THE GRADUATE INSTITUTE | GENEVA

INSTITUT DE HAUTES ETUDES
INTERNATIONALES ET DU DÉVELOPPEMENT

GRADUATE INSTITUTE OF INTERNATIONAL
AND DEVELOPMENT STUDIES

HEID Working Paper No: 08/2008

Firm Location Determinants: Empirical Evidence for France

Nadia Rocha
Graduate Institute of International Studies

Abstract

This paper analyzes the effects of local externalities on the probability of starting a new economic activity. We use firm-level data and geographic information on French zip-codes for 1993-2002. Poisson and Negative Binomial panel data models are estimated as they naturally allow for large sets of location choices with frequent zero outcomes and control for unobserved zip-code heterogeneity. To measure the geographic extent of location externalities we compute localization, urbanization and congestion variables for neighboring regions. We separately analyze the scale effects stemming from exporting and non-exporting firms. Our results show that agglomeration economies at zip-code level strongly effect the location decision of industrial establishments and find the presence of agglomeration shadows, one of the core predictions of the standard New Economic Geography formulations.

Determinants of Firm Location: Empirical Evidence for France

Nadia Rocha*

July 2008

Abstract

This paper analyzes the effects of local externalities on the probability of starting a new economic activity. We use firm-level data and geographic information on French zip-codes for 1993-2002. Poisson and Negative Binomial panel data models are estimated as they naturally allow for large sets of location choices with frequent zero outcomes and control for unobserved zip-code heterogeneity. To measure the geographic extent of location externalities we compute localization, urbanization and congestion variables for neighboring regions. We separately analyze the scale effects stemming from exporting and non-exporting firms. Our results show that agglomeration economies at zip-code level strongly effect the location decision of industrial establishments and find the presence of agglomeration shadows, one of the core predictions of the standard New Economic Geography formulations.

Keywords: Firm Location; Agglomeration Economies; Agglomeration Shadows

JEL Classification: R30; C50; L60

*Graduate Institute of International Studies, 11a, Avenue de la Paix, 1202 Geneva, Switzerland; e-mail: nadia.rocha@graduateinstitute.ch. I thank Richard Baldwin and Marc Melitz for their support and feedback. The usual disclaimer applies.

Introduction

Marshall (1920) describes three main benefits that firms obtain when locating near each other: labor market pooling, input sharing and knowledge spillovers. Agglomeration economies may also lead to congestion, pollution and other negative externalities caused by the clustering of a population of firms and people.

The benefits of firms to locate near one another not only depend on geographical distance facts but are also determined by the industrial closeness to the activities performed by firms. Hoover (1936) subdivides these agglomeration effects into localization and urbanization economies: the former refers to the concentration of similar activities (external to the firm but internal to the industry) while the latter refers to the concentration of non-related activities (external to both firm and industry).

A rich empirical literature on agglomeration economies, including papers as Sveikauskas (1975), Moomaw (1981), Henderson (1986), Nakamura (1985), Carlton (1983), Glaeser et all. (1992), Henderson et all. (1994) and Ciccone and Hall (1996), amongst others. These studies focus on whether it is specialized economic environments (localization externalities) or large and diversified cities (urbanization economies) that generate large scale effects.

In most of the cases the authors compute location externalities through the estimation of the production function. The effects of agglomeration have been measured using value added as a measure of productivity (Ciccone and Hall (1996)) and the growth of total employment in an industry (Gleaser et all. (1992) and Henderson et all. (1994)).

There are a few problems related with these kinds of approaches; inconsistencies may exist if there is not enough data on inputs. More precisely, the estimation results will be biased if there is a correlation between an omitted input in the regression and the variables representing geographical characteristics. Furthermore, endogeneity problems may arise when not only the characteristics of one location are sensitive to the externalities of that area (agglomeration effects) but also when those externalities themselves are a function of the firms located in that area.

To deal with the lack of data on inputs and with the endogeneity problems, different solutions have been proposed. Among them, the approach first developed by Rosenthal and Strange (2004) uses new establishment births in order to analyze the effects of agglomeration economies. Focusing on new entrants is appealing because the existing economic environment can be treated as exogenous such that decisions taken by new establishments are not influenced by prior choices.

In their innovative approach, Rosenthal and Strange (2004) specify a linear relationship between the number of births of new establishments and local industrial characteristics. The authors use a Tobit model to deal with the fact that a very high number of zipcodes do not experience any births for a given year.

In this context, the Tobit approach presents two limitations. First, it fails to account for the discrete nature of the dependent variable of new manufacturing firms entering into a certain location. Second, it considers the zero outcome (no births) as a result of censoring, when it is a natural outcome of the variable being modeled.

This paper investigates the effects of local externalities on firm location decisions. A novelty of our study is the testing of the presence of agglomeration shadows, one of the core predictions of the standard New Economic Geography formulations. Another contribution derives from separately analyzing the localization and urbanization effects stemming from exporting and non-exporting firms; this to investigate whether exporter status affects the intensities of local externalities.

This work is closely related with studies developed by Monseny (2005) and Arauzo-Carnod (2006), where the authors, by using count data models to analyze the scale effects of production concentration on the expected number of new entrants in a certain location (zip-code), deal with the zero outcomes and the discrete nature of the dependent variable. In these studies, together with the variables representing agglomeration and congestion economies, fixed effects at metropolitan area or labor market level are introduced to control for common characteristics of certain areas that may effect the number of new entrants.

Our paper differs from the previous approaches in that we introduce zip-code level fixed effects. Apart from the scale effects already captured with variables representing localization, urbanization and congestion economies, there may still exist some characteristics strictly related with each location that effect the productivity of potential producers.

A territorial analysis

In recent years, a greater availability of statistical information has enabled researchers to access data at a very detailed geographical level. In addition, in the 1990's, Count Data Models were developed. These econometric models have no computational constraints on the number of alternatives (larger data sets can be used) which allows them to account for one important drawback of the most popular Conditional Logit Model: the difficulty in calculating the likelihood function when there are many alternatives (sites), which is very common at a local level.

The characteristics of count data models contributed to the development of recent studies relying mainly on smaller areas such as counties (Arauzo and Manjón 2006; List 2001; Coughlin and Segev (1997) or municipalities (Arauzo and Manjón (2006); Arauzo 2005; Holl 2004a, 2004b and 2004c; Figueiredo et al. 2002; Guimares et al. 2000).

Some studies on agglomeration economies have been performed at a very detailed geographical level. For instance, Duranton and Overman (2002) and Rosenthal and Strange (2003) use UK and USA zip code level data respectively. Monseny (2005) uses local data for Catalan zip-codes.

Agglomeration economies are thought to take place at a local level but evidence of these effects has not been extensively tested to date. We address the geographic nature of localization, urbanization and congestion economies for France, by using firm-level data information at zip-code level for the time interval 1993-2002. Our results show that agglomeration economies at zip-code level strongly effect the location decision of industrial establishments, suggesting that agglomeartion effects should be studied at a very refined geographic level.

Evidence of the presence of location spill over effects between local administrative borders has been found in papers such as Henderson (2003), Rosenthal and Strange (2003) and Viladecans-Marsal (2004), amongst others. To further investigate the geographic extent of localization and urbanization economies we will consider the industrial characteristics of surrounding zip-codes.

In order to study the scale effects of agglomerative forces we will follow the Rosenthal and Strange (2003) approach based on constructing concentric rings around zip codes (1, 5, 10 and 15 miles) and analyzing how new firms are attracted to the area depending on which ring is used to measure local variables, such as employment. Our findings confirm that urbanization economies decrease with distance and find the presence of agglomeration shadows.

The Model

This section develops an analytical framework that helps to explain why the births of new establishments are concentrated in certain locations. This approach links firm location decisions to profit-maximizing and cost-minimizing strategies.

Following Rosenthal and Strange (2003) and Monseny (2005), we will assume that differences in new establishments births across locations can be explained by two factors: differences in the number of firms and differences in the probabilities of the establishments making positive profits.

In this framework it is assumed that in a certain period, the flow of new entrants into the market can be described by a Poisson distribution. It is also assumed that an entrepreneur's decision to open a new establishment is positively related with the probability of experiencing positive profits, once inputs have been optimally chosen.

Having said this, an entrepreneur, when deciding to start a new activity, will choose a vector $\mathbf{x}$ of J rows $x_1, x_2, ..., x_J$ corresponding to input levels such as land, labor, raw materials, amongst others, in order to maximize the following profit function:

$$\pi(\mathbf{x}, \mathbf{y}) = a(\mathbf{y})f(\mathbf{x})(1 + \varepsilon) - c(\mathbf{x}) \tag{1}$$

Here, output prices have been normalized to unity. $f(\mathbf{x})$ is the production function which is positive for every value of the inputs x_j. $\mathbf{y}$ is a vector of industrial characteristics of the geographical region in which the new activity will be located and $c(\mathbf{x})$ represents the input costs. The production function is shifted by the positive function $a(\mathbf{y})$; a factor related to the idea of economies of scale and network effects such as the presence of development strategies in this area of economic activity. These effects will have an impact on the average cost of production as a result of the increases or decreases in total output of a product. In order to take into account firm heterogeneity, the firm specific term ε has been included in the profit function. This last variable is independent and identically distributed across plants. For simplicity sector, time and location subscripts have been omitted.

The optimal amount of each factor hired is the level where marginal productivity equals the marginal cost of production.

$$a(\mathbf{y}) f_j(\mathbf{x})(1+\varepsilon) = w_j, \forall j = 1, ..., J \tag{2}$$

f_j represents the derivative of the production function with respect to the x_j component of the inputs. The unitary cost of this component is w_j. The profit function is:

$$V(\mathbf{x}, \varepsilon, w_1, ..., w_J) = \max_{\mathbf{x}} (a(y) f(\mathbf{y})(1+\varepsilon) - c(\mathbf{x})) \tag{3}$$

The productivity of the entrepreneur, ε, is assumed to be bounded between one and minus one and is drawn from a known distribution function $F(\varepsilon)$. We assume that the entrepreneur enters the market with the same probability with which the start up has positive profits.

For any value of the local industrial characteristics and inputs costs, there is a unique threshold value $F(\hat{\varepsilon})$ such that

$$V(\mathbf{y}, \varepsilon, w_1, ..., w_J) = 0 \tag{4}$$

Assuming that the observed values of the local characteristics and input costs in period t remain unaltered in period $t+1$, an entrepreneur starts activity at time t with probability $(1 - F(\widehat{\varepsilon}_t))$.

In order to analyze the effects of the industrial characteristics on the probability of positive profits, we evaluate (4) at $\widehat{\varepsilon}$ [1] and differentiate to get:

$$\frac{\partial \widehat{\varepsilon}}{\partial_s} = -\frac{\frac{\partial a(\mathbf{y})}{\partial y_s} f(\mathbf{x}^*)(1+\widehat{\varepsilon})}{a(\mathbf{y}) f(\mathbf{x}^*)} \tag{5}$$

$$\frac{\partial \widehat{\varepsilon}}{\partial w_j} = -\frac{-\mathbf{x}^*}{a(\mathbf{y}) f(\mathbf{x}^*)} \tag{6}$$

Given an optimal vector of inputs $\mathbf{x}^*$, the sign of expression (5) depends on the fact that characteristic y_s increases or decreases productivity. When $\frac{\partial a(\mathbf{y})}{\partial \mathbf{y}_s} < 0$, there exist benefits that firms obtain when locating near each other. Hence higher values of the component *sth* will lead to a higher probability of experiencing positive profits. More precisely, if expression (5) is a negative (positive) expression, it implies that higher values of the *sth* characteristic will lead to a higher (lower) likelihood of positive profits. Relation (6) shows that the probability of positive profits decreases in input prices. Higher wages will result in higher values of $\widehat{\varepsilon}$ hence in a lower probability of starting a new activity.

One of the implications of the previous results is that given the same number of entrepreneurs across locations, differences in the location characteristics may influence the location decisions of new entrants into the market. More precisely, locations with lower input costs and particular economic environments are more likely to experience a higher expected value of births of new establishments.

[1] It is necessary to apply the implicit theorem and then use the envelope theorem.

From a statistical point of view, after considering that becoming an entrepreneur is a rare event and assuming that probabilities across observations are independent, the number of entrepreneurs thinking about starting up an establishment will follow asymptotically a Poisson distribution:[2]

$$\Pr(N = n) = \frac{\exp(-\alpha)\alpha^n}{n!} \tag{7}$$

In the previous expression, the mean value of the distribution is given by the intensity rate α which varies across locations and across sectors.

In order to study the behavior of the expected number of new entrants into a location the following has to be considered: if the number of repetitions of a binary (zero or one) iid event (making positive profits or not) is a realization of a Poisson distribution, then the sum of these repeated binary outcomes will follow a Poisson distribution with intensity parameter $\alpha_{ikt} \times ((1 - F(\widehat{\varepsilon}_{ist}))$, where the second term represents the probability that the binary event takes the value of one (Cameron Triverdi 1998).

From the previous implication, the number of establishments entering into a certain location s, industry i and time period t, can be characterized by a Poisson distribution with a rate parameter that depends both on the entrepreneur abundance in the location and the probability of making positive profits in this location:

$$N_{ist} \sim P(\alpha_{ist}(1 - F(\widehat{\varepsilon}_{ist}))) \tag{8}$$

[2]The Poisson distribution is a discrete probability distribution that expresses the probability of a number of events occurring in a fixed period of time. One of the characteristics of this distribution is that the parameter α represents not only the mean number of occurrences , but also its variance.

Data and Variables

The Data

To perform our analysis, we merge two different data-sets: 1) French firm level data, which provides information on the entry year of plants and their zip-code location, and 2) data on territorial characteristics, including the latitude and longitude of each zip-code.

The firm level characteristics present in the analysis come from standardized annual company accounts, which were obtained from the Amadeus database maintained by Bureau van Dijk Electronic Publishing. This unique pan-European data-set constitutes a compendium of harmonized financial statements, based on registered filings with the respective national statistical offices. We constrain our sample to the unconsolidated accounts of manufacturing firms in France for the years 1993 through 2002. The available data constitute an unbalanced panel of 729,065 predominantly large and medium-sized enterprisers.

The Variables

To specify the agglomeration economies we consider the standard variables used in the literature. For location economies we use the total number of employees in the same two digit industry, for urbanization economies we use the total number of employees. We introduce industry average wages to represent the sectoral average costs of production. We also consider average wages across all industries; this variable is supposed to partly reflect urban congestion factors such as high land prices and costly commuting.

In order to measure the spatial dimension of the agglomerative effects we create a set of concentric rings for all variables. First, employment in a given region is considered to be uniformly distributed throughout the zip-code. Following a similar approach to that of Rosenthal and Strange (2003), after computing the great-circle distance[3] between each pair of zip-codes existent in the French territory, circles of radius r_i with $i = 1, 10, 15$ kilometers are drawn around the geographic centroid of each zip-code.

Calculating the differences between adjacent rings of the own industry and overall employment levels will show, respectively, the localization and urbanization variables for each of the rings.

In the case of the wage variables, it is necessary to compute separately the share of the own industry wages and the share of the overall wages for each of the concentric rings.

Descriptive Statistics

It is possible to observe in Table 1 that the manufacturing industries considered in this study account for more than half of French manufacturing employment. Furthermore, firms belonging to such sectors produce for both the domestic and the international markets.

Since our analysis focuses on intra-sectoral and inter-sectoral scale effects, it is important to have a set of industries that represent a heterogeneous set of activities, more precisely, the industries we consider represent a mix of traditional industries with established products and high technology industries, in which the innovation and the creation of new products is present.

In Table 2 for each two digit industry, the average number of firms entering into the market over the period 1993-2002 has been com-

[3]The great-circle distance is the shortest distance between any two points on the surface of a sphere measured along a path on the surface of the sphere. Let ϕ_o λ_o, ϕ_d λ_d be the geographical latitude and longitude of two points (origin and destination), respectively, $\Delta\lambda$ the longitude difference, the distance derived from Harvesine formula is given by the following expression $r 2 \arcsin(\sqrt{\sin^2(\frac{\phi_d - \phi_o}{2}) + \cos\phi_d \cos\phi_o \sin^2(\frac{\Delta\lambda}{2})})$ where r represents the Earth radius (6372.795 km).

Table 1: Employment shares

Code	Industry	Gini Index	Employment Share
15	Food products and beverages	0.008	0.105
17	Textiles	0.012	0.031
18	Wearing apparel; dressing and dyeing of fur	0.014	0.020
19	Tanning and dressing of leather	0.015	0.011
20	Wood and products of wood and cork, except furniture;	0.005	0.021
21	Pulp, paper and paper products	0.014	0.028
22	Publishing, printing and reproduction of recorded media	0.009	0.050
24	Chemicals and chemical products	0.015	0.088
25	Rubber and plastic products	0.050	0.071
26	Other non-metallic mineral products	0.024	0.039
28	Fabricated metal products, exc. machinery and equipment	0.006	0.061
29	Machinery and equipment n.e.c.	0.007	0.090
30	Office machinery and computers	0.298	0.013
31	Electrical machinery and apparatus n.e.c.	0.020	0.051
32	Radio, television, communication equipment and apparatus	0.048	0.045
33	Medical, precision, optical instruments, watches and clocks	0.022	0.033
34	Motor vehicles, trailers and semi-trailers	0.101	0.096
35	Other transport equipment	0.059	0.036

puted. In the last two columns the average over a number of years of the locations experiencing and not experiencing births during the time interval 1993-2002 is shown. It is possible to observe that for all sectors, a high number of zip-codes does not experience any new establishment births. This means that births of new establishments have been concentrated in a few number of zip-codes.

Studying the location effects only on zip-codes experiencing positive births would create a certain bias deriving from the elimination of those locations that were unable to attract at least one industrial establishment. To avoid this selection bias, we include all zip-codes whether they have firm births or not.

Empirical Implementation

Methodology

This section develops an econometric model to study the effects of local characteristics on new establishment births. One assumption

Table 2: New Entrants Spacial Distribution

Code	Industry	New Firms	Zip-codes experiencing births	Zip-codes non experiencing births
15	Food products and beverages	6091	5331	51553
17	Textiles	1193	1132	12067
18	Wearing apparel; dressing and dyeing of fur	10533	3854	8684
19	Tanning and dressing of leather	233	228	3833
20	Wood and products of wood and cork, except furniture;	481	464	16984
21	Pulp, paper and paper products	119	117	6271
22	Publishing, printing and reproduction of recorded media	12992	8786	35037
24	Chemicals and chemical products	470	422	12447
25	Rubber and plastic products	783	708	15746
26	Other non-metallic mineral products	652	542	14671
28	Fabricated metal products, exc. machinery and equipment	2188	2060	32894
29	Machinery and equipment n.e.c.	2201	2091	32629
30	Office machinery and computers	33	33	1067
31	Electrical machinery and apparatus n.e.c.	379	368	10334
32	Radio, television, communication equipment and apparatus	236	230	7106
33	Medical, precision, optical instruments, watches and clocks	932	906	17532
34	Motor vehicles, trailers and semi-trailers	134	134	6112
35	Other transport equipment	269	255	4165

of the model described above is that, in a certain location, the new establishment births follow a Poisson distribution with intensity parameter α (see equation 8).

For a given location, industry and time period, the intensity rate, and thus the expected number of births, is assumed to be given by

$$exp(z_s + \mu_{t-1} + k_i + y_{ist-1}) \tag{9}$$

The first term, z_s, represents time invariant location specific effects that account for differences across locations in the expected number of firms. This variable is very important since it also captures all omitted variables that influence the probability of positive profits e.g. being near to an airport, a port or a highway. The second term, μ_{t-1}, is a year specific effect that accounts over time for variables such as fluctuations in interests rates and in input prices, common to all locations, that might have a direct effect on firms profits. The third term k_i captures the presence of heterogeneity across two-digit sectors. The last term y_{ist-1} represents a vector of time varying local industrial characteristics that are expected to be productivity shifters and hence to have an effect on the probability of experiencing positive profits.

Considering $E(N_{sit})$ the expected number of new establishments in location s, industry i and time t, our regression is:

$$\begin{aligned} E(N_{ist}) &= z_s + y_{t-1} + k_i + \beta_1 loc_{ist}(zip-code) + \beta_2 loc_{ist}(1-5km) \\ &+ \beta_3 loc_{ist}(5-10km) + \beta_4 loc_{ist}(10-15km) + \lambda_1 urb_{st}(zip-code) \\ &+ \lambda_2 urb_{st}(1-5km) + \lambda_3 urb_{st}(5-10km) + \lambda_4 urb_{st}(10-15km) \\ &+ \gamma_1 w_{ist}(zip-code) + \gamma_2 w_{ist}(1-5km) + \gamma_3 w_{ist}(5-10km) \\ &+ \gamma_4 w_{ist}(10-15km) + \psi_1 w_{st}(zip-code) + \psi_2 w_{st}(1-5km) \\ &+ \psi_3 w_{st}(5-10km) + \psi_4 w_{st}(10-15km) + \varepsilon_{ist} \end{aligned} \tag{10}$$

We know that for our empirical analysis, the zip-codes for which the dependent variable is equal to zero (no entry) are relevant because values of the independent variables in these locations might explain why they have not been chosen by new entrants. In order to deal with the large amount of zero entries for the dependent variable we consider two non-linear estimation methods. The advantage of these models is that they take explicitly into account the nonnegativity and discreteness of our data. Moreover, the conditional versions of these models allow us to estimate a location fixed effects model.[4]

From these approaches a positive (negative) value of the coefficients in (10) implies that the expected number of new establishments increases (decreases) with the zip-code characteristic considered. Taking as an example β_1, it is possible to say that a higher level of own-indutry employment at zip-code level increases the probability of experiencing positive profits, which in turn will increase the

[4]An important departure of count data models from the linear models is that the individual specific effects are multiplicative rather than additive. Given the exponential form for $alpha_{ilt}$, individual variations can still been seen as a shift of the intercept. See Cameron and Triverdi, Chapter 9.

number of new establishments.

A group of empirical studies[5] on the effects of location characteristics on new entrants decisions, followed the approach of Guimaraes et all (2000) in which the authors show that the coefficients in (10) can be interpreted in an alternative way from the one that is derived from our model: given the equivalence between the likelihood function of a conditional Logit and a Poisson regression, the coefficients represent estimates of a conditional Logit model obtained directly from the framework of random utility maximization developed by McFaden (1974)[6].

It is important to take into account that one of the assumptions of McFaden's model is the so called Independence of Irrelevant Alternatives (IIA) which can be easily violated in this kind of context. Even with the introduction of location and time specific effects there may still exist some correlation across residuals not captured by these variables, i.e. homoskedasticity. Furthermore, these studies also assume that entrepreneurs choose a location taking into consideration only its expected profitability which may not be the case for the whole set of new establishments i.e. an entrepreneur can chose a certain location not because of its profitability but because he has an affiliation with that particular place. This last fact is explicitly accounted for in our work.

We will first perform our regression using the most popular specification of count data models: a Poisson model. This approach is useful when data at a very disaggregated territorial level are used. Nevertheless, Poisson models make an important assumptions. The

[5]List (2001), Viladecans-Marsal (2004), Holl (2004a), Arauzo Carod (2005), amongst others.

[6]The Random profit maximization problem assumes that an entrepreneur i will chose across J spatial choices the location that maximizes their expected profit function $\pi_{ij} = \beta \mathbf{z}_{ij} + \varepsilon_{ij}$, where $\mathbf{z}$ is a vector that represent location characteristics and ε_{ij} is a random term which has an iid Weibull distribution.

variance of the disturbances is proportional to the expected value of patents and weight the observations accordingly. This phenomenon, known as the over-dispersion problem, is usually violated when dealing with industrial location decisions: the concentration of entries in particular areas causes the variance to be greater than the mean. Furthermore, a drawback of Poisson Models is the inability to deal with situations where there are a large number of observations of the dependent count variable with zero values.

In order to deal with the inconsistencies of the Poisson model, a Negative Binomial specification is also used. The fixed effects Negative Binomial model that we consider is that of Palmgrem (1981) and Hausman, Hall, and Griliches(1984) where the variance is a multiple of the mean.

The Negative Binomial version of the model generalizes the Poisson specification by allowing for an additional source of variance above that due to pure sampling error. Because they make different assumptions about the variance structure they do yield different estimates of standard errors, even in the case of similar coefficients.

Results

In each regression, time and industry specific effects are included. All Tables also include the results of a Wald test on the appropriateness of the fixed effects model. Hausman tests strongly support the superiority of fixed effects over the random effects estimation. Tables 3 and 4report the estimates for the Poisson and the Negative Binomial fixed effects respectively.

The results show that there are significant localization economies at zip-code level. This means that the presence of own-industry firms encourages new firms in the same industry. Furthermore, the last column of the localization effects confirms one of the core pre-

Table 3: Poisson Fixed Effects

New Entrants	0 to 1 Km Ring	1 to 5 Km Ring	5 to 10 Km Ring	10 to 15 Km Ring
Localization Effects	0.0000437***	0.00000321	−0.00000403	−0.0000347***
	(0.00000972)	(0.00000627)	(0.00000522)	(0.00000849)
Urbanization Effects	0.0000199***	0.00000686***	0.00000103	0.00000103
	(0.00000418)	(0.00000125)	(0.000000733)	(0.000000746)
Industry average wages	0.119***	0.0630***	0.00573	0.00606
	(0.0211)	(0.0187)	(0.0276)	(0.0310)
Overall average wages	−1.182***	0.0187	0.0183	0.00915
	(0.0484)	(0.0151)	(0.0354)	(0.0378)
Log Likelihood	−16022.082			
Wald Test	1865.73***			
N. of Observations	104877			

Note: Standard errors in parentheses. */**/*** indicate significance at 10/5/1 per cent, respectively.

dictions of the standard New Economic Geography formulations:[7] the presence of agglomeration shadows. Fierce spatial price competition prevents urban areas from forming too closely to other equal or larger-size urban areas and causes population growth to be positively related to distance from (other) urban centers.

A negative and significant value of the localization economies coefficients for rings between 10 and 15 km (from the zip-code centroid) means that each zip-code is an economic driver far beyond its borders. These new patterns of agglomeration economies might reflect other economic factors such as improved transportation and communication technology, shifts in trade patterns and industry structural change.

There is evidence that zip-code level total employment has a positive and significant effect on the expected number of new entrants: the higher the number of jobs across sectors, the greater the attraction of the location. Furthermore, total employment coefficients are

[7]The concept of Agglomeration shadows was first introduced by Krugman (1993) who suggest that once a particular site is settled, its presence may skew further developments in its vicinity in its favor, via its agglomeration shadow.

Table 4: Negative Binomial Fixed Effects

New Entrants	0 to 1 Km Ring	1 to 5 Km Ring	5 to 10 Km Ring	10 to 15 Km Ring
Localization Effects	0.0000404^{***}	0.00000522	−0.00000395	-0.0000341^{***}
	(0.0000100)	(0.00000624)	(0.00000529)	(0.00000855)
Urbanization Effects	0.0000174^{***}	0.00000558^{***}	0.00000103	0.00000125
	(0.00000422)	(0.00000133)	(0.000000755)	(0.000000759)
Industry average wages	0.120^{***}	0.0610^{**}	0.00853	0.00814
	(0.0213)	(0.0191)	(0.0278)	(0.0319)
Overall average wages	-1.202^{***}	0.0183	0.0159	0.0043
	(0.0491)	(0.0156)	(0.0362)	(0.0397)
Log Likelihood	−15999.154			
Wald Test	1777.36^{***}			
N. of Observations	104877			

Note: Standard errors in parentheses. */**/*** indicate significance at 10/5/1 per cent, respectively.

smaller (but still positive) for rings that are further away from the zip-code centroid. This interesting pattern confirms the fact that urbanization effects decrease with distance.

In the literature, the results concerning the effects of urbanization economies on firm location are mixed. Due to the lack of data these studies use only one variable that acts not only as a proxy of urbanization economies but is also capturing factors such us land costs. If the coefficients representing this variable are positive the conclusion would be that the size of a location has a positive effect on firm productivity. If, on the other hand the coefficients are negative, the explanation would be that in bigger cities some congestion effects would be present, showing a negative influence on firm productivity.

In Rosenthal and Strange (2003), for example, the authors, who use our same definition of urbanization economies, do not find evidence that the zip-code local employment drives the expected number of new establishments. This is probably due to the fact that they do not control for congestion economies.

Other studies have introduced some variables to control for the negative effects of bigger locations. Very often, the square of the urbanization variables e.g. the square of the local employment, is used as a proxy of congestion effects. These kind of variables do not have a very strong explanatory power since they are based on strong assumptions of the presence of a non-linear relationship between urbanization economies and the probability of experiencing positive profits.

By introducing overall average wages as an explanatory variable, it is possible to capture the congestion effects. The negativeness and significance of its coefficients at zip-code level reflects how factors such as the price of land and the cost of commuting appear to lower the advantages of localization and urbanization effects.

Industry average wages have a positive effect on the entry of new firms. This result might reflect the presence of positive assortative matching, that is, good workers match on the labor market with good firms.[8] The existence of a positive correlation between individual skills and firm's expected profits is a key factor which is useful to explain standard aggregate measures of agglomeration economies. Our outcome can be seen as a very general[9] result of what other studies like Mion and Naticchioni (2007) find on the presence of externalities deriving from economic agent (individuals and firms) interactions.

In Tables 5 and 6 we separately analyze the effects that localization and urbanization variables of exporters and non-exporters would respectively have on the births of new establishments.

At zip code level, exporters and non-exporters employment in the own industry has a positive effect on the births of new establishments. The localization effects of exporters are statistically different

[8] See Becker (1973), Garicano, Rossi-Hansberg (2004).

[9] Due to the lack of data, the average wage variables of our study include both low and high skill workers wages, making it impossible to observe decompose the effect of different kinds of skills on expected profits.

Table 5: Poisson Fixed Effects (exporter and non-exporter agglomeration effects)

New Entrants	0 to 1 Km Ring	1 to 5 Km Ring	5 to 10 Km Ring	10 to 15 Km Ring
Localization Effects Exporters	0.0000500***	−0.000000821	−0.00000791	−0.0000399***
	(0.0000130)	(0.00000799)	(0.00000681)	(0.0000114)
Localization Effects Non Exporters	0.0000339*	0.0000159	0.00000571	−0.0000172
	(0.0000151)	(0.0000130)	(0.00000989)	(0.0000167)
Urbanization Effects Exporters	0.0000158***	0.00000763***	0.00000132	0.000000411
	(0.00000477)	(0.00000138)	(0.000000857)	(0.00000109)
Urbanization Effects Non Exporters	0.0000302***	0.00000476*	0.00000024	0.00000201
	(0.00000668)	(0.00000229)	(0.00000133)	(0.00000128)
Industry average wages	0.119***	0.0629***	0.00584	0.00647
	(0.0210)	(0.0187)	(0.0276)	(0.0310)
Overall average wages	−1.183***	0.0187	0.0186	0.00834
	(0.0485)	(0.0151)	(0.0353)	(0.0379)
Log Likelihood	−16017.88			
Wald Test	1871.94***			
N. of Observations	104877			

Note: Standard errors in parentheses. */**/*** indicate significance at 10/5/1 per cent, respectively.

and higher than those of the non-exporters. This result may derive from the evidence that has been encountered in many recent empirical studies in which exporting firms show a higher productivity than firms supplying only the domestic market.

Given the better performance of exporters, it may be that in a certain industry, due to the presence of size effects, the intra-industry technological spillovers effects of exporters located in a certain zip-code have a positive effect on the probability of experiencing positive profits, attracting in such a a way new entrants. Also in this case, for both exporters and non-exporting firms the localization effects become negative for the further concentric rings from the zip-code centroid (10km to 15km). These results, that are significant only for the former case, confirm once again the presence of agglomeration shadows.

Urbanization effects of exporters and non-exporters are positive and significant at local level and are higher for non-exporters. When

Table 6: Negative Binomial Fixed Effects (exporter and non-exporter agglomeration effects)

New Entrants	0 to 1 Km Ring	1 to 5 Km Ring	5 to 10 Km Ring	10 to 15 Km Ring
Localization Effects Exporters	0.0000450***	0.00000142	−0.00000778	−0.0000396***
	(0.0000132)	(0.00000792)	(0.00000688)	(0.0000115)
Localization Effects Non Exporters	0.0000335*	0.0000172	0.00000508	−0.0000163
	(0.0000153)	(0.0000127)	(0.00000987)	(0.0000164)
Urbanization Effects Exporters	0.0000134**	0.00000645***	0.00000121	0.000000669
	(0.00000480)	(0.00000145)	(0.000000886)	(0.00000110)
Urbanization Effects Non Exporters	0.0000275***	0.00000345	0.000000516	0.00000212
	(0.00000672)	(0.00000240)	(0.00000136)	(0.00000130)
Industry average wages	0.121***	0.0606**	0.00884	0.00854
	(0.0212)	(0.0190)	(0.0278)	(0.0319)
Overall average wages	−1.202***	0.0183	0.016	0.0036
	(0.0491)	(0.0156)	(0.0361)	(0.0397)
Log Likelihood	15995.174			
Wald Test	1786.23***			
N. of Observations	104877			

Note: Standard errors in parentheses. */**/*** indicate significance at 10/5/1 per cent, respectively.

it comes to firms that belong to other industries, it is not the concentration of exporters that is more important. This outcome may derive from the presence of supply chains in parts and components. It may be that while exporters produce more standardized products to sell abroad, non-exporters, producing more tailored goods have stronger linkages with the local producers.

The urbanization effects coefficients of both exporters and non-exporters are positive and decrease in magnitude across concentric rings. In both cases then, the urbanization effects are negatively related with distance.

As expected, the effects of the industry and overall average wages remain unaltered.

Table 7: Poisson Fixed Effects (aggregate rings)

New Entrants	1 Ring (zip-code)	1+2 Rings	1+2+3 Rings	1+2+3+4 Rings
Localization Effects	0.0000426***	0.0000167***	0.00000513	−0.000000577
	(0.00000792)	(0.00000367)	(0.00000263)	(0.00000249)
Urbanization Effects	0.0000379***	0.00000978***	0.00000392***	0.00000273***
	(0.00000369)	(0.000000876)	(0.000000363)	(0.000000268)
Industry average wages	0.116***	0.133***	0.154***	0.303***
	(0.0202)	(0.0173)	(0.0207)	(0.0273)
Overall average wages	−1.121***	−1.149***	−1.226***	−1.383***
	(0.0431)	(0.0418)	(0.0438)	(0.0504)
Log Likelihood	−16073.165	−16048.184	−16041.968	−16059.97
Wald Test	1769.82***	1825.54***	1870.11***	1840.38***
N. of Observations	104877	104877	104877	104877

Note: Standard errors in parentheses. */**/*** indicate significance at 10/5/1 per cent, respectively.

Robustness checks

In the previous regressions a number of concentric rings were created in order to study the scope of location forces. In order to control for the presence of substitution effects across rings on all coefficients, different regressions have been computed considering the first ring only, aggregating the first and the second ring, the first second and third rings and all of the rings respectively.

In Tables 7 and 8 it is possible to observe that all the coefficients still maintain their signs and consequently the results of our study still hold.

The relationship between location characteristics and new establishments births still cannot be interpreted as completely causal due to some reverse causality problems. The endogeneity of our results can be driven by to factors: first it is possible that new entrants in period $t-1$ are aware of the economic characteristic of the zip-code in which they want to locate at time t and hence might have an effect themselves on agglomeration economies. Second, it is possible

Table 8: Negative Binomial Fixed Effects (aggregate rings)

New Entrants	1 Ring (zip-code)	1+2 Rings	1+2+3 Rings	1+2+3+4 Rings
Localization Effects	0.0000391***	0.0000168***	0.00000602*	0.000000537
	(0.00000846)	(0.00000367)	(0.00000263)	(0.00000249)
Urbanization Effects	0.0000299***	0.00000775***	0.00000309***	0.00000220***
	(0.00000377)	(0.000000939)	(0.000000379)	(0.000000278)
Industry average wages	0.117***	0.136***	0.157***	0.312***
	(0.0206)	(0.0173)	(0.0209)	(0.0277)
Overall average wages	−1.142***	−1.174***	−1.254***	−1.416***
	(0.0436)	(0.0420)	(0.0442)	(0.0510)
Log Likelihood	−16037.254	−16019.063	−15999.442	−16005.19
Wald Test	1681.27***	1741.5***	1783.72***	1765.23***
N. of Observations	104877	104877	104877	104877

Note: Standard errors in parentheses. */**/*** indicate significance at 10/5/1 per cent, respectively.

that the positive effects of agglomeration economies on new firms births derives partially from the exit of the least productive firms from the market.

To deal with the first endogeneity issue we run a set of regressions considering explanatory variables that are lagged by two periods. Although the introduction of a bigger lag may not rule out completely the reverse causality bias, it can make it less severe. For the second case we run a Poisson and a Negative Binomial model considering as dependent variable the net entry defined as the total number of new establishments at time t minus the total number of firms exiting the market the same year.[10]

From the last four Tables, we can observe that the results remain largely unchanged, that is, the estimates on localization and

[10]In a few cases, for a certain year and location the number of firms exiting the market was higher that the number of new entrants. Since the count data models we use in our studies require a dependent variable that is non-negative we have dropped the zip-codes for which the dependent variable was smaller than zero.

Table 9: Poisson Fixed Effects (net entry)

Net Entry	0 to 1 Km Ring	1 to 5 Km Ring	5 to 10 Km Ring	10 to 15 Km Ring
Localization Effects	0.0000399***	0.000000861	−0.00000222	−0.0000310***
	(0.0000104)	(0.00000672)	(0.00000539)	(0.00000861)
Urbanization Effects	0.0000186***	0.00000673***	0.000000675	0.000000816
	(0.00000429)	(0.00000129)	(0.000000756)	(0.000000768)
Industry average wages	0.124***	0.0637***	0.0017	0.00527
	(0.0196)	(0.0185)	(0.0289)	(0.0309)
Overall average wages	−1.204***	0.0204	0.0219	0.0111
	(0.0479)	(0.0148)	(0.0363)	(0.0374)
Log Likelihood	−15850.172			
Wald Test	1764***			
N. of Observations	104557			

Note: Standard errors in parentheses. */**/*** indicate significance at 10/5/1 per cent, respectively.

Table 10: Negative Binomial Fixed Effects (net entry)

Net Entry	0 to 1 Km Ring	1 to 5 Km Ring	5 to 10 Km Ring	10 to 15 Km Ring
Localization Effects	0.0000358***	0.00000343	−0.000002	−0.0000299***
	(0.0000107)	(0.00000669)	(0.00000545)	(0.00000861)
Urbanization Effects	0.0000160***	0.00000518***	0.000000686	0.00000111
	(0.00000435)	(0.00000140)	(0.000000783)	(0.000000776)
Industry average wages	0.126***	0.0616**	0.00452	0.00791
	(0.0198)	(0.0188)	(0.0291)	(0.0317)
Overall average wages	−1.225***	0.02	0.0195	0.00603
	(0.0485)	(0.0153)	(0.0370)	(0.0392)
Log Likelihood	−15826.738			
Wald Test	1687.9***			
N. of Observations	104557			

Note: Standard errors in parentheses. */**/*** indicate significance at 10/5/1 per cent, respectively.

Table 11: Poisson Fixed Effects (one period lagged independent variables)

New Entrants	0 to 1 Km Ring	1 to 5 Km Ring	5 to 10 Km Ring	10 to 15 Km Ring
Localization Effects$_{t-2}$	0.0000420***	0.00000471	−0.000003	−0.0000271**
	(0.00000989)	(0.00000627)	(0.00000541)	(0.00000886)
Urbanization Effects$_{t-2}$	0.0000148***	0.00000673***	0.00000116	0.000000221
	(0.00000440)	(0.00000131)	(0.000000772)	(0.000000833)
Industry average wages$_{t-2}$	0.0137	0.0441	−0.0288	0.00168
	(0.0570)	(0.0287)	(0.0358)	(0.0445)
Overall average wages$_{t-2}$	−0.665***	0.00859	0.0501	−0.0205
	(0.0730)	(0.0224)	(0.0379)	(0.0491)
Log Likelihood	−14495.339			
Wald Test	1280.15***			
N. of Observations	92216			

Note: Standard errors in parentheses. */**/*** indicate significance at 10/5/1 per cent, respectively.

Table 12: Negative Binomial Fixed Effects (one period lagged independent variables)

New Entrants	0 to 1 Km Ring	1 to 5 Km Ring	5 to 10 Km Ring	10 to 15 Km Ring
Localization Effects$_{t-2}$	0.0000386***	0.00000667	−0.00000334	−0.0000261**
	(0.0000101)	(0.00000620)	(0.00000549)	(0.00000890)
Urbanization Effects$_{t-2}$	0.0000117**	0.00000515***	0.00000131	0.000000371
	(0.00000448)	(0.00000141)	(0.000000799)	(0.000000846)
Industry average wages$_{t-2}$	0.0142	0.038	−0.0281	0.00727
	(0.0578)	(0.0317)	(0.0379)	(0.0444)
Overall average wages$_{t-2}$	−0.679***	0.00707	0.0472	−0.0266
	(0.0744)	(0.0242)	(0.0398)	(0.0512)
Log Likelihood	−14474.229			
Wald Test	1183.63***			
N. of Observations	92216			

Note: Standard errors in parentheses. */**/*** indicate significance at 10/5/1 per cent, respectively.

urbanization economies are still positive and significant at zip-code level. The overall average wages still have a negative and significant impact on the probability of starting a new activity in a certain location. With respect to the industry average coefficients, they are still positive and significant for the regressions using net entry as a dependent variable while they become non significant after the introduction of two years lagged explanatory variables.

Concluding Remarks

This paper has focused on the impact that zip-code level characteristics have on the expected number of new manufacturing establishments. In order to capture inter-sectoral and intra-sectoral agglomeration effects, localization and urbanization variables have been computed respectively at a very detailed geographical level. We separately investigate the scale effects stemming from exporting and non-exporting firms. The impact that intra-sectoral and overall averages wages could have on location decisions has also been considered. In addition, the scope of agglomeration and congestion effects has been addressed by considering interactions between neighboring locations.

Following Monseny (2005), a simple model has been developed in order to theoretically explain the expected number of entrants into a certain location. Variations on the births of new establishments across locations derives from either variations in the abundance of firms already present in the location or variations in the expected profits of the potential incumbents. Focusing on the latter factor, from the model it derives that the expected profits increase (decrease) on location characteristics that have a positive (negative) scale effect on productivity.

In order to empirically analyze the importance of environmental location characteristics on a firm location decision it is assumed that the expected number of new establishments births follows a Poisson distribution. Having said that, we use count data models to perform our analysis. This approach enables us to deal with the discreteness and the high number of zeros of the dependent variable. Poisson and Negative Binomial regressions have been performed using French firm level data at a very detailed geographical scale (zip-code) for the time interval 1993-2002. v From the empirical results the following should be emphasized. At zip-code level, there is a positive and significant effect of localization and urbanization economies on the expected number of new establishments in a certain location. Furthermore, the localization effects generated by exporting firms are usually higher than those generated by non-exporters. For the urbanization economies, instead, the concentration of non-exporters has a positive impact on new births that is significantly higher than that of exporting firms.

Whilst the urbanization effects decrease with distance, the localization effects decrease across rings and become negative and significant for the rings that are more further away from the zip-code centroid. This last result confirms the presence of agglomeration shadows. Distance is a variable that is crucial for an entrepreneur during the decision on where to locate a new economic activity. This because the positive agglomeration effects of a certain industry take place at a certain industry-specific distance, in such a way that there will be an area between locations that is in the agglomeration shadow of the industry and thus is not going to be a profitable location.

There is a positive effect of industry average wages on the probability of experiencing positive profits while there is a negative and significant effect of overall average wages on firms location decisions. The sign of this last variable can be interpreted as the presence of congestion economies.

In the last part of our study some robustness checks are performed and our findings suggest that after controlling for reverse causality problems and for substitution effects across rings, the effects of zip-code characteristics on firm location decisions are still present and significant.

References

Arauzo-Carod, J. M. (2005): "Determinants of industrial location: An application for Catalan municipalities," *Papers in Regional Science*, 84(1), 105–120.

Arauzo Carod, J. M., and M. C. Manjón Antoln (2006): "Firm Size and Geographical Aggregation: An Empirical Appraisal in Industrial Location," Working Papers 66, Universitat Rovira i Virgili, Department of Economics.

Becker, G. S. (1973): "A Theory of Marriage: Part I," *Journal of Political Economy*, 81(4), 813–46.

Cameron, A. C., and P. Triverdi (1998): *Regression Analysis of Count Data*. New York: Cambridge University Press.

Carlton, D. W. (1983): "The Location and Employment Choices of New Firms: An Econometric Model with Discrete and Continuous Endogenous Variables," *The Review of Economics and Statistics*, 65(3), 440–49.

Ciccone, A., and R. E. Hall (1996): "Productivity and the Density of Economic Activity," *American Economic Review*, 86(1), 54–70.

Coughlin, C. C., and E. Segev (1997): "Location determinants of new foreign-owned manufacturing plants," Discussion paper.

Duranton, G., and H. G. Overman (2003): "Testing for Localisation Using Micro-Geographic Data," Royal Economic Society Annual Conference 2003 69, Royal Economic Society.

Ellison, G., and E. L. Glaeser (1994): "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," NBER Working Papers 4840, National Bureau of Economic Research, Inc.

Figueiredo, O., P. Guimaraes, and D. Woodward (2002): "Home-field advantage: location decisions of Portuguese entrepreneurs," *Journal of Urban Economics*, 52(2), 341–361.

Garicano, L., and E. Rossi-Hansberg (2004): "Inequality and the Organization of Knowledge," *American Economic Review*, 94(2), 197–202.

Glaeser, E. L., H. D. Kallal, J. A. Scheinkman, and A. Shleifer (1992): "Growth in Cities," *Journal of Political Economy*, 100(6), 1126–52.

Guimaraes, P., O. Figueiredo, and D. Woodward (2000): "Agglomeration and the Location of Foreign Direct Investment in Portugal," *Journal of Urban Economics*, 47(1), 115–135.

Hausman, J. A., B. H. Hall, and Z. Griliches (1984): "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," NBER Technical Working Papers 0017, National Bureau of Economic Research, Inc.

Henderson, J. V. (1986): "Efficiency of resource usage and city size," *Journal of Urban Economics*, 19(1), 47–70.

Henderson, J. V., A. Kuncoro, and M. Turner (1994): "Industrial Development in Cities," NBER Working Papers 4178, National Bureau of Economic Research, Inc.

Holl, A. (2004a): "Manufacturing location and impacts of road transport infrastructure: empirical evidence from Spain," *Regional Science and Urban Economics*, 34(3), 341–363.

——— (2004b): “Start-ups and relocations: Manufacturing plant location in Portugal,” *Journal of Economics*, 83(4), 649–668.

——— (2004c): “Transport Infrastructure, Agglomeration Economies, and Firm Birth: Empirical Evidence from Portugal,” *Journal of Regional Science*, 44(4), 693–712.

Hoover, E. (1936): “The Measurement of Industrial Location,” *The Review of Economics and Statistics*, 18, 162–171.

Krugman, P. (1993): “On the number and location of cities,” *European Economic Review*, 37(2-3), 293–298.

List, J. A. (2001): “US county-level determinants of inbound FDI: evidence from a two-step modified count data model,” *International Journal of Industrial Organization*, 19(6), 953–973.

Marshall, A. (1890): *Principles of Economics*. Mc. Millan New York.

Mion, G., and P. Naticchioni (2007): “The spatial sorting and matching of skills and firms,” Mpra paper, University Library of Munich, Germany.

Monseny, J. J. (2005): “On the scope of agglomeration economies: Evidence from Catalan zip codes,” Discussion paper.

Moomaw, R. L. (1981): “Productivity and City Size? A Critique of the Evidence [Are There Returns to Scale in City Size?]. [Bias in the Cross Section Estimates of the Elastic,” *The Quarterly Journal of Economics*, 96(4), 675–88.

Nakamura, R. (1985): “Agglomeration economies in urban manufacturing industries: A case of Japanese cities,” *Journal of Urban Economics*, 17(1), 108–124.

Rosenthal, S. S., and W. C. Strange (2001): "The Determinants of Agglomeration," *Journal of Urban Economics*, 50(2), 191–229.

——— (2003): "Geography, Industrial Organization, and Agglomeration," Center for Policy Research Working Papers 56, Center for Policy Research, Maxwell School, Syracuse University.

Sveikauskas, L. A. (1975): "The Productivity of Cities," *The Quarterly Journal of Economics*, 89(3), 393–413, available at http://ideas.repec.org/a/tpr/qjecon/v89y1975i3p393-413.html.

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