IEB

Institut d'Economia de Barcelona

CENTRE DE RECERCA EN FEDERALISME FISCAL I ECONOMIA REGIONAL

Document de treball 2005/4:

On the scope of agglomeration economies: Evidence from Catalan zip codes

Jordi Jofre Monseny

Institut d'Economia de Barcelona
Espai de Recerca en Economia
Facultat de Ciències Econòmiques i Empresarials
Universitat de Barcelona
Carrer del Tinent Coronel Valenzuela, 1-11
08034 Barcelona
Tel.: 93 403 46 46
Fax: 93 403 72 42
E-mail: ieb@pcb.ub.es
http://www. pcb.ub.es/ieb

ON THE SCOPE OF AGGLOMERATION ECONOMIES: EVIDENCE FROM CATALAN ZIP CODES [a]

Jordi Jofre Monseny [b, c]

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.

Keywords: Agglomeration economies, firm creation, Poisson regression.
JEL Classification: L25, R30.

RESUMEN: El objetivo de este trabajo es estudiar dos cuestiones relacionadas con el alcance de las economías de aglomeración a nivel empírico. En primer lugar, se analiza el alcance industrial de las economías de aglomeración a través de la comparación de los efectos que surgen de la colocalización de empresas que pertenecen a una misma industria (economías de localizacón) con los beneficios que aparecen en grandes concentraciones urbanas con estructuras productivas más diversificadas (economías de urbanización/diversidad). En segundo lugar, se estudia el alcance geográfico de estas economías externas. Estas cuestiones son abordadas mediante el análisis empírico de los efectos de las características industriales locales preexistentes sobre la creación local de empresas. En el trabajo se presenta un marco analítico que permite interpretar los resultados obtenidos en términos the efectos externos de escala en la producción. Las estimaciones econométricas se llevan a cabo, de forma separada, para siete sectores industriales con datos de municipios catalanes correspondientes al período 1997-2000. Los resultados indican la presencia de economías de localización así como como de urbanización/diversidad. Las economías de aglomeración tienen un alcance geográfico muy limitado.

Palabras clave: Economías de aglomeración, creación de empresas, modelo de Poisson.
Classificación JEL: L25, R30.

---

[a] Comments are welcome. The opinions expressed in the paper do not necessarily reflect the IEB's opinions.

[b] I am grateful to Elisabet Viladecans Marsal, Marco Francesconi, Pilar Sorribas Navarro and Christian Saborowski for valuable comments.

[c] Corresponding address: jjofre@pcb.ub.es
Dep. d'Economia Política i Hisenda Pública & Barcelona Institute of Economics (IEB)
Facultat de Ciències Econòmiques i Empresarials - Universitat de Barcelona
Av. Diagonal 690, Torre 4, Planta 2 - 08034 Barcelona (SPAIN)
Phone: + 34 93 402 18 12 / Fax: + 34 93 402 18 13

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) [1]. 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[2]. 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[3]. A great deal of this literature has focused on whether it is specialized economic environments (Localization/Marshall-Arrow-Romer externalities) or large and diversified cities (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 *et al.* (1992) and Henderson *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 considering interactions of agents located at different geographic distances[4]. These two studies conclude that agglomeration economies take place at a small geographic scale.

---

[1] I use agglomeration economies, external economies, external effects and scale effects without any difference in meaning.

[2] See Duranton and Puga (2004) for an extensive review.

[3] See Rosenthal and Strange (2004) for an extensive review.

[4] These are the only two papers I am aware of that perform this exercise.

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[5]. 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 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

---

[5] Studies that analyze agglomeration economies by looking at new establishment births include Carlton (1983), Coughlin *et al.* (1991), Friedman *et al.* (1992), Devereux *et al.* (2003) and Rosenthal and Strange (2003). Arauzo (2005) studied the determinants of new firm births for Catalan municipalities but the hypothesis he focuses on differ from the ones that motivate this current work.

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[6] (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

[6] Notice that this does not imply that all forces driving co-location of agents in space take place through productivity shifts. Krugman (1991), for instance, presents a model of labour market pooling where expected profits increase with the economic size of firms' location without a "productivity" effect.

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[7]. 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[8]. 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.

---

[7] Other mechanisms not mentioned by Marshall (1980) have also been proposed in the literature such as home market effects, urban consumption opportunities and rent-seeking (Rosenthal and Strange, 2004). Recent work on the micro-foundations of agglomeration economies has led to a different taxonomy of agglomeration economies. Duranton and Puga (2004) propose to classify agglomeration economies according to the sort of mechanism generating them: sharing, learning and matching economies. However, empirical work still relies on the marshallian taxonomy to a large extent.

[8] An alternative related interpretation has to do with a better matching quality. Hesley and Strange (1990) show that an increase in the number of agents trying to match in each location improves the expected quality of the match.

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 $x_1, ... x_l, ..., x_L$ are chosen to maximize the following profit function[9], π:

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

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,…); 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 ε is a firm specific term that reflects heterogeneity across firms and is identically and independently distributed (*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[10]:

$$a(y) \cdot f_l(x) \cdot (1+\varepsilon) - w_l = 0, \ \forall l = 1,...,L \tag{2}$$

[9] 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.

[10] The fact that *f(x)* is strictly concave along with the fact that *c(x)* is a linear and, thus, a convex function implies that the profit function (a sum of two concave functions) is concave and thus, the solution described by first order conditions is indeed a maximum.

where f_l 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 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[11].

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

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. ε is assumed to be bounded between minus and plus one and is distributed according to the distribution function $F(\varepsilon)$, that maps ε into the probability space. It is assumed that for any given y and w, there is a unique threshold value for ε, $\hat{\varepsilon}$, such that $V(y,\hat{\varepsilon},w_1,...w_l)=0$. Given observed values of local industrial characteristics, y, and inputs costs, w, in period t, which are supposed to remain in period $t+1$, $F(\hat{\varepsilon})$ 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 $1-F(\hat{\varepsilon})$, 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[12]), *i.e.*, this probability increases (decreases) if $\partial a(y)/\partial y_m > 0$ $(\partial a(y)/\partial y_m < 0)$ and decreases in any input price, w. To see that, evaluate the value function at $\hat{\varepsilon}$. From the definition of $\hat{\varepsilon}$, it follows that $V(y,w,\hat{\varepsilon})=0$. Applying the implicit function theorem and making use of the envelope theorem it follows that:

[11] 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.

[12] This may be explained, for instance, by non-priced congestion costs.

$$\frac{\partial \hat{\varepsilon}}{\partial y_m} = -\frac{(\partial a(y)/\partial y_m)\cdot f(x^*)\cdot(1+\hat{\varepsilon})}{a(y)\cdot f(x^*)} \tag{4}$$

$$\frac{\partial \hat{\varepsilon}}{\partial w_l} = -\frac{-x_l^*}{a(y)\cdot f(x^*)} \tag{5}$$

where x^* 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 $\hat{\varepsilon}$ 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:[13]

$$\Pr(E = e) = \exp(-\alpha)\cdot\alpha^e / e! \tag{6}$$

where the mean and variance of the distribution are given by the intensity or rate parameter, α, which is allowed to vary across locations and sectors[14]. Thus, for a given location, sector and time period, the number of entrepreneurs is a realization of a

[13] This follows from assuming that becoming an entrepreneur is a rare event. It also must be assumed that probabilities across observations are independent.

[14] The exposure time of the process has been normalized to unity.

Poisson[15] 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 $\alpha_{ijt} \cdot (1 - F(\hat{\varepsilon}_{ijt}))$, where $1 - F(\hat{\varepsilon}_{ijt})$ 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 α_{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, $1 - F(\hat{\varepsilon}_{ijt})$:

$$B_{ijt} \sim P(\alpha_{ijt} \cdot (1 - F(\hat{\varepsilon}_{ijt}))) \quad (7)$$

4. Empirical application

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[16] 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.

[15] The first and second central moments, the mean and variance, of a Poisson distribution are given by the occurrence rate (Cameron and Trivedi, 1998).

[16] 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.

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[17]. 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 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.

[17] OCDE classification of industries according to different levels of technology intensity has been used.

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

Sector	Industrial employment share	Overall employment share	Spatial Gini Index
Textiles	9.80%	2.54%	0.10
Wood and Furniture	7.02%	1.82%	0.06
Chemical products	9.80%	2.54%	0.01
Metal products except for machinery	11.99%	3.11%	0.07
Motor vehicles	9.02%	2.34%	0.02
Radio, television and communication equipments	1.45%	0.38%	0.04
Medical precision and optical instruments	1.36%	0.35%	0.03

Source: National Social Security Registry and own elaboration.

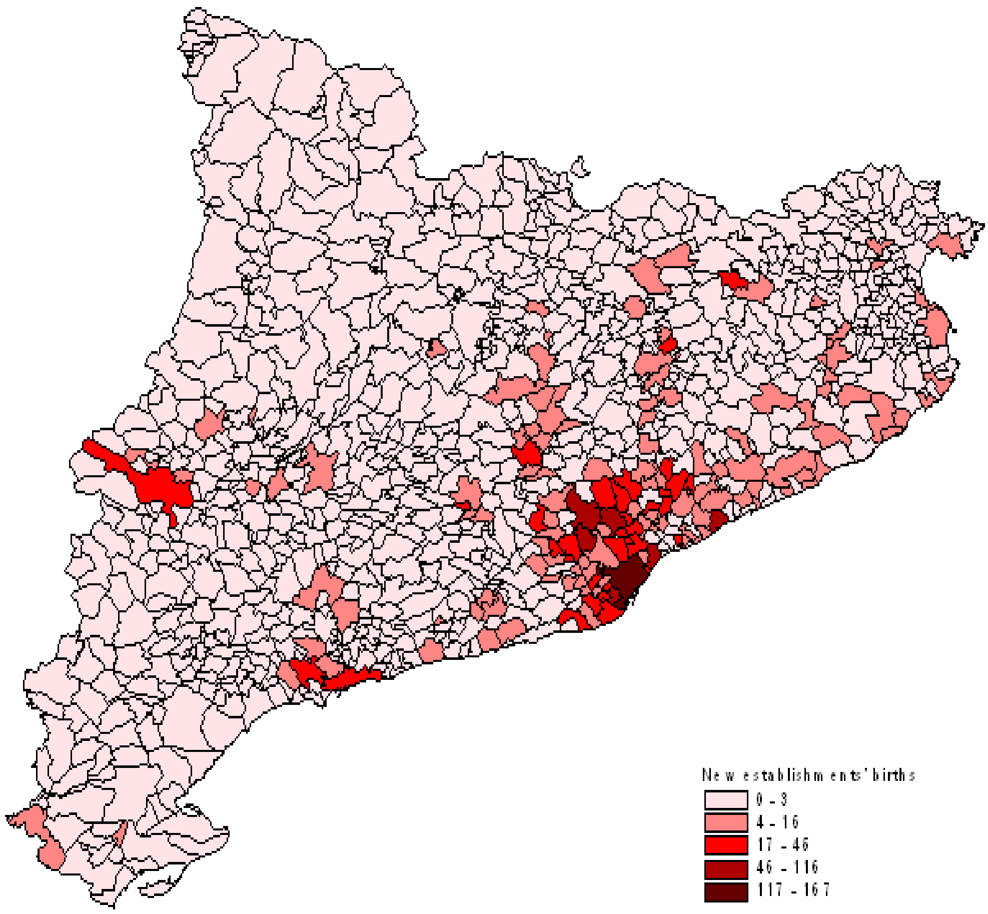
Table II: *New establishments' births summary data. 1997-2000 aggregated data.*

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

Source: Dirce and own elaboration.

Figure I shows how the birth of new establishments new (selected industries have been pooled) distribute across the Catalan geography. Most of establishments' start-ups are concentrated in some particular areas. Barcelona and its outskirts concentrate a great deal of them (the map displays Catalonia with its coast lying in the east. Barcelona can be found half way along the coast line). Other smaller clusters of Zip Codes, mostly found north-east from Barcelona, also experience some births of new establishments. There are parts of Catalonia, especially western areas that hardly see any establishment being started-up. The same analysis carried out industry by industry will show very different location patterns across industries. For instance, none of the 393 Textile start-ups for the 1997-2000 period was born in Barcelona (that accounts for roughly one quarter of industrial employment) whereas one third of all firms producing *Medical, precision and optical instruments* did.

Figure I: *Geographic distribution of new establishments' births*.



Note: The analysed sectors (*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)* are pooled.

In contrast, more than 40 % of the *Textiles* industry new establishments have been born in four Zip Codes that only account for 7 % of the Catalan employment. These are Mataró, Igualada, Terrassa and Sabadell that became specialized towns in the *Textiles* industry in the nineteenth century. This fact already points in the direction that specialized economic environments favour traditional activities, whereas large and diverse cities may be appealing when it comes to production involving high levels of technology.

Employment levels are also unequally distributed across the geographic space. The last column of Table I shows the value of the Spatial Gini Index[18] for the chosen industries. A large value of this index reflects that an industry is much more concentrated across space than the whole of economic activity. The value of this index is positive for all sectors analysed. This implies that industries' employment levels are, as new establishments' births do, highly unequally distributed across the Catalan geography, given that most of the economic activity takes place in some particular spots (for the year 2000, the first ten Zip Codes in terms of employment account for 54 % of the Catalan employment). There are differences across industrial activities. Traditional industries analysed (*Textiles and Wood and Furniture*) show the highest values for this statistic implying higher concentration levels.

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

[18] Spatial Gini Index$_j = \sum_{i=1}^{I}((L_{ij} / L_j) - (L_i / \bar{L}))^2$ where $\bar{L} = \sum_i \sum_j L_{ij}$, $L_j = \sum_i L_{ij}$ and $L_j = \sum_j L_{ij}$. Although it can not be negative it can of any positive value. This index has been used by Audrestch and Feldman (1996).

(div)[19]. 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

$$div_i = 1/\sum_j s_{ij}^2 \tag{8}$$

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.

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 i centroid have been computed. The up to 10 km concentric ring of location i includes all Zip Code locations whose Euclidean distance[20] 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

[19] Localization and urbanization economies variables are measured as in Rosenthal and Strange (2003). Congestion effects are captured as in Arauzo (2005).

[20] 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.

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 (*10-20 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[21]. Instead, the Diversity Index of the Local Labour Market (*LLM*) each Zip Code belongs to has been included[22]. Table A.1 in the Annex provides summary statistics of the data.

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 $\alpha_{ijt} \cdot (1 - F(\hat{\varepsilon}_{ijt}))$. 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 μ_{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

[21] It would require computing employment levels contained in different concentric rings for more than one hundred economic sectors.

[22] 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).

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 $t-1$.

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[23]:

$$\begin{aligned}\alpha_{it} \cdot (1 - F(\hat{\varepsilon}_{it})) = E(B_{it}) = \ & \mu_i + y_{t-1} \\ & + \beta_{11} loc(ZC)_{it-1} + \beta_{12} loc(<10\ km)_{it-1} + \beta_{13} loc(10-20\ km)_{it-1} \\ & + \beta_{21} urb(ZC)_{it-1} + \beta_{22} urb(<10\ km)_{it-1} + \beta_{23} urb(10-20\ km)_{it-1} \\ & + \beta_{31} cong(ZC)_{it-1} + \beta_{32} cong(<10\ km)_{it-1} + \beta_{33} cong(10-20\ km)_{it-1} \\ & + \beta_{41} div(ZC)_{it-1} + \beta_{42} div(LLM)_{it-1}\end{aligned} \tag{9}$$

For instance, a positive statistically significant estimate of β_{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 β_{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 $\beta^{'}$. 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[24]. Nevertheless,

[23] Since agglomeration economies' effects have been found to be very different across sectors, the regression model is estimated separately for the chosen industries.

[24] Most data is overdispersed, *i.e.*, the variance exceeds the mean of the distribution (Cameron and Trivedi, 1998).

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 given by expression (10) (Cameron and Trivedi, 1998):

$$Var[\beta_{ML}] = \left(\sum_{i=1}^{n} (\alpha_i \cdot (1 - F(\hat{\varepsilon}_i)) \cdot x_i x_i^{'} \right)^{-1} \cdot \left(\sum_{i=1}^{n} (B_i - (\alpha_i \cdot (1 - F(\hat{\varepsilon}_i))))^2 x_i x_i^{'} \right) \cdot \left(\sum_{i=1}^{n} (\alpha_i \cdot (1 - F(\hat{\varepsilon}_i))) \cdot x_i x_i^{'} \right)^{-1} \qquad (10)$$

where $x_i^{'} = [\mu_i, y, z_i^{'}]$ and time subscripts are omitted. The second term, the outer product of gradients estimator, is sandwiched by the inverse of the Hessian. Poisson pseudo-maximum likelihood is a term that denotes Maximum likelihood estimation of the coefficients and Sandwiched standard errors estimates.

The coefficient that captures differences across locations in the expected number of entrepreneurs and in time invariant profit determinants, μ_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[25] (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[26]. To solve this problem, the values of μ_i are restricted to be equal for all Zip Codes belonging to the same Local Labour Market[27]. Therefore, μ_i stands for a Local Labour Market time invariant specific effect. In the Poisson

[25] In particular, this leads to a Poisson with multiplicative fixed effects.

[26] Very high standard errors are obtained. Several reasons may explain that. In the first place, only information of Zip Codes that experience at least one establishment start-up can be used. In the second place, there is a lost of efficiency due to a decrease in the degrees of freedom. In the third place, only the within variation is being used and this variation is little over the studied four years period.

[27] A similar assumption is made in Rosenthal and Strange (2003).

regression context, Hausman *et al.* (1984) propose an easily computable statistic to check if a fixed effect can be considered to be common for a group of observations. If relevant individual specific effects have been omitted, then, the residuals of a certain individual will tend to show similar values. The idea is to construct the $T \times T$ correlation matrix of the standardized residuals. The covariance matrix is given by $\Sigma = (1/N)\sum_{n}(u_i \cdot u_i')$, where u_i is a column vector of standardized residuals belonging to observation i. A high value of this statistic would indicate that location specific time invariant effects do differ across locations within Local Labour Markets.

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[28].

Guimaraes *et al*. (2003) shows that Maximum likelihood coefficient and standard errors estimates arising from specification (9) can have an alternative interpretation to the one given in this work. The estimates obtained can be the conditional logit estimates that arise from applying the McFaden's Random Utility maximization framework to the firm location decision problem[29]. The Random Profit maximization problem assumes that a fixed number of entrepreneurs will choose the location that maximizes the expected profit function, given in this particular problem by $\pi_{it} = \eta_i + y_{t-1} + \beta' z_{it-1} + \varepsilon_{it}$, where ε_{ij} is a random term which has an *iid* Weibull distribution. This leads to the so-called independence of irrelevant alternative assumption (IIA), which has been found not to hold in many contexts (Greene, 2003). In this application, the IIA assumption leads to the Maximum Likelihood estimation of the coefficients' covariance matrix that, in most

[28] An alternative solution would be to fit a common dummy for all Local Labour Markets with no births plus one with some births. However, although using more observations this implies some restrictions across parameters that may not hold. However, results for this second specification have been obtained and do not show important differences with the preferred specification.

[29] This was first applied by Carlton (1983).

cases, will lead to incorrect statistical inference[30]. Aside from this technical issue, this framework also makes strong mobility assumptions. It is assumed that entrepreneurs choose location only taking into consideration the expected profitability of locations. This may be the case when studying firms that are not attached to a particular geographical area, like multinational firms. This is precisely the case of most of the studies following this approach (Coughlin *et al.*, 1991, Friedman *et al.*, 1992 and Devereux *et al.,* 2003). However, when considering the whole set of new establishments this may be too restrictive. Businesses can start up in a given location because this area is more profitable or, because the entrepreneur is attached to this particular location. This fact is explicitly accounted for in this work and also in Rosenthal and Strange (2003).

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[31]. 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.

[30] Although the inclusion of location and time specific dummies in the Poisson estimation of the conditional logit can mop up correlation across residuals due to unobserved year/location specific fixed effects still does not control for other sorts of *iid* violation such as homoskedasticity.

[31] For all industries analysed both sets of year and Local Labour Market specific dummies have been found to be jointly statistically significant according to Log-Likelihood ratio tests results.

Table III. *Agglomeration economies' estimates. Poisson pseudo Maximum Likelihood estimates*

	Textiles	*Wood and Furniture*	*Chemical products*	*Metal products except for machinery*	*Motor vehicles*	*Radio, television and communication equipments*	*Medical precision and optical instruments*
Localization economies							
Zip Code	0.0008392 (5.06)***	0.0011435 (3.75)***	0.0004866 (2.57)**	0.0005737 (4.26)***	0.0013187 (2.67)***	-0.0012633 (-1.22)	-0.0001398 (-0.16)
up to 10 km	0.0001951 (3.17)***	0.0000131 (0.16)	0.0000121 (0.18)	-0.0000336 (-1.15)	0.0003978 (1.85)*	-0.0005189 (-1.16)	-0.0015084 (-2.65)**
10 to 20 km	-0.0000465 (-1.31)	0.0000004 (0.01)	0.0000066 (0.31)	0.0000055 (0.39)	0.0000687 (0.57)	-0.0003862 (-0.91)	-0.0005449 (-1.50)
Urbanization economies							
Zip Code	-0.0000170 (-0.9)	0.0000203 (3.39)***	0.0000292 (4.98)***	0.0000186 (3.42)***	0.0000198 (1.51)	0.0000506 (4.08)***	0.0000419 (3.59)***
up to 10 km	-0.0000140 (-3.14)***	0.0000031 (1.14)	0.0000028 (0.75)	0.0000052 (2.76)***	-0.0000045 (-0.63)	0.0000062 (0.86)	0.0000127 (2.47)**
10 to 20 km	-0.0000048 (-2.47)**	-0.0000017 (-1.22)	-0.0000004 (-0.22)	-0.0000009 (-1.02)	-0.0000011 (-0.19)	-0.0000016 (-0.39)	-0.0000007 (-0.21)
Congestion effects							
Zip Code	$1.79 \cdot 10^{-11}$ (0.73)	$-1.27 \cdot 10^{-11}$ (-2.56)**	$-3.6 \cdot 10^{-11}$ (-2.71)***	$-8.96 \cdot 10^{-12}$ (-1.18)	$-3.29 \cdot 10^{-11}$ (-2.23)**	$-5.06 \cdot 10^{-11}$ (-2.13)**	$-4.68 \cdot 10^{-11}$ (-3.7)***
up to 10 km	$1.31 \cdot 10^{-11}$ (3.03)***	$-3.44 \cdot 10^{-12}$ (-1.35)	$-3.02 \cdot 10^{-12}$ (-0.94)	$-4.56 \cdot 10^{-12}$ (-2.73)***	$-7.95 \cdot 10^{-12}$ (-1.21)	$-2.35 \cdot 10^{-12}$ (-0.36)	$-6.04 \cdot 10^{-12}$ (-1.35)
10 to 20 km	$4.34 \cdot 10^{-12}$ (2.72)***	$1.21 \cdot 10^{-12}$ (1.06)	$2.97 \cdot 10^{-13}$ (0.19)	$7.43 \cdot 10^{-13}$ (1.02)	$-1.13 \cdot 10^{-12}$ (-0.37)	$4.05 \cdot 10^{-12}$ (1.43)	$2.36 \cdot 10^{-12}$ (0.9)
Diversity effects							
Zip Code	0.2242746 (8.72)***	0.2489029 (14.0)***	0.2365901 (6.35)***	0.2472408 (16.79)***	0.1703331 (3.22)**	0.2635343 (2.77)**	0.2886914 (4.21)***
LLM	-0.0625806 (-0.53)	0.0765863 (0.89)	0.2553995 (1.33)	-0.0252734 (-0.34)	-0.1590840 (-0.63)	0.4452644 (0.66)	0.1321790 (0.21)
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.

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 ($\beta_{11} > 0$). All sectors but two (*Textiles* and *Motor vehicles* industries) show statistically significant urbanization economies effects at the Zip Code level ($\beta_{21} > 0$). 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 (β_{21}) 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[35]. For all industries showing urbanization effects at the Zip Code level ($\beta_{21} > 0$), including the *motor vehicles* industry, these effects seem to be decreasing with the economic size of the Zip Code ($\beta_{31} < 0$). 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.

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 β_{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

[35] When not controlling for congestion effects, urbanization economies fail to show as in Rosenthal and Strange (2003)

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 $\partial E\lfloor B_{ij} \rfloor / \partial z_k = \exp(\beta' z) \cdot \beta_k$. 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 $\frac{1}{N}\sum_{i=1}^{N}\frac{\partial E[B_{it}]}{\partial x_{it-1}} = \beta' \cdot \overline{B}$, where $\overline{B}$ 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 (*div(zc)*), own industry employment level (*loc(zc)*) and overall employment level (*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*

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

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[36]. 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

[36] In Table A.I in the Annex the standard deviation for these different variables is provided.

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, *div(LLM)*, does not seem to exert any positive effect on the expected number of firm births[37]. 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[38]. 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 evidence of external effects spilling over Zip Codes comes from employment levels locations found in the up to 10 km concentric ring. Results do not show any evidence of statistically significant effects of industrial characteristics of the 10 to 20 km concentric ring employment levels on the number of firm births. For all four cases where evidence of external effects to spill over Zip Codes is found, these effects decrease by, at least, one order of magnitude. This confirms the findings of Rosenthal and Strange (2003) that agglomeration economies decrease sharply as distance increases.

Robustness checks

As already mentioned, Rosenthal and Strange (2003) run Tobit models in order to handle the fact that most locations do not experience any birth for a given year and sector. As a robustness check, Tobit estimates are presented in Table A.II in the Annex.

---

[37] None of the Local Labour Market Diversity index coefficients is statistically different from zero.

[38] In the sense that some variables reflecting up to 10 or 10 to 20 Km concentric ring employment levels have a coefficient which is statistically different from zero.

Although there are a few differences[39], results are qualitatively quite similar to the ones reported in Table III. Thus, it can be concluded that the results obtained are not too sensitive to the functional form specified.

It has been assumed that location specific effects, μ_i, can be considered to be equal across Zip Codes within Local Labour Markets. In order to check the validity of this hypothesis, within individual standardized residuals correlations are reported in Table A.III in the Annex. These correlations are not high in general terms. For six out of seven industries, correlations are lower than 0.2[40]. For the *Motor Vehicles*, *Radio, television and communication equipments* and *Medical, precision and optical instruments* industries, computed correlations are close to zero. In contrast, for the *Chemical Products* industry, some correlations are about one half indicating the presence of a misspecification problem. With the exception of the *Chemical Products* industry, the low correlations found can be interpreted as evidence that within Local Labour Market differences in entrepreneur abundances and in time invariant profit determinants are small. Thus, the assumption made regarding location specific effects seems to hold in six out of seven industries analysed.

5.-Conclusions

This paper has analysed two issues regarding the scope of agglomeration economies. First, the relative importance of intrasectoral (localization effects) and intersectoral (urbanization/diversity effects) scale economies has been addressed. Second, the geographic scope of these external economies has been studied. These questions are explored by looking at how industrial characteristics of the local environment determine the subsequent period expected number of births of new establishments. A model that explains the expected number of births is presented. It assumes that differences across locations in the number of establishments' start-ups are due to differences in entrepreneur abundances and, different probabilities of experiencing positive profits. It

[39] The *Textiles* industry show Zip Code level urbanization effects and the *Medical precision and optical instruments* do not show any statistically significant sign that own industry employment levels diminish the expected number of new establishments' births.

[40] When studying the relationship between research and development expenditures and number of patents at the firm level, Hausman *et al.* (1984) report values around 0.8 for the within individuals standardized residuals correlations, when failing to control for firm specific fixed effects.

is shown that the probability of reaping positive profits is increasing in any industrial characteristic that adds to productivity (a scale effect). A positive effect of any industrial characteristic on the number of births of new establishments can be interpreted as the existence of an external effect. Under the assumptions of the model, the number of new establishments can be described by a Poisson distribution.

The Poisson regression, with exponential mean function, has been econometrically estimated with a dataset that contains all new establishments born in Catalonia between 1997 and 2000. This Zip Code level dataset has been merged with 1996-1999 data on Zip Code employment levels. The estimation has been carried out, separately, for seven industries: *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*.

It has been found evidence of both localization and urbanization economies to exist, for five out of the seven industries analysed. The magnitude of these external 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 results are moreover robust to a Tobit specification, as similarity of the coefficients shows. The fact that the number of entrepreneurs and unobserved time invariant profit determinants are assumed to be common to all Zip Codes, that belong to the same Local Labour Market, does not seem problematic according to the residual analysis. An exception to this is the *Chemical products* industry estimates, where some caution when analysing the results is needed.

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

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

Audretsch, D.B. and Feldman, M.P. (1996), "R & D spillovers and the geography of innovation and production", *American Economic Review*, vol. 86, pp. 630-640.

Cameron, A.C. and Trivedi, P.K. (1998), *Regression analysis of count data,* Econometric Society Monograph No. 30, 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*, vol 65, pp. 440-449.

Combes, P.P. (2000). "Economic structure and local growth: France, 1984-1993", *Journal of Urban Economics*, vol 47, pp. 329-355.

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

Devereux, P., Griffith, R. and Simpson, H. (2003), "Agglomeration, regional grants and firm location", *The Institute for fiscal studies WP04/06,* December 2003.

Duranton, G.; Puga, D. (2000), "Diversity and Specialization in Cities: Why, Where and Why does it Matter?", *Urban Studies,* vol 37, pp. 533-555.

Duranton, G.; Puga, D. (2004), "Micro-foundations of urban agglomeration economies"

in Henderson, V. and J.-F. Thisse, (eds.), Handbook of Urban and Regional Economics, vol. 3, Amsterdam: North-Holland, pp. 1455-1495.

Duranton, G. and Overman, H. (2002), "Testing for localization using micro-geographic data", *CEPR discussion paper series*, No. 3379, may 2002

Ellison, G. And Glaeser, E.L. (1997), "Geographic concentration in US Manufacturing industries: A Dartboard approach", *Journal of Political Economy*, vol. 105, pp. 889-927.

Friedman, J., Gerlowski, D. and Silverman, J. (1992), “What attracts foreign multinational corporations? Evidence from branch location in the united states”, *Journal of Regional Science*, vol 32, pp. 403-418

Guimaraes, P., Woodward, D, Figuereido, O. (2003), “A tractable approach to the firm location decision”, *The Review of Economics and Statistics*, vol 84, pp. 201-204.

Glaeser, E.L., Kallal, J.A. Scheinkman and A. Shleifer (1992), “Growth in Cities”, *Journal of Political Economy,* vol 100, pp. 1126-1152.

Greene, W.H. (2003), *Econometric Analysis,* New Jersey: Prentice Hall

Hausman, J., Hall, B.H. and Griliches, Z. (1984), “Econometric models for count data with an application to the patents-R&D relationship”, *Econometrica,* 52, 909:938

Helsley, R.W. and Strange, W.C. (1990), “Agglomeration Economies and Matching in a system of Cities”, Regional science and Urban Economics, vol 20, pp. 189-212.

Henderson, J.V. (1986), “Efficiency of Resources usage and City size”, *Journal of Urban Economics*, vol. 19, pp. 47-70.

Henderson, J.V. (2003), “Marshall’s Scale Economies”, *Journal of Urban Economics* vol. 53, pp. 1-28.

Henderson, J.V., Kuncoro, A. and M. Turner, (1995), “Industrial Development in Cities”, *Journal of Political Economy*, vol. 103, pp. 1067-1085.

Hoover, E.M. (1936), “The measurement of industrial localization”, *The Review of Economics and Statistics,* vol. 18, pp. 162-171.

Jacobs, J. (1969), *The economy of cities*, Vintage: New York.

Marshall, A. (1890), *Principles of economicss*, MacMillan, New York.

Moomaw, R.L. (1983), “Is population scale a worthless surrogate for business agglomeration economies?,*”Regional science and Urban Economics,* vol. 13, pp. 525-545.

Roca, J. and Moix, M. (2004), “Cap a una nova organització territorial de Catalunya”, *Centre de política del sòl i valoracions research paper No. 5*, May 2004

Romer, P. (1986), “Increasing returns and long run growth”, *Journal of Political Economy*, vol.94, pp. 1002-1037.

Rosenthal, S. and Strange,W. (2003), "Geography, Industrial Organization and Agglomeration", *The Review of Economics and Statistics*, vol. 85, pp.377-393.

Rosenthal, S. and Strange,W. (2004), "Evidence on the Nature and Sources of Agglomeration Economics" in Henderson, V. and J.-F. Thisse, (eds.), Handbook of Urban and Regional Economics, vol. 3, Amsterdam: North-Holland, pp. 1455-1495.

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

Annex

Table A.I: *Summary statistics for new establishments' births and industrial characteristics of Zip Codes.*

Birth of new establishments	Mean	Std. Dev.	Min	Max
Textiles	0.10	0.79	0	21
Wood and Furniture	0.19	0.85	0	23
Chemical products	0.04	0.25	0	5
Metal products except for machinery	0.33	1.27	0	24
Motor vehicles	0.02	0.17	0	4
Radio, tv and communication equipments	0.01	0.09	0	2
Medical, precision and optical instruments	0.02	0.26	0	10

Employment levels		Mean	Std. Dev.	Min	Max
Textiles	Zip Code	59.15	325.45	0	5438
	Up to 10 km	734.86	1643.33	0	12164
	10 to 20 km	2100.43	3839.59	0	20561
Wood and Furniture	Zip Code	38.08	211.41	0	5910
	Up to 10 km	480.53	1224.63	0	11538
	10 to 20 km	1298.22	2545.85	0	16628
Chemical products	Zip Code	57.54	668.06	0	20689
	Up to 10 km	776.09	3008.08	0	29663
	10 to 20 km	2100.76	5890.26	0	35658
Metal products except for machinery	Zip Code	69.08	377.01	0	10281
	Up to 10 km	998.73	2803.95	0	20698
	10 to 20 km	2526.21	5565.75	0	32084
Motor Vehicles	Zip Code	47.24	721.68	0	22854
	Up to 10 km	640.85	2864.28	0	27533
	10 to 20 km	1766.47	5561.48	0	34085
Radio, tv and communication equipments	Zip Code	8.26	102.71	0	4004
	Up to 10 km	110.81	448.34	0	5569
	10 to 20 km	310.24	902.47	0	6644
Medical, precision and optical instruments	Zip Code	8.11	82.58	0	2190
	Up to 10 km	114.41	437.29	0	4173
	10 to 20 km	298.33	848.72	0	5350
Overall employment level	Zip Code	2066.85	24378.19	0	784724
	Up to 10 km	26437.59	101522.60	0	999990
	10 to 20 km	75643.89	200376.60	87	1199088
Diversity index	Zip Code	4.83	3.00	1	15.21
	LLM	14.52	4.26	3.40	21.75

Table A.II: *Agglomeration economies' estimates. Tobit regression.*

	Textiles	*Wood and Furniture*	*Chemical products*	*Metal products except for machinery*	*Motor vehicles*	*Radio, television and communication equipments*	*Medical precision and optical instruments*
Localization economies							
Zip Code	0.0023946	0.0036920	0.0008989	0.0029304	0.0015933	-0.0015707	0.0009477
	(9.64)***	(6.63)***	(3.11)**	(9.47)***	(3.34)***	(-0.96)	(0.63)
up to 10 km	0.0005160	0.0000806	-0.0000233	-0.0000793	0.0003971	-0.0004761	-0.0013003
	(5.11)***	(0.57)	(-0.25)	(-1.25)	(1.8)*	(-0.92)	(-1.29)
10 to 20 km	-0.0000322	-0.0000533	0.0000130	0.0000417	0.0001294	-0.0005322	-0.0006207
	(-0.61)	(-0.68)	(0.30)	(1.21)	(1.04)	(-1.45)	(-1.16)
Urbanization economies							
Zip Code	0.0000543	0.0000437	0.0000432	0.0000699	0.0000284	0.0000667	0.0000652
	(2.59)**	(3.61)***	(4.00)***	(4.97)***	(1.71)*	(3.31)***	(3.76)***
up to 10 km	-0.0000300	0.0000060	$5.96\cdot10^{-06}$	0.0000182	-0.0000040	0.0000036	0.0000119
	(-4.02)***	(1.50)**	(1.37)	(3.87)***	(-0.61)	(0.51)	(1.41)
10 to 20 km	-0.0000022	-0.0000022	$7.67\cdot10^{-07}$	-0.0000013	-0.0000028	-0.0000012	0.0000036
	(-0.56)	(-1.04)	(0.32)	(-0.56)	(-0.55)	(-0.26)	(0.81)
Congestion effects							
Zip Code	$-9.34\cdot10^{-11}$	$-8.42\cdot10^{-11}$	$-7.17\cdot10^{-11}$	$-1.12\cdot10^{-10}$	$-3.85\cdot10^{-11}$	$-7.5\cdot10^{-11}$	$-7.37\cdot10^{-11}$
	(-3.3)***	(-3.69)***	(-2.42)**	(-4.12)***	(-1.41)	(-1.68)*	(-3.73)***
up to 10 km	$2.68\cdot10^{-11}$	$-6.21\cdot10^{-12}$	$-5.72\cdot10^{-12}$	$-1.8\cdot10^{-11}$	$-8.07\cdot10^{-12}$	$-1.19\cdot10^{-13}$	$-6.4\cdot10^{-12}$
	(3.66)***	(-2.14)**	(-1.77)*	(-4.43)***	(-1.46)	(-0.02)	(-1.04)
10 to 20 km	$1.31\cdot10^{-12}$	$1.45\cdot10^{-12}$	$-1.01\cdot10^{-12}$	$-9.93\cdot10^{-14}$	$-1.2\cdot10^{-12}$	$4.07\cdot10^{-12}$	$-1.01\cdot10^{-12}$
	(0.42)	(0.90)	(-0.55)***	(-0.06)	(-0.41)	(1.25)	(-0.32)
Diversity effects							
Zip Code	0.3453889	0.3143415	0.0174103	0.3911351	0.2265275	0.2065579	0.2466707
	(7.25)***	(12.19)***	(1.03)	(13.84)***	(4.27)***	(2.36)**	(3.39)***
LLM	-0.0079522	-0.2316130	0.2094929	0.0867859	-0.1606212	0.3485371	0.1973371
	(-0.04)	(-0.22)	(2.08)**	(0.71)	(-0.66)	(0.58)	(0.38)
N	2756	3624	2532	3440	2152	844	1852
LR-Test	569.51***	1039.8***	349.7***	1500.7***	134.86***	71.8***	186.57***

Notes: 1. Figures in parenthesis are *t*-statistics. 2. *, **, ***: statistically significant at the 90%, 95% and 99% levels, respectively.

Table A.III: *Within individuals standardized residuals correlations.*

Textiles	1997	1998	1999	2000
1997	1.000	0.112	0.150	0.096
1998	0.112	1.000	0.111	0.035
1999	0.150	0.111	1.000	0.074
2000	0.096	0.035	0.074	1.000

Wood and Furniture	1997	1998	1999	2000
1997	1.000	0.129	0.051	0.187
1998	0.129	1.000	0.051	0.010
1999	0.051	0.051	1.000	0.073
2000	0.187	0.010	0.073	1.000

Chemical products	1997	1998	1999	2000
1997	1.000	0.460	0.602	0.342
1998	0.460	1.000	0.475	0.243
1999	0.602	0.475	1.000	0.392
2000	0.342	0.243	0.392	1.000

Metal products except for machinery	1997	1998	1999	2000
1997	1.000	0.145	0.175	0.178
1998	0.145	1.000	0.255	0.157
1999	0.175	0.255	1.000	0.160
2000	0.178	0.157	0.160	1.000

Motor vehicles	1997	1998	1999	2000
1997	1.000	-0.006	0.094	-0.003
1998	-0.006	1.000	0.058	-0.035
1999	0.094	0.058	1.000	-0.032
2000	-0.003	-0.035	-0.032	1.000

Radio, television and communication equipments	1997	1998	1999	2000
1997	1.000	-0.020	-0.066	-0.084
1998	-0.020	1.000	-0.044	-0.060
1999	-0.066	-0.044	1.000	0.041
2000	-0.084	-0.060	0.041	1.000

Medical precision and optical instruments	1997	1998	1999	2000
1997	1.000	-0.046	-0.032	0.191
1998	-0.046	1.000	-0.016	-0.030
1999	-0.032	-0.016	1.000	-0.029
2000	0.191	-0.030	-0.029	1.000

SÈRIE DE DOCUMENTS DE TREBALL DE L'IEB

2000

2000/1 - Esteller, A.; Solé, A., "Vertical Income Tax Externalities and Fiscal Interdependence: Evidence from the US"

2000/2 - Castells, A., "The role of intergovernmental finance in achieving diversity and cohesion: the case of Spain"

2000/3 - Costa, M.T.; Segarra, A. (U. Rovira i Virgili)**; Viladecans, E.,** "Pautas de localización de las nuevas empresas y flexibilidad territorial"

2000/4 - Costa, M.T.; Duch, N.; Lladós, J. (U. Autònoma de Barcelona), "Determinantes de la innovación y efectos sobre la competitividad: el caso de las empresas textiles"

2000/5 - Solé, A., "Determinantes del gasto público local: necesidades de gasto vs. capacidad fiscal"

2000/6 - Barberán, R. (U. de Zaragoza); **Bosch, N.; Castells, A.; Espasa, M.**, "The redistributive power of the Central Government Budget"

2001

2001/1 - Espasa, M., "The territorial redistribution of the EU budget. Empirical evidence at national and regional level"

2001/2 - Viladecans, E., "La concentración territorial de las empresas industriales: un estudio sobre la unidad geográfica de análisis mediante técnicas de econometría espacial"

2001/3 - Castells, A., "La descentralización de las políticas sociales en el Estado del Bienestar"

2001/4 - Bosch, N.; Pedraja, F. (U. de Extremadura); **Suárez-Pandiello, J.** (U. de Oviedo), "The influence of Environmental Variables in Measuring the Efficiency of Refuse Collection Services: An Application to the Spanish Municipalities"

2001/5 - Solé, A., "Budget spillovers in a metropolitan area: typology and empirical evidence"

2001/6 - Sanromà, E.; Ramos, R. (U de Barcelona i AQR), "Local human capital and external economies: evidence for Spain"

2001/7 - Leonida, L. (U. della Calabria); **Montolio, D.**, "Convergence and Inter-Distributional Dynamics among the Spanish Provinces. A Non-parametric Density Estimation Approach"

2001/8 - García Quevedo, J., "University research and the location of patents in Spain"

2001/9 - Esteller, A.; Solé A., "Tax Setting in a Federal System: The Case of Personal Income Taxation in Canada"

2001/10 - **Durán J.M.; Gispert, C. de**, "Fiscalidad medioambiental sobre la energía: propuestas para España "

2001/11 - Álvarez, M., "España y la senda de desarrollo de la inversión directa: una aproximación"

SÈRIE DE DOCUMENTS DE TREBALL DE L'IEB

2002

2002/1 - Bosch, N.; Espasa, M.; Sorribas, P., "La capacidad redistributiva y estabilizadora del presupuesto del Gobierno Central Español"

2002/2 - García Quevedo, J., "The location of innovation. Universities and technological infrastructure in Spain"

2002/3 - Viladecans Marsal, E., "The growth of cities: Does agglomeration matter?"

2002/4 - Pons Novell, J.; Tirado Fabregat, D.A. (U. de Barcelona)**,** "Discontinuidades en el crecimiento económico en el periodo 1870-1994: España en perspectiva comparada"

2002/5 - Bosch, N.; Espasa, M.; Sorribas, P., "The redistributive, stabiliser and insurance effects at territorial level of "federal" government budgets"

2002/6 - Callejón, M. (U. de Barcelona)**; García Quevedo, J.,** "Las ayudas públicas a la I+D empresarial. Un análisis sectorial"

2003

2003/1 - Solé Ollé, A.; Viladecans Marsal, E., "Fiscal and growth spillovers in large urban areas"

2003/2 - Gual, J. (IESE)**; Trillas, F.,** "Telecommunications Policies: Determinants and Impact"

2003/3 – Leonida, L. (U. Della Calabria)**; Montolio, D.,** “Public Capital, Growth and Convergence in Spain. A Counterfactual Density Estimation Approach”

2003/4 – Álvarez, M., “FDI Determinant Factors: The Case of Catalan Multinational Manufacturing Firms”

2003/5 – Álvarez, M., “Wholly-Owned Subsidiaries Versus Joint Venture: The Determinant Factors in the Catalan Multinational Manufacturing Case”

2003/6 – Pelegrín, A., “Regional Distribution of Foreign Manufacturing Investment in Spain. Do agglomeration economies matter?”

2004

2004/1 – Sánchez Hugalde, A., "Movilidad intergeneracional de ingresos y educativa en España (1980-90)"

2004/2 – Esteller, A., "Tax Evasion in Interrelated Taxes"

2004/3 – Castells, A.; Esteller, A.; Vilalta, M., "Full Characterisation of the Political Economy of Fiscal Adjustment: Evidence from Spanish Municipalities"

2004/4 – Lambert, P.J. (U. of Cork)**,** "Income Taxation and Equity"

2004/5 – Bosch, N.; Solé, A., ″Yardstick competition and the political costs of raising taxes: An empirical analysis os Spanish municipalities″

2004/6 – Maggioni, M. (DISEIS i Catholic University of Milan)**,** "The rise and fall of industrial clusters: Technology and the life cycle of regions″

SÈRIE DE DOCUMENTS DE TREBALL DE L'IEB

2004/7 – Frey, B.S. (Institute for Empirical Economic Research of the University of Zurich i CREMA)**; Torgler, B.,** (Yale Center for International and Area Studies i CREMA) "Taxation and Conditional Cooperation″

2004/8 – Alonso, J. (U. de Vigo i RGEA)**; Freire-Serén, M.J.** (U. de Vigo)**; Manzano, B.** (U. de Vigo)**,** "Rentabilidad social de la inversión pública española en infraestructruras″

2005

2005/1 - Arocena, P. (U. Pública de Navarra), "The measurement of scope, scale and diversification economies: How (in)efficient is electricity restructuring and unbundling?"

2005/2 – Solé Ollé, A., "The effects of party competition on budget outcomes: Empirical evidence from local governments in Spain "

2005/3 – Solé Ollé, A., "Expenditure spillovers and fiscal interactions: Empirical evidence from local governments in Spain"