Agglomeration economies in manufacturing industries: the case of Spain [†]

Olga Alonso-Villar[*]
José-María Chamorro-Rivas
Xulia González-Cerdeira

Universidade de Vigo

October 2001

Abstract:
This paper analyses the extent of geographical concentration of Spanish industry between 1993 and 1999, and study the agglomeration economies that could underlie that concentration. The results confirm that there is major geographic concentration in a number of industries with widely varying characteristics, including high-tech businesses and those linked to the provision of natural resources as well as traditional industries. The analysis of the scope of spillovers behind this agglomeration supports the idea that transportation costs may induce plants in some industries to locate near their customers and suppliers. However, we cannot conclude this is a common fact for all industries. This paper also shows that the higher the technological level of an industry, the higher the agglomeration it experiences. This result implies the importance of the labour market, informational spillovers and producer services location for the agglomeration of these industries.

Keywords: geographical concentration, knowledge spillovers, transport costs, industry

[†] We are grateful for financial aid received from the Spanish Ministry of Science and Technology via DGICYT PB98-0613-C02-01 and SEC1999-1236-C02-01, and from the Regional Government of Galicia via PGIDT00PXI30001PN.
[*] Address for correspondence: Departamento de Economía Aplicada, Universidade de Vigo, Campus Lagoas-Marcosende s/n, 36200 Vigo, Spain. Tel. 34 986 81 25 07; Fax. 34 986 81 24 01; E-mail: ovillar@uvigo.es.

1. Introduction

Concentration of economic activity appears as one of the most significant modern features. In general, companies and individuals are not distributed uniformly in space, but rather in some places agglomerate with higher intensity than in others. In recent years a great number of works, involved in what has been called "the new economic geography", have been focused on analysing these agglomeration processes. See, for example, Glaeser *et al*. (1992), Henderson *et al*. (1995) and Ellison & Glaeser (1997) in the American case; Haaland *et al.* (1999) in the European case; or Maurel and Sédillot (1999) in the French case. This new line of research continues the work of earlier economic geographers,[1] while tackling this question with a more rigorous and formal approach.[2]

The geographic concentration of production may arise from different sources. On one hand, agglomeration allows a labour market pooling for workers with specialised skills. On the other hand, access to a large market allows reductions in transport costs. This means that an upstream industry is attracted to locations where there are many downstream firms and firms in the downstream industry will reduce costs by locating where there are many upstream firms.[3] Also, proximity between producers facilitates rapid diffusion of technology and greater opportunities to exchange information (knowledge spillovers).[4] In fact, face-to-face contacts have been emphasised as an important factor driving concentration of economic activity, both through formal and informal channels (Jacobs, 1969; Saxenian, 1996). These are some of the causes behind industrial location pointed out by the existing literature.

This paper seeks to analyse the extent of geographic concentration of Spanish industries between 1993 and 1999, and study the agglomeration economies, that is, the advantages for firms clustering in the same location, which may be behind that concentration. First of all we look at whether localisation patterns vary widely from one industry to

[1] Marshall (1890), Christaller (1933), Lösch (1940) and Pred (1966), among others.
[2] Important theoretical contributions in the field are Krugman (1991) and Venables (1996), among others. A review of this literature can be seen in Schmutzler (1999). Also, Fujita *et al.* (2000) offer a thorough analysis of the main contributions.
[3] This topic has been analysed in a formal model by Venables (1996).
[4] These three reasons had already been identified by Marshall (1890) at the end of the 19th century.

another.[5] Second, we focus on the industrial scope of spillovers fostering industrial agglomeration.[6] In this vein, we first analyse if spillovers are highly restricted in nature and are found only among firms in the same industry, or whether they also affect firms in related industries.[7] Next, we study whether industries vertically linked tend to choose the same location and also whether industrial agglomeration depends on the technological intensity of the industries involved. On one hand, this will allow us to analyse whether both proximity to demand and supply are important factors for industrial location. On the other hand, high-technology industries need specialised workers and producer services, and depend strongly on exchange of information, as in the famous clusters of Silicon Valley and Boston's Route 128. This topic is also addressed in this paper inasmuch as location patterns can differ among industries depending on their technological intensity.

We use the approach proposed in Maurel & Sédillot (1999), which enables us not only to determine the degree of concentration of each industry but also to analyse the spillovers involved. The paper by Maurel & Sédillot (1999) discusses the similarities between the index proposed by its authors (which is referred to here as M-S) and that put forward earlier in Ellison & Glaeser (1997) (referred to here as E-G). However, it does not analyse the differences between the two indices, so it is hard to perform an empirical analysis to determine why they do not always coincide. We therefore analyse the differences between these indices here and indicate the aspects of concentration on which each places most emphasis. We also compare these indices with the Gini index, which is also widely used in literature.

The paper is structured as follows. Section 2 defines the M-S index, discusses the differences between M-S and E-G, and compares both indices with that of Gini. Section 3 presents the data used and studies the concentration of industry in Spain between 1993 and 1999. The scope of spillovers between plants is shown in Section 4,

[5] In contrast to other works that study the geographic concentration of industry in Spain (Callejón & Costa, 1995, Paluzie *et al*., 2000, and Viladecans, 2000), we not only use the Gini index but also the concentration indices proposed by Ellison & Glaeser (1997) and Maurel & Sédillot (1999), which allows us to control for the size distribution of plants.

[6] When using the term spillovers we actually mean agglomeration economies in production.

[7] A distinction is drawn in the relevant literature between localisation economies (across businesses in the same industry) and urbanisation economies (across businesses in different industries). See Henderson *et al*. (1995) and Glaeser *et al*. (1992) among others.

where a distinction is drawn between industries, sub-industries, upstream-downstream relationships and technological intensity. Section 5 provides the main conclusions.

2. Spatial concentration indices

Maurel & Sédillot approach

This section defines the concentration index, $\hat{\gamma}$, which we use throughout this paper to attempt to determine the degree of spatial concentration of Spanish industries. Concentration is analysed industry by industry, i.e. a concentration index must be defined for each industry considered. In what follows we therefore assume that business units belong to the same industry.

Taking Ellison & Glaeser (1997) as a reference, Maurel & Sédillot (1999) proposes a model of industrial location according to which plants in a particular industry decide to locate in particular geographical regions either because of the natural conditions of those regions or because of spillovers, broadly defined as advantages due to proximity between plants.[8]

An outline of the most significant elements of the probabilistic model proposed in Maurel & Sédillot (1997) are presented below. The random variable U_{ij} is defined, which takes a value of 1 if plant *j* is located at location *i*, and 0 otherwise. It is assumed that all pairs of plants *j* and *k* in the industry have the same joint distribution for their binary responses (U_{ij}, U_{ik}), such that:[9]

$$
\begin{aligned}
&E(U_{ij}) = E(U_{ik}) = x_i, \\
&prob(U_{ij} = U_{ik} = 1) = x_i^2 + x_i(1-x_i)\gamma, \\
&prob(U_{ij} = U_{ik} = 0) = (1-x_i)^2 + x_i(1-x_i)\gamma, \\
&prob(U_{ij} = 1, U_{ik} = 0) = prob(U_{ij} = 0, U_{ik} = 1) = x_i(1-x_i)(1-\gamma).
\end{aligned}
$$

[8] This model does not discriminate between these two possible causes of the location decision.
[9] This 2-dimensional random variable is made up of two non-independent Bernouilli variables.

This means that all the plants in the industry have the same probability, denoted by x_i, of locating at a particular location *i.* [10] Moreover, from the foregoing expressions it can be deduced that $\gamma = corr(U_{ij}, U_{ik})$ for $j \neq k$, i.e. the correlation between the locations of plants *j* and *k* is precisely γ, a parameter which shows both the interdependence of plant location decisions due to their interests in natural advantages and the existence of spillovers between them, $\gamma \in [-1,1]$.

As can be deduced from the above probability distribution, the probability that any two plants in the industry will choose the same location, *p*, [11] can be written as a linear function of the parameter γ so that by proposing an estimator for *p* an estimator can be obtained for γ, which is what ultimately interests us, as will be shown below. The Maurel & Sédillot (1999) paper proposes an estimator for *p* which leads to an estimator for γ as follows:[12]

$$\hat{\gamma} = \frac{\dfrac{\sum_i s_i^2 - \sum_i x_i^2}{1 - \sum_i x_i^2} - H}{1 - H},$$

where *i* denotes the location,[13] s_i is the proportion of employment in the industry accounted for by location *i,* x_i is the proportion of industrial employment at location *i,* and *H* is the Herfindahl index for the industry, which is given by $H = \sum_j z_j^2$, where z_j is the proportion of employment in the industry accounted for by plant *j.* *H* thus shows concentration of output, i.e. whether the industry's output is concentrated in just a few plants. If all the employment in the industry is concentrated in one plant, *H* takes a value of 1, and if there are many plants of similar sizes it is close to 0.

---

[10] This probability depends on the size of the location, measured in terms of aggregate industrial employment there, so that if one location has twice as much employment as another, the probability of a plant in the industry analysed choosing to locate there is twice as high as at the other location. In other words, x_i is the proportion of industrial aggregate employment at *i*.

[11] This probability, *p*, is precisely $\sum_i prob(U_{ij} = 1, U_{ik} = 1) = \sum_i x_i^2 + \gamma\left(1 - \sum_i x_i^2\right)$.

[12] $\hat{p} = \sum_i \frac{\sum_{j,k \in i} z_j z_k}{\sum_{j,k} z_j z_k}$, where $j,k \in i$ denotes the plants in the industry that choose to locate at location *i.*

[13] At empirical level the location may be a natural district, department, province, region, state, etc.

This index is similar to that proposed previously by Ellison & Glaeser (1997), which is expressed as

$$\hat{\gamma}_{EG} = \frac{\dfrac{\sum_i (s_i - x_i)^2}{1 - \sum_i x_i^{\,2}} - H}{1 - H}.$$

Both indices are non-biased estimators of parameter γ, though M-S has the advantage that it comes from a simpler probabilistic location model. On an empirical level there are also differences between the two indices, as they do not necessarily emphasise the same points in assessing concentration. This is discussed later.

Now let us look at why both indicators can be used as concentration indices. Firstly, the first terms of the numerators of both $\hat{\gamma}$ and $\hat{\gamma}_{EG}$ can be interpreted as primary indices (according to their terminology) of geographic concentration, insofar as they measure the differences between spatial distribution in the industry (given by s_i) and the industrial aggregate (given by x_i). As Maurel & Sédillot (1999) show, the expectations of both primary indices can be written as $H + \gamma(1 - H)$.[14] Thus, γ is actually showing the excess of primary concentration, i.e. that part of the geographic concentration which is above the concentration of production (given by *H*). Moreover, using either $\hat{\gamma}$ or $\hat{\gamma}_{EG}$, if a industry is randomly distributed throughout the different geographic units, or if there are no spillovers across the various plants in a industry, these indices average zero, regardless of how concentrated production is in a small number of plants. However this is not true if we directly use the primary spatial concentration index, as deduced from the mean value given above. The fact that indices $\hat{\gamma}$ and $\hat{\gamma}_{EG}$ have this property makes them especially suitable for measuring spatial concentration.

This method can be adapted, as proposed in Maurel & Sédillot (1999), not only to study the concentration of an industry but also to check for spillovers across firms, as will see later.

[14] Note that the fraction of employment in the industry at a location can be written in terms of random variables, and hence the primary indices can also be considered as random variables. $s_i = \sum_j z_j U_{ij}$.

Indices comparison

Although we concentrate in this paper on the use of the M-S index, we compare the results with those obtained with the E-G and Gini indices. This section presents the similarities and differences between these indices. All three aim to measure the geographic concentration of a industry, taking industrial aggregate as their point of reference.

The first two indices differ basically in the way in which their primary indices are obtained. M-S calls for the calculation of the differences between $\sum_i {s_i}^2$ and $\sum_i {x_i}^2$, which are taken as reflecting the divergences between the territorial location of the industry in terms of employment and that of the industrial aggregate. $\sum_i (s_i - x_i)^2$ appears in the calculation of the primary index of E-G, which also takes into account the differences between what happens on the industrial level and in the industrial aggregate, though in this case these differences are calculated location by location. The Gini concentration index also measures the extent to which the spatial distribution of an industry differs from that of the industry as a whole.[15]

For any of these indices the concentration will therefore show the divergences between what happens on an industrial level and on an aggregate level, so that if the geographical distribution of a particular industry coincides with that of the industry as a whole, that industry is said not to be concentrated. As shown before, both the M-S and E-G indices measure geographical concentration beyond the concentration of output in

[15] The Gini index is calculated by ordering the various units of territory in accordance with the Hoover-Balassa index, which measures the ratio s_i / x_i. The x-axis represents the cumulative proportions of industrial employment as a whole, and the y-axis the cumulative proportions for the industry under study. The Gini index measures the quotient between the area within the corresponding Lorenz curve and the 45-degree line and the area below this line. Specifically, the Gini index would take the form $\sum_{i=1}^{n-1}(p_i - q_i) / \sum_{i=1}^{n-1} p_i$, where p_i denotes the cumulative proportion of employment in the industry and q_i the cumulative proportion in industry for the first *i* units of territory in the ranking obtained via the Hoover-Balassa index. A particular industry which is distributed in a way similar to the industry as a whole gives a value for the Gini index of zero. We have also calculated the Gini index, taking population distribution as a reference, and the results are very similar. Correlation between the two indices is high.

just a few plants (measured by the Herfindahl index), which means they have advantages over the Gini index.
It should be noted that the first two indices differ in the degree of importance which they allocate to divergences between the industry analysed and industrial activity as a whole. A location in which the percentage of the industry is greater than that of total industrial activity contributes with a positive factor in the M-S index, while one in which the contrary is true contributes with a negative factor (note that the first term of the numerator of the M-S index can be written as $\sum_i (s_i - x_i)(s_i + x_i)$). Moreover, if the location has a high level of aggregate industrial employment and an even higher level in terms of the industry, its contribution to the index is very great, while if it has little industrial activity, even though the weight of the particular industry in question is greater, its contribution is positive but small (though higher than its contribution to the E-G index, since in the former case it would contribute $(s_i - x_i)(s_i + x_i)$ and in the latter it would contribute $(s_i - x_i)(s_i - x_i)$).

We can therefore conclude that the M-S index takes on high values when the industry is located in the most industrialised areas, as shown in some industries discussed below, while if the industry is situated at locations with little industrial weight the index shows lower concentration. However the E-G index takes into account the divergences between the industrial percentage and the industrial aggregate in each location, regardless of the sign of the difference, and the contribution to the index value is the same in both cases. Moreover, if a location has more employment in one industry than for the whole industrial activity, its contribution to this index is lower than its contribution to M-S.[16] This difference makes the M-S index quite interesting, so long as it is more sensitive to spatial distributions where firms are located in the most industrialised areas, which is actually what these indices try to measure: whether an industry has a higher employment rate in a location than that of the industrial aggregate.

[16] The correlation of the array obtained from the M-S index with that obtained from the E-G index in the Spanish case is around 56%. With the Gini index it is 68%. Using French data, the paper by M-S finds higher correlations on the order of 90% for the first two indices.

3. Concentration of Spanish industry between 1993 and 1999

The Data

The data used in the analysis are taken from the *Encuesta Industrial de Empresas* (EI) provided by the INE (Spanish National Institute of Statistics). The EI is a rich survey that provides information on different firm characteristics.[17] In particular, this survey provides data on employment according to two geographical subdivisions (17 regional autonomous communities, denoted by CCAA, and 50 provinces) and two industry classifications (2- and 3-digit of the CNAE-93 classification). Table 1 shows the industries available. The analysis covers the period 1993-1999. [18] The results are presented in detail only for 1999, since the performance of the industries was observed to be similar throughout the period, as will be shown later.

It should be noted that there is a trade-off between locational fineness and industrial fineness so that a comprehensive breakdown in both industry and territory is not available. For this reason, this analysis on the one hand uses information at regional level with a breakdown to 2- and 3-digit, and on the other hand information at provincial level with a breakdown to only 2-digit industries.[19] The analysis performed using 2-digit classification takes in a total of 30 industries, 5 of which were eliminated on grounds of lack of information in practically all locations with positive values, while the 3-digit classification takes in 118 industries.[20]

In Section 4, we also use the last input-output matrix of the Spanish economy provided by the INE. This matrix includes 34 industries which correspond to some of the 2- and 3-digit industries in the CNAE classification.

---

[17] The EI covers all population for firms over 20 workers, while for smaller firms an estimation based on a representative sample is undertaken.
[18] In 1993 the survey was modified in two important points: the survey unit changed from establishments to firms, and the CNAE-93 industrial classification was adopted. The period analysed begins in 1993 so that homogenous data are available for the full period.
[19] The INE provides no information on industries in a localisation (province or CCAA) when there are less than 4 plants.
[20] Industries 11, 12, 13, 16 and 23 have been eliminated. These industries cover part of the *mining and extraction* industry, *tobacco* and *coke plants/ oil refineries/ nuclear waste treatment*. In industries 11 and 12 the INE provides no information for the industry. In industries 13, 16 and 23 the number of CCAA in which data are not available is 9, 11 and 13 out of 17, respectively.

Concentration in 1999

We now go on to discuss the geographical concentration of 2-digit industries performed at provincial level. In grouping industries according to their degree of concentration we have followed the M-S concentration index, but the results are also compared with those obtained via the E-G and Gini indices. To enable us to compare our concentration figures with those presented in Maurel & Sédillot (1999) for France and those obtained by Ellison & Glaeser (1997) for the USA, we consider their critical values: index values (both M-S and E-G) lower than 0.02 are taken as low concentration, values from 0.02 to 0.05 represent intermediate concentration and values higher than 0.05 are taken as high concentration.

Table 2 shows the M-S, E-G and Gini concentration indices with the corresponding ranking of industries obtained with each of them, plus the Herfindahl index. The most highly concentrated industries according to M-S are the following: *Preparation, tanning & finishing of leather* (19),[21] *Office machinery & computer equipment* (30), *Textiles* (17), *Electronic materials, radio, TV & communications* (32), *Mining & extraction of anthracite, coal, lignite and peat* (10), *Publishing & graphic arts* (22), *Medical, precision and optical instruments & watch-making* (33), and the *Chemical industry* (24). These industries are characterised by the concentration of most of their activity in just a few provinces, generally Barcelona and Madrid.

The results seem fairly robust, in view of the degree to which the three indices used coincide. In fact, the E-G and Gini indices also place these industries among the most highly concentrated, though there are exceptions: industry 24 is considered as having intermediate level concentration under E-G and low under Gini. When the employment level of an industry in an industrialized area is only a bit higher that of the industrial aggregate, the M-S index has a higher value that the E-G index. Employment in industry 24 is concentrated in industrialized provinces such us Barcelona and Madrid, although at a much lower degree than in the above industries, this explains the divergence between the two indices.

[21] In this case the Herfindahl index informs us that employment is distributed across many plants, but in spite of this a high spatial concentration is observed.

The industries which show up as having low concentration under all three indices used are: *Foodstuff & beverage industry* (15), *Manufacture of metal products other than machinery and equipment* (28), *Manufacture of furniture, other manufacturing industries (toys, jewellery, musical instruments & sports articles)* (36) and *Production & distribution of electricity, gas, steam and hot water* (40).[22]

However, there are some industries in which the indices analysed show certain contradictions in classification. For instance *Mining & extraction of non-metallic minerals (stone, sand, minerals for fertiliser and salts)* (14) is the least concentrated industry according to M-S, but has an intermediate concentration under E-G.[23] Similar divergences are found in industries 20, 26, 27 and 35.

A more in-depth look at the causes of these discrepancies shows that they are due to different nuances in the definitions of the two indices, along the lines of those mentioned in the previous section. For example a look at the distribution of employment in industry 14 leads us to deduce that it is not heavily concentrated in the most highly-industrialised provinces, and thus is rated lower under M-S than under E-G. Barcelona and Madrid account for less than 12% of the employment in this particular industry, while they account for 33% of aggregate industry employment. Similar patterns can be found in the other four industries, i.e. the divergences between the M-S and E-G indices for these industries are due to the relatively high number of locations in less industrialised provinces.

Finally we have a group of industries which are classified as being of intermediate concentration under M-S and low under E-G. The industries involved are 18, 21, 25 and 31. A more exhaustive analysis reveals that the concentration of these four industries does not appear to be low, given that a major part of the employment in them is located in provinces with a considerable industrial weight and in provinces known for their specialisation in these industries. This is why M-S yields higher values than E-G.

[22] It should be noted that the Gini index does not classify industry 40 among the lowest. This is probably due to the high degree of concentration of output at a small number of plants, as can be deduced from the Herfindahl index.
[23] The Gini index in this case does not show very high figures.

The analysis has also been performed at regional level. We have observed that index values are slightly higher at CCAA level (see Table 3). [24] The rankings resulting at regional and provincial level do not differ substantially, except in *Manufacturing of motor vehicles, trailers and semi-trailers* (34), which ranks higher at provincial level and has a higher index value. Observation of data enables us to state that employment in this industry is distributed across several CCAA's, but within them is located in only one or two provinces.

Concentration in the period 1993-1999

To analyse the geographic concentration of industry in Spain between 1993 and 1999, we also calculate the M-S index of 2-digit industries at provincial level. We are particularly interested in analysing how industrial agglomeration has evolved throughout the period, and whether there exists a tendency to a greater or lower geographical concentration. As we can see in Table 4 it seems not to be a general tendency to concentration. We observe that both the mean and median values are quite similar throughout the period. The stability in agglomeration levels observed in most Spanish industries is a pattern common among other countries (see for example Dumais *et al*., 1997, for the US and Devereux et al., 1999, for UK).[25] However, data show that some industries have experienced remarkable changes in their levels of concentration (Figure 1).

[insert Figure 1]

Figure 1 shows the evolution of the M-S index between 1993 and 1999 for each manufacturing industry. The x-axis represents the index value in 1993, with respect to the median, and the y-axis the average rate of change. Industries 30 and 19 have strongly increased their concentration during the period. An exhaustive analysis of industry 30 allows us to observe that Madrid has gained employment, whereas Valencia has lost an amount of employment similar to that gained by Madrid. It seems therefore that there has been a relocation of the industry and, since Madrid had a high share of manufacturing in 1999 (11%), it is reasonable to expect this change in the industrial

[24] A thorough analysis on this subject can be found in Alonso Villar *et al* (2001).
[25] Callejón (1997) also shows that the E-G index does not change significantly between 1981 and 1992.

location to be associated with a higher value of the index. With respect to industry 19, we observe that Alicante, which already had a high share of firms in 1993, has experienced a remarkable employment increase.

4. The scope of spillovers

So far the approach used considered that spatial concentration was the result of natural advantages or spillovers that only affected plants in the same industry. However, spillovers may also affect plants belonging to different industries. Consider, then, a group consisting of L different industries, grouped by any criterion one can think of. To analyse the scope of these spillovers we define the following coagglomeration index based on E-G: [26]

$$\hat{\gamma}_o = \frac{\dfrac{\sum_i s_i^2 - \sum_i x_i^2}{1-\sum_i x_i} - H - \sum_{l=1}^{L} \hat{\gamma}_l w_l^2 (1-H_l)}{1-\sum_{l=1}^{L} w_l^2}, \tag{1}$$

with $\hat{\gamma}_l$ being the concentration index for industry *l*, w_l the proportion of employment accounted for by industry *l* within the group and H_l the Herfindahl index for industry *l*.

An estimate of $\hat{\gamma}_0 = 0$ means that there is no more agglomeration of plants in the group than there is in each industry separately. That is, plants in any industry of the group have no particular interest in locating near other plants of another industry in the same group. On the contrary, a high value of $\hat{\gamma}_0$ means that spillovers benefit firms in all industries, so that plants in the group tend to choose the same locations.

On the other hand, it can be shown that

$$\hat{\gamma} = \frac{\overbrace{\sum_l \hat{\gamma}_l w_l^2 (1-H_l)}^{\text{intra}} + \overbrace{\hat{\gamma}_0 \left(1-\sum_l w_l^2\right)}^{\text{inter}}}{1-\sum_l w_l^2 H_l}, \tag{2}$$

[26] See the Appendix.

i.e. concentration due to spillovers between plants in a group, $\hat{\gamma}$, can be written as a weighted mean of the concentration due to spillovers between plants in the same industry, denoted by $\hat{\gamma}_l$ for each industry *l*, and the coagglomeration between plants in different industries, $\hat{\gamma}_0$.

Using this expression, we can calculate what part of the concentration in the group, $\hat{\gamma}$, is due to intra-industry spillovers (within the same industry), and what part is due to inter-industry spillovers (between plants in different industries of the group).

In what follows, we first discuss whether spillovers are specific for any industry or they also affect plants in related industries. To this end, a distinction is drawn within each 2-digit industry between spillovers affecting plants within the same 3-digit industry and those affecting plants between them. Secondly, we study whether geographic proximity is important for industries vertically linked. We use the input-output matrix of the Spanish economy provided by the INE to capture the interdependence of different industries due to supply/demand relations. Thirdly, we focus on whether different technological intensities can lead to different location patterns, so that high-technology industries choose different locations from those of low-technology. We use the classification of the OECD to group industries according to their technology intensity.[27] Given the lack of information at provincial level for some of the 3-digit industries, the analysis is restricted solely to CCAA level.[28]

Industries and sub-industries

We are now interested in knowing whether spillovers only affect plants in a particular industry or whether they also extend to plants in related industries. To this end each group is composed of the 3-digit industries belonging to each 2-digit industry. The results are shown in Table 5. They suggest that in industries such as the *Textiles* (17) and the *Chemical* industry (24), concentration is due more to spillovers across companies belonging to different sub-industries (but all within the same 2-digit industry) than to those within each sub-industry. In industry 17 this may be due to the

[27] This classification excludes industries 10-14 and 40-41.
[28] Both the industrial classification given by the input-output matrix and the OECD include 3-digit industries.

input-output relationship between different sub-industries (*Preparation and spinning of textile fibres*, *Manufacture of fabrics*, *Finishing of textile products*, etc.), while in industry 24 it may be due to the use of skilled labour or research facilities common to various sub-industries (*Basic chemicals, Pesticides, Paint, Pharmaceuticals, Soaps*, etc.).

In other words, the degree of inter-relation between sub-industries in industries 17 and 24 could be greater than that of other industries, thus resulting in spillovers between sub-industries having more weight than those within a sub-industry. Similar results for *Textiles* and part of the *Chemical* industry are also observed in France (Maurel & Sédillot, 1999). However in other industries, such as *Tanning & leather*, *Precision instruments & watch-making*, higher spillovers across companies in the same sub-industry are observed in both countries.

Upstream-Downstream Relations

To analyse vertical linkages between industries we have constructed two classifications. The first one pairs each industry with its main customer (sector-customer classification). This will allow us to analyse the spillovers due to proximity to demand. The second list matches each industry with its main seller (sector-seller classification), so that supply linkages are analysed. These pairs have been defined using the last input-output matrix of the Spanish economy provided by the INE in 1995. This matrix includes 34 industries which correspond with some of the 2- and 3-digit industries in the CNAE. As was mentioned above, 4 industries have been excluded from the analysis for lack of information in the EI.

In the industry-customer classification, 18 out of the 30 industries analysed are paired with themselves, so that the analysis of the coagglomeration is instead that of agglomeration. The results show that these industries are highly concentrated: the average value of the M-S index is 0.07, even though only 8 of them have an index value above 0.05. The coagglomeration of the rest of the industries are shown in Table 6. As we observe, only pairs (17-18) and (21-22) are highly coagglomerated (above 0.05). This means that demand linkages may induce the *Textile* and *Garment-making* industries to locate together. The same applies for *Paper* and *Publishing* industries, since the second industry is the main customer of the former. In both cases, the main customer represents, respectively, 38.8% and 36.9% of their output. However, other pairs, where the main customer represents higher values of output (above 80%), have much lower coagglomeration levels, such as in pairs (265-266 to 268) and (37-27). This may be the result of the high geographical dispersion shown by employment in these industries, which makes the first term in expression (1) to be small, so that this leads to a low coagglomeration index.

With respect to the importance of proximity to suppliers, we find that 15 of the 30 industries analysed are paired with themselves. These industries have an average concentration index of 0.09, which means that they are highly agglomerated, although just 6 out of 15 have a concentration value over 0.05. For the rest of pairs, which are shown in Table 7, we observe that pairs (18-17), (22-21), (25-24) and (33-32) show a high coagglomeration level. Moreover, some pairs in this classification had also been entered in the industry-customer classification, as pairs (18-17) and (22-21). In both cases, their main suppliers represent, respectively, 66.1% and 58.8% of total inputs. [29]

It should be noted that most of the above pairs of industries are located in the most industrialised regions (Catalonia, Madrid and Valencia). So, industries 32 and 33, which are the most coagglomerated industries in the industry-supplier classification, have about two thirds of their employment in Madrid and Barcelona (Catalonia).

Technological intensity

We are now interested in studying industrial concentration taking into account the technology intensity of the industries involved. So, we focus on whether the advantages of geographical proximity are greater for high-technology industries than for the low ones. In order to do this, we have classified industries into four groups, according to their technological intensity. Table 8 shows the industries in each of the groups. For each group we calculate the M-S index and use the breakdown given in expression (2). The results are shown in Table 9. As we can see, only industries in the first group (high technology) have a high concentration level. In this group, spillovers between its

[29] We cannot undertake a deeper analysis on the different sub-industries of *Textiles*, since the input-output matrix does not allow it.

different industries are higher than those within industries (61.4% of concentration is due to inter-industry spillovers, while intra-industry spillovers are just 38.6%). The scope of spillovers for the rest of the groups is less significant since they show low or intermediate concentration levels. However, it should be noted that the second group (intermediate-high technology) has higher concentration than the third (low-intermediate technology) and fourth (low technology) groups. Moreover, inter-industry spillovers represent 86.2% of total concentration for the second group. In other words, the higher the technological intensity of the group, the higher the industrial agglomeration and inter-industry spillovers. [30]

The explanations of this result can be threefold: the market pooling for specialised labour, informational spillovers and producer services location. Firms in industries with rapidly changing production technologies open and close relatively easy. This makes them cluster together to quickly fill their job vacancies. Besides, this means that the search costs for those workers are also lower, so that they tend to choose the same locations. This concentration also facilitates the information flow, both through formal and informal channels (see Saxenian, 1996). Also, it should be noted that services and manufacturing firms are engaged in an input-output structure that makes each sector benefit from proximity to the other. [31] Concentration of services in most industrialised regions may induce manufacturing industries to follow the same pattern so long as transport cost are not negligible (see Alonso-Villar & Chamorro-Rivas, 2001). [32]

5. Conclusions

This paper analyses the extent of geographic concentration of the Spanish industry, using the approach proposed by Maurel and Sédillot (1999). The results confirm the interdependence which exists among firms as regards location decisions in a large

[30] Most high-technology industries are located in Catalonia and Madrid. In fact, *Pharmaceutical goods* (244), which represents 44.8% of total employment in the first group, has 54% of its output in Catalonia and 28% in Madrid.

[31] High technology industries strongly depend on producer services. As Hansen (1994) comments, only 10-15% of the value of an IBM computer comes from the manufacturing process, the rest coming from services such as research, design, engineering, maintenance, or sales.

[32] Coffey and Polèse (1989) support evidence of the centralization in producer services in countries such as Canada, UK, France and USA.

number of industries. This is reflected in a major geographic concentration of the output of those industries.

The industries which show up as being most highly concentrated include especially those for which geographic location is strongly determined by access to raw materials (*mining & extraction*); traditional industries (*textiles* and *leather*), those based on high technology (*IT*, *medical instruments* and *electronics*), for which knowledge spillovers seem to be important, and those which require specialised labour (e.g. the *chemical* industry or *publishing & graphic arts*). The *textiles* and *leather* industries are also highly concentrated in other countries, such as France, UK and the USA, as evidenced by papers such as Maurel & Sédillot (1999), Devereux *et al.* (1999) and Ellison & Glaeser (1997). From this it can be inferred that these industries tend to concentrate to a greater extent than others. A comparison between Spain, France and UK shows similarities also in *mining & extraction*, while both in Spain and France, *electronics* and *publishing* are concentrated.

In Spain, as in France and the USA, the least concentrated industries include *the manufacturing of furniture* and *metal products*. Other industries with low concentration levels in Spain include *foodstuffs & beverages* and *production & distribution of energy*, which are less dispersed in other countries, or for which no information is available in the aforementioned papers.

With respect to the industrial scope of spillovers that could underlie the above concentration patterns, we have found the following results. Input-output relationships seem to be a possible explanation for coagglomeration of some industries. This is the case, for example, of *textiles* and g*arment-making* industry; that of *paper* and *publishing & graphic arts* industries; that of *rubber goods & plastics* and *chemical* industry; and also that of *medical & precision equipment*, and *electronic material*, where coagglomeration is quite high. This means that transportation costs may induce plants in these industries to locate near to their customers and suppliers. However, we cannot conclude that this is a common fact for all industries, as also suggested by Dumais *et al* (1997).

Our results also suggest that in the *textile* and *chemical* industries concentration is due more to spillovers across companies belonging to different sub-industries than to those within each of them. In the *textile* industry this may be the result of the input-output relationships between its different sub-industries (*preparation and spinning of textile fibres*; *manufacture of fabrics*; *finishing of textile products*, etc.), while in the *chemical* industry it may be due more to the use of skilled labour or research facilities common to various sub-industries (*basic chemicals; pesticides; paint; pharmaceuticals; soaps*, etc.). Similar results for *textiles* and part of the *chemical* industry are observed in France.

This paper has also shown that the higher the technological level of an industry, the higher the agglomeration it experiences. This result implies the importance for agglomeration of the labour market, informational spillovers and producer services location. Firstly, high-technology industries require highly skilled labour, so that firms locate near one another to share workers. Secondly, informational spillovers are especially significant for those industries in which technological advances are rapid, both through informal communications and collaborative practices. Finally, it should be noted that these industries strongly depend on producer services, so that they tend to choose locations where such services are already situated, which are precisely the most industrialised regions.

References

Alonso-Villar, O. and Chamorro-Rivas, J.M. (2001): "How do producer services affect the location of manufacturing firms? The role of information accessibility", *Environment and Planning A*, forthcoming.

Alonso Villar, O., Chamorro-Rivas, J.M. and González-Cerdeira, X. (2001): "An analysis of the geographical concentration of industry in Spain", Documento de traballo 0103, Departamento de Economía Aplicada, Universidade de Vigo.

Callejón, M. (1997): "Concentración geográfica de la industria y economías de aglomeración", *Economía Industrial*, 317, 61-68.

Callejón, M. and Costa, M.T. (1995): "Economías externas y localización de las actividades industriales", *Economía Industrial*, 305, 75-86.

Christaller, W (1933): *Diezentralen Orte in Süddeutschland*, Jena: Fisher. *Central Places in Southern Germany* translated by C. W. Baskin, 1966, London: Prentice-Hall.

Coffey, W. and Polèse, M. (1989): "Producer services and regional development: a policy-oriented perspective", *Papers of the Regional Science Association,* 67, 13-27.

Devereux, M., Griffith, R. and Simpson, H. (1999): "The geographic distribution of production activity in the UK", WP 26/99, Institute for Fiscal Studies.

Dumais, G., Ellison, G. and Glaeser, E. (1997): "Geographic concentration as a dynamic process", NBER Working Paper 6270.

Ellison, G. and Glaeser, E. (1997): "Geographic concentration in U.S. manufacturing industries: a dartboard approach", *Journal of Political Economy*, 105, 889-927.

Fujita, M, Krugman, P., and Venables, A. (2000): *The Spatial Economy: Cities, Regions and International Trade*, 2nd printing, Cambridge: MIT Press.

Glaeser, E.; Kallal, H.; Scheinkman, J. and Shleifer, A. (1992): "Growth in cities", *Journal of Political Economy*, 100, 1126-1152.

Haaland, J., Kind, H., Midelfart Knarvik, K., and Torstensson, J. (1999): "What determines the economic geography of Europe", CEPR Discussion Paper 2072.

Hansen, N. (1994); "The strategic role of producer services in regional development", *International Regional Science Review*, 16, 187.195.

Henderson, V., Kunkoro, A. and Turner, M. (1995): "Industrial development in cities", *Journal of Political Economy*, 103, 1067-1090.

Jacobs, J. (1969): *The Economy of Cities*, New York, Vintage Books.

Krugman, P. (1991): “Increasing returns and economic geography”, *Journal of Political Economy*, 99, 483-499.

Lösch, A. (1940): *Die räumliche Ordnung der Wirtschaft*, Jena: Fisher. *The Economics of Location*, translated by W. Waglam, New Haven: Yale University Press.

Marshall, A. (1890): *Principles of Economics*, 8th printing 1986, Basingstoke: Macmillan.

Maurel, F. and Sédillot, B. (1999): “A measure of the geographic concentration in French manufacturing industries”, *Regional Science and Urban Economics*, 29, 575-604.

Paluzie, E., Pons, J. and Tirado, D. (2001): “Regional integration and specialisation patterns in Spain”, *Regional Studies*, 35(4), 285-296.

Pred, A. (1966): *The Spatial Dynamics of US Urban-Industrial Growth*, Cambridge: MIT Press.

Saxenian, A. (1996): *Regional Advantages: Culture and Competition in Silicon Valley and Route 128*, Cambridge: Harvard University Press .

Schmutzler, A. (1999): “The new economic geography”, *Journal of Economic Surveys*, 13, 355-379.

Venables, A. (1996): “Equilibrium locations of vertically linked industries”, *International Economic Review*, 37, 341-359.

Viladecans, E. (2000): “Efectos de las economías externas en la concentración geográfica de la industria: un modelo aplicado a los municipios españoles”, mimeo, Universitat de Barcelona.

Table 1: 2-digit industries and number of 3-digit industries included in each of them

Industry	2-digit industries	Number of 3-digit industries
Mining, extraction & agglomeration of anthracite, coal, lignite and peat	10	3
Extraction of crude oil and natural gas; activities in services related to oil and gas fields other than prospecting.	11	2
Uranium and thorium ore extraction	12	1
Metal mineral ore extraction	13	2
Extraction of non-metallic and energy-destined mineral ores	14	5
Foodstuff products and beverage industry	15	9
Tobacco industry	16	1
Textile industry	17	7
Garment-making and fur industry	18	3
Preparation, tanning & finishing of leather; manufacture of leather and travel goods; accessories and footwear.	19	3
Wood and cork industry other than furniture, basket-making and mat making	20	5
Paper industry	21	2
Publishing, graphic arts & reproduction of recorded media	22	3
Coke, oil-refining and nuclear fuel treatment plants	23	3
Chemical industry	24	7
Manufacture of rubber goods and plastics	25	2
Manufacture of other non-metallic mineral products	26	8
Metallurgy	27	5
Manufacture of metal products other than machinery & equipment	28	7
Machinery and mechanical equipment construction industry	29	7
Manufacture of office machinery and computer equipment	30	1
Manufacture of electrical material and machinery	31	6
Manufacture of electronic material, manufacture of radio, TV and communication equipment and sets	32	3
Manufacture of medical & surgical, precision and optical equipment and instruments and watch-making	33	5
Manufacture of motor vehicles, trailers and semi-trailers	34	3
Manufacture of other transport material	35	5
Manufacture of furniture, other manufacturing industries	36	6
Recycling	37	2
Production & distribution of electricity, gas, steam and hot water, catchment, treatment and distribution of water	40 41	3 1

Table 2: Concentration indices in 1999

Industry	M-S		E-G		Gini		Herfindahl	
14	-0.044	(1)	0.036	(15)	0.378	(10)	0.0022	(12)
40	-0.033	(2)	-0.003	(1)	0.475	(14)	0.0348	(23)
15	-0.032	(3)	0.012	(8)	0.192	(1)	0.0005	(3)
20	-0.028	(4)	0.022	(13)	0.273	(3)	0.0005	(4)
26	-0.024	(5)	0.032	(14)	0.310	(6)	0.0008	(7)
27	-0.007	(6)	0.046	(18)	0.528	(16)	0.0116	(17)
35	-0.004	(7)	0.041	(17)	0.577	(18)	0.0146	(18)
36	0.001	(8)	0.015	(11)	0.316	(7)	0.0004	(2)
28	0.009	(9)	0.004	(4)	0.202	(2)	0.0002	(1)
41	0.010	(10)	0.013	(9)	0.441	(12)	0.0250	(21)
34	0.019	(11)	-0.003	(2)	0.481	(15)	0.0255	(22)
29	0.020	(12)	0.011	(6)	0.309	(5)	0.0010	(10)
18	0.031	(13)	0.014	(10)	0.387	(11)	0.0008	(8)
25	0.032	(14)	0.007	(5)	0.321	(8)	0.0034	(14)
21	0.034	(15)	0.004	(3)	0.280	(4)	0.0029	(13)
31	0.040	(16)	0.011	(7)	0.348	(9)	0.0041	(15)
37	0.058	(17)	0.017	(12)	0.582	(19)	0.0173	(19)
24	0.111	(18)	0.037	(16)	0.451	(13)	0.0018	(11)
33	0.123	(19)	0.062	(20)	0.599	(20)	0.0076	(16)
22	0.128	(20)	0.056	(19)	0.531	(17)	0.0007	(6)
10	0.178	(21)	0.320	(25)	0.944	(25)	0.1714	(25)
32	0.179	(22)	0.086	(22)	0.747	(22)	0.0218	(20)
17	0.182	(23)	0.087	(23)	0.676	(21)	0.0009	(9)
30	0.221	(24)	0.074	(21)	0.906	(24)	0.1342	(24)
19	0.235	(25)	0.292	(24)	0.790	(23)	0.0005	(5)

Table 3: M-S index at regional level

Industry	M-S index
40	-0.067
14	-0.025
15	-0.023
20	-0.020
34	0.002
26	0.004
36	0.007
35	0.007
27	0.013
28	0.013
41	0.014
18	0.030
29	0.032
25	0.041
31	0.041
21	0.053
37	0.070
22	0.100
33	0.103
24	0.141
32	0.152
10	0.176
30	0.182
17	0.262
19	0.262

Table 4: M-S index in the period 1993-1999

Year	Mean	Median	Deviation
1993	0.053	0.026	0.085
1994	0.058	0.028	0.091
1995	0.060	0.027	0.087
1996	0.064	0.029	0.096
1997	0.061	0.027	0.093
1998	0.059	0.024	0.088
1999	0.058	0.031	0.086

Table 5: Intra- and inter-industries spillovers[33]

Industry	γ	Intra-	Inter-
14	-0.025	-0.194	1.194
15	-0.023	0.207	0.793
17	0.262	0.102	0.898
18	0.030	2.916	-1.916
19	0.262	0.938	0.062
20	-0.020	-0.257	1.257
21	0.053	0.625	0.375
22	0.099	1.782	-0.782
24	0.141	0.312	0.688
25	0.041	0.842	0.158
26	0.004	2.198	-1.194
27	0.013	0.848	0.152
28	0.013	1.103	-0.103
29	0.032	0.691	0.309
31	0.041	0.223	0.777
32	0.152	0.907	0.093
33	0.103	1.364	-0.364
34	0.002	2.604	-1.604
35	0.007	11.71	-10.71
36	0.007	18.36	-17.36
37	0.070	0.627	0.373
40	-0.067	-0.501	1.501

[33] These spillovers have been written as a percentage, i.e. the ratio between intra- (or inter-) spillovers and γ.

Table 6: Coagglomeration index between each industry and its main customer

Industry		Main customer		γ_0.	%[34]
14	Extraction of non-metallic and energy-destined mineral ores	266-268	Other non-metallic mineral products	-0.034	46.0
402-403	Production and distribution of electricity, gas and hot water	262-264	Ceramic products	-0.038	15.5
41	Catchment, treatment and distribution of water	24	Chemical industry	0.023	13.1
17	* Textile industry	18	Garment-making and fur industry	0.083	38.8
21	* Paper industry	22	Publishing, graphic arts & reproduction of recorded media	0.053	36.9
25	Manufacture of rubber goods and plastics	34	Motor vehicles, trailers and semi-trailers	0.023	28.9
265	*Processing of cement, lime and plaster	266-268	Other non-metallic mineral products	-0.020	89.4
261	Processing of glass and glass products	159	Beverage industry	0.041	32.6
262-264	Ceramic products	27	Metallurgy	-0.064	60.6
27	* Metallurgy	28	Manufacture of metal products other than machinery & equipment	-0.016	27.7
28	Manufacture of metal products other than machinery & equipment	27	Metallurgy	-0.016	19.0
29	Machinery and mechanical equipment construction industry	28	Manufacture of metal products other than machinery & equipment	0.021	14.6
37	Recycling	27	Metallurgy	0.003	82.2

Note: Industries marked with * are those common for both tables 6 and 7.

[34] Percentage that the main customer represents.

Table 7: Coagglomeration index between each industry and its main seller

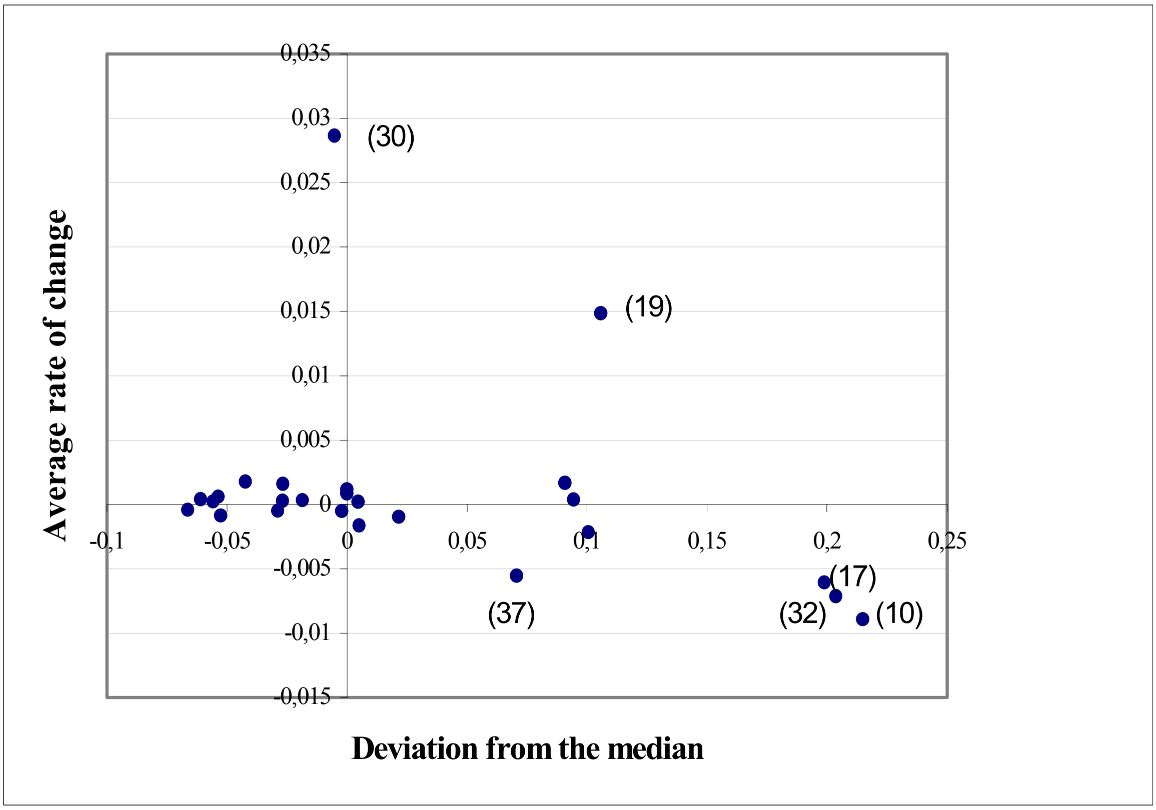
Industry		Main seller		γ_{0}.	%
14	Extraction of non-metallic and energy-destined mineral ores	401	Production and distribution of electricity	-0.041	27.7
41	Catchment, treatment and distribution of water	401	Production and distribution of electricity	-0.020	43.8
18	* Garment-making and fur industry	17	Textile industry	0.083	66.1
22	* Publishing, graphic arts & reproduction of recorded media	21	Paper industry	0.053	58.8
25	Manufacture of rubber goods and plastics	24	Chemical industry	0.072	44.9
265	Processing of cement, lime and plaster	401	Production and distribution of electricity	-0.038	27.5
261	Processing of glass and glass products	24	Chemical industry	0.041	21.7
262-264	Ceramic products	14	Extraction of non-metallic and energy-destined mineral ores	-0.022	18.7
266-268	* Other non-metallic mineral products	265	Processing of cement, lime and plaster	-0.020	34.0
28	* Manufacture of metal products other than machinery & equipment	27	Metallurgy	-0.016	50.1
29	Machinery and mechanical equipment construction industry	27	Metallurgy	-0.009	32.2
31	Manufacture of electrical material and machinery	27	Metallurgy	-0.035	37.2
33	Manufacture of medical & surgical, precision and optical equipment and instruments and watch-making	32	Manufacture of electronic material, radio, TV and communication equipment and sets	0.133	33.3
35	Manufacture of other transport material	27	Metallurgy	-0.032	21.7
36	Manufacture of furniture, other manufacturing industries	20	Wood and cork industry other than furniture, basket-making and mat making	-0.014	32.4
37	Recycling	28	Manufacture of metal products other than machinery & equipment	0.041	56.7

Table 8: Industries included in each technological group

		3-digit CNAE classification
	Pharmaceutical goods	244
High technology group	Office machinery and computer equipment	300
	Electronic components and apparatus	223, 333, 321, 322, 323
	Aircraft and space equipment	353
	Man-made and synthetic fibres	247
	Other chemical industries	241, 242, 243, 245, 246
Intermediate-high technology group	Machinery and mechanical equipment	292 to 296
	Machinery and electrical equipment	297, 311 to 316
	Motor vehicles	341, 342, 343
	Railway equipment	352
	Other transport equipment	354, 355
	Precision instruments	331 to 335
	Metallurgy	271 to 274
Low-intermediate technology group	Mineral products other than metal	261 to 268
	Metal products	275, 281 to 287
	Shipbuilding	351
	Rubber and plastics	251, 252, 372
	Other manufacturing industries	362 to 366
	Foodstuff, beverages and tobacco	160, 151 to 159
Low technology industries group	Textiles	171, 172, 173, 175, 176, 177
	Leather	191, 192
	Footwear and garment-making	174, 193, 181 to 183
	Wood, furniture and cork	361, 201 to 205
	Paper, graphic arts and publishing	372, 251, 252

Table 9: Spillovers intra- and inter- industries

		Intra-	Inter-
High technology industries	0.172	0.386	0.614
Intermediate-high technology industries	0.035	0.138	0.862
Low-intermediate technology industries	-0.008	-0.529	1.529
Low technology industries	0.005	1.594	-0.594



Average rate of change
Deviation from the median
0,035
0,03
0,025
0,02
0,015
0,01
0,005
0
-0,005
-0,01
-0,015
-0,1
-0,05
0
0,05
0,1
0,15
0,2
0,25
(30)
(19)
(37)
(17)
(32)
(10)

Figure 1

Appendix

Lemma 1. Let p be the probability that two plants in an industry locate in the same area, and $\gamma = corr\left[U_{ij}, U_{ik}\right]$, where j and k represent two plants of sector r, $j \neq k$. It can be shown that

$$p = \gamma\left(1 - \sum_i x_i^{\,2}\right) + \sum_i x_i^{\,2}.$$

Besides, $\hat{p} = \sum_i \frac{\sum\limits_{\substack{j,k \in i \\ j,k \in r}} z_j z_k}{\sum\limits_{j,k \in r} z_j z_k}$ is an estimator of p.[35] From which it follows that

$$\hat{\gamma} = \frac{\hat{p} - \sum_i x_i^2}{1 - \sum_i x_i^2}$$

is an estimator of γ.[36]

Proof. See Maurel & Sédillot (1999).

Lemma 2. Let us assume that sector r has two subsectors, l and l'. Then,

$$\sum_{\substack{j \in l \\ k \in l'}} z_j z_k = 1 - \left(\sum_{j \in l} z_j\right)^2 - \left(\sum_{k \in l'} z_k\right)^2,$$

where z_j is the employment share of plant j in sector r.

Proof. Taking into account that $\sum\limits_{j \in r} z_j = 1$, we can write that

$$1 = \left(\sum_{j \in r} z_j\right)^2 = \sum_{j \in l} z_j^{\,2} + \sum_{j \in l'} z_j^{\,2} + \sum_{\substack{j \in l \\ k \in l'}} z_j z_k + \sum_{j,k \in l} z_j z_k + \sum_{j,k \in l'} z_j z_k. \qquad \text{(A1)}$$

A straightforward calculation shows that

$$\sum_{\substack{j \in l \\ k \in l'}} z_j z_k = 1 - \left[\sum_{j \in l} z_j^{\,2} + \sum_{j,k \in l} z_j z_k\right] - \left[\sum_{j \in l'} z_j^{\,2} + \sum_{j,k \in l'} z_j z_k\right],$$

---

[35] By $j,k \in i$ we mean that j and k locate in the same geographic area i.

[36] Analogous expressions can be found when $j \in l, k \in l'$, l and l' being two subsectors in sector r, $l \neq l'$. In this case, we denote by $\gamma_0 = corr\left(U_{ij,} U_{ik}\right)$. Othewise, that is, if $j,k \in l$, we denote by $\gamma_l = corr\left(U_{ij}, U_{ik}\right)$.

which leads to the expression we wanted to obtain.

Proposition 1. We propose an estimator of the probability of two plants in industry r choosing the same location as given by the following expression

$$\hat{p}_0 = \frac{\sum_i s_i^2 - H - \sum_{l\in r} w_l^2 (1-H_l)\hat{\gamma}_l + \sum_i x_i^2 \sum_{l\in r} w_l^2 (1-H_l)(\hat{\gamma}_l - 1)}{1-\sum_{l\in r} w_l^2},$$

where s_i is the employment share of sector r in location i, H is the Herfindahl index in sector r, w_l is the share of subsector l in sector r employment, H_l is the Herfindahl index in subsector l, $\hat{\gamma}_l$ is the geographic concentration index in sector l, and x_i is the proportion of the whole manufacturing employment in location i.

Proof. The estimator of p_0 we use is analogous to the one proposed by Maurel & Sédillot (1999) for the case in which plants belong to the same sector[37]

$$\hat{p}_0 = \sum_i \frac{\sum_{\substack{j,k\in i \\ j\in l, k\in l'}} z_j z_k}{\sum_{j\in l, k\in l'} z_j z_k},$$

where z_j denotes the share of plant j in employment sector r.

Step 1. We first prove that $\sum_{j\in l, k\in l'} z_j z_k = 1 - \sum_{l\in r} w_l^2$.

Using Lemma 2 when more than two subsectors exist, we can write

$$\sum_{j\in l, k\in l'} z_j z_k = 1 - \sum_{l\in r}\left(\sum_{j\in l} z_j\right)^2. \qquad \text{(A2)}$$

Denoting by z_{jl} the proportion of plant j's employment with respect to subsector l, it follows that $z_j = w_l z_{jl}$. Introducing this expression in equation (A2) we obtain that

$$\sum_{j\in l, k\in l'} z_j z_k = 1 - \sum_{l\in r}\left(\sum_{j\in l} w_l z_{jl}\right)^2 = 1 - \sum_{l\in r}\left(w_l \sum_{j\in l} z_{jl}\right)^2 = 1 - \sum_{l\in r} w_l^2 \left(\sum_{j\in l} z_{jl}\right)^2 = 1 - \sum_{l\in r} w_l^2.$$

---

[37]In order to simplify notation, by j and k we mean two plants belonging to the same sector r without making it explicit in the equation. By $j\in l, k\in l'$ we mean that j belongs to subsector l, while k does to subsector l', l and l' being two subsectors of sector r.

Step 2. Now we are going to prove that the numerator in $\hat{p}_0$ can be written as

$$\sum_i \sum_{\substack{j,k\in i \\ j\in l,k\in l'}} z_j z_k = \sum_i s_i^2 - H - \sum_{l\in r} w_l^2 \left(1-H_l\right)\left[\hat{\gamma}_l\left(1-\sum_i x_i^2\right)+\sum_i x_i^2\right].$$

Analogously to the steps followed to obtain (A1) we have that

$$\sum_{\substack{j,k\in i \\ j\in l,k\in l'}} z_j z_k = \left(\sum_{j\in i} z_j\right)^2 - \sum_{j\in i} z_j^2 - \sum_l \sum_{\substack{j,k\in i \\ j,k\in l}} z_j z_k \ .$$

Using the above expression and taking into account that $\sum_{j\in i} z_j = s_i$, and $\sum_i \sum_{j\in i} z_j^2 = H$, we have that

$$\sum_i \sum_{\substack{j,k\in i \\ j\in l,k\in l'}} z_j z_k = \sum_i s_i^2 - H - \sum_i \sum_{l\in r} \sum_{\substack{j,k\in i \\ j,k\in l}} z_j z_k \ .$$

Since $z_j = w_l z_{jl}$, we can write

$$\begin{aligned}\sum_i \sum_{\substack{j,k\in i \\ j\in l,k\in l'}} z_j z_k &= \sum_i s_i^2 - H - \sum_i \sum_{l\in r} \sum_{\substack{j,k\in i \\ j,k\in l}} w_l^2 z_{jl} z_{kl} \\ &= \sum_i s_i^2 - H - \sum_i \sum_{l\in r} w_l^2 \sum_{\substack{j,k\in i \\ j,k\in l}} z_{jl} z_{kl} \\ &= \sum_i s_i^2 - H - \sum_{l\in r} w_l^2 \sum_i \sum_{\substack{j,k\in i \\ j,k\in l}} z_{jl} z_{kl}.\end{aligned}$$

The estimator of the probability, p_l, of two plants in subsector l choosing the same location is

$$\hat{p}_l = \frac{\sum_i \sum_{\substack{j,k\in i \\ j,k\in l}} z_{jl} z_{kl}}{\sum_{j,k\in l} z_{jl} z_{kl}} \ .$$

Using this estimator the above expression can be written as

$$\sum_i \sum_{\substack{j,k\in i \\ j\in l,k\in l'}} z_j z_k = \sum_i s_i^2 - H - \sum_{l\in r} w_l^2 \left(\hat{p}_l \sum_{j,k\in l} z_{jl} z_{kl}\right).$$

Besides,

$$1 = \left(\sum_{j\in l} z_{jl}\right)^2 = \sum_{j\in l} z_{jl}^2 + \sum_{j,k\in l} z_{jl} z_{kl} \ .$$

Hence

$$\sum_i \sum_{\substack{j,k\in i \\ j\in l, k\in l'}} z_j z_k = \sum_i s_i^2 - H - \sum_{l\in r} w_l^2 \hat{p}_l \left(1 - \sum_{j\in l} z_{jl}^2\right).$$

Since $H_l = \sum_j z_{jl}^2$, it follows that

$$\sum_i \sum_{\substack{j,k\in i \\ j\in l, k\in l}} z_j z_k = \sum_i s_i^2 - H - \sum_{l\in r} w_l^2 \hat{p}_l \left(1 - H_l\right).$$

Using Lemma 1 in subsector l we have that

$$\hat{p}_l = \hat{\gamma}_l \left(1 - \sum_i x_i^2\right) + \sum_i x_i^2,$$

from which we get to Step 2

$$\sum_i \sum_{\substack{j,k\in i \\ j\in l, k\in l}} z_j z_k = \sum_i s_i^2 - H - \sum_{l\in r} w_l^2 \left(1 - H_l\right)\left[\hat{\gamma}_l \left(1 - \sum_i x_i^2\right) + \sum_i x_i^2\right].$$

Step 3. Finally, we use Steps 1 and 2 in $\hat{p}_0$ and after

$$\hat{p}_0 = \frac{\sum_i s_i^2 - H - \sum_{l\in r} w_l^2 \left(1 - H_l\right)\left[\hat{\gamma}_l \left(1 - \sum_i x_i^2\right) + \sum_i x_i^2\right]}{1 - \sum_{l\in r} w_l^2}$$

$$= \frac{\sum_i s_i^2 - H - \sum_{l\in r} w_l^2 \left(1 - H_l\right)\hat{\gamma}_l + \sum_{l\in r} w_l^2 \left(1 - H_l\right)\hat{\gamma}_l \sum_i x_i^2 - \sum_{l\in r} w_l^2 \left(1 - H_l\right) \sum_i x_i^2}{1 - \sum_{l\in r} w_l^2}$$

$$= \frac{\sum_i s_i^2 - H - \sum_{l\in r} w_l^2 \left(1 - H_l\right)\hat{\gamma}_l + \sum_i x_i^2 \sum_{l\in r} w_l^2 \left(1 - H_l\right)\left(\hat{\gamma}_l - 1\right)}{1 - \sum_{l\in r} w_l^2}.$$

Theorem 1. The estimator of the correlation between the location decision of two plants belonging to different subsectors of the same sector, $\gamma_0 = corr\left(U_{ij}, U_{ik}\right)$, can be written as

$$\hat{\gamma}_0 = \frac{G - H - \sum_{l\in r} w_l^2 \left(1 - H_l\right)\hat{\gamma}_l}{1 - \sum_{l\in r} w_l^2},$$

where $G = \dfrac{\sum_i s_i^2 - \sum_i x_i^2}{1 - \sum_i x_i^2}$.

Proof. Using Lemma 1 when plants belong to different subsectors, we have that $\hat{\gamma}_0 = \dfrac{\hat{p}_0 - \sum_i x_i^2}{1-\sum_i x_i^2}$, where $\hat{p}_0$ is the estimator of the probability of two plants in different subsectors choosing the same location. Using expression $\hat{p}_0$ in Proposition 1, we can rewrite $\hat{\gamma}_0$ as

$$\hat{\gamma}_0 = \frac{1}{1-\sum_i x_i^2}\left[\frac{\sum_i s_i^2 - H - \sum_l w_l^2(1-H_l)\hat{\gamma}_l + \sum_i x_i^2 \sum_{l\in r} w_l^2(1-H_l)(\hat{\gamma}_l - 1)}{1-\sum_{l\in r} w_l^2} - \sum_i x_i^2\right]$$

$$= \frac{1}{1-\sum_i x_i^2}\left[\frac{\sum_i s_i^2 - H - \sum_{l\in r} w_l^2(1-H_l)\hat{\gamma}_l + \sum_i x_i^2 \sum_{l\in r} w_l^2(1-H_l)(\hat{\gamma}_l - 1) - \sum_i x_i^2\left(1-\sum_{l\in r} w_l^2\right)}{1-\sum_{l\in r} w_l^2}\right]$$

$$= \frac{1}{1-\sum_i x_i^2}\left[\frac{\sum_i s_i^2 - \sum_i x_i^2 - H - \sum_{l\in r} w_l^2(1-H_l)\hat{\gamma}_l + \sum_i x_i^2 \sum_l w_l^2(1-H_l)(\hat{\gamma}_l - 1) + \sum_i x_i^2 \sum_l w_l^2}{1-\sum_l w_l^2}\right]$$

$$= \frac{\dfrac{\sum_i s_i^2 - \sum_i x_i^2}{1-\sum_i x_i^2} + \dfrac{1}{1-\sum_i x_i^2}\left[-H - \sum_{l\in r} w_l^2(1-H_l)\hat{\gamma}_l + \sum_i x_i^2 \sum_{l\in r} w_l^2(1-H_l)(\hat{\gamma}_l - 1) + \sum_i x_i^2 \sum_{l\in r} w_l^2\right]}{1-\sum_{l\in r} w_l^2}$$

$$= \frac{1}{1-\sum_{l\in r} w_l^2}\left\{G + \frac{1}{1-\sum_i x_i^2}\left[\left(-1+\sum_i x_i^2\right)\sum_{l\in r} w_l^2(1-H_l)\hat{\gamma}_l\right]\right.$$

$$\left. + \frac{1}{1-\sum_i x_i^2}\left[-H - \sum_i x_i^2 \sum_{l\in r} w_l^2(1-H_l) + \sum_i x_i^2 \sum_{l\in r} w_l^2\right]\right\}$$

$$= \frac{1}{1-\sum_{l\in r} w_l^2}\left\{\left[G - \sum_{l\in r} w_l^2(1-H_l)\hat{\gamma}_l\right] + \frac{1}{1-\sum_i x_i^2}\left[-H + \sum_i x_i^2 \sum_{l\in r} w_l^2 H_l\right]\right\}$$

$$= \frac{G - H - \sum_{l\in r} w_l^2(1-H_l)\hat{\gamma}_l}{1-\sum_{l\in r} w_l^2}.$$