On the scope of agglomeration economies:
Evidence from Catalan Zip Codes

Jordi Jofre Monseny
Universitat de Barcelona & Institut d’Economia de Barcelona (IEB)
Octubre, 2005

ABSTRACT: This paper aims at studying the scope of agglomerations economies empirically. In particular, two issues are explored. First, the industrial scope of agglomeration economies is analysed, by comparing the effects arising from co-localization of same industry firms (localization economies) to the benefits derived from large and diversified economic environments (urbanization/Jacobs diversity effects). Second, the geographic scope of these external effects is studied. These issues are addressed by studying the effects of local industrial characteristics on the one number of births of new establishments in the subsequent period. A theoretical framework is used to interpret regression results in terms of scale effects (productivity shifters). Econometric estimations are carried out, separately, for seven industries for Catalonia, which is a Spanish region, using 1997-2000 data. Evidence of localization, urbanization and diversity effects has been found. Agglomeration economies seem to work at a very local level.

Key words: agglomeration economies, firm creation, Poisson regression

Jel Codes: L25, R30

Contact Address:
Avenida Diagonal 690, Torre 4, Plant 2ª / 08034 Barcelona
e-mail: jjofre@pcb.ub.es

*I am grateful to Elisabet Viladecans Marsal, Marco Francesconi, Pilar Sorribas Navarro and Christian Saborowski for valuable comments.
1.-Introduction

External effects exist when the economic scale of the geographical location, a firm is located in, enhances its productivity (Rosenthal and Strange, 2004). There is a rich literature aiming at explaining why firms co-locate in space and how this fact results in productivity differences across firms found in different locations\(^1\). The existence of these external scale effects has important policy implications. A good understanding of these phenomena can help in designing policies aiming at fostering particular industries at the local and regional level and, also at guiding more general policies on local and regional growth.

The empirical literature on agglomeration economies is very large\(^2\). A great deal of this literature has focused on whether it is specialized economic environments (Localization/\textit{Marshall-Arrow-Romer} externalities) or large and diversified cities (\textit{urbanization/Jacobs} diversity effects) that generate larger scale effects. Empirical studies have found results pointing in different directions (Rosenthal and Strange, 2004). Hence, this question remains unsolved. Much less applied work has analyzed which geographic scope these external effects have, since data at a geographically detailed level has not been available until recent times. Seminal papers of Glaeser \textit{et al.} (1992) and Henderson \textit{et al.} (1995) both use data at the USA Metropolitan Statistical Area (MSA) level. Two examples of recent work on the empirics of agglomeration economies at a more local level are Duranton and Overman (2002) and Rosenthal and Strange (2003), who use United Kingdom and United States Zip Code level data, respectively. By means of mapping software, these authors have georeferenced their datasets and are, thus, able to study how external economies’ effects vary when

\(^1\) See Duranton and Puga (2004) for an extensive review.

\(^2\) See Rosenthal and Strange (2004) for an extensive review.
considering interactions of agents located at different geographic distances\textsuperscript{3}. These two studies conclude that agglomeration economies take place at a small geographic scale.

If external effects are productivity shifters, the most straightforward way to quantify these effects is by means of estimation of production functions. However, this approach requires very detailed data on the inputs which is very rarely available. Any omitted input, that turned out to be correlated with some variable summarizing industrial environment characteristics, would lead to a bias in the estimation of agglomeration economies’ effects on productivity. The size of this bias has been found to be huge in the literature (Moomaw, 1983). Many other approaches to study the empirics of agglomeration economies, which do not require data on the usage of inputs, have been proposed (Rosenthal and Strange, 2004). The analysis of the determinants of births of new establishments is one of them\textsuperscript{4}. Focusing on new establishments is appealing because it enables to treat the existing economic environment as given and decisions taken by new establishments are not influenced by prior choices (Rosenthal and Strange, 2003). The paper by Rosenthal and Strange (2003) is the most closely related one to the analysis presented here. These authors study the impact of pre-existing local industrial characteristics on the number of firm births. Controlling for differences in entrepreneur abundances, a positive effect of a certain local industrial characteristic on the birth of new establishments is taken as evidence of existing agglomeration economies. The authors specify a linear relationship between the number of births of new establishments and local industrial characteristics. They use the Tobit model to deal with the fact that a very high share of Zip Codes does not experience any

\textsuperscript{3} These are the only two papers I am aware of that perform this exercise.

birth for a given year. However, the Tobit approach presents two limitations in this context. First, it fails to account for the discrete nature of the dependent variable. Second, it considers the zero outcome as a result of censoring, when it is a natural outcome of the variable being modelled. Guimares et al. (2003) consider the Tobit approach to be difficult to justify in this context.

The objectives of this paper are twofold. In the first place, establish a theoretically driven econometric model that: i) takes into consideration that different locations can have different entrepreneur abundances; ii) deals, in a natural way, with the zero outcome and the integer nature of the dependent variable (number of births); and iii) makes possible to interpret the statistically significant positive effect of an industrial local characteristic on the number of new establishments’ births as existence of a scale effect. Secondly, carry out an empirical application to shed some light on the scope of agglomeration economies. In particular, two issues will be addressed. First, the relative importance of external effects arising from same industry or different industries co-localization of firms is studied. Particular attention is drawn to differences these effects may exert in different types of industries. Second, by georeferencing the database used, the geographic scope of agglomeration economies is analysed.

This paper studies Zip Code level data on new establishments’ births. The analysis is restricted to Catalonia (a Spanish region) and establishments being born between 1997 and 2000. The industries analysed are: Textiles, Wood and furniture, Chemical products, Fabricated metals except for machinery, Motor vehicles Manufacture of radio, television and communication equipments and Medical, precision and optical instruments.

This paper is organized as follows. After this introduction, Section 2 reviews the strand of the literature that has studied the nature and industrial scope of agglomeration
economies. In Section 3, the model that backs the econometric analysis is presented. Section 4 deals with the empirical analysis. After describing the data and variables (4.1 and 4.2), the chosen econometric specification (4.3) is explained and justified. Then, results (4.4) are presented and discussed. Section 4 finishes with robustness analysis. Finally, concluding remarks are provided in Section 5.

2.-Agglomeration economies

As mentioned in the introduction, external economies exist when the scale of the urban environment adds to productivity (Rosenthal and Strange, 2004). This is to say that agglomeration economies emerge as a consequence of summing up individual external effects stemming from the interaction of firms located in the same geographical environment. Many mechanisms that explain the rationale for firms to co-locate have been proposed in the literature. A very well known typology is the one inspired in the work of Marshall. Marshall (1890) points out three main advantages stemming from the co-localization of agents: Labour market pooling, input sharing and knowledge spillovers. What is meant by the labour market pooling externality is that the co-localization of industrial activity in the same geographical area enables both firms and workers to share risks of demand fluctuations at the individual level. Input sharing refers to benefits arising from the fact that concentrations of firms from a particular industry may promote specialized input industries to flourish. Finally, knowledge spillovers external effects occur because geographic proximity fosters knowledge transmission amongst firms. However, incentives for agents to disperse may appear as city sizes increases. Agglomeration of economic activity may increase competition for immobile factors of production, raising the price of production inputs (Devereux et al.,
2003). Other incentives for firms to disperse may include non-priced congestion costs such as traffic congestion and pollution.

The benefits for two firms to localize close in space are very likely to vary depending not only on the geographic distance, but also on the industrial closeness of their activities. Addressing inter-firm industrial closeness implies defining what industrial proximity is. This definition will always, to some extent, be arbitrary. Ellison and Glaeser (1997) defined the concept of coagglomeration and derived a measure that can easily be computed. Nevertheless, that is only one possibility. Probably being explained by this conceptual difficulty, most studies treat industrial distance in a binary fashion, i.e., firms belonging to the same industry or not. This leads to the localization and urbanization economies distinction first proposed by Hoover (1934). Localization economies are externalities arising between firms belonging to the same industrial activity. The term urbanization economies stands for external effects taking place between firms producing loosely connected products, as well as the advantages derived from city size as, for instance, the development of financial and commercial services.

Romer (1986) places knowledge spillovers and learning by doing at the core of economic growth. Glaeser et al. (1992) aim at testing some growth implications at the local level. This paper stresses the role of knowledge spillovers as a mechanism explaining why cities form and grow. As the distinction between localization and urbanization economies found in the more static marshallian approach, the distinction between intrasectoral and intersectoral effects has also been an issue in this dynamic externalities literature. The Marshall-Arrow-Romer (MAR) externalities concern knowledge spillovers amongst firms within an industry. MAR economies imply that sectors, which are overrepresented in a city, should experience higher growth rates than the average since technology levels raise as industry size grows. The, somehow,
opposite vision that it is not specialization but industry diversity that promotes innovation and growth is usually identified with Jane Jacobs’ hypothesis. Jacobs (1969) claims and presents some evidence that it is the interaction amongst not very related industries that foster growth through cross-fertilization of ideas.

Empirical work has not been conclusive with respect to the relative importance of intersectoral or intrasectoral external effects. Applied work on the industrial scope of agglomeration economies has shown that the effects of localization/MAR and urbanization/diversity economies are very different between industrial sectors. Although not overwhelming, there is evidence that localization/MAR economies have stronger effects for low and middle levels of sectoral technology intensity whereas urbanization and diversity economies are particularly relevant when considering high-tech industries. Henderson et al. (1995) first stressed this result. Similar evidence has been found by Combes (2000) and Viladecans-Marsal (2004) for France and Spain, respectively.

3.-The model

This section aims at providing an analytical framework that explains why some geographical locations experience more births of new establishments than others. It is assumed that differences in new establishments’ births across locations can be explained by two phenomena: differences in the number of entrepreneurs and differences in the probability of establishments to experience positive profits. By entrepreneur in a given location, sector and time period, denoted \( i, j \) and \( t \), respectively, I refer to a person thinking about opening up an establishment in this particular location, sector and time period. It is assumed that the number of entrepreneurs is a random outcome that can be reasonably well described by a Poisson distribution. In a context of uncertainty about the individual efficiency level of the establishment, it is assumed that an entrepreneur
will randomize over the decision to open up an establishment according to the probability of experiencing positive profits, when inputs are chosen optimally. Then, relying on one property of the Poisson distribution, it is argued that the number of new establishments’ start-ups is also Poisson distributed.

If an entrepreneur decides to settle a new industrial establishment, input levels \( x_1, \ldots, x_L \) are chosen to maximize the following profit function\(^5\), \( \pi \):

\[
\pi(x; y) = a(y) \cdot f(x) \cdot (1 + \varepsilon) - c(x)
\]  

where the output prices have been normalized to one and time, sector and location subscripts are omitted; \( x \) is a vector of \( L \) rows accounting for the inputs chosen by the entrepreneur (land, labour, capital, raw materials,\ldots); \( y \) is a vector of \( M \) rows that summarizes the industrial characteristics of the geographical location of the firm; \( f(x) \) is the production function which is supposed to take positive values if \( \forall l, \; x_l > 0 \); \( c(x) \) is a positive linear function of the unitary input costs; \( a(y) \) is a positive function that shifts the production function; and \( \varepsilon \) is a firm specific term that reflects heterogeneity across firms and is identically and independently distributed (\textit{iid}). This last term enable some establishments, using the same technology and input levels, to produce more than others, reflecting different managerial abilities. The solution of the problem yields the following \( L \) first order conditions that, at the optimum, must equal zero:

\[
a(y) \cdot f_i(x) \cdot (1 + \varepsilon) - w_l = 0, \; \forall l = 1, \ldots, L
\]  

where \( f_i \) denotes the partial derivative of \( f(x) \) with respect to \( l \) and is supposed to be a decreasing function; and \( w_l \) denotes the unitary cost of the \( l^{th} \) input. The first order

\(^5\) The literature has considered agglomeration economies to be a supply shifter (Rosenthal and Strange, 2004). Henderson (1986) found some evidence in favour of the hypothesis that agglomeration economies are Hicks-neutral, implying that the ratio between marginal productivities is held constant regardless of the industrial environment the firm is found.
conditions imply that any factor is hired up to a positive level where its marginal productivity equals its marginal cost. By substituting the optimal input choices back into the profit function, the value function which only depends on parameters is obtained. This expression resembles the one proposed by Rosenthal and Strange (2003) in the way the managerial ability and external effects enter the profit function\(^6\).

\[
V(y, \varepsilon, w_1, \ldots, w_i) = \text{Max}_x \{a(y) f(x)(1 + \varepsilon) - c(x)\} \tag{3}
\]

The entrepreneur does not know the managerial ability of her establishment before starting up the business. However, she knows that this managerial ability is randomly drawn from a known distribution. The assumption this work relies on is that the entrepreneur will decide to create a new establishment with the exact probability with which the start-up will experience positive profits. \(\varepsilon\) is assumed to be bounded between minus and plus one and is distributed according to the distribution function \(F(\varepsilon)\), that maps \(\varepsilon\) into the probability space. It is assumed that for any given \(y\) and \(w\), there is a unique threshold value for \(\varepsilon\), \(\hat{\varepsilon}\), such that \(V(y, \varepsilon, w_1, \ldots, w_i) = 0\). Given observed values of local industrial characteristics, \(y\), and inputs costs, \(w\), in period \(t\), which are supposed to remain in period \(t+1\), \(F(\varepsilon)\) is the probability that the managerial ability in period \(t+1\) will be lower than the threshold level required to obtain positive profits. Thus, an entrepreneur will start-up an establishment with probability \(1 - F(\varepsilon)\), which is nothing but the probability of experiencing positive profits. This probability is increasing (decreasing) in any industrial characteristic of the local environment that shifts the production function upward (downward\(^7\)), i.e., this probability increases (decreases) if \(\frac{\partial a(y)}{\partial y_m} > 0\) (\(\frac{\partial a(y)}{\partial y_m} < 0\)) and decreases in any input price, \(w\). To see that,

---

\(^6\) Rosenthal and Strange (2003) take a similar approach. They, however, assume rather than show that an increase in any characteristic shifting the production function upward imply a higher probability for an establishment to experience positive profits.

\(^7\) This may be explained, for instance, by non-priced congestion costs.
evaluate the value function at \( \hat{e} \). From the definition of \( \hat{e} \), it follows that \( V(y, w, \hat{e}) = 0 \). Applying the implicit function theorem and making use of the envelope theorem it follows that:

\[
\frac{\partial \hat{e}}{\partial y_m} = -\frac{\partial a(y) / \partial y_m \cdot f(x^\ast) \cdot (1 + \hat{e})}{a(y) \cdot f(x^\ast)}
\]

(4)

\[
\frac{\partial \hat{e}}{\partial w_i} = -\frac{-x^*_i}{a(y) \cdot f(x^\ast)}
\]

(5)

where \( x^\ast \) denotes the vector of optimal inputs. Given the assumptions of the model, it follows that (5) is a positive expression implying that higher input prices will result in a higher value of \( \hat{e} \) and, thus, in a lower probability of an entrepreneur to decide to start-up an establishment. The sign of expression (4) depends on the fact that the \( m^{th} \) local characteristic increases or decreases productivity. If it increases (decreases) productivity then (4) is a negative (positive) expression implying that higher values for the \( m^{th} \) characteristic will lead to a higher (lower) likelihood of experiencing positive profits.

The implications of these results are that differences in costs and in the economic characteristics of geographical locations can cause that, given the same number of entrepreneurs, locations with lower costs and particular economic environments to experience more births of new establishments than others.

It is assumed that the fact that a person becomes an entrepreneur and, thus, considers to start-up an establishment happens to people with certain probability. If this probability is low and the set of people who can become entrepreneurs is large, then it follows that the number of entrepreneurs considering to start up an establishment, \( E \), will follow, asymptotically, a Poisson distribution:

\[
\Pr(E = e) = \exp(-\alpha) \cdot \alpha^e / e!
\]

(6)

\(^8\) This follows from assuming that becoming an entrepreneur is a rare event. It also must be assumed that probabilities across observations are independent.
where the mean and variance of the distribution are given by the intensity or rate parameter, $\alpha$, which is allowed to vary across locations and sectors. Thus, for a given location, sector and time period, the number of entrepreneurs is a realization of a Poisson distribution with intensity parameter $\alpha_{ijt}, E_{ijt} \sim P(\alpha_{ijt})$. A known result in statistics is that if the number of repetitions ($E_{ijt}$) of a binary (zero or one) identically and independently distributed (iid) event (to experiment positive profits or not) is a realization of a Poisson distribution, then, the value of the sum of this $E_{ijt}$ repeated binary outcome will follow a Poisson distribution with intensity parameter $\alpha_{ijt} \cdot (1 - F(\varepsilon_{ijt}))$, where $1 - F(\varepsilon_{ijt})$ is the probability that the binary event takes the value of one (Cameron and Trivedi, 1998). In this particular problem, this implies that the number of births taking place in a given location, sector and time period, $B_{ijt}$, can be characterized by a Poisson distribution with rate parameter that depends on the entrepreneur abundance of the location, since $\alpha_{ijt}$ determines the expected value of $E_{ijt}$, as well as the probability of reaping positive profits if starting up a business in this location, $1 - F(\varepsilon_{ijt})$:

$$B_{ijt} \sim P(\alpha_{ijt} \cdot (1 - F(\varepsilon_{ijt})))$$

(7)

### 4.-Empirical application

#### 4.1.-Data

The data set used in this work has been obtained from two different sources: the Spanish National Social Security Registry and the Central Directory of Firms. The former
contains data on Zip Code\textsuperscript{9} employment levels at the two digit sectoral classification. The latter records all new establishments born in Spain and contains establishment level information, including the two digit sectoral classification and the Zip Code geographical location. The analysis is carried out for the period 1997-2000 period and is restricted to Catalonia.

As already mentioned, analysing the different role that intrasectoral and intersectoral external effects may play across industries is one of the goals of this paper. Therefore, industries have been chosen in a way to represent heterogeneous industrial activities. The importance of the industrial sector in terms of employment has also been considered. The industries chosen account for almost half of the industrial employment (see Table I). Textiles and Wood and furniture are the low technology industries chosen since their employment shares (9.8 \% and 7\% of total industrial employment, respectively) are the highest within their type\textsuperscript{10}. Sectors showing an intermediate technology intensity include Chemical products, Fabricated metals except for machinery and Motor vehicles that account, as a whole, for almost 30\% of industry employment. Regarding high-tech industries, only data for Manufacture of radio, television and communication equipments and Medical, precision and optical instruments industries are rich enough to be worth analysing. Summed up, their employment levels do not even reach 3\% of industrial employment. Table I highlights some features of industries’ employment for the analysed industries and Table II summarizes data on births of new establishments.

As can be seen in Table II, a striking feature of data on the birth of new establishments is that for all sectors, a high number of Zip Codes do not experience any

\textsuperscript{9} In Catalonia, except for Barcelona, the Zip Code equals the municipality. Currently, there are 946 Catalan municipalities. Instead the current analysis is restricted to 945 Zip Codes because one split during 1997 and 2000.

\textsuperscript{10} OCDE classification of industries according to different levels of technology intensity has been used.
new establishment’s birth (third column of Table II). In fact, the inhabitants of 385 out of 945 Zip Codes have not seen any industrial establishment being started-up in their Zip Code during the 1997-2000 period. By comparing, for each industry, the number of new establishments born and the number of Zip Codes experiencing births (first and second columns in Table II) it can be inferred that establishments’ start-ups have to be concentrated on some locations.

**Table I** Industrial and Overall employment shares and Spatial Gini Index for selected sectors. Employment data for 2000.

<table>
<thead>
<tr>
<th>Sector</th>
<th>Industrial employment share</th>
<th>Overall employment share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Textiles</td>
<td>9.80%</td>
<td>2.54%</td>
</tr>
<tr>
<td>Wood and Furniture</td>
<td>7.02%</td>
<td>1.82%</td>
</tr>
<tr>
<td>Chemical products</td>
<td>9.80%</td>
<td>2.54%</td>
</tr>
<tr>
<td>Metal products except for machinery</td>
<td>11.99%</td>
<td>3.11%</td>
</tr>
<tr>
<td>Motor vehicles</td>
<td>9.02%</td>
<td>2.34%</td>
</tr>
<tr>
<td>Radio, television and communication equipments</td>
<td>1.45%</td>
<td>0.38%</td>
</tr>
<tr>
<td>Medical precision and optical instruments</td>
<td>1.36%</td>
<td>0.35%</td>
</tr>
</tbody>
</table>

*Source*: National Social Security Registry and own elaboration.

**Table II**: New establishments’ births summary data. 1997-2000 aggregated data.

<table>
<thead>
<tr>
<th>Sector</th>
<th>New establishments</th>
<th>Zip Codes experiencing births</th>
<th>Zip Codes not experiencing births</th>
</tr>
</thead>
<tbody>
<tr>
<td>Textiles</td>
<td>393</td>
<td>123</td>
<td>822</td>
</tr>
<tr>
<td>Word and Furniture</td>
<td>732</td>
<td>250</td>
<td>695</td>
</tr>
<tr>
<td>Chemical products</td>
<td>164</td>
<td>92</td>
<td>853</td>
</tr>
<tr>
<td>Metal products except for machinery</td>
<td>1237</td>
<td>254</td>
<td>691</td>
</tr>
<tr>
<td>Motor vehicles</td>
<td>82</td>
<td>57</td>
<td>888</td>
</tr>
<tr>
<td>Radio, television and communication equipments</td>
<td>30</td>
<td>25</td>
<td>920</td>
</tr>
<tr>
<td>Medical, precision and optical instruments</td>
<td>69</td>
<td>35</td>
<td>910</td>
</tr>
</tbody>
</table>

*Source*: Dirce and own elaboration.
4.2.-Variables

The dependent variable, $B_{ijt}$, is the number of new establishments’ births that occur in each Zip Code for a given industry and time period. The relevant industrial characteristics for sector $j$ and location $i$ are assumed to be industry $j$’s local employment level ($loc$), overall local employment level ($urb$), the square of the overall local employment level ($cong$), and a proxy of the local degree of sectoral diversity ($div$)$^{11}$. Following Glaeser et al. (1992), Henderson et al. (1995) and Rosenthal and Strange (2003), industry $j$’s employment level aims at capturing localization economies (or Marshall-Arrow-Romer economies in a dynamic context) whereas overall employment level is expected to reflect the advantages of city size (i.e., urbanization economies). The square of the overall employment level is expected to capture congestion effects as done in the work of Arauzo (2005). Locations with similar overall levels of employment can show very different economic environments and, thus, a diversity index is introduced to better characterize intersectoral external effects. Besides, this diversity index will enable us to test some hypothesis associated with Jacobs (1969). The diversity index used is nothing but the inverse of a Hirshmann-Herfindahl index. This index has been used in Duranton and Puga (2000) and Rosenthal and Strange (2003), among others. This index is given by

$$div_i = 1/\sum_j s_{ij}^2$$

where $s_{ij}$ denotes the share of overall employment in location $i$ that is devoted to industry $j$. The larger the value of the index, the more diverse the described economic environment is.

$^{11}$ Localization and urbanization economies variables are measured as in Rosenthal and Strange (2003). Congestion effects are captured as in Arauzo (2005).
As mentioned in the introduction, agglomeration economies are thought to take place at a local scale but, evidence of these effects to spill over between local administrative borders has been found (Henderson, 2003; Rosenthal and Strange, 2003; and Viladecans-Marsal, 2004). In order to study the geographic scope of agglomeration economies, industrial characteristics of surrounding Zip Codes are also considered. Following a similar approach to that of Rosenthal and Strange (2003), own industry and overall employment levels contained in two different concentric rings from Zip Code’s centroid have been computed. The up to 10 km concentric ring of location \( i \) includes all Zip Code locations whose Euclidean distance\(^{12} \) between its centroid and location \( i \)'s centroid is inferior to 10 km. In the same way, the 10 to 20 km concentric ring of location \( i \) includes all Zip Codes whose described distance with location \( i \) is between 10 and 20 Kilometres. Thus, localization and urbanization economies, as well as congestion effects for location \( i \), are characterized by three different variables, namely, Zip Code employment levels (\( ZC \)), up to 10 km concentric ring employment levels (\(<10 km\)) and 10 to 20 km concentric ring employment levels (\(10\text{-}20 \ km\)). Regarding diversity effects, also the effects of surrounding Zip Codes have been considered. However, to compute the Diversity Index for employment contained in different concentric rings would have been very cumbersome. Instead, the Diversity Index of the Local Labour Market (\( LLM \)) each Zip Code belongs to has been included\(^{13} \). Table A.1 in the Annex provides summary statistics of the data.

\(^{12} \) Euclidean distances have been computed using UTM \( xy \) coordinates. The \( xy \) coordinates for each Zip Code’s centroid have been obtained through ArcView mapping software.

\(^{13} \) The aggregations of Zip Codes used here have been constructed with a slightly different methodology than the one used to obtain the British Local Labour Markets. For details see Roca and Moix (2004).
4.3.-Econometric specification

This section sets up an econometric model that enables us to quantify and test relationships between data described above. The model outlined in Section 3 is the starting point of the econometric analysis and, becomes a conceptual framework that enables us to interpret results in a causal way. As outlined above, the observed number of new establishments births taking place in a given location and sector is supposed to be a realization of a Poisson process, with intensity parameter $\alpha_{ijt} \cdot (1 - F(\varepsilon_{ij}))$. The regression model is obtained by assuming that this intensity rate varies across observations according to observable and unobservable variables. For a given location and time period, the intensity rate and, thus, the expected number of births is assumed to be given by $\exp(\mu_{ij} + y_{jt-1} + \beta'z_{ijt-1})$, where $\mu_{ij}$ is a time invariant location specific effect. This term accounts for differences across locations in the expected number of entrepreneurs and in time invariant profit determinants such as cost differentials; $y_{jt-1}$ is a year specific effect, which reflects variation over time in variables that are common to all locations. This accounts for changes in variables such as interest rates, economic downturns and raw materials prices, which are thought to drive both the expected number of entrepreneurs and the probability of experiencing positive profits; $z_{ijt-1}$ is a $k^{th}$ column vector of time varying local industrial characteristics that are expected to be productivity shifters and thus, drive the probability to experience positive profits; and $\beta'$ is a $k^{th}$ row vector of unknown coefficients. Notice that time varying covariates are one period lagged. This follows from the model outlined in Section 3. Entrepreneurs assume that profit determinants in period $t$ will be given by the ones observed in period $t-1$. 

16
Given the database used here, the intensity rate characterizing the Poisson distribution, which \( B_{it} \) is supposed to follow, is given, within an industry, by:

\[
a_{it} \cdot (1 - F(\hat{e}_{it})) = E(B_{it}) = \mu_i + y_{t-1} \\
+ \beta_{11} \text{loc}(ZC)_{t-1} + \beta_{12} \text{loc}( < 10 \text{ km})_{t-1} + \beta_{13} \text{loc}(10 - 20 \text{ km})_{t-1} \\
+ \beta_{21} \text{urb}(ZC)_{t-1} + \beta_{22} \text{urb}( < 10 \text{ km})_{t-1} + \beta_{23} \text{urb}(10 - 20 \text{ km})_{t-1} \\
+ \beta_{31} \text{cong}(ZC)_{t-1} + \beta_{32} \text{cong}( < 10 \text{ km})_{t-1} + \beta_{33} \text{cong}(10 - 20 \text{ km})_{t-1} \\
+ \beta_{41} \text{div}(ZC)_{t-1} + \beta_{42} \text{div}(LLM)_{t-1}
\]

(9)

For instance, a positive statistically significant estimate of \( \beta_{11} \) implies that the expected number of new establishments being born increases with the Zip Code own industry employment level. Given the model outlined in Section 3, this can be interpreted as follows. A higher level of own industry local employment shifts the productivity of the local establishments and, given a fixed expected number of entrepreneurs, this will result in a higher number of births, since the probability of experiencing positive profits is higher. Thus, a positive and statistically significant estimate for \( \beta_{11} \) can be interpreted as evidence for the existence of localization economies at the Zip Code level.

Maximum likelihood is the standard procedure to estimate the vector of unknown parameters \( \hat{\beta} \). As mentioned before, the mean and the variance for a Poisson distribution are assumed to be the same. This is the so-called equidispersion property of the Poisson distribution. Most of the data do not satisfy this assumption. Nevertheless, the consistency of the coefficients' estimates does not rely on this assumption and will hold, as long as the conditional mean is correctly specified (Cameron and Trivedi, 1998). However, if the conditional variance does not equal the conditional mean, the maximum likelihood covariance matrix estimator will be inconsistent, leading to incorrect statistical inference. If the conditional mean is correctly specified, a consistent estimate of the covariance matrix of the coefficients, when it is evaluated at the
maximum likelihood ones, can be obtained through a robust Sandwich estimator (Cameron and Trivedi, 1998):

The coefficient that captures differences across locations in the expected number of entrepreneurs and in time invariant profit determinants, \( \mu_i \), can, in principle, be different for each Zip Code. This would lead to a Poisson regression with year and Zip Code specific dummies. This is equivalent to a two-way Poisson fixed effects model\(^{14}\) (Cameron and Trivedi, 1998). Given the panel structure of the data set used, this estimation can be carried out. However, results are not satisfactory due to a poor efficiency of the estimates. To solve this problem, the values of \( \mu_i \) are restricted to be equal for all Zip Codes belonging to the same Local Labour Market. Therefore, \( \mu_i \) stands for a Local Labour Market time invariant specific effect. Given that some Local Labour Markets do not experience any birth for the whole period, a dummy for these Local Labour Markets cannot be fitted. For this reason, only the observations belonging to Local Labour Markets with some births have been considered and, thus, the regression for each industry has been estimated with a different number of observations.

4.4.-Results

Table III shows the Poisson Pseudo-Maximum likelihood estimates arising from expression (9). Estimates for Local Labour Market and year dummies are not reported to save space. The last row of Table III reports log-likelihood ratio tests for the null hypothesis that the model is jointly statistically not significant. For all sectors analysed the null can be rejected at very high confidence levels.

\(^{14}\) In particular, this leads to a Poisson with multiplicative fixed effects.
All sectors but two (Radio, television and communication equipments and Medical precision and optical instruments industries) show statistically significant localization economies’ effects at the Zip Code level ($\beta_{1i} > 0$). All sectors but two (Textiles and Motor vehicles industries) show statistically significant urbanization economies effects at the Zip Code level ($\beta_{2i} > 0$). This is in line with the results obtained by Henderson et al. (1995), Combes (2000) and Viladecans-Marsal (2004). Localization effects fail to show in high-tech industries (Radio, television and communication equipments and Medical precision and optical instruments) whereas urbanization effects have the smallest size in low-tech industries like the traditional type Textiles industry. Although not significant at the 5% level of significance, the urbanization economies’ coefficient ($\beta_{2i}$) for the Motor Vehicles industry, shows a $t$-statistic that is close to denote statistical significance. In contrast, Rosenthal and Strange (2003) do not find evidence that the Zip Code local employment level drives the expected number of new establishments. This is probably due to the fact that they do not control for congestion effects. By introducing, as an explanatory variable, the squared of the local employment level, the relationship between urbanization economies and new establishments’ births is allowed to be non-linear. This turns out to be quite successful.\textsuperscript{15} For all industries showing urbanization effects at the Zip Code level ($\beta_{2i} > 0$), including the motor vehicles industry, these effects seem to be decreasing with the economic size of the Zip Code ($\beta_{3i} < 0$). An interpretation is that productivity increases with the local employment level but, as this raises, congestion effects appear and lower the advantages of city size.

\textsuperscript{15} When not controlling for congestion effects, urbanization economies fail to show as in Rosenthal and Strange (2003)
### Table III. Agglomeration economies’ estimates. Poisson pseudo Maximum Likelihood estimates

<table>
<thead>
<tr>
<th>Localization economies</th>
<th>Textiles</th>
<th>Wood and Furniture</th>
<th>Chemical products</th>
<th>Metal products except for machinery</th>
<th>Motor vehicles</th>
<th>Radio, television and communication equipments</th>
<th>Medical precision and optical instruments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zip Code</td>
<td>0.0008392</td>
<td>0.0011435</td>
<td>0.0004866</td>
<td>0.0005737</td>
<td>0.0013187</td>
<td>-0.0012633</td>
<td>-0.0001398</td>
</tr>
<tr>
<td>**</td>
<td>(5.06)***</td>
<td>(3.75)***</td>
<td>(2.57)**</td>
<td>(4.26)***</td>
<td>(2.67)***</td>
<td>(-1.22)</td>
<td>(-0.16)</td>
</tr>
<tr>
<td>up to 10 km</td>
<td>-0.0000465</td>
<td>0.0000004</td>
<td>0.0000066</td>
<td>0.0000055</td>
<td>0.0000687</td>
<td>-0.0003862</td>
<td>-0.0005449</td>
</tr>
<tr>
<td>10 to 20 km</td>
<td>(-1.31)</td>
<td>(0.01)</td>
<td>(0.31)</td>
<td>(0.39)</td>
<td>(0.57)</td>
<td>(-0.91)</td>
<td>(-1.50)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Urbanization economies</th>
<th>Textiles</th>
<th>Wood and Furniture</th>
<th>Chemical products</th>
<th>Metal products except for machinery</th>
<th>Motor vehicles</th>
<th>Radio, television and communication equipments</th>
<th>Medical precision and optical instruments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zip Code</td>
<td>-0.0000170</td>
<td>0.0000203</td>
<td>0.0000292</td>
<td>0.0000186</td>
<td>0.0000198</td>
<td>0.0000506</td>
<td>0.00000419</td>
</tr>
<tr>
<td>**</td>
<td>(0.9)</td>
<td>(3.39)***</td>
<td>(4.98)***</td>
<td>(3.42)***</td>
<td>(1.51)</td>
<td>(4.08)***</td>
<td>(3.59)***</td>
</tr>
<tr>
<td>up to 10 km</td>
<td>-0.0000140</td>
<td>0.0000031</td>
<td>0.0000028</td>
<td>0.0000052</td>
<td>-0.0000045</td>
<td>0.0000062</td>
<td>0.0000127</td>
</tr>
<tr>
<td>10 to 20 km</td>
<td>-0.0000048</td>
<td>-0.000017</td>
<td>-0.000004</td>
<td>-0.000009</td>
<td>-0.000011</td>
<td>-0.000016</td>
<td>-0.0000007</td>
</tr>
<tr>
<td></td>
<td>(-1.22)</td>
<td>(-0.22)</td>
<td>(-1.02)</td>
<td>(-0.19)</td>
<td>(-0.39)</td>
<td>(-1.35)</td>
<td>(-2.21)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Congestion effects</th>
<th>Textiles</th>
<th>Wood and Furniture</th>
<th>Chemical products</th>
<th>Metal products except for machinery</th>
<th>Motor vehicles</th>
<th>Radio, television and communication equipments</th>
<th>Medical precision and optical instruments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zip Code</td>
<td>1.79·10^{-11}</td>
<td>-1.27·10^{-11}</td>
<td>-3.6·10^{-11}</td>
<td>-8.96·10^{-12}</td>
<td>-3.29·10^{-11}</td>
<td>-5.06·10^{-11}</td>
<td>-4.68·10^{-11}</td>
</tr>
<tr>
<td>**</td>
<td>(0.73)</td>
<td>(-2.56)**</td>
<td>(-2.71)**</td>
<td>(-1.18)</td>
<td>(-2.23)**</td>
<td>(-2.13)**</td>
<td>(-3.7)**</td>
</tr>
<tr>
<td>up to 10 km</td>
<td>1.31·10^{-11}</td>
<td>-3.44·10^{-12}</td>
<td>-3.02·10^{-12}</td>
<td>-4.56·10^{-12}</td>
<td>-7.95·10^{-12}</td>
<td>-2.35·10^{-12}</td>
<td>-6.04·10^{-12}</td>
</tr>
<tr>
<td>10 to 20 km</td>
<td>4.34·10^{-12}</td>
<td>1.21·10^{-12}</td>
<td>2.97·10^{-13}</td>
<td>7.43·10^{-13}</td>
<td>-1.13·10^{-12}</td>
<td>4.05·10^{-12}</td>
<td>2.36·10^{-12}</td>
</tr>
<tr>
<td></td>
<td>(2.72)**</td>
<td>(1.06)</td>
<td>(1.02)</td>
<td>(4.37)</td>
<td>(1.43)</td>
<td>(0.9)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Diversity effects</th>
<th>Textiles</th>
<th>Wood and Furniture</th>
<th>Chemical products</th>
<th>Metal products except for machinery</th>
<th>Motor vehicles</th>
<th>Radio, television and communication equipments</th>
<th>Medical precision and optical instruments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zip Code</td>
<td>0.2242746</td>
<td>0.2489029</td>
<td>0.2365901</td>
<td>0.2472408</td>
<td>0.1703331</td>
<td>0.2653343</td>
<td>0.2886914</td>
</tr>
<tr>
<td>**</td>
<td>(8.72)***</td>
<td>(14.0)***</td>
<td>(6.35)***</td>
<td>(16.79)***</td>
<td>(3.22)**</td>
<td>(2.77)***</td>
<td>(4.21)***</td>
</tr>
<tr>
<td>LLM</td>
<td>-0.0625806</td>
<td>0.0765863</td>
<td>0.2553995</td>
<td>-0.0252734</td>
<td>-0.1590840</td>
<td>0.4452644</td>
<td>0.1321790</td>
</tr>
<tr>
<td>**</td>
<td>(-0.53)</td>
<td>(0.89)</td>
<td>(1.33)</td>
<td>(-0.34)</td>
<td>(-0.63)</td>
<td>(0.66)</td>
<td>(0.21)</td>
</tr>
</tbody>
</table>

| N                      | 2756      | 3624               | 2532               | 3440                               | 2152          | 844                                           | 1852                                      |
| LR-Test                | 1499.4*** | 1882***            | 390.6***           | 3588***                           | 158.4**       | 69.3***                                       | 348***                                    |

**Notes:** 1. Figures in parenthesis are t-statistics. 2. *, **, ***: statistically significant at the 90%, 95% and 99% confidence levels, respectively.
The negative and statistically significant coefficient that overall employment levels exert on the expected number of Textiles new establishments’ births ($\beta_{22} < 0$ and $\beta_{23} < 0$), can be due to the fact that benefits stemming from local employment levels are overcome by non-priced congestion costs. The estimates for the Medical, precision and optical instruments industry show a negative, statistically significant and unexpectedly large $\beta_{12}$ coefficient. This implies that higher levels of the up to 10 Km own industry employment ($loc(<10\,km)$) produce lower expected numbers of new establishments’ births. This result cannot be explained arguing that congestion costs overcome the benefits of agglomeration, since congestion is not likely to be caused by own industry levels, particularly. Diversity effects at the Zip Code level are positive and statistically significant determinants of productivity for all sectors analysed and, the size of the coefficients is similar across industrial sectors analysed ($\beta_{41} > 0$). This is consistent with Jacobs (1969) hypothesis who stresses the benefits arising from a diversified economic environment. Glaeser et al. (1992) and Rosenthal and Strange (2003) also find that sectoral diversity exerts an important external effect. In the Poisson regression with exponential mean function, coefficients do not have a marginal effect interpretation which, for the $k^{th}$ variable is given by $\partial E[B_{ij}]/\partial z_k = \exp(\beta'z) \cdot \beta_k$. Notice that variables aiming at capturing localization and urbanization effects have a common scale (number of workers). This implies that if one coefficient is 10 times larger than another one, the marginal effect is also ten times larger, given a unit change in both variables (Cameron and Trivedi, 1998). For all sectors showing both urbanization and localization economies (Wood and Furniture, Chemical products, Metal products except for machinery and Motor vehicles) the coefficient measuring localization effects is, at least, of one order of magnitude larger than the urbanization effects’ one. This is consistent with Rosenthal and Strange (2003). However, this does
not help to get a sense of how large these effects are, neither to compare the effects of diversity with localization and urbanization economies. A problem when evaluating the marginal effects is that these change across individuals due to different characteristics, \( z_{it-1} \). In the Poisson regression with exponential mean function the average response for the \( j^{th} \) variable and a given industry is given by

\[
\frac{1}{N} \sum_{i=1}^{N} \frac{\partial E[B_{it}]}{\partial x_{it-1}} = \beta \cdot \bar{B},
\]

where \( \bar{B} \) denotes the sample average of \( B_{it} \) (Cameron and Trivedi, 1998). Using this expression, marginal averaged effects have been calculated for one standard deviation increase for the Zip Code level variables: diversity index (\( \text{div}(zc) \)), own industry employment level (\( \text{loc}(zc) \)) and overall employment level (\( \text{urb}(zc) \)). The same exercise has been performed but considering instead, one hundred extra workers increase in the employment levels (this can not be performed for the diversity index).

### Table IV: Averaged marginal effect of one standard deviation and one hundred workers increase on the expected number of new establishments’ births

<table>
<thead>
<tr>
<th>Sector</th>
<th>Localization economies one std.dev</th>
<th>Localization economies 100 workers</th>
<th>Urbanization economies one std.dev</th>
<th>Urbanization economies 100 workers</th>
<th>Diversity effects one std.dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Textiles</td>
<td>0.0087</td>
<td>0.0284</td>
<td>-0.0431</td>
<td>-0.0002</td>
<td>0.0700</td>
</tr>
<tr>
<td>Wood and Furniture</td>
<td>0.0221</td>
<td>0.0468</td>
<td>0.0958</td>
<td>0.0004</td>
<td>0.1447</td>
</tr>
<tr>
<td>Chemical products</td>
<td>0.0021</td>
<td>0.0141</td>
<td>0.0309</td>
<td>0.0001</td>
<td>0.0308</td>
</tr>
<tr>
<td>Metal products except for machinery</td>
<td>0.0188</td>
<td>0.0708</td>
<td>0.1484</td>
<td>0.0006</td>
<td>0.2429</td>
</tr>
<tr>
<td>Motor vehicles</td>
<td>0.0029</td>
<td>0.0206</td>
<td>0.0105</td>
<td>0.00004</td>
<td>0.0111</td>
</tr>
<tr>
<td>Radio, television and communication equipments</td>
<td>-0.0010</td>
<td>-0.0010</td>
<td>0.0098</td>
<td>0.00004</td>
<td>0.0063</td>
</tr>
<tr>
<td>Medical, precision and optical instruments</td>
<td>-0.0003</td>
<td>-0.0002</td>
<td>0.0186</td>
<td>0.0001</td>
<td>0.0158</td>
</tr>
</tbody>
</table>

Averaged marginal effects are small and have an order of magnitude similar to the one found by Rosenthal and Strange (2003). The Metal products except for
machinery industry show the largest averaged marginal effects for both urbanization and localization economies, given a hundred extra workers increase in the relevant variable levels. For this industry, one hundred extra workers increase the expected number of firm births by 0.07, if these workers belong to the same industry and 0.0006, otherwise. A different picture of the size of these effects across industrial sectors is obtained when considering, instead, a one standard deviation increase in employment levels. This is due to the fact that standard deviations are not insensitive to the scale of variables and aggregate employment levels are, for obvious reasons, larger than for single industries. Regarding the effects of diversity, averaged marginal effects for this variable seem to be larger than both urbanization and localization economies when considering a one standard deviation change in variable levels. The Wood and Furniture industry shows the highest marginal estimated effect. This result is, in a way, surprising since intuitively one would expect diversity effects to be more relevant in industries with a more intensive use of technology.

All localization, urbanization and diversity effects seem to have a very local scope. There is no evidence of diversity effects to spill over Catalan Zip Codes. The Local Labour Market diversity index, \( \text{div}(LLM) \), does not seem to exert any positive effect on the expected number of firm births. This result is reasonable since diversity effects are thought to stem from knowledge spillovers and these require face to face contacts. Only four, out of seven sectors analysed, provide strong evidence that either localization or urbanization effects spill over neighbouring Zip Codes. Localization effects do not die out within Catalan municipalities for the Textiles and Motor Vehicles industries \((\beta_{12} > 0)\). Higher neighbouring Zip Codes overall employment levels imply higher expected numbers of new establishments’ births for the Metal products except for machinery and the Medical, precision and optical instruments \((\beta_{22} > 0)\). All
证据表明外部效应的溢出主要来自就业水平较高的区域，位于1到10公里的同心圆区域内。结果未显示任何统计显著性的证据表明工业特征对10至20公里同心圆区域内就业水平的影响。对于所有四例中发现的外部效应，这些效应至少下降了一个数量级。这证实了罗斯内尔和斯特兰格（2003）的发现，即集聚经济效应随着距离的增加而急剧下降。

5. 结论

本论文分析了两个关于集聚经济影响范围的问题。首先，针对本地化（本地化影响）和区域化（都市化/多样性影响）规模经济相对重要性的讨论。其次，研究了这些外部经济的地理范围。这些问题通过观察工业特征如何影响所在环境确定后续期间预期的新生企业数量得到探索。提出了一种解释新生企业数量的模型。它假定不同地点的企业建立差异源于企业家丰度的不同以及经历正向利润的概率差异。它表明任何增加生产力的工业特征都可能带来正向影响，这可以解释为外部效应的存在。在这个假设下，新生企业的数量可以描述为泊松分布。

已经发现证据表明存在本地化和都市化经济，对于七个行业中五个行业。这些外部
effects is limited. There is weak evidence of localization economies to be relevant in high technology sectors (Radio, television and communication equipments and Medical precision and optical instruments industries). Higher current overall employment levels diminish the one period ahead expected number of Textiles new establishments’ births. This last result may reflect the presence of congestion costs that overcome the benefits arising form co-localization. The more diverse the economy of a Zip Code is, the higher the expected number of births of new establishments, other things being equal, for all industries analysed.

Localization economies seem to exert a bigger effect than urbanization economies at the Zip Code level, when these effects are evaluated by the impact of a hundred extra workers in pre-existing employment levels on the expected number of births (marginal average effects). However, a one standard deviation change in the Zip Code overall employment level and in the diversity index (also an intersectoral effect), yield a far larger effect on the expected number of births of new establishments than a one standard deviation change in the own industry Zip Code employment level. External effects have a very limited geographic scope. In most cases, no sign of external effects to spill over Zip Codes is found. When this is not the case, the size of these scale effects decrease in at least on order of magnitude.

The fact that agglomeration economies work at a very local scale along with the relatively important role that a diversified economy plays on firms’ productivity, points in the direction that knowledge spillovers play a crucial role in explaining the geographic concentration of industrial activity. Notice that if input sharing and labour market pooling were the main sources underlying agglomeration economies, then it would be difficult to explain why benefits of agglomeration die out so soon and why Zip Code diversity matters so much.
References


