INV) and dummies representing the choices. This enables us to explore the impact of the individual-site dimension on the probability of choosing a particular site, conditioned to the appropriate sufficient statistic.

[Insert Table V about here]

The results displayed in Table V support the main conclusions of the multinomial model regarding the relevance of the size of the firm. It therefore appears that location decisions are not independent of firm size. In particular, the attractiveness of Barcelona with respect to the other sites (e.g., Girona, Lleida and Tarragona in the specification using provinces as sites) increases with the number of employees of the new establishment and decreases with the investment made by the new establishment. Labour-intensive firms seem more likely to be located in the city of Barcelona whereas capital-intensive firms seem more likely to do it outside Barcelona. Notice also that in the multinomial logit (see Table IV) dis-aggregation produced heterogeneous outcomes for the *comarques*. Once the site heterogeneity is controlled, results are much more consistent.

The characteristics of the site are also statistically significant and this suggests that results from the multinomial logit model are somehow flawed. The coefficients of urbanisation and dis-urbanisation economies as well as that of the density have the expected signs in both specifications - i.e. provinces and *comarques*. However, if we compare the values of the two choice sets, we are tempted to conclude that the aggregation causes an overestimation of the site effects. In any case, the concentration of the economic activity attracts new firms up to a point at which it impairs their costs. Moreover, crowded areas are not appealing locations for Catalan industrial firms. A by-product of this is that the preferred environment of households and industrial establishments is not necessarily the same.

[Insert Table VI about here]

Table VI shows results of the Poisson specification (4) based on the Case 1 model of Guimarães *et al.* (2003). This means that the dependent variable is the number of establishments that have chosen a particular site and the explanatory variables are site-

specific. Industrial diversity is measured by a Hirshmann-Herfindahl index ($DIV_j = \sum_{s=5}^{13} h_{sj}^2$, where *s* is the corresponding industrial sector - see footnote 10); human capital (*HC* = number of people with medium and high levels of education per km^2); and location economies (*LOC5* to *LOC13* = number of workers per km^2 in each industrial sector, 1986). Other covariates are urbanisation economies, urbanisation dis-economies and density of population per km^2 - as in the conditional logit specification. The data set includes the municipalities that received no new entrants during the period of analysis. The dependent variable for these municipalities was zero[14]. Generally speaking, the results for *comarques* and municipalities were similar. However, some differences are worth noting.

Urbanisation economies have a positive effect on the location decision and, as expected, urbanisation dis-economies and population density act in the opposite way. These results broadly agree with those from the conditional logit. However, notice that the impact of these variables in the Poisson model is more powerful at the *comarques*' level. This can be interpreted in the sense that new firms are prone to locate near the incumbents (i.e. in the same *comarca*), albeit at a distance far enough to avoid the 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 hand, 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 mobility between municipalities of the same *comarca* is not expensive.

Human capital shows a negative coefficient in both geographical areas. This is probably explained by the specialisation of the Catalan manufacturing sector in products/processes that do not demand highly qualified labour. Alternatively, there might be some kind of spatial mismatch between industrial firms and high-skilled workers. Moreover, a comparison of the estimates for the location economies suggests that in most sectors the employment density reduces the expected number of new establishments - i.e. the probability of choosing a particular site in the equivalent conditional logit specification. This effect, however, is bigger in the *comarques*

specification. This means that sectorial dis-economies are more powerful at the supra-local level (*comarca*). By the same token, entrants prefer a specialised industrial environment in the *comarca* and a more diversified one at the local level.

6. Conclusions

In this paper we argue that empirical investigations into the determinants of start-up concerns need to apply a discriminating criterion based on their size. Empirical evidence shows that larger firms are guided by more objective decision-making reasons, whereas smaller ones are mostly oriented by the entrepreneur's preferences. In practice, the range of alternatives open to the latter is frequently reduced to the nearest geographical areas. Although the literature has traditionally focused on the role of territorial factors, the personal characteristics of the entrepreneur do influence the location decision of some new firms.

We have also investigated the effects of territorial aggregation. Location factors, contrary to what is sometimes assumed, do not act uniformly over broad geographical areas. Therefore, appropriate territorial units are required for constructing the data set. Otherwise, the analytical foundations of the investigation might be undermined. Moreover, the lack of availability of local data could cause important biases. This suggests that comparing results from several territorial levels (cities, counties, regions, etc.) may be a good strategy in empirical studies.

Statistical tests should be consistent with these ideas. As an illustration, we have analysed the hypothesis that differences in size stretch to the sectorial distribution of employment but we have concluded that in Catalonia there is no evidence of this phenomenon. We have also presented econometric models for the determinants of industrial location in Catalonia. We used alternative specifications (a multinomial logit, a conditional logit and a Poisson model), proxies for the size of the establishment (number of employees and investment) and three levels of administrative aggregation (municipalities, *comarques* and provinces). Our results clearly show it pays to consider size and dis-aggregation.

Estimates from the Poisson model, for example, show that the effects of some variables are more apparent in the *comarques* than in the municipalities. A plausible interpretation is that the entrants' site of reference is mainly the *comarca* and that the differences between municipalities in the same *comarca* are less crucial. The implication for economic policy is that local and regional authorities might be misguided and should not limit their activities to their own administrative borders. Analogous and interesting insights can be gained from the multinomial logit specifications (individual-specific covariates) and the conditional logit specifications (site-specific covariates). In particular, both types of specification emphasise the role played by the size of the new establishment when deciding where to locate. Labour-intensive start-ups, for example, seem more attracted by the city of Barcelona than capital-intensive ones. Generally, small and large establishments tend to behave differently. However, the reasons for this remain unclear and should be a matter for future research.

7. 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, (2000), Infrastructure publique et localisation des enterprises à Bruxelles et en Wallonie, *in* M. Beine & F. Docquier, eds, "La Politique de Dévelopement Local et L'infrastructure Publique: Bruxelles et Wallonie", Bruxelles.

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. 13-50.

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

Coombes, M., Green, A. & 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.

Costa, M., Segarra, A. & Viladecans, E. (2000), "Business dynamics and territorial flexibility". Paper presented at the 40th European Congress of the Regional Science Association, Barcelona.

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. 171-214.

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.

Figueiredo, O., Guimarães, P. & Woodward, D. (2002), "Home-field advantage: Location decision of Portuguese entrepreneurs", *Journal of Urban Economics* **52**, 341-361.

Friedman, J., Gerlowski, D. & Silberman, J. (1992), "What attracts foreign multinational corporations?, *Journal of Regional Science* **32**, 403-418.

Geroski, P. (1995), "What do we know about entry?", *International Journal of Industrial Organitation* **100**, 1126-1152.

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

Guimarães, P., Figueiredo, O. & Woodward, D. (2000), "Agglomeration and the location of foreign direct investment in Portugal", *Journal of Urban Economics* **47**, 115-135.

Guimarães, P., Figueiredo, O. & Woodward, D. (2000), "A tractable approach to the firm location decision problem", *Review of Economics and Statistics* **Forthcoming.**

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.

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

McFadden, D. (1978), Modelling the choice of residential location, *in* A. Karquist, L. Lundqvist, F. Snickars & J. Weibull, eds, "Spatial Interaction. Theory and Planning Models", North-Holland, Amsterdam, pp. 75-96.

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". Paper presented at the 40th European Congress of the Regional Science Association, Barcelona.

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.

Acknowledgements: We are grateful to an anonymous referee, W.J. Meester, P. McCann and seminar participants at the 8th ERES meeting (UA, Alacant), *I Encuentro de Economía Industrial* (UPF, Barcelona), *XXVII Reunión de Estudios Regionales* (Madrid) and the Workshop on Demography of Firms and Industries (UB, Barcelona) for helpful comments. This research was partially funded by the CICYT (SEC2000-0882-C02-02). Usual caveats apply.

Table I. Descriptive statistics

Variables	N	Mean	Std. Dev.	Min	Max
NE	17719	7.821425	64.5342	0	8190
INV	17719	14510.37	141544	0	1.02×10^{-7}
URB	Mun (721)	108.9943	375.4553	0.0000	6429.2970
	Com (41)	168.9683	794.2876	1.3386	5112.3240
	Pro (4)	59.4355	77.0866	9.3121	174.3736
DIS	Mun (721)	152650.7	1661933	0.0000	0.4130×10^{-8}
	Com (41)	644055.4	4080722	1.7918	0.2610×10^{-8}
	Pro (4)	7989.3350	14947.7500	86.7145	30406.1400
DEN	Mun (721)	412.5006	1625.9120	1.2795	20571.9800
	Com (41)	554.9435	2572.0980	4.0209	16538.6200
	Pro (4)	195.4654	263.2176	29.6004	588.5511
DIV	Mun (721)	0.3918	0.2110	0.0000	1.0000
	Com (41)	0.2445	0.0973	0.1477	0.6911
	Pro (4)	0.1750	0.0149	0.1581	0.1938
CH	Mun (721)	15.5351	80.9649	0.0000	1825.7940
	Com (41)	38.9682	209.2513	0.2318	1344.6350
	Pro (4)	10.9519	16.0983	1.5018	35.0525
LOC5	Mun (721)	3.1253	11.5334	0.0000	137.6366
	Com (41)	2.6189	9.1842	0.0007	57.8465
	Pro (4)	1.1900	1.6751	0.1355	3.6918
LOC6	Mun (721)	4.5747	19.3921	0.0000	249.5256
	Com (41)	6.5360	31.3087	0.0033	200.5523
	Pro (4)	2.1130	3.2402	0.0772	6.9336
LOC7	Mun (721)	14.0393	56.7590	0.0000	750.5859
	Com (41)	15.5785	69.1212	0.0249	442.4751
	Pro (4)	5.8295	8.7020	0.5034	18.8392
LOC8	Mun (721)	2.1724	14.8593	0.0000	276.0360
	Com (41)	5.5089	30.9555	0.0000	198.5015
	Pro (4)	1.4248	2.5218	0.0395	5.2043
LOC9	Mun (721)	5.0214	16.0398	0.0000	151.0439
	Com (41)	4.7221	17.2008	0.0240	110.8511
	Pro (4)	2.2097	2.1688	0.6022	5.3971
LOC10	Mun (721)	12.3270	47.8660	0.0000	783.7347
	Com (41)	10.9184	38.8140	0.0082	243.5250
	Pro (4)	5.3457	8.3640	0.3264	17.8199
LOC11	Mun (721)	2.5757	7.5704	0.0000	105.6229
	Com (41)	2.6052	9.9802	0.0328	64.3921
	Pro (4)	1.1945	1.2217	0.1854	2.9622
LOC12	Mun (721)	3.5938	17.9637	0.0000	244.5111
	Com (41)	5.6454	29.9478	0.0000	192.2186
	Pro (4)	1.6540	2.6959	0.1118	5.6924
LOC13	Mun (721)	2.9274	12.2355	0.0000	117.1840
	Com (41)	3.3008	14.5891	0.0007	93.4357
	Pro (4)	1.1934	1.8695	0.0469	3.9876

Table II. New Establishments in Catalonia (1987-1996): Average Number of Employees.

Sector	Size 1 [c]	t-test [a]	Size 2	t-test	Size 3	t-test	Size 4	t-test	Size 5	t-test	Size 6	Total	Total / 1986 cohort
5	4,72	-1.31** [b]	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	9,85	-1.03	23,27	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	-

[a] *t-test* is the *t* statistic for the alternative hypothesis H_1: $\mu_m^s < \mu_{m'}^s$ when variances are unequal and unknown; *s* = sector; *m* = *Size 1*, ... , *Size 5*; $m' = m+1$.

[b] * and ** denote 1%, 5% and 10% significance.

[c] *Size 1* denotes municipalities with less than 2001 inhabitants; *Size 2* between 2001 and 10000; *Size 3* between 10001 and 50000; *Size 4* between 50001 and 100000; *Size 5* between 100001 and 1000000; *Size 6* more 1000000 (Barcelona city).

[d] *Source: Own calculations with data from REI, IDESCAT and Encuesta Industrial (1986 cohort).*

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

Sector	TOTAL WORKERS	ENT < 10 [c]	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	ENT > 99
	1986	**1987**	**1987**	**1987**	**1987**	**1988**	**1988**	**1988**	**1988**	**1989**	**1989**	**1989**	**1989**
5	26821	302	168	0	0	367	183	50	0	350	196	0	120
6	46183	231	67	75	0	141	102	121	0	179	25	65	8190
7	112833	2213	2313	209	793	2227	2347	358	170	2102	1779	380	0
8	43614	286	131	82	932	225	166	0	205	333	161	60	0
9	57614	517	164	0	0	502	215	121	0	452	75	0	0
10	99449	1204	1409	273	0	1190	1652	234	323	1181	1123	151	246
11	18140	731	387	123	0	702	373	112	0	715	482	124	0
12	27423	514	335	205	0	444	463	99	112	510	574	0	409
13	24979	434	401	0	0	354	363	51	130	409	495	0	0
TOTAL	457056	6432	5375	967	1725	6152	5864	1146	940	6231	4910	780	8965

z [a]	**z**	**z**	**z**	**z**	**z**	**z**	**z**	**z**	**z**	**z**	**z**
3,980* [b]	8,527*	7,764*	10,369*	-0,323	8,920*	2,166*	7,655*	0,838	5,573*	6,973*	18,199
17,279*	21,520*	2,421*	13,922*	20,289*	21,218*	-0,509	10,278*	18,894*	22,290*	1,640	-237,321
-17,919*	-30,922*	2,215*	-20,440*	-20,759*	-26,993*	-5,137*	4,690*	-16,424*	-18,631*	-15,543*	54,039
13,860*	17,696*	1,124	-62,285*	15,664*	17,452*	10,994*	-12,772*	11,233*	14,905*	1,757	30,721
10,983*	21,061*	11,808*	15,769*	10,456*	20,572*	2,086*	11,642*	12,672*	23,356*	10,605*	35,910
5,871*	-7,864*	-4,872*	21,891*	4,563*	-11,811*	1,098	-9,351*	5,333*	-1,880	1,623	43,479
-29,805*	-12,009*	-13,893*	8,443*	-29,352*	-9,289*	-10,034*	6,233*	-29,784*	-20,722*	-17,008*	19,241
-6,664*	-0,714	-19,832*	10,492*	-3,988*	-6,063*	-3,755*	-7,621*	-7,197*	-16,622*	7,056*	5,689
-4,486*	-6,385*	7,476*	9,985*	-0,990	-2,425*	1,510	-11,254*	-3,785*	-14,096*	6,715*	22,753

[a] z is the statistic for the null hypothesis of the equality of proportions under the assumption of normality in the distributions.
[b] * denotes 5% significance.
[c] ENT<10 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.
[d] *Source: own calculations from REI (entrants) and Encuesta Industrial (1986* cohort*).*

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

	Provinces [b, c]			***Comarques*** [b, c]				
	Girona	**Lleida**	**Tarragona**	**Baix Camp**	**Baix Llobregat**	**Gironès**	**Vallès Occidental**	**Solsonès**
ENT < 10	.7545 (.0947)* [a]	.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	$.40\times10^{-6}$ ($.16\times10^{-6}$)*	$.22\times10^{-6}$ ($.28\times10^{-6}$)	$.30\times10^{-6}$ ($.17\times10^{-6}$)*	14.20×10^{-6} (1.60×10^{-6})*	14.00×10^{-6} (1.57×10^{-6})*	14.90×10^{-6} (1.57×10^{-6})*	14.20×10^{-6} (1.56×10^{-6})*	14.70×10^{-6} (1.60×10^{-6})*
χ^2	919.44*			3890.40*				
Log likelihood	-13408.99			-47440.67				

[a] * and ** significant at the 5% level and, respectively, at the 10% level. Standard errors in brackets.

[b] Barcelona (province) and Barcelonès ("comarca") are the comparison group.

[c] Dummy sectorial variables are included as explanatory variables (results are available from the authors upon request).

Table V. Conditional logit models (choices = provinces and *comarques*).

Provinces [b]		***Comarques*** [b]	
URB	.0431 (.0113)* [a]	URB	.0168 (.0006)*
DIS	-36.8×10^{-6} (-21.1×10^{-6})**	DIS	-1.84×10^{-6} $(.03\times10^{-6})$*
DEN	-.0063 (.0038)	DEN	-.0021 (.0001)*
NE × Girona	-.0459 (.0051)*	NE × Baix Camp	-.0261 (.0054)*
NE × Lleida	-.0349 (.0056)*	NE × Baix Llobregat	-.0046 (.0017)*
NE × Tarragona	-.0142 (.0030)*	NE × Gironès	-.1138 (.0094)*
		NE × Vallès Occidental	-.0085 (.0017)*
		NE × Solsonès	-.0143 (.0163)*
INV × Girona	1.89×10^{-6} $(.34\times10^{-6})$*	INV × Baix Camp	19.6×10^{-6} (1.65×10^{-6})*
INV × Lleida	1.60×10^{-6} $(.34\times10^{-6})$*	INV × Baix Llobregat	17.8×10^{-6} (1.58×10^{-6})*
INV × Tarragona	1.44×10^{-6} $(.28\times10^{-6})$*	INV × Gironès	22.4×10^{-6} (1.64×10^{-6})*
		INV × Vallès Occidental	18.3×10^{-6} (1.58×10^{-6})*
		INV × Solsonès	22.3×10^{-6} (1.64×10^{-6})*
χ^2	21578.74*	χ^2	30816.03*
Log likelihood	-13774.37	Log likelihood	-50392.76

[a] ** and ** means significant at the 5% and 10% level, respectively. Standard errors in brackets.*

[b] *Barcelona (province) and Barcelonès (comarca) are the comparison group.*

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

	Comarques	**Municipalities**
URB	.1990 (.0070)* [a]	.0061 (.0001)*
DIS	-2.39×10^{-6} (5.51×10^{-7})*	-4.64×10^{-7} (9.79×10^{-9})*
DEN	-.0233 (.0010)*	-.0002 (7.50×10^{-6})*
DIV	.6369 (.1509)*	-2.0740 (.0521)*
HC	-.3813 (.0214)*	-.00218 (.0002)*
LOC5	-.3537 (.0296)*	-.0195 (.0011)*
LOC6	-.3790 (.0163)*	-.0008 (.0005)
LOC7	-.0076 (.0056)	-.0033 (.0002)*
LOC8	.7792 (.0536)*	-.0003 (.0004)
LOC9	-.0079 (.0164)	-.0116 (.0005)*
LOC10	-.0280 (.0048)*	-.0023 (.0001)*
LOC11	-1.0449 (.0603)*	-4.89×10^{-6} (.0013)
LOC12	-.1650 (.0280)*	-.0144 (.0006)*
LOC13	-.0045 (.0208)	.0088 (.0009)*
CONS	4.1355 (.0538)*	2.9190 (.0194)*
χ^2	30466.65*	43508.12*
Log likelihood	-1322.83	-11803.45

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

[1] The size of the firm is often an omitted aspect in theoretical studies. Therefore, the matter of "why should the behaviour of small firms not mirror that of large ones" (which, incidentally, was raised by one referee) does not have a supportive theoretical framework. Empirical and casual evidence, however, can be found in Mueller and Morgan (1962), Johnson and Cathcart (1979), Chapman and Walker (1991), Cotorruelo and Vázquez (1997), Figueiredo *et al.* (2002) and Meester (2000).

[2] Catalonia is a Spanish administrative region (comunidad autónoma) in the northeast of Spain. It has a population of about 6 million people (15% of the population of Spain) and it covers 31895 km^2. The GDP of Catalonia is approximately 19% of the GDP of Spain. The data in this study are 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.

[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 area 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, the *comarca* of Barcelona).

[4] Following Greene (2000: chap. 19) it is useful to distinguish between aspects that are specific to the individual (the characteristics of the establishment) and attributes of the choices (sites) that may also vary across 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".

[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 business location decision. (...) 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] Other studies 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.* 2000).

[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, condition) the rise of these entrepreneurs.

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

[10] The municipalities data set actually has 17718 observations. The missing value corresponds to a new municipality created during the period of analysis (L'Ampolla). The establishments are grouped according to the old Spanish sectorial classification CNAE-74 on the basis of the following categories (for comparative purposes we also include the current CNAE-93 code):

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 manufacturing industries.	25

[11] We have analysed the sensitivity of the results to the use of other clusters - e.g. rural (less than 10000 people), urban (more than 100000 people) and mixed (10000 to 100000 people) - and the conclusions remained unaltered. We also grouped concerns by size using the categories of Table III (under 10 employees, 10 to 49, 50 to 99, and over 99) to construct a new contingency table. The Pearson χ^2 test rejected the hypothesis that the variables were independent (χ^2 = 65.02 using Table II categories and $\chi^2 = 19.47$ using rural/urban/mixed categories).

[12] For example, Intercooled Stata sets the limit in 50 alternatives and Limdep limits the number of parameters in the models to 150.

[13] 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. (2000).

[14] In the 221 municipalities not actually chosen in our sample, the corresponding (categorical, binary or count) dependent variable will always be zero. In the conditional logit model these alternatives drop out of the probability and their coefficient cannot be identified in the likelihood function. In practice, J = 721. (The underlying selection problem is beyond the scope of this paper. On this see e.g. Woodward 1992 and Guimarães et al. 2003). In the Poisson specification, however, this information is relevant for the estimation. Thus, J = 942.