The mechanisms of agglomeration: Evidence from the effect of inter-industry relations on the location of new firms

Jordi Jofre-Monseny Raquel Marín-López Elisabet Viladecans-Marsal

Universitat de Barcelona &
Institut d'Economia de Barcelona (IEB)

April, 2011

ABSTRACT: The objective of this paper is to explore the relative importance of each of Marshall's agglomeration mechanisms by examining the location of new manufacturing firms in Spain. In particular, we estimate the count of new firms by industry and location as a function of (pre-determined) local employment levels in industries that: 1) use similar workers (labor market pooling); 2) have a customer-supplier relationship (input sharing); and 3) use similar technologies (knowledge spillovers). We examine the variation in the creation of new firms across cities and across municipalities within large cities to shed light on the geographical scope of each of the three agglomeration mechanisms. We find evidence of all three agglomeration mechanisms, although their incidence differs depending on the geographical scale of the analysis.

Keywords: agglomeration economies, coagglomeration, labor market pooling, input sharing, knowledge spillovers

JEL codes: L60, R30.

1. Introduction

The term "agglomeration economies" is used to denote the mechanisms that drive employees and firms to co-locate geographically. Many papers have tested and quantified the importance of these economies[1]. Some analyze their influence on the geographical concentration of economic activities, whereas others test their effect on wages. Despite the accumulation of a substantial body of literature, further empirical work is needed to understand more precisely the mechanisms through which agglomeration economies work (Puga, 2010; and Glaeser and Gottlieb, 2009). The classification of agglomeration mechanisms which is most often used in the (empirical) literature is due to Marshall (1890), who described three mechanisms: labor market pooling, input sharing and knowledge spillovers[2]. A densely-populated local labor market (labor market pooling) facilitates the flows of workers across firms in the presence of firm-specific shocks (Krugman, 1991) and enhances employer-employee matches (Hesley and Strange, 1990). The concentration of firms in a geographical area also enables firms to share input suppliers (input sharing) and facilitates the transmission of knowledge (knowledge spillovers).

One of the first papers to empirically analyze the sources of agglomeration economies was Rosenthal and Strange (2001). These authors try to identify the characteristics of an industry that determine its degree of geographical concentration, using proxies of the three agglomeration mechanisms described by Marshall. If labor market pooling is a relevant agglomeration theory, then industries that use workers who are less mobile across industries should be spatially concentrated. If input sharing is a relevant agglomeration theory, then industries that make more intensive use of inputs should be spatially concentrated. Finally, the observation that knowledge-intensive industries are more spatially concentrated would be indicative of the presence of knowledge spillovers. Rosenthal and Strange (2001) find that labor market pooling is the most important agglomeration mechanism at work and that knowledge spillovers also seem to contribute to industry agglomeration, but only at the local level.

Ellison et al (2010) ingeniously twists the methodology developed by Rosenthal and Strange (2001) and re-defines the dependent variable, making it the tendency of two industries to co-locate ("co-agglomerate" is the term they use). An index that measures the co-agglomeration of an industry pair is then regressed on measures of the extent to which an industry pair use the

[1] See Glaeser and Gottlieb (2009) and Puga (2010) for two extensive reviews of the research on the economics of agglomeration.

[2] Duranton and Puga (2004) provide an alternative, more theoretically driven, classification. These authors propose to classify agglomeration mechanisms as sharing, matching or learning mechanisms. Agglomeration can be beneficial as a means to share facilities and infrastructures, input suppliers, the gains of individual specialization and a labor pool. Matching and learning can be enhanced in a more economically dense environment.

same type of workers (labor market pooling), have a customer-supplier relationship (input sharing) and use the same technologies (knowledge spillovers). Although they find positive and statistically significant evidence of the existence of all three mechanisms, they find input sharing to be the most important.

The studies by Dumais et al (1997) and Glaeser and Kerr (2009) also use inter-industry relations to shed light on the sources of agglomeration. Dumais et al (1997) seeks to explain industry employment growth as a function of the local employment levels in industries that use similar workers, that have a customer-supplier relationship, and that use similar technologies[3]. The results suggest that labor market pooling is the most important agglomeration mechanism. Their contribution is, however, limited by the fact that their data are aggregated at the two-digit industry level, masking many of the inter-industry relations that take place within this level. Glaeser and Kerr (2009) study the local determinants of manufacturing firm entry. They conclude that firm entry in a given industry is higher in cities where the industries that employ similar workers are more prevalent. It is also found that the strong presence of the relevant input supplier industries spurs firm entry, especially if the average size of these input supplier firms is small[4].

The objective of this paper is to shed more light on the relative importance of each of Marshall's agglomeration mechanisms by examining the location of new manufacturing firms in Spain. In particular, we estimate the count of new firms by industry and location as a function of (pre-determined) local employment levels in industries that: 1) use similar workers (labor market pooling); 2) have a customer-supplier relationship (input sharing); and 3) use similar technologies (knowledge spillovers).

In Ellison et al (2010), the dependent variable is defined as the tendency of a pair of industries to co-locate. Note that the random co-location of an industry pair could induce the firms involved to use the same type of workers, to start a customer-supplier relationship or to use the same type of new technologies; if so, industrial relations may be the result and not the cause of co-location. We follow Glaeser and Kerr's (2009) study and use the count of new firms as the dependent variable which partially addresses this identification problem. As recognized by Rosenthal and Strange (2003), from the viewpoint of an entrepreneur, location attributes are fixed at the time of the start-up, alleviating concerns about simultaneity. That is, if sharing

[3] Dumais et al (1997) contains different analyses. Here, we refer to the one developed in Section 6; this does not appear in Dumais et al (2002), the published version of the paper.

[4] More generally, Glaeser et al (2010) and Rosenthal and Strange (2010) are two studies that find that firm size is a very strong predictor of firm entry.

workers were the result and not the cause of co-location, the location of new firms would not react to the (pre-determined) geographical distribution of industries that use similar workers. A second contribution of this paper is that we use a novel measure of knowledge flows between industries. In the literature, information flows have been proxied using patent citations data (patents in industry *i* that are cited in patents of industry *j*) or Scherer's (1984) technology matrix which measures R&D activity flows between industries. Ellison et al (2010) use measures based on both of these approaches, accepting that they only reflect flows of ideas at the highest level, whereas Glaeser and Kerr (2009) focus on patent citation data. We use a survey conducted by Statistics Spain asking manufacturing firms about the use of new technologies in their production processes. This allows us to measure the extent to which two industries use the same new technologies in their productions. We replicate our analysis at two different geographical levels, the rationale being that different agglomeration mechanisms may operate at different geographical scales. We examine variation in the creation of new firms across cities and across municipalities within large cities to shed light on the geographical scope of each of the three agglomeration mechanisms. Since municipalities in Spain are very small[5], this paper studies the relative importance of the different agglomeration mechanisms within a very narrow geographical scope[6], a question that is left unexplored in Ellison et al (2010) and Glaeser and Kerr (2009). This constitutes the third contribution of this paper.

Our main results can be summarized as follows. The creation of new firms in a given industry is higher in areas with a strong presence of industries that use similar workers. The results also indicate that a strong presence of the relevant input suppliers also favors the creation of new firms. Hence, our results indicate that labor market pooling and input sharing are relevant agglomeration theories, and that the relative importance of these two mechanisms is roughly the same. These effects show up when we examine variation in the creation of new firms both across cities and across municipalities within large cities. In the latter case, we also find some evidence that new firms locate in areas with the presence of industries that use similar technologies, although this effect is relatively small. This suggests that the knowledge spillovers may be relevant but, most likely, only operate at a limited geographical scale.

[5] Spanish municipalities average 60 square kilometres, being much smaller than US zip codes. In the sample of metropolitan US zip codes used in Rosenthal and Strange (2003), the zip code average surface is 200 sq. km. (more than three times larger than the average Spanish municipality).

[6] A number of papers have shown that there are agglomeration effects that have a very limited geographical scope, including Rosenthal and Strange (2003) using US data and Viladecans-Marsal (2004) and Jofre-Monseny (2009) using Spanish data.

To our knowledge, Dumais et al (1997), Glaeser and Kerr (2009) and Ellison et al (2010) are the only other studies that use inter-industry relations to shed light on the sources of agglomeration. However, our paper also relates to a number of studies that have tested the existence of a particular agglomeration mechanism. Fallick et al (2006) show that workers' mobility between firms is higher in specialized areas. Overman and Puga (2010) find that industries with more risk are more geographically concentrated. Thus, these two studies provide evidence that, in a thick labor market, firms and workers are in a better position to face firm-specific shocks. Costa and Khan (2000) and Andersson et al (2007) have shown that employee-employer matches are better in densely populated areas. Other studies have tested the relevance of the input sharing mechanism, including Bartlesman et al (1994), Holmes (1999), Holmes and Stevens (2002) and Li and Lu (2009). Their results indicate that the co-location of firms reduces transportation costs in purchasing inputs and selling outputs. It is more difficult to test for the existence of knowledge spillovers. The most direct test of their existence is provided by studies showing that inventors are more likely to cite other inventors who are geographically closer (Jaffe et al, 1993; and Agrawal et al, 2008 and 2010).

After this introduction, the rest of the paper is organized as follows. In Section 2 we introduce the firm-level database used to construct the count of new firms by industry and location. This count constitutes the dependent variable of this paper, and is also described in this section. In Section 3 we explain the way in which we measure inter-industry relations along the three different agglomeration theories. In Section 4 we discuss the econometrics of the paper and in Section 5 we present the results. Section 6 concludes.

2. The location of new firms

Previous work has shown that the strength of different agglomeration mechanisms may differ at different geographical scales[7]. We therefore perform our analysis at two different geographical levels. First, we work with Spanish cities, which are aggregations of municipalities built on the basis of commuting patterns[8]. There are 806 such cities in Spain, although we only consider those with more than 10,000 inhabitants in order to exclude primarily rural areas. Finally we work with 477 cities which in 2001 contained 95% of the Spanish population and employment. Sometimes we will use the term 'between-cities analysis' to refer to the regression analysis in which we

[7] See Kerr and Kominers (2010) for a theoretical foundation and Rosenthal and Strange (2004) and Arauzo-Carod et al (2010) for two reviews of the relevant empirical literature.

[8] The cities we use were built by Boix and Galleto (2006) by aggregating municipalities to obtain self-contained local labor markets. There were 8,108 municipalities in Spain in 2001. The municipalities are political and administrative units. We exclude the municipalities of the regions of Ceuta and Melilla (the two Spanish enclaves in North Africa).

explain variation in the creation of new firms across these 477 cities. Alternatively, our aim will be to explain variation in new firm creations across municipalities within large cities (within-cities analysis), in order to explore the agglomeration sources that are relevant across small geographical units within economically dense areas. To capture this, we select the 19 cities whose central municipality has more than 200,000 inhabitants. There are 755 municipalities in these 19 cities, which contained 45% of the Spanish population and employment in 2001.

The dependent variable is constructed using SABI, the Iberian part of the (Bureau Van Dijk's) Amadeus database, which contains the annual accounts of more than 1 million Spanish firms. In 2002, the firms in this database represented 80 percent of the firms in the Spanish Social Security Register [9]. This firm-level database contains the location (municipality) of the firm, the year the firm was created, and its industry. Our dependent variable is defined as the count of firms created in 2002, 2003 and 2004 by industry and location. 17,600 new manufacturing firms were created in Spain in this three-year period. The industry definition that we use corresponds to the three-digit level of the 1993 National Classification of Economic Activities (NACE 93 Rev.1). In our regressions we exclude those industries with less than 15 creations in the estimation sample; this leaves us with 75 and 62 three-digit industries in the between-cities and the within-cities analyses respectively. The distribution of counts of new firms per city and industry is summarized in Table 1a.

[Insert Table 1 here]

We report the maximum and the average count of new firms per industry and city for the five industries with most creations, the median industry in terms of creations, and the five industries with fewest creations. The figure reported in the last column of the table is the share of cities with zero births in the industry and reflects the geographical concentration of the variable. The Manufacture of luggage and handbags (CNAE 192) industry has the median number of new firm creations (73). The city with the highest count of creations in this industry (13) is Ubrique-Elda, one of the leather clusters in Spain. Table 1b shows the analogous figures for the count of new firms per municipality and industry. The Manufacture of parts and accessories for motor vehicles and their engines (CNAE 343) has the median number of new firm creations (39), the

[9] To explore the representativeness of the SABI database in terms of the geographical and industrial distribution of its firms, we have computed different correlations comparing the SABI and the Social Security Register. In terms of the count of firms per municipality (province), the correlation between the SABI and the Social Security Register distributions is 0.902 (0.943). Regarding the count of firms per (2-digit) industry, the correlation between these two distributions is 0.942. Hence, the coverage (and the geographical and industrial representativeness) of the SABI database seem reasonably good.

municipality of Madrid being the location with the highest count of new firms in this industry (5)[10].

In all regressions we include the pre-determined own industry employment level as a control variable. In Graph 1 we pool the observations across all industries and plot the count of new firms per industry and location as a function of the local own-industry employment level. The top and bottom panels illustrate the data at the city and municipality levels respectively.

[Insert Graph 1 here]

Each dot in Graph 1 represents the average count of new firms in each (pre-determined) employment cell. Each of the first 50 cells represents only one employment value (1 to 50). Beyond this point, each cell contains a percentile of the remaining observations (49 and 27 observations respectively). The dashed lines depict the 10th and 90th firm birth percentiles within each cell. The differences between these percentiles indicate that there is a fair amount of variation in the creation of new firms within each employment cell. It is precisely this variation that will be used in the analysis.

3. Inter-industry relations and agglomeration theories

Inter-industry relations are the basis for identifying the sources of agglomeration economies. Our strategy is to construct measures of the extent to which two industries 1) use the same type of workers (labor market pooling); 2) have a customer-supplier relationship (input sharing); or 3) share technology and knowledge (knowledge spillovers). Once we have these measures for all industry pairs, we construct weighted sums of (pre-determined) employment levels by industry and location, where higher weights are assigned to industries with stronger relationships throughout the three different dimensions. These industry-specific weighted sums of employment can thus be interpreted as the employment in industries that: 1) use workers with the same occupations as those used by industry *i* ($labor_{ic}$); 2) supply inputs to industry *i* ($input_{ic}$); 3) buy the outputs of industry *i* ($output_{ic}$); and 4) use the new production technologies ($techno_{ic}$) used by industry *i*.

Labor market pooling: Labor market pooling denotes the advantages that firms and employees obtain from locating in a thick labor market. If labor market pooling is a relevant agglomeration theory, then industries that use similar workers should co-locate because of the higher workers' mobility between these industries. Following Dumais et al (1997), Glaeser and

[10] Table 1 shows that, for several industries, Madrid ranks first in terms of new firm creations. As shown below, the results reported throughout this paper are robust to the exclusion of Madrid.

Kerr (2009) and Ellison et al (2010) we look at the distribution of workers by industry and occupation. We consider all the manufacturing workers contained in the second quarters of the 2001 and 2005 waves of the Spanish Labor Force Survey (EPA). Workers are classified in 207 different occupations which correspond to the three-digit level of the 1994 National Classification of Occupations listed in Table A1 in the Annex. The variable *labor similarity*$_{ij}$ measures the extent to which the distribution of workers by occupation in industry *i* is similar to that in industry *j*:

$$labor\ similarity_{ij} = 1 \Big/ \frac{1}{2} \sum_{o} \left| \frac{L_{oi}}{L_i} - \frac{L_{oj}}{L_j} \right| \tag{1}$$

where *o* indexes occupation and *L* denotes number of workers. Notice that *labor similarity*$_{ij}$ is the inverse of a Duncan and Duncan (1955) dissimilarity index. This index is bounded between 0 and 1 and, in this application, can be interpreted as the share of workers in industry *j* that need to change occupation to mimic the distribution of occupations in industry *i*. Hence, the variable *labor similarity*$_{ij}$ takes positive values that are greater than one and is computed for all industry pairs (including those in the agriculture and the services sectors). We rank all *J* industries in descending order based on their labor similarity with industry *i* and construct the following industry-specific weights:

$$\begin{aligned} W_{ij}^{L} &= 0 && \text{if } r > 10 \\ W_{ij}^{L} &= \frac{labor\ similarity_{ij}}{\sum_{r=1}^{10} labor\ similarity_{ij}} && \text{if } r \leq 10 \end{aligned} \tag{2}$$

where *r* identifies the r^{th} closest industry in this labor market pooling metric. To increase the weights assigned to the closest industries, we only consider the 10th closest. This is the number of industries whose value in the *labor similarity*$_{ij}$ metric typically exceeds its average value by more than one standard deviation. The highest weight in our sample corresponds to the Manufacture of rubber products (CNAE 251) and the Manufacture of plastic products (CNAE 252) industry pair. Based on this industry-specific set of weights we construct the variable *labor*$_{ic}$:

$$labor_{ic} = \sum_{j \neq i} (W_{ij}^{L} \cdot L_{cj}) \tag{3}$$

which is a weighted sum of industry *(j)* and location *(c)* employment levels where industries that use workers who are more similar to those used by industry *i* are given higher weights. Hence, *labor*$_{ic}$ is a measure of the local employment in the industries that use the same workers as those used by industry *i*.

Input sharing: The concentration of firms in a geographical area enables them to share a larger base of suppliers and, at the same time, to be closer to customers. Following previous work, we use data from Input-Output Tables to characterize customer-supplier relations. In particular, we use data from the 2001 Catalan Input-Output Table built by Statistics Catalonia (IDESCAT)[11]. We use this regional table instead of the Spanish one because it enables us to characterize customer-supplier relations for narrowly defined industries[12]. We construct the two following sets of industry-specific weights:

$$W_{ij}^{I} = \frac{inputs_{i \to j}}{total\ inputs_{i}} \tag{4}$$

$$W_{ij}^{O} = \frac{outputs_{i \to j}}{total\ outputs_{i}} \tag{5}$$

W_{ij}^{I} is the share of the inputs that industry *i* purchases from industry *j* (including those in the agriculture and the services sectors). Conversely, W_{ij}^{O} is the share of the outputs produced by industry *i* that are purchased by industry *j*. The most intense dependence on a single input supplier industry is that shown by the producers of Manufacture of articles of paper and paperboard (CNAE 212) which obtain 66% of their inputs from producers of Manufacture of pulp, paper and paperboard (CNAE 211). The most intense dependence on a single customer is that shown by the producers of Manufacture of prepared animal food (CNAE 157) which sell 96% of their output to the producers in Agriculture, hunting and related service activities (CNAE 100). Based on these two industry-specific sets of weights we construct the variables $input_{ic}$ and $output_{ic}$:

$$input_{ic} = \sum_{j \neq i} \left(W_{ij}^{I} \cdot L_{cj} \right) \tag{6}$$

$$output_{ic} = \sum_{j \neq i} \left(W_{ij}^{O} \cdot L_{cj} \right) \tag{7}$$

which are weighted sums of industry *(j)* and location *(c)* employment levels where industries that have stronger customer-supplier relationships are given higher weights. Notice that $input_{ic}$ measures the local employment in the industries that are industry *i*'s main input supplier. Likewise, $output_{ic}$ measures the local employment in the industries that are industry *i*'s main customers.

[11] Catalonia is a region in the north-east of Spain. In 2001, the population of Catalonia (6,361,365 inhabitants) represented 15.5% of the Spanish population, 17.5% of its employment and 24% of its manufacturing employment.

[12] The Catalan (Spanish) Input-Output table enables us to characterize the supplier-customer relations for 122 (71) industry pairs. However, $input_{ic}$ and $output_{ic}$ do not vary at the three digit level in all cases as the Input-Output products can only be grouped into 54 manufacturing industries. We address this mismatch by clustering the standard errors at the two-digit industry and location in all the estimations.

Knowledge spillovers: Marshall (1890) considered that knowledge and ideas flow more easily between firms and employees located nearby (knowledge spillovers). If firms co-locate to share knowledge and ideas, industries that use similar knowledge should be co-located. Knowledge spillovers are difficult to measure. In the literature, information flows between industries have been proxied using patent citations data (patents in industry *i* that are cited in patents of industry *j*) or Scherer's (1984) technology matrix which measures R&D activity flows between industries. Ellison et al (2010) use measures based on both approaches, accepting that they only reflect flows of ideas at the highest level, whereas Glaeser and Kerr (2009) focus on patent citation data. The construction of measures of information flows between industries using patent citations data or Scherer's (1984) technology matrix seems especially hard to justify in the Spanish context. The Spanish economy has low levels of innovation: innovation expenditure accounts for only 1.35% of GDP, compared with 2.77% in the US. The picture that emerges from patent data is even more striking: 0.005 patents per one thousand inhabitants in Spain, compared with 0.048 in the US. In the light of these figures, we propose an alternative approach to measure the extent to which different industries share knowledge.

We use a survey conducted by Statistics Spain asking manufacturing firms about their use of different new technologies in their production processes. Around 11,000 firms were interviewed within this survey entitled "Use of new technologies in manufacturing" which was carried out in 1998 as part of the broader "Innovation in Companies Survey". This survey details the use of 26 new technologies in production and is representative at the industry level for the population of firms with at least one employee[13]. The classification of these technologies follows OECD guidelines and is listed in Table A2. In principle, this classification has been designed to cover a wide range of the elements related to innovation activities and knowledge diffusion in the manufacturing sector[14].

The variable $\textit{technology similarity}_{ij}$ measures the extent to which industry *i* and *j* use the same new technologies in their production processes:

$$\textit{technology similarity}_{ij} = 1 \Big/ \frac{1}{2} \sum_{n} \left| \frac{NT_{ni}}{NT_i} - \frac{NT_{nj}}{NT_j} \right| \tag{8}$$

[13] An interesting exercise is to compare, at the industry level, the use of the new production technologies listed in Table A2 with a measure of innovation effort like R&D expenditures over sales. We have data on this innovation effort measure for the year 2000 at the (extended) two-digit level (NACE 93 Rev.1). To make the comparison operational we compute, at the industry level, the average use of the 26 new technologies which summarizes the use of new production technologies in the industry. The correlation between this measure and the innovation effort is 0.67. The following four industries (*Manufacture of radio, TV and communication, Manufacture of electric equipment, Pharmaceutical products* and *Manufacture of electronic components*) rank amongst the top five industries in both metrics

[14] This classification has also been used in surveys conducted in Australia, Canada and the US.

where n indexes new technologies in production and NT_{ni}/NT_i denotes the share of firms in industry i which, using at least one new technology in production, use technology n. The Manufacture of electronic valves and tubes (CNAE 321) and the Manufacture of television and radio transmitters (CNAE 322) represent one of the closest industry relations in terms of sharing new technologies. We rank all J industries in descending order based on their technology similarity with industry i and construct the following industry-specific weights:

$$
\begin{aligned}
W_{ij}^{T} &= 0 && \text{if } r > 3 \\
W_{ij}^{T} &= \frac{technology\ similarity_{ij}}{\sum_{r=1}^{3} technology\ similarity_{ij}} && \text{if } r \leq 3
\end{aligned}
\tag{9}
$$

where r identifies the r^{th} closest industry in this knowledge spillovers metric. We set r=3 as a means of increasing the weight assigned to the closest industries. This may seem inconsistent with the weights defined to characterize proximity in the labor market pooling metric where we set r=10. However, note that there are fewer industry pairs to consider here. First, only the manufacturing industries were surveyed on their use of new technologies in production. Second, this survey is only available for an aggregation of the three-digit industry classification (29 manufacturing industries). As a matter of fact, three coincides with the number of industries whose value in the *technology similarity*$_{ij}$ metric typically exceeds its average value by more than one standard deviation.

Based on this industry-specific set of weights we construct the variable $techno_{ic}$:

$$
techno_{ic} = \sum_{j \neq i} \left(W_{ij}^{T} \cdot L_{cj} \right) \tag{10}
$$

which is the weighted sum of industry *(j)* and location *(c)* employment levels where industries that use more similar new technologies in their production processes are given higher weights. Hence, $techno_{ic}$ is a measure of the local employment in the industries that share knowledge and ideas with industry *i*.

Note that there are alternative ways to characterize the local industry mix other than computing the employment levels along different vectors of industry needs. To test the labor market pooling hypothesis, Glaeser and Kerr (2009) measure the extent to which the (nationwide) distribution of workers by occupation in industry i is similar to the analogous (nationwide) distributions of the industries that are more prevalent in location *c*. As for the input sharing hypothesis, these authors measure the extent to which the (nationwide) industry mix of the purchased inputs of industry i is similar to the industry mix of location *c*. Finally, their knowledge spillovers metric measures the extent to which the (nationwide) industry mix of the patents cited by industry *i* is similar to the industry mix of location *c*. Unlike ours, the metrics developed in Glaeser and Kerr (2009) are orthogonal to city size by construction.

4. Econometric specification and identification issues

Model Specification: We use the random profit maximization approach (Carlton, 1983) to formalize the location decisions of new firms. A linearized expected profit function can be written as:

$$\pi_{kic} = a_{ic}'\beta + \delta_i \cdot emp_{ic} + x_{ic}'\gamma_i + \varepsilon_{kic} \tag{11}$$

where π_{kic} denotes the profit level that firm *k*, belonging to industry *i*, would obtain in geographical unit *c*. This profit level is determined by local agglomeration economies that are relevant for industry *i*, a_{ic}. This vector contains the log-employment in industries that: 1) use workers with the same occupations as those used by industry *i* ($labor_{ic}$); 2) supply inputs to industry *i* ($input_{ic}$); 3) buy the outputs of industry *i* ($output_{ic}$); and 4) use the new production technologies ($techno_{ic}$) used by industry *i*. The variable emp_{ic}, captures the own-industry employment in location *c* whereas x_{ic} is a vector of control variables, which will be described below. ε_{kic} is an unobservable random term which varies across firms and locations.

If firms locate where profits are the highest and ε_{kic} follows an (iid) Extreme Value Type II distribution, the probability that firm *k* locates in geographical unit *c* has a Conditional Logit form:

$$Pr(\text{firm } k \text{ locates in } c) = \frac{exp(a_{ic}'\beta + \delta_i \cdot emp_{ic} + x_{ic}'\gamma_i)}{\sum_c exp(a_{ic}'\beta + \delta_i \cdot emp_{ic} + x_{ic}'\gamma_i)} \tag{12}$$

Guimarães et al (2003) have shown that the Conditional Logit coefficients can be equivalently estimated using the Poisson regression with exponential mean function:

$$E(N_{ict}) = exp(a_{ic}'\beta + \delta_i \cdot emp_{ic} + x_{ic}'\gamma_i) \tag{13}$$

where the dependent variable, N_{ict}, is the count of new firms in industry *i* that locate in geographical unit *c*. This implies that Poisson estimates can be given a Random Profit Maximization interpretation.

In a different vein, Becker and Henderson (2000) considered a situation in which each location has a latent pool of geographically immobile entrepreneurs. This pool of entrepreneurs will result in more or less new firms being created in industry *i* (as opposed to firms being created in other industries or firms not being created at all) depending on the expected profits of doing so (demand side) and the number of 'latent' entrepreneurs in the area (supply side). Hence, the number of firms being created in industry *i* and location *c* is determined by local variables that shift firms' profits (like local agglomeration economies) and the potential pool of local entrepreneurs (the size of the local economy). Hence, the estimates of (13) can also be interpreted as the outcome of geographically immobile entrepreneurs creating more or less firms in response to local conditions. Brülhart et al (2007) label these two observationally equivalent models as the 'Footloose Startup' and the 'Latent Startup' models.

In our empirical specification the dependent variable, N_{ic}, is the number of new firms created in industry *i* and geographical unit *c* (cities in the between-cities analysis and municipalities in the within-cities analysis) between 2002 and 2004. The explanatory variables correspond to 2001 (to avoid simultaneity). Since the explanatory variables are measured in logs, the estimated coefficients can be interpreted as elasticities[15]. All specifications include the own-industry employment as a control variable where a separate parameter is estimated for each industry, allowing the strength of the so-called localization economies to be industry-specific (i.e. $\delta_i \cdot emp_{ic}$)[16]. Our most parsimonious specification includes industry fixed effects (a_i) and two additional controls, the urban surface of the geographical unit of analysis ($land_c$) and a set of fixed effects for some aggregation of the geographical units of analysis (a_r). Hence, the baseline specification (whose results are reported in the first column of Tables 2 and 3) is:

$$E(N_{ic}) = \exp\left(a'_{ic}\beta + \delta_i \cdot emp_{ic} + \gamma_l \cdot land_c + a_r + a_i\right) \tag{14}$$

where $land_c$ will be the (log) land area of the city (in the between-cities analysis) or that of the municipality (in the within-cities analysis) and is included following Bartik (1985), who emphasized that geographical units with more available land are 'mechanically' more likely to be chosen. In the between-cities analysis, the term a_r corresponds to 17 European NUTS-2 fixed effects which control for location determinants that are common to all locations within a region such as the market potential (in terms of consumers)[17], regional policies, or the remoteness of an area. In the within-cities analysis, the term a_r corresponds to (aggregate) city fixed effects. In terms of the Random Profit Maximization Framework (the 'Footloose Entrepreneurial' model),

[15] Given that these variables are zero for some industries and municipalities, we follow Crépon and Duguet (1997) and sum one to the observations that are zero to take the log of this transformed variable. Additionally, we include a dummy variable that indicates whether the original variable was zero. For instance, $\beta_{labor} \cdot log\left(labor_{ic} + 1\left[labor_{ic} = 0\right]\right) + d_{labor} \cdot 1\left[labor_{ic} = 0\right]$ corresponds to the way in which $labor_{ic}$ enters the specification.

[16] Given that the employment level is also zero in some industries and municipalities, we apply to this variable the transformation proposed by Crépon and Duguet (1997) described in footnote 15.

[17] In a paper that examines the effect of consumer market potential on the location of multinational firms across European regions, Head and Mayer (2004) consider that Spain comprises 7 NUTS 1 regions.

one can think of location choices as being made in two sequential steps: Mobile entrepreneurs first choose the city and then, in the second step, decide in which municipality to locate within the chosen city. Our estimates can be interpreted as estimates of location determinants driving this second decision.

In the second specification, we additionally include the overall employment level excluding that of industry *i* (emp_{-ic}) in order to control for the so-called urbanization economies (the effect of the size of the local economy on firms' profitability) and for the fact that larger local economies have more latent entrepreneurs. Hence, the specification whose results correspond to the second column of Tables 2 and 3 is:

$$E(N_{ic}) = exp(a'_{ic}\beta + \delta_i \cdot emp_{ic} + \delta_{urb} \cdot emp_{-ic} + \gamma_l \cdot land_c + a_r + a_i) \tag{15}$$

In a third specification, using the fact that the variables of interest vary across industries and locations, we include location-specific fixed effects (city-fixed effects in the between-cities analysis and municipality-fixed effects in the within cities analysis). This implies that variables that only show variation across locations (e.g. $land_c$) are no longer identified[18]. The specification whose results are reported in the third column of Tables 2 and 3 is:

$$E(N_{ic}) = exp(a'_{ic}\beta + \delta_i \cdot emp_{ic} + a_c + a_i) \tag{16}$$

where a_c is the location fixed effect. This is our preferred specification since it effectively controls for location determinants (i.e. natural advantages) that are not always easy to measure, such as wages, the composition of the labor force, rents, business climate, land-use regulations, proximity to airports and major infrastructures.

In these analyses, there is one observation for each industry in every city (or municipality), implying that city (or municipality) shocks would generate correlated error terms. Failing to account for this group component of the error term can result in estimated standard errors that are too small (Moulton, 1990). Besides, as mentioned above: a) the match between the classification of products (Input-Output Tables) and industries is not perfect; and b) the Survey of the Use of New Technologies in Production was not carried out at the three-digit level in all industries. This implies that for some industries, the variables of interest ($input_{ic}$, $output_{ic}$ and $techno_{ic}$) take the same values for some three-digit industries within the two-digit industry classification, generating an additional source of (grouped-structure) correlation in the error term. In order to produce valid statistical inference, we cluster the standard errors at the city and two-

[18] In fact, the overall employment level does vary by industry, since it excludes own-industry employment. However, the variation is small and its inclusion generates problems of convergence in the estimation routines.

digit industry level in the between-city analysis (and at the municipality and two-digit industry level in the within-cities analysis).

Identification issues: Ellison et al (2010) explains the co-location of industry pairs as a function of the extent to which industry pairs use the same type of workers, have a customer-supplier relationship, and use the same new technologies. This approach, which exploits purely cross-sectional data variation, faces two important identification difficulties: simultaneity, and the presence of omitted variables with the potential of confounding the effects of interest. The co-location of an industry pair driven by natural advantage (e.g. the presence of a port) could induce firms in this industry pair to use the same type of workers, to start a customer-supplier relationship or to use the same type of new technologies, implying that industrial relations may be the result and not the cause of co-location (a simultaneity bias). Concerns regarding identification would not entirely disappear if one were willing to assume that inter-industry relations are the cause and not the result of co-location: it could be that industries that co-locate due to a common dependence on an unobserved natural advantage turn out to employ similar workers, use similar technologies or have a customer-supplier relationship (an omitted variables bias). For instance, two industries that turn out to use similar workers may locate in the same area not in order to share workers but attracted by the proximity to a hub airport (a location factor omitted by the researcher).

In order to minimize the potential confounding effect of natural advantages, Ellison et al (2010) construct an estimated spatial distribution of industries based on the 16 natural advantages studied in Ellison and Glaeser (1999). Using this estimated spatial distribution of industries, they construct an index which reflects co-agglomeration due to natural advantage and introduce this index as a control variable in the regressions. However, this control is not perfect, given the difficulties found in measuring some natural advantages. To deal with the simultaneity bias (the fact that inter-industry relations are the result and not the cause of agglomeration), Ellison et al (2010) resort to an instrumental variables approach, using UK data to construct measures of inter-industry relations which are then used to instrument their US counterparts. However, as the authors concede, these instruments will only mitigate this simultaneity bias if there are similarities in the ways in which natural advantage drives industry co-location in the US and in the UK[19].

Using the count of new firms as the dependent variable partially addresses the omitted variable and the simultaneity biases. Regarding the potential bias due to unobserved natural

[19] As an alternative set of instruments, Ellison et al (2010) measure inter-industry relations in areas in the US where pairs of industries are not co-agglomerated. The main results of the paper turn out to be similar using either the UK instruments or this alternative set.

advantage, the approach we follow allows us to condition the count of new firms in year *t* on the stock of own-industry employment in year *t*-1. Notice that the omitted factors that drive the location of new firms in year *t* are very likely to have driven the location decisions of new firms in the past. To give an example, in an industry where proximity to airports is particularly important, the geographical distribution of its old firms will be very strongly correlated with the geographical distribution of its new firms. Hence, the stock of employment in year *t*-1 acts as a catch-all control variable for sector-specific location determinants (either observed or unobserved)[20]. As pointed out by Rosenthal and Strange (2003), a study that examines the location decisions of new firms in the US, location attributes are fixed at the time of the start-up. In other words, the characteristics of cities (or municipalities) are seen as fixed from the viewpoint of a single entrepreneur, alleviating concerns about simultaneity. If sharing workers were the result and not the cause of co-location, the location of new firms would not react to the (pre-determined) geographical distribution of industries that use similar workers. Notice, however, that this is only true if there are no confounding unobserved location determinants (i.e. natural advantages). In this respect, we emphasize that besides including the stock of employment in year *t*-1 as a catch-all control for sector-specific location determinants, our preferred econometric specification, described in (6), contains location-specific fixed effects. These fixed effects control for all the observed and unobserved location determinants that do not vary by industry, including wages, the composition of the labor force, rents, business climate, land-use regulations, proximity to airports and major infrastructures.

Despite this, note that a local shock in the creation of firms in industry *i* might be correlated with local shocks affecting the (pre-determined) employment levels in industries that use similar workers, that have a customer-supplier relationship and that use similar technologies. Such a correlation could arise if there are cluster policies implemented at the local level to promote the creation of firms in specific industries in areas where the (pre-determined) employment levels in the relevant industries are already high. Therefore, one should interpret the results reported throughout this paper as partial correlations rather than as causal effects.

5. The results

Between-cities evidence: We first report and discuss the baseline results obtained when we analyze variation in new firms across (aggregated) cities. The first column in Table 2 shows

[20] Becker and Henderson (2000) argue that if location determinants are very persistent over time, conditioning the count of new firms in year *t* on the stock of pre-existing firms is essentially equivalent to introducing location- and sector-specific fixed effects.

the results of the specification described in (14), where new firms in industry *i* are regressed on the variables of interest (namely, $labor_{ic}$- employment in industries that use workers with the same occupations as those employed by industry *i*, $input_{ic}$- employment in industry *i*'s input suppliers, $output_{ic}$- employment in industry *i*'s customers *i* and $techno_{ic}$- employment in industries that use the same new production technologies as those used in industry *i*) and a set of control variables: own-industry employment, the urban surface of the city, and industry and regional fixed effects.

[Insert Table 2 here]

The fact that the explanatory variables are measured in logarithms coupled with the Poisson exponential mean specification implies that the coefficient estimates in Table 2 can be interpreted as elasticities. The estimates reported in the first column imply that a 1% increase in the city employment in industries that use workers with the same occupations as those used by industry *i* increases new firms' creation in this industry by 0.11%. Likewise, a 1% increase in the city employment in industries that provide the inputs to industry *i* increases new firms' creation in this industry by 0.27%. Employment increases in industry *i*'s customers and employment increases in industries that use the same new technologies in production as those used in industry *i* do not affect on the creation of new firms in this industry. Hence, the between-cities results suggest that labor market pooling and input sharing seem to be relevant agglomeration mechanisms at the city level, whereas knowledge spillovers do not.

The results of the second specification, described in (15), are reported in the second column of Table 2. This specification includes the employment level in the city (excluding that of industry *i*) as an additional control. This has implications for the way in which the estimates of interest are interpreted. Notice that an employment increase in a given industry, keeping the overall employment level constant, implies an employment reduction in another industry. Hence, the estimates reported in column 2 imply that a 1% employment increase in industry *i*'s input suppliers drawn from other industries in the city increases new firms' creation in this industry by 0.33%. Likewise, an analogous employment increase in industries that use similar workers as those used by industry *i* increases new firm creations in this industry by 0.13%. The negative coefficient estimate for the overall employment implies that more employment deters firm births, holding constant the employment in industry *i* and in those industries that are especially relevant for industry *i* (industries that use workers with the same occupations, have a customer or supplier relationship or use the same new production technologies). This suggests that the crowding effects associated with this employment increase (increased wages, rents and congestion) more than offsets the benefits of agglomeration. Notice that the positive effects of employment increases in specific industries can thus be interpreted as net effects of agglomeration (agglomeration benefits offsetting crowding or congestion costs).

Specification 3, described by (16), whose results are reported in the third column of Table 2 includes city fixed effects. The estimates imply that a 1% employment increase in industries that use similar workers to those used by industry *i* increases firm births by 0.12%. Likewise, a 1% increase in city employment in industries that provide inputs to industry *i* increases new firm births in this industry by 0.26%. Overall, the results are relatively similar in all three specifications and indicate that input sharing and labor market pooling are relevant agglomeration mechanisms, whereas we find no evidence supporting the relevance of the knowledge spillover theory.

<u>Within-cities evidence:</u> Different agglomeration mechanisms may operate at different intensities at different geographical scales. Table 3 shows the results of our analysis of variation in the creation of new firms across municipalities within large (aggregated) cities. In our baseline specification, we restrict our sample to (aggregated) cities where the central city has more than 200,000 inhabitants. The results shown in the three columns in Table 3 correspond to the specifications discussed in Table 1, adapted to the geographical unit of analysis in question (i.e. the municipality). In the first column, new firms are regressed on the variables of interest, own-industry employment, the urban surface of the municipality and city fixed effects. The results reported in the second column are those of a specification in which the overall (outside industry) employment is included as an additional control variable. In the third and last specification, there are municipality specific fixed effects which imply that identification comes from the variation in the creation of new firms across industries within municipalities.

[Insert Table 3 here]

The results reported in the first column in Table 3 imply that a 1% increase in municipal employment in industries that use similar workers to those used by industry *i* increases firm births in this industry by 0.065%. Likewise, an analogous employment increase in industries that provide inputs to industry *i* increases firm births in industry *i* by 0.22%. Employment increases in industries that buy the outputs of industry *i* and employment increases in industries that use the same new technologies as those used in industry *i* do not have an effect on firm births in industry *i*. The results in the second specification imply that increasing the overall employment in municipality *i* (holding constant the employment in industry *i* and in those industries that are especially relevant for industry *i*) reduces the creation of new firms in this industry by 0.41%. The comparison of the coefficient estimates in the first and second columns in Table 3 indicates that keeping employment size fixed increases the estimated effects of interest. The effect of an employment increase in industries that use workers with the same occupations rises from 0.065% to 0.115% whereas the effect of an employment increase in industries that supply inputs rises from 0.22% to 0.42%. Finally, an employment increase in industries that use the same new technologies in production as those used by industry *i* increases firm births by 11%, suggesting that knowledge spillovers may also be relevant. The results obtained in the third specification (which includes municipality fixed effects) are similar to those reported in the second column, although the effect of an employment increase in industries that buy the outputs of industry *i* increases and becomes (weakly) statistically significant. Overall, the results in Table 3 suggest that all agglomeration theories are relevant. There are many reasons why knowledge spillovers appear as a relevant agglomeration theory in the within-cities analysis (Table 3) and not in the between-cities analysis (Table 2). Nevertheless, we stress one of them: the geographical scope of knowledge spillovers is probably very limited and the municipality may be a more appropriate geographical unit to capture these effects.

The relative importance of different agglomeration mechanisms: All the reported coefficients have the interpretation of elasticities which are meaningful in themselves. However, in the interests of comparability across the size of the coefficient estimates (and the relative importance of different agglomeration mechanisms), we report the average marginal effect of increasing 1,000 employees in each of the variables of interest[21]. The results, based on the location specific fixed effects specification (results shown in the third columns of Tables 2 and 3), are shown in Table 4.

[Insert Table 4 here]

In the between-cities analysis, the estimates imply that an increase of 1,000 employees in industries that use workers with the same occupations as those used by industry *i* creates 2.24 new firms (over a 3-year period). Likewise, an increase of 1,000 employees in the industries that supply inputs to industry *i* creates 1.42 new firms over the same time period. Hence, taking these estimates at their face values implies that labor market pooling is a more relevant agglomeration theory than input sharing. Labor market pooling and input sharing seem to have the same order of magnitude when we examine variation in the creation of new firms across municipalities

[21] For the X variable, the marginal effect for individual *i* is given by $(\beta_X / X_i) \cdot exp(\cdot)$. We average the marginal effect across all observations.

within large cities. An increase of 1,000 employees in industries that use workers with the same occupations as those used by industry *i* creates 1.56 new firms, whereas the same employment increase in the industries that supply inputs to industry *i* creates 1.45 new firms. Much smaller is the implied effect of an equal increase in the employment of industries that use the same new technologies in production as those used by industry *i* (0.6 births). More employment in industries that buy the outputs of industry *i* has a tiny effect on the births of firms in this industry.

It is also interesting to compare the estimates across the two columns (between vs. within city evidence) since this may shed some light on the relevance of different agglomeration mechanisms at different geographical scales. The results indicate that an increase of 1,000 employees in industries that use workers with the same occupations as those used by industry *i* generates a higher impact if this increase is at the city level (2.24 new firms) than at the municipality level (1.56). This is consistent with the intuition that labor market pooling operates at the city-level (a self-contained labor market), implying that estimates based on within-city comparisons underestimate the labor market pooling effects by failing to internalize spillovers occurring between municipalities within cities. In contrast, an increase of 1,000 employees in industries that use the same new technologies as those used in industry *i* has a much larger effect if this increase is at the municipality level (0.6 new firms) rather than at the city-level (0.1 new firms) suggesting that in order to generate firm births, the activities using similar technologies must be concentrated in a given municipality within the city. The effects of an increase of 1,000 employees in industries that are the input suppliers of industry *i* are similar if they take place at the city or at the municipality level (about 1.4 new firms).

In order to better contextualize the order of magnitude of the results presented here, we compute the analogous marginal effects corresponding to an increase in the own-industry employment level. The results based on between-cities comparisons (third column of Table 2) imply that a 1,000 employees increase in a given industry increases own industry firm births by 8.81[22]. The corresponding figure for the analysis that exploits variation in the creation of firms across municipalities within large cities (third column of Table 3) is slightly smaller, 6.71. As mentioned above, the stock of employment in year *t*-1 will tend to be strongly correlated with (unobserved) sector-specific location determinants and, therefore, these effects will generally be biased upwards. Nevertheless, note that the effects of interest reported in Table 4 are relatively

[22] This effect is the average across industries since we estimate industry-specific coefficients for the effect of the own-industry employment.

small compared to these own-industry effects. The same conclusion is reached if instead of comparing marginal effects one compares elasticities, i.e. coefficient estimates. The average elasticity for the between-cities analysis is 0.42 and 0.35 for the within-cities analysis.

Robustness checks: As a first robustness check, we assess the extent to which our results are sensitive to the somewhat arbitrary definition of the local employment level in the industries that share workers ($labor_{ic}$) and knowledge ($techno_{ic}$) with industry *i*. $labor_{ic}$ ($techno_{ic}$) are weighted sums of industry (*j*) and location (*c*) employment levels where industries that use workers (new technologies) more similar to those used in industry *i* are given higher weights. Industries that are not among the ten closest in terms of sharing workers are given a weight of zero. Likewise, industries that are not among the three closest in terms of sharing knowledge are given a weight of zero. Among the 10(3) closest industries, the closer the industry is, the higher the weight assigned to this industry – see expressions (2) and (9) for a formal definition of these weighting schemes. Notice that 10(3) corresponds to the number of industries whose value in the $labor\ similarity_{ij}$ ($techno\ similarity_{ij}$) metric typically exceeds its average value by more than one standard deviation.

As a first alternative measure, we apply the scheme just described to the 15(5) closest industries. Formally, this amounts to setting r=15 in the labor market pooling metric, expression (2), and r=5 in the knowledge spillovers metric, expression (9). The results are shown in the second column of Tables 5 (between-cities) and 6 (within-cities). The second alternative that we consider can be described as follows. Industries that are not among the 10(3) closest are given a weight of zero but the 10(3) closest industries are all given the same weight. The results of this second exercise are shown in the third column of Tables 5 (between-cities) and 6 (within-cities). In all these specifications, the number of industries considered in the technology spillovers metric is smaller than that of the labor market pooling metric for the reasons detailed above. In columns 4 to 6 (in Tables 5 and 6) we report the results of the schemes described by expressions (2) and (9) but considering the same number of industries (3, 5 and 10) in both metrics.

[Insert Table 5 here]

[Insert Table 6 here]

There are no major differences across the results of the different specifications presented in Table 5 (between-cities analysis) although the labor market pooling effect becomes somewhat smaller in column 4, where we consider the closest 3 industries in both metrics. More significant differences appear in Table 6 (within-cities analysis). Here, the labor market pooling effect also becomes smaller the fewer industries we consider. Conversely, the knowledge spillovers effect becomes smaller the more industries we consider. In fact, the results reported in the last column of Table 6 indicate that this effect completely vanishes when the most similar industries (in the relevant dimension) are not given a high enough weight. These results show one limitation of the analysis. Our key explanatory variables are only proxies of the local employment levels in industries that use similar workers, that have a customer-supplier relationship and that use similar technologies. Hence, the results reported above need not imply that one agglomeration mechanism is more important than another one but rather that is measured with less noise.

The within-cities evidence is based on examining variations in the creation of new firms across municipalities within the largest cities in the country. In particular, we select the 19 cities whose central municipality has more than 200,000 inhabitants. In order to explore whether the results reported in Table 6 are sensitive to this particular cutoff, we replicate this within-cities evidence for the largest 6 and 31 cities in Spain (the number of cities whose central municipality has more than 500,000 and 150,000 inhabitants, respectively). The results are shown in the second and third columns in Table 7.

[Insert Table 7 here]

The overall tenor of the results does not change across the columns in Table 7, although the coefficient estimates that correspond to $techno_{ic}$ (the local employment level in the industries that share knowledge) change significantly across the specifications. The results suggest that when examining firm locations across municipalities within large cities, knowledge spillovers become increasingly important as one restricts the attention to increasingly large (and dense) cities. The estimates imply that a 1,000 employees increase in the industries that use the same new technologies in production as those used by industry *i* increases from 0.44, 0.60 and 1.01 if we focus on the 31, 19 and 6 largest cities. This suggests that knowledge spillovers are especially relevant in the densest economic environments.

Madrid ranks first in terms of new firm creations in several industries (See Table 1). As a final robustness check, we redo the analysis excluding Madrid. Table A3 replicates the (between-cities) estimates reported in Table 2 excluding the (aggregated) city of Madrid. Table A4 replicates the (within-cities) estimates in Table 3 excluding the municipality of Madrid. The qualitative results remain virtually unchanged.

[Insert Tables A3 and A4]

Discussion of the results: Our results corroborate those of previous studies in the literature which support the empirical relevance of the Marshallian agglomeration economies reviewed in the introduction. In fact, we find evidence for each of the three agglomeration mechanisms (labor market pooling, input sharing and technological spillovers). In this respect, our results are similar to those found by Dumais et al (1997), Glaeser and Kerr (2009) and Ellison et al (2010), the other studies that use inter-industry relations to assess the relative importance of different agglomeration mechanisms. Our results suggest that labor market pooling is the most important agglomeration mechanism (especially in the between-cities analysis). The same result has been found in Dumais et al (1997), Rosenthal and Strange (2001) and Glaeser and Kerr (2009) but not in Ellison et al (2010), who concluded that input sharing is the most relevant agglomeration mechanism.

The results of this paper suggest that knowledge spillovers may be relevant but, in any case, only at a very local level. This is consistent with Rosenthal and Strange's (2001) study that concludes that industries that are more knowledge-intensive are more spatially concentrated but only at the zip code level. It is also consistent with the results of a more recent study conducted by the same authors which indicates that human capital spillovers (measured by the effect of the local abundance of college graduates on local wages) are important and attenuate sharply with distance (Rosenthal and Strange, 2008)[23].

Even the estimates that make a stronger case for the knowledge spillovers mechanism (third column in Table 7) imply that sharing knowledge is less important than sharing workers or having a customer-supplier relationship to explain the co-location of industry pairs. Similar results appear in Dumais et al (1997), Glaeser and Kerr (2009) and Ellison et al (2010), probably related in some way to the difficulties found in measuring inter-industry knowledge flows. In fact, Ellison et al (2010) consider that part of the inter-industry knowledge flows may take place through the mobility of workers between industries (labor market pooling) or through customer-supplier relationships involving transmission of knowledge embodied in products or machinery (input sharing).

Our results also indicate that the labor market pooling mechanism may be more important in explaining agglomeration between cities than within cities. This is consistent with the intuition that labor market pooling should operate at the local labor market level. In contrast, the input sharing mechanism seems to act with the same strength in the between- and within-cities analyses. This is somewhat surprising since transport costs are not expected to be particularly high across locations within a city. As just explained, one possibility is that inter-industry customer-supplier relations partly capture knowledge flows between industries.

[23] In a related paper, Doms et al (2010) find that more educated areas have higher firm entry rates.

6. Conclusions

This paper contributes to the literature of the micro-foundations of agglomeration economies and quantifies the relative importance of each of Marshall's agglomeration mechanisms by examining the location of new manufacturing firms in Spain. We find evidence of the three Marshallian mechanisms (labor market pooling, input sharing and knowledge spillovers) but their incidence differs depending on the geographical scale of the analysis (variation in firm births across cities vs. variation in firm births across municipalities within large cities). Taken at face value, our estimates imply that the most important mechanism is the labor market pooling (especially in the between-cities analysis), followed by the input sharing. The knowledge spillovers mechanism seems to be much less important and, in any case, only relevant at a very local level (within-cities analysis). The findings of this paper are broadly in line with the US evidence reported in Dumais et al (1997), Glaeser and Kerr (2009) and Ellison et al (2010). Given the differences in the underlying population of firms in Spain and the US, it is tempting to conclude that the results obtained here may have some validity in other (developed) countries.

We would like to emphasize that ranking different agglomeration theories in terms of their quantitative importance is an exercise to be interpreted with some caution. On a conceptual basis, it is not obvious that different agglomeration theories can always be identified separately. In particular, knowledge spillovers may occur through the mobility of workers between industries or through customer-supplier relationships (Ellison et al, 2010). On a more practical vein, the variables measuring the local employment levels that are relevant according to the different agglomeration theories are only (noisy) proxies. Hence, the (estimated) relative importance of these theories will not be independent of the amount of noise in their proxies.

There are several extensions of the analysis presented here that may be worth pursuing. First, it would be interesting to examine if the effects found in this paper are heterogeneous across industries. This would allow answering questions such as: Is the employment in industries using similar new technologies more important for those industries that are more knowledge-intensive? Is employment in the relevant input supplier industries particularly important for those industries that make more intensive use of inputs? A second interesting extension would be to explore if the effects found in this paper are heterogeneous with respect to local average firm sizes. This would allow testing the Chinitz (1961) hypothesis, which states that the smaller the firms size in the input supplier industries, the highest the firm entry rate. Finally, it would be equally interesting to explore if the effects found in this paper are heterogeneous with respect to entrant's size, as suggested by the results in Holmes and Stevens (2002). Finally, the present analysis could also be extended to the geography of firm births in the services sector. We leave these extensions for future research.

This paper contributes to the empirical literature that sheds light on the agglomeration mechanisms that shape the geography of economic activities. A better understanding of these mechanisms can be important for ultimately applying wise local development policies. The finding that knowledge spillovers may have a very limited geographical scope can be of special interest for local policy makers aiming to promote knowledge-based activities.

Acknowledgements

We gratefully acknowledge the helpful comments of Albert Solé-Ollé, Jan Brueckner, Diego Puga, Clement Bosquet, Jorge de la Roca and many participants at the Urban Economics Sessions of the North American Regional Science Association Conference in Denver, November 2010. This research has received financial support from projects SEJ2007-65086 and ECO2010-16934 (*Ministerio de Educación y Ciencia*) and 2009SGR102 (*Generalitat de Catalunya*).

References

Agrawal, A., Cockburn, I., Rosell, C. 2010 Not invented here? Innovation in company towns, Journal of Urban Economics, 67, 78-89.

Agrawal, A., Devesh, K., McHale, J., 2008. How do spatial and social proximity influence knowledge flows? Evidence from patent data, Journal of Urban Economics, 64, 264-269.

Andersson, F., Burgess, S., Lane, J.I., 2007. Cities matching and the productivity gains of agglomeration, Journal of Urban Economics, 61, 112-128.

Arauzo-Carod, J.M., Liviano, D., Manjón, M., 2010. Empirical studies in industrial location: an assessment of their methods and results. Journal of Regional Science, 50, 685–711.

Bartik, T., 1985. Business Location Decisions in the United States: Estimates of the effects of Unionization, taxes, and other characteristics of states, Journal of Business and Economic Statistics, 3, 14-22.

Bartlesman, E., Caballero R., Lyons R., 1994. Customer and supplier driven externalities, American Economic Review*,* 84, 1075-1084.

Becker, R., Henderson, J.V., 2000. Effects of air quality regulations on polluting industries, Journal of Political Economy*,* 108, 379– 421.

Boix, R., Galletto, V., 2006. Sistemas Locales de Trabajo y Distritos Industriales Marshallianos en España, Economía Industrial, 359, 165-184.

Brülhart, M., Jametti, M., Schmidheiny, K., 2007. Do agglomeration economies reduce the sensitivity of firm location to tax differentials?, Discussion Paper 6606, CEPR.

Carlton, D. 1983. The Location and Employment Choices of New Firms: An Econometric Model with Discrete and Continuous Endogenous Variables, Review of Economics and Statistics, 65, 440-449.

Chinitz, B., 1961. Contrasts in Agglomeration: New York and Pittsburgh, American Economic Review, 51, 279-289.

Costa, D., Kahn, M., 2001. Power Couples, Quarterly Journal of Economics, 116, 1287-1315.

Crépon, B., Duguet, E., 1997. Estimating the innovation function from patent numbers: GMM on count panel data, Journal of Applied Econometrics, 12, 243-263.

Doms, M., Lewis, E., Robb, A., 2010. Local labor force education, new business characteristics, and firm performance, Journal of Urban Economics, 67, 61-77.????

Dumais, G., Ellison, G., Glaeser, E., 1997. Geographic concentration as a dynamic process, NBER Working Paper 6270.

Dumais, G., Ellison, G., Glaeser, E., 2002. Geographic concentration as a dynamic process, The Review of Economics and Statistics, 84, 193-204.

Duncan, O., Duncan, B., 1955. A Methodological Analysis of Segregation Indexes, American Sociological Review, 20, 210-17.

Duranton G,. Puga, D., 2004. Micro-foundations of urban agglomeration economies, in: Henderson, J.V., Thisse, J.F. (Eds.), Handbook of urban and regional economics, vol. 4, North-Holland, Amsterdam, pp. 2063-2117.

Ellison, G., Glaeser, E., 1999. The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?, American Economic Review, 89, 311-316.

Ellison, G., Glaeser, E., Kerr, W., 2010. What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns, American Economic Review, 100, 1195-1213.

Fallick, B., Fleischman, C., Rebitzer, J.B., 2006. Job hopping in the Sillicon Valley: the micro-foundations of an industry cluster, Review of Economics and Statistics, 88, 472-481.

Glaeser, E.L., Gottlieb, J.D., 2009. The wealth of cities: Agglomeration economies and spatial equilibrium in the United States, Journal of Economic Literature, 47, 983-1028.

Glaeser, E.L., Kerr, W.R. 2009. Local industrial conditions and entrepreneurship: How much of the spatial distribution can we explain?, Journal of Economics & Management Strategy, 18, 623-663.

Glaeser, E.L, Kerr, W.R., Ponzetto, G.A.M. 2010 Clusters and entrepreneurship, Journal of Urban Economics, 67, 150-168.

Guimarães, P., Figueiredo, O., Woodward, D., 2003. A tractable approach to the firm location decision problem, Review of Economics and Statistics, 85, 201–204.

Head, K., Mayer, T., 2004. Market Potential and the Location of Japanese Firms in the European Union, Review of Economics and Statistics, 86, 959-972.

Hesley, R.W., Strange, W.C., 1990. Agglomeration Economies and Matching in a System of Cities, Regional Science and Urban Economics, 20, 189-212.

Holmes, T., 1999. Localization of Industry and Vertical Disintegration, Review of Economics and Statistics, 81, 314-325.

Holmes, T., Stevens, J., 2002. Geographic Concentration and Establishment Scale, Review of Economics and Statistics, 84, 682-691.

Jaffe, A., Trajtenberg, M., Henderson, R., 1993. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations, Quarterly Journal of Economics, 108, 577-598.

Jofre-Monseny, J., 2009. The scope of agglomeration economies: Evidence from Catalonia, Papers in Regional Science, 88, 575-590.

Kerr, W.R., Kominers, S.D., 2010. Agglomerative forces and cluster shapes. Working Paper 11-061. Harvard Business School.

Krugman, P.R., 1991. Geography and Trade, MIT Press, Boston.

Li, B., Lu, Y. 2009. Geographic concentration and vertical disintegration: Evidence from China, Journal of Urban Economics, 65, 294-304.

Marshall, A., 1890, Principles of Economics, MacMillan, London.

Moulton, B. R., 1990. An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units, The Review of Economics and Statistics, 72, 334-338.

Overman, H.G., Puga, D., 2010, Labour pooling as a source of agglomeration: An empirical investigation, in: Edward L Glaeser (ed.) Agglomeration Economics. Chicago, IL: Chicago University Press.

Puga, D., 2010. The magnitude and causes of agglomeration economies, Journal of Regional Science, 50, 203-219.

Rosenthal, S., Strange, W.C., 2001. The determinants of agglomeration, Journal of Urban Economics, 50, 191-229.

Rosenthal, S., Strange, W.C., 2003. Geography, industrial organization and agglomeration, The Review of Economics and Statistics, 85, 377-393.

Rosenthal, S., Strange, W.C., 2004. Evidence on the Nature and Sources of Agglomeration Economics, in: Henderson, J.V., Thisse, J.F. (Eds.), Handbook of urban and regional economics, vol. 4. North-Holland, Amsterdam, 2119-2172.

Rosenthal, S., Strange, W.C., 2008. The attenuation of human capital spillovers, Journal of Urban Economics, 64, 373-389.

Rosenthal, S., Strange, W.C., 2010 Small establishments/big effects: Agglomeration, industrial organization and entrepreneurship, in Glaeser, E.L. (Ed.), The Economics of agglomeration. NBER. University Chicago Press, Chicago, 277-302.

Scherer, F. M., 1984. Using Linked Patent Data and R&D Data to Measure Technology Flows, in: Griliches, Z. (Ed.), R & D, Patents and Productivity,: The University of Chicago Press. Chicago, IL, pp. 417-464.

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

Table 1a. New firms in Spain. City-level. Years 2002-2004. 75 three-digit manufacturing industries

Industry	New firms	New firms (%)	Mean	Maximum	Cities with zero births (%)
The five industries with the highest number of new firms					
Manufacture of structural metal products (CNAE 281)	2,188	15.65%	4.587	167 (Madrid)	26.21%
Printing and service activities related to printing (CNAE 222)	1,159	8.29%	2.430	294 (Madrid)	61.64%
Manufacture of furniture (CNAE 361)	1,108	7.92%	2.323	101 (Valencia)	49.06%
Publishing (CNAE 221)	971	6.94%	2.036	329 (Madrid)	73.38%
Manufacture of other wearing apparel and accessories (CNAE 182)	593	4.24%	1.243	86 (Madrid)	69.81%
Median					
Manufacture of luggage, handbags and the like, saddlery and harness (CNAE 192)	73	0.52%	0.153	13 (Ubrique - Elda)	94.76%
The five industries with the lowest number of new firms					
Manufacture of motor vehicles (CNAE 341)	19	0.14%	0.040	3 (Barcelona - Zaragoza)	96.86%
Manufacture of grain mill products, starches and starch products (CNAE 156)	18	0.13%	0.377	2 (Madrid)	96.44%
Manufacture of sports goods (CNAE 364)	17	0.12%	0.356	6 (Barcelona)	97.90%
Manufacture of leather clothes (CNAE 181)	16	0.11%	0.335	4 (Madrid)	97.48%
Manufacture of insulated wire and cable (CNAE 313)	16	0.11%	0.335	3 (Barcelona - Zaragoza)	97.69%

Table 1b. New firms in Spain. Municipalities within largest cities. Years 2002-2004. 62 three-digit manufacturing industries

Industry	New firms	New firms (%)	Mean	Maximum	Municipalities with zero births (%)
The five industries with the highest number of new firms					
Manufacture of structural metal products (CNAE 281)	836	14.32%	1.107	45 (Barcelona)	66.75%
Publishing (CNAE 221	721	12.35%	0.955	241 (Madrid)	86.23%
Printing and service activities related to printing (CNAE 222)	721	12.35%	0.955	148 (Madrid)	79.47%
Manufacture of furniture (CNAE 361)	402	6.89%	0.532	25 (Madrid)	81.19%
Manufacture of other wearing apparel and accessories (CNAE 182)	312	5.34%	0.413	62 (Madrid)	88.34%
Median					
Manufacture of parts and accessories for motor vehicles and their engines (CNAE 343)	39	0.67%	0.516	5 (Madrid)	96.82%
The five industries with the lowest number of new firms					
Manufacture of electricity distribution and control apparatus (CNAE 312)	17	0.29%	0.225	2 (Madrid)	98.01%
Dressing and dyeing of fur; manufacture of articles of fur (CNAE 183)	16	0.27%	0.212	7 (Barcelona)	98.81%
Manufacture of diverse non-metallic mineral products (CNAE 268)	16	0.27%	0.212	2 (Murcia)	98.01%
Manufacture of vegetable and animal oils and fats (CNAE 154)	15	0.26%	0.199	3 (Madrid)	98.68%
Manufacture of pulp, paper and paperboard (CNAE 211)	15	0.26%	0.199	3 (Barcelona)	98.28%

Source: Bureau van Dijk Electronic Publishing (SABI).

Table 2. Agglomeration economies estimates (between-cities evidence). Poisson estimates. The dependent variable is the count of new firms created by industry and city.

	I	II	III
Agglomeration mechanisms			
Labor Market Pooling; $labor_{ic}$	0.107*** (0.022)	0.125*** (0.022)	0.118*** (0.019)
Input Sharing;			
$input_{ic}$	0.268*** (0.042)	0.326*** (0.046)	0.264*** (0.043)
$output_{ic}$	-0.036 (0.044)	-0.007 (0.043)	0.042 (0.042)
Knowledge Spillovers; $techno_{ic}$	-0.025 (0.019)	-0.001 (0.020)	0.010 (0.020)
Controls			
City employment (excluding that of industry i)		-0.143*** (0.037)	n.i.
Own industry city employment	Yes	Yes	Yes
City land area	Yes	Yes	n.i.
City fixed effects	No	No	Yes
Regional fixed effects (17 NUTS 2 regions)	Yes	Yes	n.i.
Industry fixed effects	Yes	Yes	Yes
No. of industries	75	75	75
No. of cities	477	477	477
No. of observations	35,775	35,775	35,775

Notes: 1) Robust standard errors (in parentheses) clustered at the two-digit industry and city level; 2) ***, ** and * statistically significant at 1, 5 and 10 percent; 3) All the explanatory variables measured in its logarithmic form (those with zero values have been transformed as detailed in the text); 4) $labor_{ic}$, $input_{ic}$, $output_{ic}$ and $techno_{ic}$ are (weighted) sums of the employment in different industries in each city. The weights are industry-specific and reflect the intensity of inter-industry relationships. $labor_{ic}$ is the employment in industries that use workers with the same occupations as those employed in industry *i*. $input_{ic}$ is the employment in industry *i*'s input suppliers. $output_{ic}$ is the employment in industry *i*'s customers and $techno_{ic}$ is the employment in industries that use the same new production technologies as those used in industry *i*; 5) n.i. indicates that the variable is not identified because it does not vary across industries in a given city.

Table 3. Agglomeration economies estimates (within-cities evidence). Poisson estimates. The dependent variable is the count of new firms created by industry and municipality.

	I	II	III
Agglomeration mechanisms			
Labor Market Pooling;	0.065^{*}	0.115^{***}	0.099^{***}
$labor_{ic}$	(0.033)	(0.034)	(0.033)
Input Sharing;			
$input_{ic}$	0.225^{***}	0.424^{***}	0.369^{***}
	(0.067)	(0.073)	(0.060)
$output_{ic}$	-0.027	0.026	0.095^{*}
	(0.059)	(0.061)	(0.055)
Knowledge Spillovers;	0.016	0.113^{***}	0.130^{***}
$techno_{ic}$	(0.030)	(0.034)	(0.034)
Controls			
Overall municipality employment (excluding that of industry *i*)		-0.412^{***} (0.060)	n.i.
Own industry employment in the municipality	Yes	Yes	Yes
Municipality land area	Yes	Yes	n.i.
Municipality fixed effects	No	No	Yes
City fixed effects	Yes	Yes	n.i.
Industry fixed effects	Yes	Yes	Yes
No. of industries	62	62	62
No. of municipalities	775	775	775
No. of cities	19	19	19
No. of observations	48,050	48,050	48,050

Notes: 1) Robust standard errors (in parentheses) clustered at the two-digit industry and municipality level; 2) ***, ** and * statistically significant at 1, 5 and 10 percent; 3) All the explanatory variables measured in its logarithmic form (those with zero values have been transformed as detailed in the text); 4) $labor_{ic}$, $input_{ic}$, $output_{ic}$ and $techno_{ic}$ are (weighted) sums of the employment in different industries in each municipality. The weights are industry-specific and reflect the intensity of inter-industry relationships. $labor_{ic}$ is the employment in industries that use workers with the same occupations as those employed in industry *i*. $input_{ic}$ is the employment in industry *i*'s input suppliers. $output_{ic}$ is the employment in industry *i*'s customers and $techno_{ic}$ is the employment in industries that use the same new production technologies as those used in industry *i*; 5) n.i. indicates that the variable is not identified because it does not vary across industries in a given municipality.

Table 4. The effect on the creation of new firms of increasing employment (by one thousand employees) in industries that share workers, have a customer-supplier relationship and share knowledge

Agglomeration mechanisms	Between-cities evidence	Within-cities evidence
Labor Market Pooling;		
$labor_{ic}$	2.238***	1.561***
Input Sharing;		
$input_{ic}$	1.421***	1.449***
$output_{ic}$	0.059	0.139*
Knowledge Spillovers;		
$techno_{ic}$	0.103	0.603***

Notes: 1) Effects implied by the estimates reported in the third columns of Table 2 and 3 (Between-cities and within-cities evidence); 2) The marginal effect is computed for each observation and averaged across all observations; 3)***, ** and * statistically significant at 1, 5 and 10%.

Table 5. Robustness checks. Alternative weight schemes. Poisson estimates (between-cities evidence). The dependent variable is the count of new firms created by industry and city.

	Closest 10 (labor) and 3 (techno) industries, weighted	**Closest 15 (labor) and 5 (techno) industries, weighted**	**Closest 10 (labor) and 3 (techno) industries, unweighted**	**Closest 3 (labor) and 3 (techno) industries, weighted**	**Closest 5 (labor) and 5 (techno) industries, weighted**	**Closest 10 (labor) and 10 (techno) industries, weighted**
Agglomeration mechanisms						
Labor Market Pooling;	0.118***	0.111***	0.119***	0.0516***	0.0771***	0.123***
labor$_{ic}$	(0.019)	(0.023)	(0.019)	(0.0145)	(0.0156)	(0.0187)
Input Sharing;						
input$_{ic}$	0.264***	0.248***	0.264***	0.270***	0.263***	0.259***
	(0.043)	(0.043)	(0.043)	(0.0427)	(0.0417)	(0.0424)
output$_{ic}$	0.042	0.0676	0.042	0.0470	0.0514	0.0554
	(0.042)	(0.042)	(0.042)	(0.0419)	(0.0416)	(0.0412)
Knowledge Spillovers;	0.010	0.0208	0.010	0.0215	0.0200	-0.0616
techno$_{ic}$	(0.020)	(0.022)	(0.020)	(0.0202)	(0.0222)	(0.0389)
Controls						
City employment (excluding that of industry i)	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.
Own industry city employment	Yes	Yes	Yes	Yes	Yes	Yes
City land area	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.
City fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Regional fixed effects (17 NUTS 2 regions)	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of industries	75	75	75	75	75	75
No. of cities	477	477	477	477	477	477
No. of observations	35,775	35,775	35,775	35,775	35,775	35,775

Notes: 1) Robust standard errors (in parentheses) clustered at the two-digit industry and city level; 2) ***, ** and * statistically significant at 1, 5 and 10%; 3) All the explanatory variables measured in its logarithmic form (those with zero values have been transformed as detailed in the text); 4) *labor*$_{ic}$, *input*$_{ic}$, *output*$_{ic}$ and *techno*$_{ic}$ are (weighted) sums of the employment in different industries in each city. The weights are industry-specific and reflect the intensity of inter-industry relationships. *labor*$_{ic}$ is the employment in industries that use workers with the same occupations as those employed in industry *i*. *input*$_{ic}$ is the employment in industry *i*'s input suppliers. *output*$_{ic}$ is the employment in industry *i*'s customers and *techno*$_{ic}$ is the employment in industries that use the same new production technologies as those used in industry *i*; 5) n.i. indicates that the variable is not identified because it does not vary across industries in a given city.

Table 6. Robustness checks. Alternative weight schemes. Poisson estimates (within-cities evidence). The dependent variable is the count of new firms created by industry and municipality.

	Closest 10 (labor) and 3 (techno) industries, weighted	**Closest 15 (labor) and 5 (techno) industries, weighted**	**Closest 10 (labor) and 3 (techno) industries, unweighted**	**Closest 3 (labor) and 3 (techno) industries, weighted**	**Closest 5 (labor) and 5 (techno) industries, weighted**	**Closest 10 (labor) and 10 (techno) industries, weighted**
Agglomeration mechanisms						
Labor Market Pooling;	0.099***	0.091**	0.105***	0.0353*	0.0461*	0.120***
$labor_{ic}$	(0.033)	(0.041)	(0.034)	(0.021)	(0.025)	(0.033)
Input Sharing;						
$input_{ic}$	0.369***	0.372***	0.370***	0.383***	0.376***	0.342***
	(0.060)	(0.061)	(0.060)	(0.060)	(0.062)	(0.061)
$output_{ic}$	0.095*	0.128**	0.093*	0.0929*	0.123**	0.140**
	(0.055)	(0.056)	(0.055)	(0.055)	(0.055)	(0.055)
Knowledge Spillovers;	0.130***	0.099***	0.115***	0.139***	0.103***	0.0257
$techno_{ic}$	(0.034)	(0.038)	(0.034)	(0.033)	(0.038)	(0.073)
Controls						
Overall municipality employment (excluding that of industry *i*)	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.
Own industry employment in the municipality	Yes	Yes	Yes	Yes	Yes	Yes
Municipality land area	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.
Municipality fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
City fixed effects	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of industries	62	62	62	62	62	62
No. of municipalities	755	755	755	755	755	755
No. of cities	19	19	19	19	19	19
No. of observations	48,050	48,050	48,050	48,050	48,050	48,050

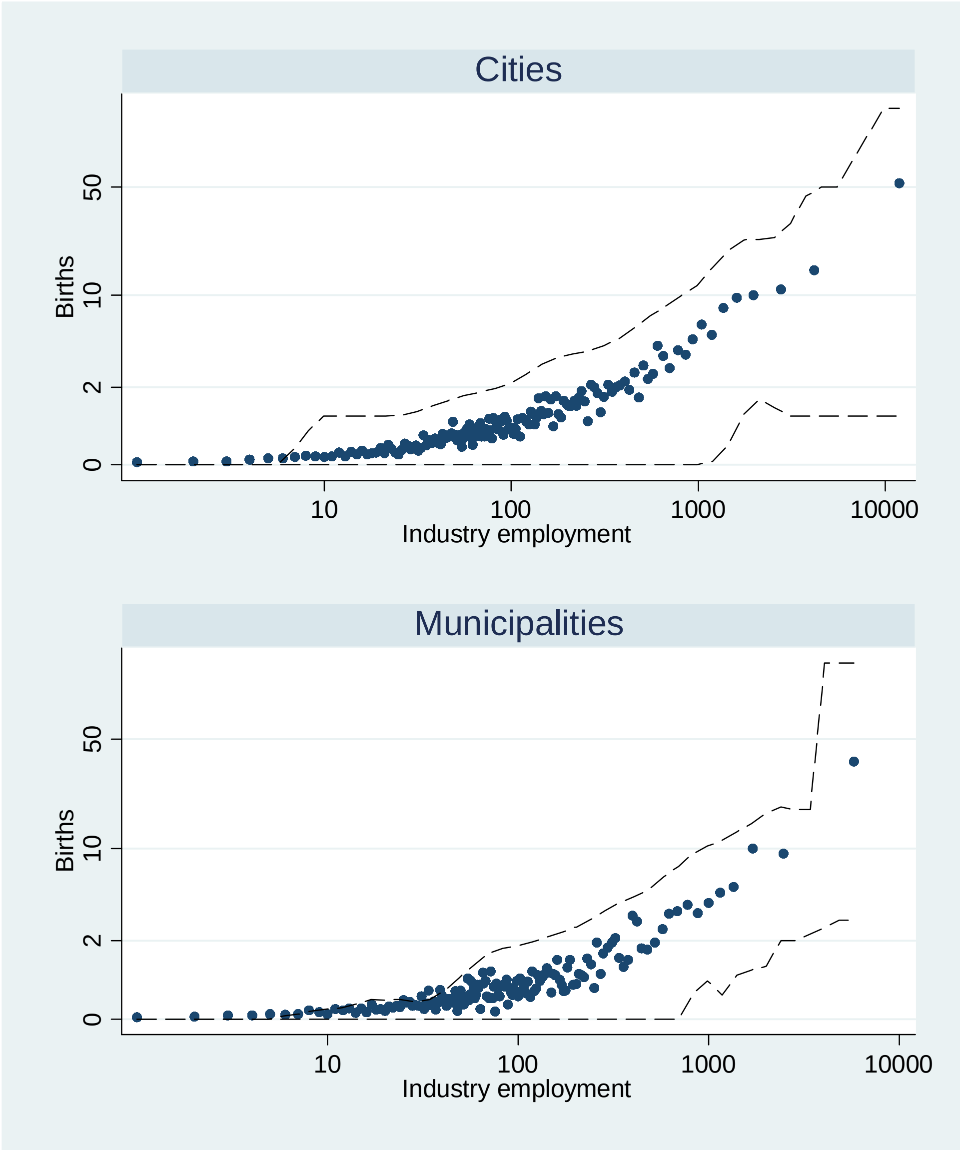
Notes: 1) Poisson estimates based on the third specification of Table 3 (replicated in the first column); 2) Robust standard errors (in parentheses) clustered at the two-digit industry and municipality level; 3) ***, ** and * statistically significant at 1, 5 and 10%; 4) All the explanatory variables measured in its logarithmic form (those with zero values have been transformed as detailed in the text); 5) $labor_{ic}$, $input_{ic}$, $output_{ic}$ and $techno_{ic}$ are (weighted) sums of the employment in different industries in each municipality. The weights are industry-specific and reflect the intensity of inter-industry relationships. $labor_{ic}$ is the employment in industries that use workers with the same occupations as those employed in industry *i*. $input_{ic}$ is the employment in industry *i*'s input suppliers. $output_{ic}$ is the employment in industry *i*'s customers and $techno_{ic}$ is the employment in industries that use the same new production technologies as those used in industry *i*; 6) n.i. indicates that the variable is not identified because it does not vary across industries in a given municipality.

Table 7. Robustness checks. Alternative definitions of large city defined by the population of the largest municipality in the city. Poisson estimates. The dependent variable is the count of new firms created by industry and municipality.

	>500,000 inhabitants	**>200,000 Inhabitants**	**>150,000 Inhabitants**
Agglomeration mechanisms			
Labor Market Pooling;	0.126^{***}	0.099^{***}	0.088^{***}
labor$_{ic}$	(0.043)	(0.033)	(0.028)
Input Sharing;			
input$_{ic}$	0.360^{***}	0.369^{***}	0.377^{***}
	(0.077)	(0.060)	(0.053)
output$_{ic}$	0.066	0.095^{*}	0.103^{**}
	(0.072)	(0.055)	(0.050)
Knowledge Spillovers;	0.237^{***}	0.130^{***}	0.105^{***}
techno$_{ic}$	(0.0539)	(0.034)	(0.029)
Controls			
Overall municipality employment (excluding that in industry *i*)	n.i.	n.i.	n.i.
Own industry employment in the municipality	Yes	Yes	Yes
Municipality land area	n.i.	n.i.	n.i.
Municipality fixed effects	Yes	Yes	Yes
City fixed effects	n.i.	n.i.	n.i.
Industry fixed effects	Yes	Yes	Yes
No. of industries	62	62	62
No. of municipalities	348	755	1,421
No. of cities	6	19	30
No. of observations	21,576	48,050	88,102

Notes: 1) Poisson estimates based on the third specification of Table 3 (replicated in the first column); 2) Robust standard errors (in parentheses) clustered at the two-digit industry and municipality level; 3) ***, ** and * statistically significant at 1, 5 and 10%; 4) All the explanatory variables measured in its logarithmic form (those with zero values have been transformed as detailed in the text); 5) *labor*$_{ic}$, *input*$_{ic}$, *output*$_{ic}$ and *techno*$_{ic}$ are (weighted) sums of the employment in different industries in each municipality. The weights are industry-specific and reflect the intensity of inter-industry relationships. *labor*$_{ic}$ is the employment in industries that use workers with the same occupations as those employed in industry *i*. *input*$_{ic}$ is the employment in industry *i*'s input suppliers. *output*$_{ic}$ is the employment in industry *i*'s customers and *techno*$_{ic}$ is the employment in industries that use the same new production technologies as those used in industry *i*; 6) n.i. indicates that the variable is not identified because it does not vary across industries in a given municipality.

Graphic 1. Count of new firms by industry and location as a function of the own-industry employment level.



Notes: Each dot represents the average number of firm births in each cell; The dashed lines represent the 10th and 90th percentiles within each cell. Each of the first 50 cells represents only one employment value (1 to 50). Beyond this point, each cell contains a percentile of the remaining observations (49 and 27 observations respectively).

Online Appendix. Table A1. 1994 National Classification of Occupations (CNO 94)

Code	Title	Code	Title
001	Armed forces: top officers	232	Judges
002	Armed forces: middle officers	239	Legal professionals not elsewhere classified
003	Armed forces: regular officers	241	Business professionals
101	Legislators	242	Economists
102	Senior government officials	243	Social science and related professionals
103	Heads of villages & towns	251	Writers and creative or performing artists
104	Senior officials of special-interest organisations	252	Archivists, librarians and related information professionals
111	Directors and chief executives	253	Other professionals in the Public Administration
112	Production and operations department managers	261	Physical and engineering science technicians
113	Other department managers	262	Mathematics and statistics technicians
121	General managers in wholesale trade with less than 10 employees	263	Computer associate professionals
122	General managers in retail trade with less than 10 employees	264	Architecture technicians
131	General managers of hotels with less than 10 employees	265	Engineering technicians (e.g., ship and aircraft technicians)
132	General managers of restaurants with less than 10 employees	271	Life science technicians and related associate professionals
140	General managers not elsewhere classified with less than 10 employees	272	Nursing and midwifery associate professionals
151	General managers in wholesale trade with no employees	281	Primary and pre-primary education teaching associate professionals
152	General managers in retail trade with no employees	282	Special education teaching associate professionals
161	General managers of hotels with no employees	283	Other teaching associate professionals
162	General managers of restaurants with no employees	291	Accountants, personnel and careers professionals, and other
170	General managers not elsewhere classified with no employees	292	Archivists, librarians and related information professionals
201	Physicists, chemists and related professionals	293	Social science and related professionals
202	Mathematicians, statisticians and related professionals	294	Religious professionals
203	Computing professionals	295	Government professionals
204	Architects, town and traffic planners	301	Draughtspersons
205	Engineers	302	Physical, chemical and engineering science technicians
211	Life science professionals	303	Computer assistants
212	Medical doctors & dentists	304	Optical and electronic equipment operators
213	Veterinarians	305	Ships' engineers, deck officers and pilots
214	Pharmacists	306	Aircraft pilots, air traffic controllers and safety technicians
219	Health professionals (except nursing) not elsewhere classified	307	Safety, health and quality inspectors
221	College, university and higher education teaching professionals	311	Life science technicians
222	Secondary education teaching professionals	312	Medical assistants
223	Other teaching professionals	313	Modern health associate professionals (except nursing) not elsewhere classified
231	Lawyers	321	Pre-primary and special education teaching associate professionals
232	Judges	322	Other teaching associate professionals

Online Appendix. Table A1. 1994 National Classification of Occupations (CNO 94) (Continuation)

Code	Title	Code	Title
331	Finance and sales associate professionals	532	Shop salespersons and demonstrators
332	Technical and commercial sales representatives	533	Stall and market salespersons
341	Administrative associate professionals	601	Self-employed market gardeners and crop growers
342	Customs, tax and related government associate professionals	602	Employed market gardeners and crop growers
351	Business services agents and trade brokers	611	Self-employed market-oriented animal producers and related workers
352	Police inspectors and detectives	612	Employed market-oriented animal producers and related workers
353	Social work associate professionals	621	Self-employed market-oriented crop and animal producers
354	Artistic, entertainment and sports associate professionals	622	Self-employed forestry and related workers
355	Religious clerks	623	Employed market-oriented crop and animal producers
401	Numerical clerks	624	Employed forestry and related workers
402	Material-recording and transport clerks	631	Self-employed fishery workers, hunters and trappers
410	Library, mail and related clerks	632	Employed fishery workers, hunters and trappers
421	Secretaries and keyboard-operating clerks	701	Foremen of building frame and related trades workers
422	Data entry operators	702	Foremen of building finishers and related workers
430	Other office clerks with no contact with customers	703	Painters, building structure cleaners and related trades
440	Other office clerks with contact with customers	711	Bricklayers and stonemasons
451	Client information clerks	712	Concrete placers, concrete finishers and related workers
452	Travel agency, receptionists and information clerks, and related	713	Carpenters and joiners
460	Cashiers, tellers and related clerks	714	Building frame and related trades workers not elsewhere classified
501	Cooks	721	Plasterers
502	Waiters, waitresses and bartenders	722	Plumbers and pipe fitters
503	Restaurant and bar maitresses	723	Building and related electricians
511	Institution and home-based personal care workers	724	Painters, varnishers and related painters and workers
512	Other personal care and related workers	725	Building structure cleaners
513	Hairdressers, barbers, beauticians and related workers	729	Building finishers and related trades workers not elsewhere classified
514	Travel attendants and related workers	731	Shopfloor foremen of metal moulders, welders, sheet-metal workers
515	Housekeepers and related workers	732	Shopfloor foremen of motor vehicle mechanics and fitters
519	Other personal services workers	733	Shopfloor foremen of machinery and aircraft engine mechanics
521	Paramilitary police officers	734	Shopfloor foremen of electrical and electronic equipment
522	Police officers	741	Foremen of miners, shotfirers, stone cutters and carvers
523	Fire-fighters	742	Miners, shotfirers, stone cutters and carvers
524	Prison guards	751	Metal moulders, welders, sheet-metal workers, structural metal workers
525	Private guards	752	Blacksmiths, tool-makers and related trades workers
529	Protective services workers not elsewhere classified	761	Machinery mechanics and fitters
531	Fashion and other models	762	Electrical and electronic equipment mechanics and fitter

Online Appendix. Table A1. 1994 National Classification of Occupations (CNO 94) (Continuation)

Code	Title	Code	Title
771	Precision workers in metal and related materials	834	Wood-products machine operators
772	Printing and related trades workers	835	Printing-, binding- and paper-products machine operators
773	Potters, glass-makers and related trades workers	836	Textile-, fur- and leather-products machine operators
774	Handicraft workers in wood, textile, leather and related workers	837	Food and related products machine operators
780	Food processing and related trades workers	841	Assemblers
791	Wood treaters and related trades workers	849	Other machine operators and assemblers
792	Joiners and cabinet-makers	851	Locomotive-engine drivers and related workers
793	Textile, garment and related trades workers	852	Foremen of agricultural and other mobile-plant operators
794	Pelt, leather and shoemaking trades workers	853	Agricultural and other mobile-plant operators
801	Foremen of mining- and mineral-processing-plant operator	854	Other agricultural and other mobile-plant operators not elsewhere classified
802	Foremen of metal-processing-plant operators	855	Ships' deck crews and related workers
803	Foremen of glass, ceramics and related plant operators	861	Car, taxi and van drivers
804	Foremen of wood-processing- and papermaking-plant operators	862	Bus and tram drivers
805	Shopfloor foremen of chemical-processing-plant operators	863	Heavy truck and lorry drivers
806	Shopfloor foremen of power-production and related plant operators	864	Motor-cycle drivers
807	Shopfloor foremen of automated-assembly-line and industr	900	Street vendors and related workers
811	Mining- and mineral-processing-plant operators	911	Domestic helpers and cleaners
812	Metal-processing-plant operators	912	Helpers and cleaners in offices, hotels and other establishments
813	Glass, ceramics and related plant operators	921	Building caretakers, window and related cleaners
814	Wood-processing- and papermaking-plant operators	922	Watchpersons
815	Chemical-processing-plant operators	931	Shoe cleaning and other street services elementary occupations
816	Power-production and related plant operators	932	Doorkeepers and related workers
817	Automated-assembly-line and industrial-robot operators	933	Messengers, package and luggage porters and deliverers
821	Foremen of metal- and mineral-products machine operators	934	Vending-machine money collectors, meter readers and related workers
822	Foremen of chemical-products machine operators	935	Garbage collectors and related labourers
823	Foremen of rubber- and plastic-products machine operators	941	Agricultural labourers
824	Foremen of wood-products machine operators	942	Catlle, hunting and trapping labourers
825	Shopfllor foremen of printing-, binding- and paper-production	943	Farm-hands and labourers
826	Foremen of textile-, fur- and leather-products machine operators	944	Forestry labourers
827	Foremen of food and related products machine operators	945	Fishery labourers
828	Foremen of assemblers	950	Mining and quarrying labourers
831	Metal- and mineral-products machine operators	960	Building and other construction and maintenance labourer
832	Chemical-products machine operators	970	Manufacturing labourers
833	Rubber- and plastic-products machine operators	980	Transport labourers and freight handlers

Source: National Statistics Institute (INE)

Online Appendix. Table A2. New technologies in manufacturing classification (1998 Technological Innovation in Companies Survey)

1.1	Computer Assisted Design (CAD) and/or computer assisted engineering (CAE)
1.2	Computer assisted design applicable to the monitoring of the production of machinery (computer assisted manufacturing) CAD/CAM
1.3	Use of the digital output of the CAD for buying or provisioning activities
2.1	NC/CNC autonomous machines
2.2	Flexible manufacturing cells or systems (FMC/FMS)
2.3	Laser for the treatment of material
2.4	Advanced technologies other than those using laser
2.5	Pick & Place robots
2.6	Other more complex robots
3.1	Automatic storage and recovery systems
3.2	Automatic guided vehicle systems
4.1	Inspection based on automated sensor and/or test equipment conducted in the input of materials or during the process
4.2	Inspection based on automated sensor and/or test equipment conducted on the final product
5.1	Local area network computers for technical information
5.2	Local area network computers for use in factory
5.3	Information network between companies connecting the factory with subcontractors, suppliers and/or clients
5.4	Internet/electronic mail
5.5	Programmable logic controllers
5.6	Industrial control computers
6.1	Total quality control
6.2	Just in time systems
6.3	Planning of material needs
6.4	Planning of manufacturing resources
7.1	Manufacturing integrated by computer
7.2	Entry and supervision of production data
7.3	Artificial intelligence and/or expert systems

Source: National Statistics Institute (INE)

Online Appendix. Table A3. Agglomeration economies estimates (between-cities evidence) excluding the (aggregated) city of Madrid. Poisson estimates. The dependent variable is the count of new firms created by industry and city.

	I	II	III
Agglomeration mechanisms			
Labor Market Pooling;	0.097^{***}	0.121^{***}	0.115^{***}
$labor_{ic}$	(0.022)	(0.023)	(0.019)
Input Sharing;			
$input_{ic}$	0.306^{***}	0.380^{***}	0.320^{***}
	(0.043)	(0.045)	(0.041)
$output_{ic}$	-0.086^{*}	-0.046	0.001
	(0.045)	(0.045)	(0.044)
Knowledge Spillovers;	-0.022	0.008	0.025
$techno_{ic}$	(0.019)	(0.020)	(0.020)
Controls			
City employment (excluding that of industry i)			n.i.
Own industry city employment	Yes	Yes	Yes
City land area	Yes	Yes	n.i.
City fixed effects	No	No	Yes
Regional fixed effects (17 NUTS 2 regions)	Yes	Yes	n.i.
Industry fixed effects	Yes	Yes	Yes
No. of industries	75	75	75
No. of cities	476	476	476
No. of observations	35,700	35,700	35,700

Notes: 1) Robust standard errors (in parentheses) clustered at the two-digit industry and city level; 2) ***, ** and * statistically significant at 1, 5 and 10 percent; 3) All the explanatory variables measured in its logarithmic form (those with zero values have been transformed as detailed in the text); 4) $labor_{ic}$, $input_{ic}$, $output_{ic}$ and $techno_{ic}$ are (weighted) sums of the employment in different industries in each city. The weights are industry-specific and reflect the intensity of inter-industry relationships. $labor_{ic}$ is the employment in industries that use workers with the same occupations as those employed in industry *i*. $input_{ic}$ is the employment in industry *i*'s input suppliers. $output_{ic}$ is the employment in industry *i*'s customers and $techno_{ic}$ is the employment in industries that use the same new production technologies as those used in industry *i*; 5) n.i. indicates that the variable is not identified because it does not vary across industries in a given city.

Online Appendix. Table A4. Agglomeration economies estimates (within-cities evidence) excluding the municipality of Madrid). Poisson estimates. The dependent variable is the count of new firms created by industry and municipality.

	I	II	III
Agglomeration mechanisms			
Labor Market Pooling;	0.060^{*}	0.108^{***}	0.096^{***}
$labor_{ic}$	(0.033)	(0.033)	(0.033)
Input Sharing;			
$input_{ic}$	0.238^{***}	0.430^{***}	0.369^{***}
	(0.065)	(0.072)	(0.060)
$output_{ic}$	-0.035	0.015	0.0958^{*}
	(0.059)	(0.061)	(0.056)
Knowledge Spillovers;	0.0153	0.108^{***}	0.115^{***}
$techno_{ic}$	(0.030)	(0.034)	(0.034)
Controls			
Overall municipality employment (excluding that of industry *i*)		-0.395^{***} (0.059)	n.i.
Own industry employment in the municipality	Yes	Yes	Yes
Municipality land area	Yes	Yes	n.i.
Municipality fixed effects	No	No	Yes
City fixed effects	Yes	Yes	n.i.
Industry fixed effects	Yes	Yes	Yes
No. of industries	62	62	62
No. of municipalities	774	774	774
No. of cities	19	19	19
No. of observations	47,988	47,988	47,988

Notes: 1) Robust standard errors (in parentheses) clustered at the two-digit industry and municipality level; 2) ***, ** and * statistically significant at 1, 5 and 10 percent; 3) All the explanatory variables measured in its logarithmic form (those with zero values have been transformed as detailed in the text); 4) $labor_{ic}$, $input_{ic}$, $output_{ic}$ and $techno_{ic}$ are (weighted) sums of the employment in different industries in each municipality. The weights are industry-specific and reflect the intensity of inter-industry relationships. $labor_{ic}$ is the employment in industries that use workers with the same occupations as those employed in industry *i*. $input_{ic}$ is the employment in industry *i*'s input suppliers. $output_{ic}$ is the employment in industry *i*'s customers and $techno_{ic}$ is the employment in industries that use the same new production technologies as those used in industry *i*; 5) n.i. indicates that the variable is not identified because it does not vary across industries in a given municipality.