



Título artículo / Títol article: Combined effect of human capital, temporary employment and organizational size on firm Performance

Autores / Autors Vicente Roca-Puig, Inmaculada Beltrán-Martín, Mercedes Segarra Cipres

Revista: Personnel Review, Vol. 41 No. 1, 2012, pp. 4-22.

Versión / Versió: Pre-print

Cita bibliográfica / Cita bibliogràfica (ISO 690): ROCA PUIG, Vicente, BELTRÁN MARTÍN, Inmaculada, SEGARRA CIPRES, Mercedes. Combined effect of human capital, temporary employment and organizational size on firm Performance. *Personnel Review*. 2012, Vol. 41, No 1, p. 4-22.

url Repositori UJI: <http://hdl.handle.net/10234/67725>

Spatial location patterns of Spanish manufacturing firms^{*}

José M. Albert[†], Marta R. Casanova[†] and Vicente Orts[‡]

Abstract

In this paper, we evaluate the spatial location patterns of Spanish manufacturing firms in each industry. We use a distance-based method, Ripley's K function, which allows us to treat space as continuous. With 'complete spatial randomness' as benchmark, every sector presents significant concentration whatever the distance considered. If we use the locations of all manufacturing firms as benchmark, we find dispersion in some sectors and concentration in others, finding also differences in the spatial scale at which clustering occurs. Finally, the use of cities as benchmark reveals that not only is 'first nature' among the location determinants of manufacturing firms but they also include the self-reinforcing advantages of 'second nature'.

Keywords: distance-based method, Ripley's K function, Spanish manufacturing firms, spatial location patterns.

JEL classification: C15, C40, C60, R12.

^{*} We wish to thank G. Duranton, J. Mateu, J.C. Pernías, D. Tirado, the participants of the 49th Congress of the European Regional Science Association and two anonymous referees for their helpful comments and suggestions. Furthermore, we gratefully acknowledge financial support from the Spanish Ministry of Science and Innovation (project ECO2008-06057/ECON), Generalitat Valenciana (BFPI/2007/204 and PROMETEO/2009/068) and Fundació Caixa Castelló-Bancaixa (P1-1B2010-17).

[†] Department of Economics, Universitat Jaume I, Av. Sos Baynat s/n, 12071 Castellón, Spain.

[‡] Department of Economics and Institute of International Economics, Universitat Jaume I.

Corresponding author: Marta R. Casanova (mroig@eco.uji.es).

1. Introduction

Economic activity is distributed across space heterogeneously and the tendency for different industries to cluster in certain areas has been well known for a long time. Some traditional and well-known examples of this geographic concentration are high-tech firms in Silicon Valley, the auto industry in Detroit or the carpet industry in Dalton (USA) and, going back further in time, the textile industry in Lancashire (UK). In Spain, the tile industry in Castellón or the leather and footwear industry in Elche are also good examples of this tendency towards the geographic concentration of economic activity.

Heterogeneity of the spatial distribution of activity can be caused by multiple and very different factors. The pioneering work by Marshall (1890) put special emphasis on the role of specialised local markets, positive externalities and linkages, as sources of agglomeration. A century later, the ‘new economic geography’, initiated by Krugman (1991a, b), highlighted the role of economies of scale and transport costs as the main agglomeration forces, which interact with the dispersion forces, immobile factors and product market competition to determine the location of economic activity. This strand of the literature has made significant advances in the identification of the forces and factors that determine the choice of location by firms and the resulting patterns of agglomeration of economic activity.¹

However, economists are not only concerned about the determinants of geographic concentration, but also how to measure and characterise the patterns of geographic concentration of firms and industries in space; this paper deals with this second aspect. In fact, its main objective will be to measure and characterise the nature and the physical scale of location patterns of Spanish manufacturing

¹ For further details see Ottaviano and Puga (1998), Fujita et al. (1999), Puga (1999, 2002), Overman et al. (2003), Venables (1995, 2006), Ottaviano and Thisse (2004) Fujita and Thisse (2009) or Redding (2009).

sectors by improving the application of some recent methods developed to assess the geographic distribution of economic activity.

Specifically, we are going to use Ripley's K function², which is a spatial distance-based statistical method for measuring the spatial distribution of any set of points, in our case, the geographic location of firms. Consequently, the *method* employed will *treat space as continuous* in order to obtain a proper analysis of spatial location patterns, instead of sticking to administrative-scale data. Thus, by avoiding the inconvenience of geographic scale, our paper follows the *third generation* of measures to test localisation, according to the classification developed by Duranton and Overman (2005). In addition, our approach satisfies the five essential requirements that these authors stressed that any test for measuring concentration should fulfil, namely: (1) be comparable across industries, (2) control for the overall agglomeration of manufacturing, (3) control for industrial concentration, (4) be unbiased with respect to scale and aggregation, and (5) give an indication of the significance of the results.

In Spain, many authors are also interested in this topic, the most notable being Callejón (1997), Viladecans (2001), Alonso-Villar et al. (2003, 2004) or Paluzie et al. (2004). From a different perspective, other papers have also used this kind of indices in the analysis of the determinants of industrial localisation in Spain, as is the case of Paluzie et al. (2001) or Tirado et al. (2002). It should be borne in mind that all of these Spanish studies have a common drawback: they *treat space as being discrete*. They restrict the spatial distribution to just *one scale* and analyse the distribution of activity over discrete geographic units, which do not necessarily coincide with the relevant scale from the economic point of view. In this way, the spatial scale chosen is a key decision that may alter the results and conclusions reached.

² For further details, see Ripley (1976, 1977, 1979).

In addition, we must emphasise another aspect that differentiates our empirical analysis from others: our location of firms is very accurate, because we know the geographic coordinates (longitude and latitude) of every establishment. Thus, we employ these coordinates to situate each firm as a dot on the map, without taking administrative borders into account.³

The remainder of the paper is organised as follows. In Section 2.1, we present an overview of the literature on the different methods used in the empirical analysis of the spatial location of economic activity, in order to put our approach into perspective. In Section 2.2, we outline the methodology employed and the key improvements incorporated into our application. In Section 3, we describe our data set. In Section 4, we present and discuss the main results achieved, comparing them with those from other authors. Finally, Section 5 contains the most important conclusions reached and an outline of future research.

2. Methodology

2.1. An overview of the literature

As a starting point from which to understand our methodological approach and put it into perspective, in the following we present a brief summary of the literature on the empirical measurement of economic agglomeration.

First, it is important to note that this literature has been influenced by two very different traditions, economic geography and spatial statistics, and therefore has followed two different paths. Over time, these paths have converged and the positions of the two approaches have gradually got closer to each other. However, there remain some significant differences. So, our paper attempts to help unify the two approaches by providing an empirical method that highlights the interactions between them.

³ Until now most studies have used the postcodes of the manufacturing firms and the geographic coordinates associated with each postcode to locate them in space. See, for example, Marcon and Puech (2003) and Duranton and Overman (2005).

From the perspective of economic geography, the ‘first generation’ of measures of localisation is based on indicators such as the Herfindahl or Gini index. This type of measures did not take space into consideration and could not distinguish whether concentration of activity is due to the presence of a few huge firms in a specific area or many small firms in the same specific area; in other words, they did not control for the level of industry concentration. Krugman (1991a), Brülhart (2001), Amiti (1997) or Paluzie et al. (2001), among others, used the Gini index to measure the concentration of economic activity.

The ‘second generation’ of measures, as it was called by Duranton and Overman (2005), was initiated by Ellison and Glaeser (1997) and tried to solve some of the drawbacks of the previous generation. It began to take space into account and to control for the overall agglomeration of manufacturing and for industrial concentration. In addition, the Ellison and Glaeser index provided a null hypothesis which could be used to test for the presence of concentration. Thus, this type of indices allowed us to compare concentration between industries properly. However, this generation of measures had a disadvantage: they *treated space as being discrete*. According to these indices, an industry was said to be concentrated if a large part of the production or the workers were concentrated in a small number of regions, regardless of whether those regions were neighbours or not. Consequently, although the spatial dimension started to gain importance in this approach, there remained a downward bias when dealing with industries that cross the boundaries of the administrative units considered. Furthermore, this bias increases as the size of these administrative units is reduced. Therefore, by treating space as discrete, they restrict the spatial distribution to just *one administrative scale*. In the words of Duranton and Overman (2005), ‘they transform points on a map into units in boxes’. Thus, as already mentioned, the spatial scale that is chosen is a critical decision that will have important effects on the final results of the analysis. For instance, Viladecans (2001) and Alonso-Villar et al. (2003) used more than one geographic level in their analysis to determine

the most suitable administrative unit in each case⁴. Many other authors, such as Devereux et al. (2004), Rosenthal and Strange (2001) or Maurel and Sédillot (1999), have developed alternative indices, with similar characteristics, and used them to measure the geographic concentration of activity in their respective countries, i.e. UK, USA and France. Other applications of these indices to the Spanish economy can be found in Callejón (1997), for Spanish provinces, Alonso-Villar et al. (2003, 2004) or Paluzie et al. (2004).

The ‘third generation’, put forward by Duranton and Overman (2005), introduces the treatment of space as something that is continuous. Results can then be compared across different scales, thereby enabling their measurement to be unbiased with respect to arbitrary changes in the spatial units. Moreover, their localisation test provides an indication of the statistical significance of the results. Duranton and Overman (2005) popularised ‘*K*-density’, which was a density function of bilateral distances between pairs of firms. Additionally, they also developed a new way to test for localisation and answered questions about issues such as the spatial scale at which localisation occurs and its sectoral scope. Obviously, their approach meets the five aforementioned requirements that any test that measures concentration must fulfil.

The second path, with its origins in the tradition of spatial statistics, was not at first of an economic nature. That is one of the reasons why we decided to treat the evolution of the research as two different paths, in addition to the differences in the methodologies used.

A tool was needed to analyse spatial point process data and, at the very beginning, in the analysis of a point pattern only the nearest neighbour was taken into account. Ripley (1976) introduced the theory of the second-order property of point processes, which considered all the neighbours rather than just the nearest. This methodology was named ‘*K* function’ in Ripley (1977) and, from then on, it

⁴ Viladecans (2001) used the municipal and provincial level, corresponding to NUTS 5 and 3 and Alonso-Villar et al. (2003) used the provincial and regional level, corresponding to NUTS 3 and 2.

has been a very common measure in spatial statistics and very widely used to analyse the spatial structures of point patterns. Ripley's K function has been followed, modified and improved over time by many authors, such as Diggle (1983), Cuzick and Edwards (1990), Diggle and Chetwynd (1991) (who tried to deal with inhomogeneity), Cressie (1993) or Marcon and Puech (2003b), who tried to introduce the size of firms.

In terms of its application, this function was first introduced into fields such as forestry (Duncan, 1993; Goreaud and Pélissier, 1999), regional science (Sweeney and Feser, 1998), or disease cases (Diggle and Chetwynd, 1991). Only more recently was it incorporated into the analysis of the geographic concentration of economic activity. In fact it was first introduced into economics by Arbia and Espa (1996) and later made more popular by Marcon and Puech (2003a), followed by Quah and Simpson (2003) or by Arbia et al. (2008). Since then, the use of this measure in economic fields has not stopped growing. Finally, we should highlight the fact that Ripley's K function is a cumulative measure, instead of being a density function of bilateral distances, as is the case of the K -density used by Duranton and Overman (2005).

The aim of our paper, beyond the specific application to Spanish manufacturing sectors, is also to help unify both perspectives, by the direct use of a statistical function and trying to make our approach meet the five requirements of Duranton and Overman (2005).

Thus, like Marcon and Puech (2003a) and other previously mentioned authors, the distance-based method we are going to use to measure the spatial distribution of activity in Spain is Ripley's K function, which offers important advantages over traditional concentration indices. Indeed, by means of this method, we can know whether concentration exists, what its intensity is and at what distance, or spatial scale, its highest level is obtained.⁵ Furthermore, in contrast to Marcon and

⁵ Marcon and Puech (2003a) have previously used this concentration index, in France, to measure the spatial distribution of economic activity.

Puech (2003a), and like Duranton and Overman (2005), we controlled for the general tendency of the manufacturing industry to agglomerate and for the industrial concentration of each sector.

2.2. A framework for measuring economic agglomeration: Ripley's K function

Ripley's K function, $K(r)$, is a distance-based method that measures concentration by counting the average number of neighbours each firm has within a circle of a given radius, 'neighbours' being understood to mean all firms situated at a distance equal to or lower than the radius (r). From here on, firms will be treated as points.

The $K(r)$ function describes characteristics of the point patterns at many and different scales simultaneously, depending on the value of ' r ' we take into account, that is,

$$K(r) = \frac{1}{\lambda N} \sum_{i=1}^N \sum_{j=1, i \neq j}^N w_{ij} I(d_{ij})$$

$$I(d_{ij}) = \begin{cases} 1, & d_{ij} \leq r \\ 0, & d_{ij} > r \end{cases}$$

where d_{ij} is the distance between the i^{th} and j^{th} firms; $I(x)$ is the indicator function; N is the total number of points observed in the area of the study region; $\lambda = N/A$ represents its density, A being a rectangular or alternatively a polygonal area covering the study region⁶ and w_{ij} is the weighting factor to correct for border effects.⁷ The indicator function, $I(d_{ij})$, takes a value of 1 if the distance between the i^{th} and j^{th} firms is lower than r , or 0 otherwise, and w_{ij} will be equal to the area

⁶ The general tendency of the spatial location patterns of each sector does not vary qualitatively when we introduce the 'polygonal boundary', and only the K value is reduced somewhat. Nevertheless, when some dissimilarity appears due to a different area being employed, it will be remarked.

⁷ These border-effect corrections should be incorporated to avoid artificial decreases in $K(r)$ when r increases, because the increase in the area of the circle under consideration is not followed by the increase of firms (outside the study area there are no firms).

of the circle divided by the intersection between the area of the circle and the area of study. Regarding the area of study, an initial analysis is performed using a rectangular bounding box in order to continue on from the literature by Marcon and Puech (2003). We then implement a later analysis by using a polygonal bounding box, which allows a better fit to the land area of the country.

Finally, using the definition of λ , the $K(r)$ function can be rewritten as:

$$K(r) = \frac{A}{N^2} \sum_{i=1}^N \sum_{j=1, i \neq j}^N w_{ij} I(d_{ij})$$

Therefore, the $K(r)$ function shows the average number of neighbours in an area of radius (r), divided by the density of the whole study region (λ).

The next step in the evaluation of the location patterns of economic activity is to determine the null hypothesis and compare it with our results. The null hypothesis is usually a kind of randomly distributed set of locations in the area of study. Thus, if firms were located in the study area random and independently from each other, we would have a location pattern known as *Complete Spatial Randomness (CSR)*. This is our first benchmark and, as long as we assume *CSR*, the K function will be equal to πr^2 .

Consequently, we define $M_{CSR}(r)$ as the value that quantifies the difference between the empirical K value of the real point pattern of each sector and the theoretical K value,⁸ that is:

$$M_{CSR}(r) = K(r) - \pi r^2$$

If the empirical K value, $K(r)$, is higher than the theoretical K value, πr^2 , this indicates concentration of our point pattern distribution, since the real density is greater than that of the benchmark. Lower values indicate dispersion and if $K(r)$ is equal to πr^2 , then this means that our points are independently distributed.

⁸ Marcon and Puech (2003) used the normalised function, $L(r)$, because they considered that ‘a practical limitation of Ripley’s K function is the need to compare any value to πr^2 ’. However, we do this directly thanks to the statistical software employed.

Figures 1 and 2 show two spatial distributions of points, both with the same number of points (100) inside the same area. However, the points in Figure 1 are distributed at random and independently from each other, while the points in Figure 2 show some tendency to cluster.

INSERT FIGURE 1 AND 2 ABOUT HERE

Figures 3 and 4 show the empirical and theoretical K functions and the corresponding M_{CSR} functions of the point patterns that appeared in Figures 1 and 2.

INSERT FIGURE 3 ABOUT HERE

As we can see, on the left-hand side of Figure 3 there are two lines. The dashed line represents the empirical K value, that is to say, the K value of the observed point pattern in Figure 1, and the continuous line takes a value of πr^2 , since this represents the CSR benchmark. We can observe that the values of these two lines are almost the same, no matter what radius we take into account ($K(r) \approx \pi r^2$). The right-hand graph in Figure 3 shows the value of the M_{CSR} function. Consequently, it can be seen that an independent and random distribution of points produces an almost flat M curve, with a value of around zero.

The graphs in Figure 4 give us the same information as the previous ones, but refer to the spatial distribution of the points in Figure 2. Here, in the left-hand graph we can observe that $K(r) > \pi r^2$ at all distances of ‘ r ’ considered, which means that the point pattern in Figure 2 presents concentration at all distances. Moreover, from the M_{CSR} curve we can determine the distance at which the highest level of concentration is reached. This curve also indicates whether the point pattern observed is concentrated or dispersed relative to the benchmark under consideration, depending on its positive or negative value. Therefore, the M_{CSR} curve provides concise information with which to analyse different location patterns and we will use this function throughout the paper to analyse the spatial location patterns of Spanish manufacturing firms.

INSERT FIGURE 4 ABOUT HERE

In actual fact, considering firms to be randomly and independently distributed from each other within a particular area is not completely correct, because economic activity cannot be located in a random and independent way and the results are sensitive to the specific area considered. Economic activities are spatially concentrated for other reasons, very different to economic factors, for example because of dissimilarities in such natural features as mountains, rivers or harbours, that is, ‘first nature’. Additionally, with CSR as our benchmark we cannot isolate the idiosyncratic tendency of each sector to locate itself in accordance with the general tendency of manufacturing firms to agglomerate.

Consequently, as an alternative scenario, we use the ‘whole of manufacturing’⁹ as a benchmark, thus minimising the aforementioned drawbacks. Indeed, we can compare the spatial distribution of each sector with the overall tendency of manufacturing industry to agglomerate, that is:

$$M_{TM}(r) = K(r) - K_{TM}(r)$$

Here, $M_{TM}(r)$ is the difference between the K -value of each sector under consideration and the K -value of the total manufacturing at radius r . In this case, the null hypothesis will be that the spatial distribution of firms in a given sector is the same as the one we would get by choosing a random sample, of the same size, from the set of all manufacturing firms, that is to say, ‘spatial randomness conditional on both industrial concentration and the overall agglomeration of manufacturing’¹⁰. Relative localisation or dispersion will appear within a particular sector depending on whether its K -value is higher or lower than K -value of the total manufacturing. In such a case, our claim is that this sector is concentrated or dispersed relative to the whole of the manufacturing industry.

⁹ Henceforth TM .

¹⁰ Duranton and Overman (2005), p. 1078.

Now, to evaluate the statistical significance of departures from the null hypothesis of each scenario in a robust way, we should construct a confidence interval for M_{CSR} and M_{TM} . The traditional technique used to construct this confidence interval is the Monte Carlo method, which involves generating a large number of independent random simulations. It should be noted that the construction of the confidence interval will be different in the two scenarios, M_{CSR} and M_{TM} . In the first scenario, we simulate Poisson patterns with the same number of points as in the real distribution of each sector and then locate them randomly within the area of study. In the second scenario, we also simulate random distributions with the same number of firms as in each of the sectors under consideration, but this time the location of these hypothetical firms is restricted to the sites where we can currently find firms from the whole manufacturing sector.¹¹ Both scenarios are generated by running 100 simulations and both allow us to reject the non-significant values. A confidence interval of 95% was utilised. In this way, the construction of the confidence interval allows us to control for industrial concentration, as well as to assess the significance of departures from randomness, in the first scenario, and from the overall pattern of location of the manufacturing industry, in the second scenario.

The use of TM as a benchmark and the way of testing for localisation, of course, could give rise to some controversy for two distinct reasons. On the one hand, if all points are used as the ‘alternative’, the in-sector points are repeated. This is an obvious fact, but our very reason for using ‘the whole manufacturing’ as a benchmark is to be able to compare across industries – one of the five aforementioned requirements that we want our approach to meet. In fact, the use of ‘ $TM-i$ ’¹² as a benchmark does not allow us to make a comparison across industries because this benchmark is different in each case. Even so, we carried

¹¹ This procedure is the same as the one used by Duranton and Overman (2005) in the construction of their counterfactuals and again minimises the shortcomings associated to the specificity of the considered area.

¹² The whole of manufacturing minus the sector itself.

out a parallel analysis using $TM-i$ as our benchmark, with negligible differences in the results.¹³ Moreover, we should emphasise that by using the whole of the manufacturing industries as our benchmark, we are trying to bear in mind the path followed in the economics field and compare the spatial location patterns of firms in an industry with that of the whole manufacturing industry.

On the other hand, using the TM as a benchmark and comparing the deviations from TM of each sector ' i ' with those of hypothetical industries with the same number of firms randomly distributed across all manufacturing sites (making Monte Carlo simulations) prevents us from using the class of spatial clustering methods proposed by authors like Cuzick and Edwards (1990) or Diggle and Chetwynd (1991). The reason for this is that, as Cuzick and Edwards (1990) pointed out, these methods are appropriate if the population density is unknown in advance. In other words, we do not need to use a previously selected group of controls, or representative sample of the entire population, because the entire population is known to us.

Finally, it must be emphasised that the most popular papers that have used distance-based methods to assess the geographical concentration of activity, probably Marcon and Puech (2003a) and Duranton and Overman (2005), followed similar, but not identical, approaches. In fact, several differences can be found if we analyse them in depth.

On the one hand, Duranton and Overman (2005) emphasised that their test fulfils all the five requirements that any measure of concentration should satisfy. However, Duranton and Overman's method has a disadvantage with regard to Ripley's K function, that is to say, it is not possible to quantify the concentration or dispersion but only detect the proportion of sectors that are concentrated.

On the other hand, if we revise Marcon and Puech (2003a), we can see that their application of the K function, like our M_{CSR} function, is very sensitive to the

¹³ Results available upon request.

study area considered¹⁴ and does not satisfy two of the five above-mentioned requirements, since it does not control for the overall tendency of manufacturing firms to agglomerate or for industrial concentration.

Now, on examining the properties of the measure of concentration used after improving the benchmark employed and making more robust the construction of the confidence interval, that is, the M_{TM} function, it can be observed that fulfils all the five requirements established by Duranton and Overman and allows us to obtain a measure that quantifies the concentration or dispersion of all our point patterns, thus minimising the drawbacks associated to the use of rectangular areas in the analysis.

Obviously, our concentration measure based on Ripley's K function is comparable across sectors. Additionally, we control for the overall agglomeration of manufacturing by considering the whole set of manufacturing firms as our benchmark and, when constructing the confidence interval, by randomly locating the firms across the set of locations of all the manufacturing firms. The third requirement, control for industrial concentration, is satisfied by considering hypothetical sectors with the same number of firms as in each existing sector. Fourth, as we use a continuous distance method to measure spatial concentration and not an administrative scale, information about characteristic features of the patterns of localisation at different scales is known and our test is unbiased with respect to scale and aggregation. Lastly, statistical significance is also satisfied, since the confidence interval allows us to know whether the observed distribution is significantly different from conditional randomness. In this case, randomness is conditioned to the industrial concentration of each sector and to the location of the overall manufacturing.

¹⁴ Marcon and Puech (2003) restricted their area of study and did not analyse the whole of France, but instead an industrial area of 40 x 40 km around Paris and a larger rectangular area of France measuring 550 x 630 km.

Finally, we would like to stress that by using alternative benchmarks we could redefine the null hypothesis and see to what extent the spatial distribution of each sector diverges from the spatial distribution of the proposed benchmark, as well as to what spatial scale and with what intensity this occurs. We could then attempt to explain the possible nature behind the different localisation behaviours. An interesting possibility is to use the spatial distribution of towns and cities in the same area as a benchmark.¹⁵ This benchmark will be denominated ‘municipalities’ (M) and will allow us to discriminate between ‘first nature’ and ‘second nature’, since the settlement of towns in space has usually been conditioned by some distinctive natural advantage (proximity to a river or the sea, availability of natural resources or productive soil, etc.).¹⁶ Additionally, we will use it as a test of the robustness of the tendency of industries to cluster for reasons other than the map of cities in the Spanish territory. In this case, we can compare the spatial distribution of each sector with the spatial distribution of Spanish ‘cities and towns’, that is:

$$M_M(r) = K(r) - K_M(r)$$

$M_M(r)$ being the difference between the K -value of each sector and the K -value of the distribution of ‘cities and towns’ at radius r . If the location patterns do not coincide, that is to say, if the K -values are not the same at each distance, we will be able to determine not only the fact that the spatial distribution of firms is conditioned by ‘first nature’, but also the self-reinforcing advantages of ‘second nature’.¹⁷

The different benchmarks used provide us with distinct counterfactuals with which to compare the K of each sector. With CSR we do not control for anything, with M we control for first nature and with TM we control for first nature and also

¹⁵ We would like to thank one of the reviewers for this suggestion.

¹⁶ Throughout the paper we use the terms ‘municipalities’ and ‘cities and towns’ as synonyms and when we say ‘municipality’ we are referring to a single settlement (a city, town or village) and not to an administrative area that may contain multiple settlements.

¹⁷ See Krugman (1993).

for the general tendency of manufacturing firms to agglomerate.¹⁸ In this way, the information that each one provides us with is more detailed and, thus, the comparison between the different values of M gives us clues about the different causes or factors that lead industries to cluster.

3. Data

Our empirical analysis uses current establishment level data, for the year 2007, from the Analysis System of Iberian Balances database,¹⁹ which contains detailed information about Spanish and Portuguese companies. We restrict our database to Spanish manufacturing establishments, using the National Classification of Economic Activities²⁰ and analysing sectors at the two-digit level. Furthermore, we add another two requirements to our database. First, we ensure that our database contains only Spanish manufacturing firms on the peninsula, without including firms from the Canary and Balearic Islands, Ceuta and Melilla. Second, we restrict our analysis just to firms employing at least ten workers. Finally, once these requirements have been applied, our database contains exactly 43,087 firms.

At this point the comparison between our restriction, related to the number of employees, and Marcon and Puech (2003)'s restriction must be considered. In fact, they used French manufacturing firms employing at least twenty workers. This difference, concerning the number of employees, is due to the fact that SMEs (small and medium-sized enterprises) are predominant in Spain. Consequently, too many firms would be left out if we only considered those with twenty or more than twenty workers, as Marcon and Puech did.

Spanish manufacturing activities are classified into 23 sectors according to 'NACE 93 - Rev. 1' and these are as follows: (15) Food products and beverages, (16) Tobacco products, (17) Textiles, (18) Wearing apparel and dressing, (19) Tanning and dressing of leather, (20) Wood and products of wood, (21) Pulp,

¹⁸ Note that $M_{TM}(r) = [K(r) - K_M(r)] - [K_{TM}(r) - K_M(r)]$

¹⁹ SABI

²⁰ NACE 93 - Rev. 1

paper and paper products, (22) Publishing, printing and recorded media, (23) Coke, refined petroleum products, (24) Chemical and chemical products, (25) Rubber and plastic products, (26) Other non-metallic mineral products, (27) Basic metals, (28) Fabricated metal products, (29) Other machinery and equipment, (30) Office machinery and computers, (31) Electrical machinery, (32) Radio, televisions and other appliances, (33) Instruments, (34) Motor vehicles and trailers, (35) Other transport equipment, (36) Furniture and other products, (37) Recycling.

Table A1, situated in *Appendix 1*, shows us a brief descriptive analysis of the above-mentioned sectors including additional information, such as the number of firms, the number of employees, or the technological intensity of each one. Thus, as we can see in this table, there are great differences in the number of firms, depending on the sector we are referring to. Hence, three of the twenty-three sectors considered (16, 23 and 30) will not be analysed because they are too small as far as their number of establishments is concerned.

INSERT FIGURE 5 ABOUT HERE

Figure 5 shows the spatial distribution of four Spanish manufacturing sectors (15, 17, 20 and 35). Here, each dot corresponds to an establishment and we can see that there are great differences in the spatial distributions of these sectors.²¹

We know the *precise location* of every establishment through its geographic coordinates (longitude and latitude), which enables us to minimise the margin of error in each case.²² These geographic coordinates are transformed into UTM²³ coordinates, or flat coordinates, as they are also called. This transformation was carried out by means of the method proposed by Morton (2003). This procedure converts latitude and longitude coordinates into Easting and Northing coordinates,

²¹ The area of study is normalised, (0, 1), but the real measures would be 1075×882 km.

²² Marcon and Puech (2003) and Duranton and Overman (2005) only know the postcode of each manufacturing firm and thus obtain a location error of between 100 m and about 2 km. However, if we had used the Spanish postcodes to locate the firms, our margin of error would have been higher, since our postcodes cover larger areas than those from the United Kingdom or France.

²³ Universal Transverse Mercator.

on a Transverse Mercator projection, the UTM coordinates being expressed in metres. We should highlight the fact that the construction of this system allowed us to move away from the Equator with hardly any distortions, because any point is a long way away from the central meridian of its zone.

Finally, in order to introduce the ‘municipalities’ as a benchmark we obtained the location data on 7938 Spanish ‘towns and cities’ from MUNI97.ARJ. *Centro Nacional de Información Geográfica* (Spain). This database contains information about the geographic coordinates of Spanish cities and towns, which we transformed into UTM so as to be able to carry out our analysis, in the same way as we did with the geographic coordinates of the firms.

4. Empirical Results and Discussion

Following the order in which we have developed our discussion of the methodology, first we will discuss the results coming from the ‘CSR benchmark’, after that, those coming from the ‘TM benchmark’, and finally the results considering the set of cities and towns, the ‘M benchmark’.²⁴

Figure 6 shows the M_{CSR} curves of the four sectors presented earlier (15, 17, 20 and 35) and their corresponding confidence intervals.²⁵

INSERT FIGURE 6 ABOUT HERE

If we examine the confidence intervals, we can observe that they are very narrow. This is due to the way they are constructed, since the difference between the Poisson simulations and the theoretical K -value is almost zero for every r considered. Consequently, although there are differences between them, all the

²⁴ We used the free statistical software ‘R’ to conduct our research. This software is downloadable from the following website: <http://www.r-project.org/>. Specifically, we used the package ‘Spatstat’ (kest) in our analysis. Other alternatives, as the package Splanacs (khat), should be used to analyse the spatial clustering methods proposed by authors like Cuzick and Edwards (1990) or Diggle and Chetwynd (1991).

²⁵ The results obtained from computing the M_{CSR} function for each sector are available from the authors upon request.

sectors present concentration relative to complete spatial randomness, whatever the distance we consider. But a different result was not really expected. As mentioned in Section 2.2, complete spatial randomness ignores the heterogeneity of the space in which firms are located. However, in actual fact, the spatial location of firms is also influenced by ‘first nature’, that is, by dissimilarities in such natural features as mountains, rivers or harbours, and even by the general tendency of manufactures to agglomerate.

In order to correct these shortcomings, our second benchmark, as explained in the methodology, is constructed taking into account the location of overall manufacturing. Figure 7 shows the M_{TM} curves of the four sectors presented earlier (15, 17, 20 and 35), together with their associated confidence intervals.²⁶ At first glance, we observe that the location patterns of these sectors differ considerably from the first analysis. In fact, as is well known, this type of analysis is sensitive to the benchmark being considered. Two aspects related to the confidence intervals also stand out. First, they are not as narrow as in the first case and, second, the smaller the sector is, the wider the confidence interval will be.

We are going to use sector 35, *Other transport equipment*, to exemplify the different location patterns coming from the two benchmarks. In Figure 5d, we can see the spatial distribution of firms from sector 35, in the Spanish territory. It can be seen that the establishments in this sector are distributed in small clusters around Madrid, Barcelona, Vigo and the Basque Country, as well as along coastal areas in general.

If we compare the outcomes obtained using both benchmarks (Figures 6d and 7d), it becomes clear that the resulting location patterns are very different. In fact, the location pattern shown in Figure 7d has characteristic features that are more consistent with the real distribution of firms (shown in Figure 5d) than Figure 6d. The apparent concentration of sector 35 that appears in Figure 6d is partly due to

²⁶ It should be noted that the graphs of the M_{TM} curves of all the sectors analysed appear in *Appendix 2*.

an inappropriate benchmark. Indeed, when the M_{CSR} function is used, sector 35 presents significant concentration up to a radius length of 186 km. However, when we control for the ‘first nature’ factors and for the general tendency of economic activity to agglomerate, using the M_{TM} function, we achieve results that are more realistic. In this case significant concentration is observed up to a radius of 25 km and the maximum significant peak is reached at a distance of 9 km. Furthermore, significant dispersion is found from kilometre 57 onwards. Thus, we note that the results obtained from the M_{TM} function describe the real distribution of establishments in space in a more accurate way.

INSERT FIGURE 7 ABOUT HERE

The rest of the sectors analysed using the second benchmark also show different location patterns, and this result is independent of the bounding box used (rectangular or polygonal). Table 1 summarises the results obtained from computing the M_{TM} function for each sector by using a rectangular boundary, while Table 2 summarises the same results using a polygonal boundary.

INSERT TABLE 1 ABOUT HERE

INSERT TABLE 2 ABOUT HERE

If we focus on the second and third columns in Table 1 and 2, we can see that not every sector presents concentration compared to manufacturing firms as a whole, whatever the length of the radius considered, as happens with the CSR benchmark. Nevertheless, in terms of the *intensity* reached when we use the rectangular boundary and TM as a benchmark, the sectors with a higher concentration remain the same. In this way, if we pay attention to the fourth column in Table 1, we can see that the most highly concentrated sectors, i.e. those that reach the highest M value, are the same as the ones in the first analysis and in the same order: (19) Tanning and dressing of leather, (17) Textiles, (32) Radio, televisions & other appliances, (22) Publishing, printing and recorded media, (33) Instruments, (24) Chemical and chemical products, and (31) Electrical machinery.

Obviously, the M_{TM} value of these sectors is not as high as the M_{CSR} value, because in this case we compare each sector with the location of overall manufacturing and not with complete spatial randomness. When we apply the polygonal boundary, these sectors remain the same but in a different order.

It may be interesting to compare the above-mentioned results with those obtained by Duranton and Overman (2005) in the UK. Surprisingly, they are very similar. On the one hand, they find that the localised sectors in the UK are 17, 18, 19, 22, 30, 31 32 and 33, which almost coincide with the most concentrated sectors in Spain. On the other hand, the least concentrated sectors in Spain match those non-localised sectors in the UK perfectly, i.e. 15, 20 and 26. Therefore, it can be observed that manufacturing sectors tend to follow similar patterns of location between countries; at least, this seems to be the case between Spain and the UK. Additionally, the most concentrated sectors in France and the USA are Textile (17) and Leather products (19),²⁷ these results also coinciding with the most concentrated sectors in Spain and the UK.

Alonso-Villar et al. (2004) also studied the geographical concentration of Spanish industry, between 1993 and 1999,²⁸ and concluded that the most highly concentrated industries, according to the Maurel and Sédillot index, are 19, 17, 32, 22, 33 and 24. As we can see, these results are the same as ours. Hence, thanks to the results obtained by these authors, we can also deduce that, broadly speaking, the spatial location patterns of Spanish manufacturing sectors have not varied significantly in recent years, since the highest concentrated sectors are still the same as in 1999, thus showing some temporal persistence.

In spite of this regularity, every Spanish manufacturing sector possesses its own singularities and its patterns of location differ meaningfully from one to another. In fact, the *persistence* of the agglomeration in space, i.e. the spatial scale

²⁷ See Maurel and Sédillot (1999) and Ellison and Glaeser (1997).

²⁸ Nevertheless, they only presented results for 1999, since not many differences were observed throughout the whole period.

dimensions of the cluster, varies depending on the sector considered. Hence, we classify all of the sectors analysed into five groups. First, type 0 sectors are those formed by industries that do not present statistically significant concentration or dispersion in comparison to the total manufacturing industry. Second, type 1 sectors include industries that show a greater tendency to concentrate than the manufacturing as a whole at all distances under consideration. Type 2 sectors consist of those industries that are systematically less concentrated than the overall manufacturing industry, that is, they show just relative dispersion patterns. Type 3 industries are relatively concentrated at low distances and dispersed at large distances. Finally, type 4 sectors include those industries that are relatively dispersed at low distances, while at long distances they are more concentrated than manufacturing as a whole.

The classification of the different sectors appears in the last column of Table 1, 'Type of cluster'. On the one hand, it can be seen that 12 sectors (17, 19, 21, 22, 24, 25, 27, 29, 31, 32, 33 and 34), more than half of the total number that were considered, present a stronger tendency to cluster than manufacturing firms as a whole at all distances (Type 1), whereas just one sector presents only dispersion in relation to the manufacturing industry as a whole at all distances (Type 2). Finally, it should be noted that sectors 18, 28 and 37 do not present significant divergence in their tendency to cluster with regard to manufacturing sectors as a whole (Type 0).

On the other hand, we can see that the rest of the sectors present different location patterns depending on the spatial scale chosen, that is to say, depending on the length of the radius being considered. Hence, only sector 35 show concentration patterns, as compared to manufacturing as a whole, at low distances and relative dispersion at large distances, that is to say, concentration takes place on a relatively small scale (Type 3). Consequently, we may deduce that the establishments in this sector are distributed in small clusters, presenting dispersion when the distances become longer. Finally, sectors 20, 26 and 36 are relatively

dispersed at low distances and more concentrated than manufacturing as a whole at long distances (Type 4). In the case of sectors 20 and 26, however, this tendency to cluster in relation to overall manufacturing at long distances is quite weak and, in the three cases, the distance at which this tendency appears is very long. As we can see in Table 1, we only find relatively significant concentration patterns in these sectors from 137, 145 and 96 km onwards, respectively, and so these three sectors can be considered to be relatively dispersed.

When we introduce the ‘polygonal boundary’ (Table 2), the concentration intensity of the sectors is reduced. Thus, sectors such as 17 and 19 stop showing a greater tendency to concentrate than the manufacturing as a whole at all distances considered (Type 1) and become sectors which are concentrated at low distances and dispersed at large distances, as compared to manufacturing as a whole (Type 3). Moreover, the number of sectors that do not present a statistically significant concentration or dispersion in comparison to the total manufacturing industry is also increased, since there are three sectors when we use the rectangular area and five sectors by using the polygonal area.

Taking into account the M_{TM} value from the fourth column in Table 1 and 2, we can deduce that dispersion happens on different scales and the most highly dispersed sectors in Spain are (15) Food products and beverages, (20) Wood and products of wood, and (26) Other non-metallic mineral products. These dispersed sectors, which have an elevated dependence on natural resources, are related to food or to the primary sector. Moreover, these sectors are also likely to be made up of specialised manufacturers that disperse their establishments to supply the different markets in the best possible way.

Alternatively, if we pay attention to the sectors that have been classified as the most highly concentrated, it can be seen that the agglomeration forces that explain this concentration do not exhibit particular characteristics. On the one hand, the two sectors that show a higher level of concentration in Spain, 19 and 17, are clearly low-tech. We presume that the geographical concentration presented by

these sectors is wholly due to *historical trends*.²⁹ On the other hand, *technological spillovers* seem to be the main reason for location in sectors 31, 32 and 33. Finally, sectors 22 and 24 are the ones for which the search for *skilled labour* appears to play a decisive role in their decision to concentrate. This indicates that ‘knowledge spillovers’ are not the only factors determining the concentration of activity. Indeed, there are also many other factors such as local labour pooling, natural advantages, tradition, transport costs or upward and forward linkages³⁰ that can be factors determining this concentration and which play an important role in the patterns of localisation of each sector.

Hence, although according to the literature high-tech sectors may be the most highly concentrated, this is not so in our analysis. In fact, it does not occur in the UK, France or in the USA either. Thus, Devereux et al. (2004) stressed that ‘*the most geographically concentrated industries appear to be relatively low-tech*’ and Maurel and Sédillot (1999) found that the most concentrated sectors in France are textile and leather products, two of the most traditional and low-tech sectors. Similar results can also be found for the United States. Thus, in Chapter 2 of Krugman (1991a) it is said that the sectors with a higher degree of concentration are not high-tech, but rather they are sectors related to the textile industry. Nevertheless, we are not trying to say that the high-tech sectors are not concentrated; we just want to draw attention to the fact that low-tech sectors can also be concentrated. As a result, the diversity of forces that, according to the theory, can cause agglomeration seems to be present with different degrees of intensity in each sector.

It can be observed that quite a lot of sectors show concentration patterns at shorter distances, i.e. less than 90 km (13 out of 20). The percentages of manufacturing firms and workers employed in these concentrated sectors are 65% and 67% respectively, whereas 35% of the manufacturing firms and 33% of

²⁹ They have probably settled and clustered in the same area since the Industrial Revolution.

³⁰ Emphasised in this way by Krugman (1991a).

manufacturing workers belong to dispersed sectors. Therefore, we can state that a larger proportion of manufacturing employees work in sectors with concentration patterns and that these sectors have a larger proportion of firms.

Finally, the agglomerative strengths belonging to each sector, which pull economic activities together, may weaken at long distances and determine the differences in the size of the clusters and in their spatial sequence. The specific location of firms is the result of the trade-off between centripetal (external economies, scale economies, technological spillovers, specialised factor markets, and so on) and centrifugal forces (diseconomies of agglomeration, immobile factors, land rents, etc.) and this is what will determine the differences in the size of each cluster. At first, the increase in firms located in an area creates a self-reinforcing process of agglomeration, which leads to a progressive increase in the centripetal forces associated with this location. Given the intensity of centripetal forces, an increase in the radius (r) and the corresponding increase in distance may reduce the incentive to locate in a particular cluster. Thus, this trade-off between the centripetal and the centrifugal forces does not necessarily increase or decrease monotonically with distance. For this reason, it is interesting not only to analyse the average values of the M function at each radius, but also its variations when we change the radius, that is, $\Delta M/\Delta r$ (the marginal M_{TM} value at each distance). This marginal M_{TM} value informs us about the increase in the number of neighbours in each sector when r becomes higher as compared to the increase in neighbours of the overall manufacturing industry.

In Appendix 2, we can see this information for every sector that was considered; nevertheless, we are going to give a detailed description of the location patterns of four sectors, which present a higher level of concentration. Thus, in Figure 8, we can see the M_{TM} value (continuous line) and the marginal M_{TM} value (dashed line) for sectors 17 (textiles), 19 (tanning and dressing of leather), 24 (chemical and chemical products) and 32 (radio, televisions and other appliances). On observing these graphics, some questions come to mind: *What is*

the distance at which the greatest increases in the relative density of neighbours are produced? Do these increases present any kind of regularity?

The distances at which we can find the maximum increase in neighbours (maximum $\Delta M/\Delta r$) vary considerably for each sector, depending on the agglomerative strengths that are specific to each sector. These distances will be called ‘*distance of maximum increase*’ and are summarised in Table 3. In this table, we can compare the distances at which an absolute maximum in M_{TM} is found (absolute M_{TM}) with the distances at which maximum increases in M_{TM} take place (marginal M_{TM}). It can also be observed how the two values complement each other in order to give us more detailed information about the location patterns of each sector.

INSERT TABLE 3 ABOUT HERE

INSERT FIGURE 8 ABOUT HERE

In sector 17, we can see that the marginal M_{TM} value presents four noticeable peaks at the distances of 2, 8, 17 and 25 km. These increases in the number of neighbours at short distances can be interpreted as the existence of nearby clusters (or micro-clusters) in the textile industry. Moreover, we find new peaks when we increase the distance and the radius measures 62, 69 and 73 km, respectively. This can be due to the existence of new clusters, which are separated from the previous ones, in other regions or cities. Sector 19 also shows multiple nearby clusters up to a similar distance (27 km). In fact, we find significant increases in the number of neighbours when r measures 2, 9, 24, 26 and 27 km. Thus, sectors 17 and 19 allow us to observe similar location patterns at short distances, and the distances at which the clusters appear almost coincide with one another. However, as you can see in *Appendix 2*, the increase in neighbours in the textile sector is more regular than in sector 19, since two great increases in the number of neighbours appear in the latter at distances of 2 and 27 km.

The centripetal forces of the previously analysed sectors, 17 and 19, may be different from those of sectors 24 and 32, since the initial spatial scale at which multiple nearby clusters appear in these sectors is smaller (16 and 18 km, respectively, instead of 25 and 27 km). Indeed, sectors 17 and 19 are traditional, low-tech sectors whereas 24 and 32 are high-tech sectors. Therefore, we can assume that the technological spillovers may reduce the distance at which the clusters are produced. Thus, the nature of the centripetal and centrifugal forces of each sector will determine the differences in the size of the clusters and the sequence in which clusters appear within the territory.

Finally, we will discuss the results and highlight some of the most important conclusions of considering the set of ‘municipalities’ as a benchmark.

INSERT TABLE 4 ABOUT HERE

Although the location of the ‘cities and towns’ is a clear constraint on the location of manufacturing, and therefore the existence of agglomerations of people will be an obvious reason for finding a concentration of manufacturing firms in space (as shown in Table 4), the different industrial sectors in Spain are concentrated for reasons other than just the spatial distribution of Spanish ‘cities and towns’. In general terms, the relative concentration intensity of each Spanish manufacturing sector is higher if we compare it with the distribution of the Spanish ‘municipalities’ instead of comparing it with the distribution of the whole of manufacturing. Obviously, this means that the ‘Spanish municipalities’ present a more dispersed spatial pattern than the ‘whole of manufacturing’. Thus, we will be able to determine that the spatial distribution of firms is conditioned not only by ‘first nature’, but also by the self-reinforcing advantages of ‘second nature’.

However, although all sectors have a higher relative concentration level if we compare them with the distribution of the Spanish ‘cities and towns’, the location patterns of each one vary if they are analysed in detail. The highest intensity in the concentration appears mostly in the initial distances of ‘ r ’. This means that

Spanish ‘municipalities’ are more spatially dispersed than the whole of manufacturing at small distances of ‘ r ’, but when this distance becomes larger, it changes.³¹ Therefore, when ‘ r ’ is made larger, the relative level of concentration of the sectors decreases more quickly if we compare it with the distribution of Spanish ‘municipalities’ instead of comparing it with the distribution of the whole of manufacturing. This evidence has two implications: first, it confirms that the self-reinforcing advantages of ‘second nature’ at work are different for each sector and, second, in the case of Spain, these reinforcing advantages have a smaller scope than those of the ‘first nature’.

5. Conclusions

This paper analyses the spatial location patterns of manufacturing firms in Spain. To do this, we use a distance-based method, which allows us to consider space as continuous and avoids the drawbacks of the administrative scale, and thus geographic concentration can be measured at different scales. Therefore, this method enables us to know the intensity of concentration or dispersion of each Spanish manufacturing sector, the distance at which its maximum level is obtained, and the spatial sequence of the increases in the said intensity. Moreover, we can detect whether the departures from randomness are statistically significant.

The characteristics of the location patterns of the Spanish manufacturing sectors can be attributed, in our first analysis, to various forces acting simultaneously. Thus, the location of the activity may be due, first of all, to the dissimilarities in such natural features as mountains, rivers or harbours, that is, ‘first nature’. Secondly, it may be due to the general tendency of manufacturing firms to agglomerate and, thirdly, to the idiosyncratic tendency of each particular sector to concentrate. In fact, by using ‘*complete spatial randomness*’ as the benchmark, we find that in most of the sectors a general tendency to concentrate predominates, whatever the distance considered.

³¹ Detailed results are available upon request from the authors.

In our second analysis, we control for the first nature and for the general tendency of manufacturing firms to agglomerate, which allows us to isolate the specificities of each sector and their idiosyncratic tendency to concentrate. In this way, although the most highly concentrated sectors coincide in both analyses, not every sector presents concentration as compared to the overall manufacturing industry. Our results show that about 70% of Spanish manufacturing sectors and a similar proportion of manufacturing employees are relatively concentrated. The most highly concentrated sectors are both traditional (predominated by textile-related industries) and high-tech industries. This also occurs in the UK, France or the USA, as we have already commented, and it coincides with the Spanish results obtained in 1999. From this, we can conclude that these location patterns have not varied significantly in Spain in the last few years and that manufacturing sectors in different countries tend to follow similar patterns of location. These findings lead us to believe that this is more likely to be due to idiosyncratic features of the sectors (technology, input structure, transport costs, and so forth) than to characteristics of the countries themselves.

With regard to the specific location patterns of each sector, more than half of them present a stronger tendency to cluster than the manufacturing industry as a whole, no matter what the distance is; in fact just one sector presents dispersion at every radius measured and the rest present concentration or dispersion depending on the spatial scale that is chosen. However, by employing a polygonal boundary instead of a rectangular bounding box, the sectors that present relative concentration patterns at all distances considered are reduced, while those that present some degree of relative dispersion increase. Finally, using ‘*total manufacturing*’ as the benchmark, our index is comparable across industries, it controls for the overall agglomeration of manufacturing and for industrial concentration, it is unbiased with respect to scale and aggregation, and it gives an indication of the significance of the results.

By means of the marginal M_{TM} value ($\Delta M/\Delta r$), which informs us about the increase in neighbours when r becomes higher in every sector relative to the increase in neighbours in the overall manufacturing, it can be seen that the distances at which we can find the maximum increase in neighbouring firms vary considerably among sectors. In all probability, the differences in the size of the clusters and the sequence in which clusters appear in the territory are determined by the nature of the centripetal and centrifugal forces of each sector. As a result, we find multiple nearby clusters in the most highly concentrated sectors, but the initial spatial scale at which we find these multiple micro-clusters is smaller in high-tech sectors (i.e. 16 km) than in traditional, low-tech sectors, where it is 25 km. This may be because the centripetal forces are stronger in high-tech sectors. Finally, when the spatial scale considered becomes larger, new clusters appear and this indicates the existence of new agglomerations of firms in different regions.

By using the spatial distribution of ‘cities and towns’ as a benchmark, we can go a step forward. Thus, although all sectors have a higher relative concentration level if we compare them with the distribution of Spanish municipalities, this tendency is stronger at small distances. This result indicates a clear tendency for industries to form clusters more than Spanish municipalities at small distances, thus confirming that the self-reinforcing advantages of ‘second nature’ at work have a smaller scope than those of ‘first nature’.

Finally, before ending, we should just add that we are aware of the aggregation issue in our analysis, given that we are dealing with sectors that are aggregated at the two-digit level. This aspect may generate a clear compensation effect between the different branches of each sector, since the most aggregated and the most dispersed ones can compensate each other. Consequently, one of the next steps in our analysis could be to try to find out whether the location patterns presented by the different branches of each sector are similar to or different from those displayed by the sector itself.

References

- Alonso-Villar O, Chamorro-Rivas JM, González-Cerdeira X (2003) Spillovers geográficos y sectoriales de la industria española. *Revista de Economía Aplicada* 32: 77-95
- Alonso-Villar O, Chamorro-Rivas JM, González-Cerdeira X (2004) Agglomeration economies in manufacturing industries: the case of Spain. *Applied Economics* 36: 2103-2116
- Amiti M (1999) Specialisation patterns in Europe. *Review of World Economics (Weltwirtschaftliches Archiv)* 135: 573-593
- Arbia G, Espa G (1996) *Statistica Economica Territoriale*, Cedam, Padua.
- Arbia G, Espa G, Quah D (2008) A class of spatial econometric methods in the empirical analysis of clusters of firms in the space. *Empirical Economics* 34: 81-103.
- Baddeley A, Turner R (2005) Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software* 12 (6): 1-42. ISSN: 1548-7660. URL: www.jstatsoft.org
- Brühlhart M (2001) Evolving geographical concentration of European manufacturing industries. *Review of World Economics (Weltwirtschaftliches Archiv)* 137: 215-243
- Callejón M (1997) Concentración geográfica de la industria y economías de aglomeración. *Economía Industrial* 317: 61-68
- Cressie NA (1993) *Statistics for Spatial Data*. New York: Wiley
- Cuzick J, Edwards R (1990) Spatial clustering for inhomogeneous populations. *Journal of the Royal Statistical Society, Series B* 52: 73-104
- Devereux MP, Griffith R, Simpson H (2004) The geographic distribution of production activity in the UK. *Regional Science and Urban Economics* 34: 533-564
- Diggle PJ (1983) *Statistical Analysis of Spatial Point Patterns*. London: Academic Press
- Diggle PJ, Chetwynd AG (1991) Second-order analysis of spatial clustering for inhomogeneous populations. *Biometrics* 47: 1155-1163
- Duncan RP (1993) Testing for life historical changes in spatial patterns of four tropical tree species in Westland, New Zealand. *Journal of Ecology* 81: 403-416
- Duranton G, Overman H (2005) Testing for Localization using Micro-Geographic Data. *Review of Economic Studies* 72: 1077-1106
- Ellison G, Glaeser E (1997) Geographic concentration in U.S. manufacturing industries: a dartboard approach. *Journal of Political Economy* 105: 889-927

- Ellison G, Glaeser E (1999) The geographic concentration of industry: does natural advantage explain agglomeration? *American Economic Review* 89: 311-316
- Fujita M, Krugman P, Venables A (1999) *The Spatial Economy: Cities, Regions and International Trade*. MIT Press, Cambridge, MA
- Fujita M, Thisse JF (2009) New Economic Geography: An appraisal on the occasion of Paul Krugman's 2008 Nobel Prize in Economic Sciences. *Regional Science and Urban Economics* 39: 109-119
- Goreaud and Pélissier (1999) On explicit formulas of edge effect correction for Ripley's *K*-function. *Journal of Vegetation Science* 10: 433-438
- Krugman P (1980) Scale Economies, Product Differentiation and the Pattern of Trade. *American Economic Review* 70: 950-959
- Krugman P (1991a) *Geography and Trade*. MIT Press, Cambridge, MA
- Krugman P (1991b) Increasing returns and economic geography. *Journal of Political Economy* 99: 483-499
- Krugman P (1993) First Nature, Second Nature, and Metropolitan Location. *Journal of Regional Science*, 33 (2): 129-144
- Marcon E, Puech F (2003a) Evaluating the Geographic Concentration of Industries using Distance-Based Methods. *Journal of Economic Geography* 3: 409-428
- Marcon E, Puech F (2003b) Measures of the Geographic Concentration of Industries: Improving Distance-Based Methods. *Cahiers de la MSE* 18: 409-428
- Marshall A (1890) *Principles of Economics*. MacMillan, London
- Maurel F, Sédillot B (1999) A measure of the geographic concentration in French manufacturing industries. *Regional Science and Urban Economics* 29: 575-604
- Morton A (2003) Workbook from Alan Morton. Electronic publication. Distribution mapping software (DMAP), <http://www.dmap.co.uk>
- Ottaviano G, Puga D (1998) Agglomeration in the Global Economy: A Survey of the 'New Economic Geography'. *The World Economy* 21: 707-731
- Ottaviano G, Thisse JF (2004) Agglomeration and Economic Geography. In: Henderson JV, Thisse JF (eds) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland, 2563-2608

- Overman H, Redding S, Venables AJ (2003) The Economic Geography of Trade Production and Income: a Survey of Empirics. In: Choi EK, Harrigan J (eds) *Handbook of International Trade*. Blackwell Publishing Ltd., 353-387
- Paluzie E, Pons J, Tirado, D (2001) Regional Integration and Specialization Patterns in Spain. *Regional Studies* 35: 285-296
- Paluzie E, Pons J, Tirado, D (2004) The geographical concentration of industry across Spanish regions, 1856-1995. *Review of Regional Research* 24: 143-160
- Puga D (1999) The rise and fall of regional inequalities. *European Economic Review* 43: 303-334
- Puga D (2002) European regional policies in light of recent location theories. *Journal of Economic Geography* 2: 373-406
- Quah D (1996) Regional convergence clusters across Europe. *European Economic Review* 40: 951-958
- Quah D, Simpson H (2003) Spatial cluster empirics. London School of Economics Working Paper Series
- R Development Core Team (2007) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Redding SJ (2009) Economic Geography: A Review of the Theoretical and Empirical Literature. CEPR Discussion Papers No. 7126
- Ripley BD (1976) The second-order analysis of stationary point processes. *Journal of Applied Probability* 13: 255-266
- Ripley BD (1977) Modelling Spatial Patterns. *Journal of the Royal Statistical Society - Series B (Methodological)* 39: 172-192
- Ripley BD (1979) Test of 'randomness' for spatial patterns. *Journal of the Royal Statistical Society - Series B (Methodological)* 41: 368-374
- Rosenthal SS, Strange WC (2001) The Determinants of Agglomeration. *Journal of Urban Economics* 50: 191-229
- Rowlingson B, Diggle P, adapted, packaged for R by Bivand R, pcp functions by Petris G and goodness of fit by Eglen S (2007). SplanCs: Spatial and Space-Time Point Pattern Analysis. R package version 2.01-23. <http://www.maths.lancs.ac.uk/~rowlings/SplanCs/>
- SABI. System of Iberian Balances Analysis.

Sweeney SH, Feser EJ (1998) Plant size and clustering of manufacturing activity. *Geographical Analysis* 30: 45-64

Tirado D, Paluzie E, Pons J (2002) Economic integration and industrial location: the case of Spain before World War I. *Journal of Economic Geography* 2: 343-363

Venables AJ (1995) Economic Integration and the Location of Firms. *American Economic Review* 85: 296-300

Venables AJ (2006) Shifts in Economic Geography and Their Causes. Federal Reserve Bank of Kansas City - *Economic Review*, Fourth Quarter: 61-85

Viladecans E (2001) La concentración territorial de las empresas industriales: un estudio sobre la unidad geográfica de análisis mediante técnicas de econometría espacial. Document de treball 2001/2, Institut d'Economia de Barcelona.

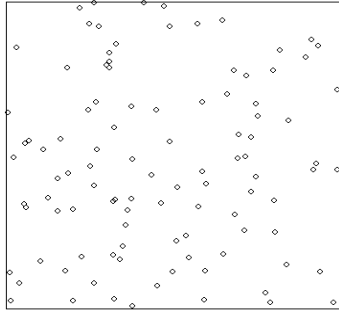


Figure 1. Independent distribution.

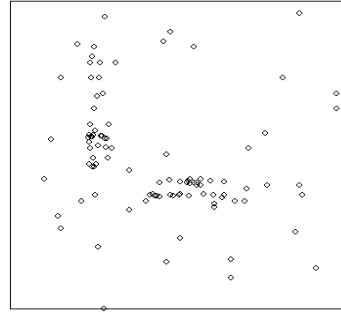


Figure 2. Concentrated distribution.

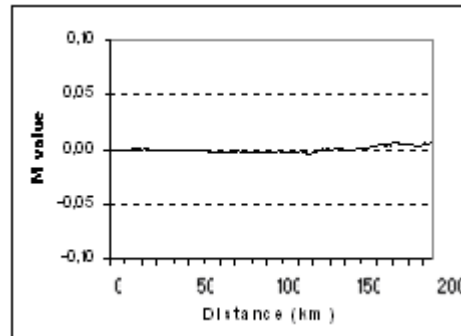
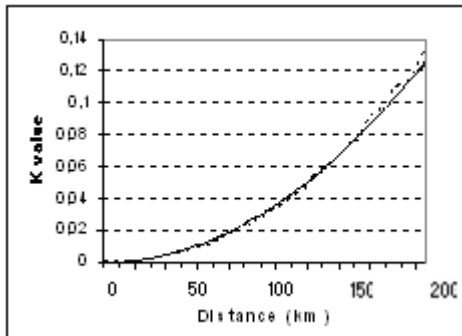


Figure 3. Ripley's K function (theoretical and empirical) and M function corresponding to the point pattern from Figure 1.

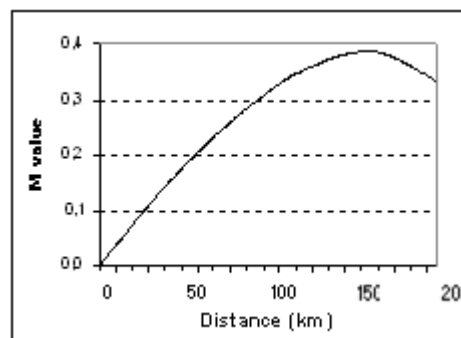
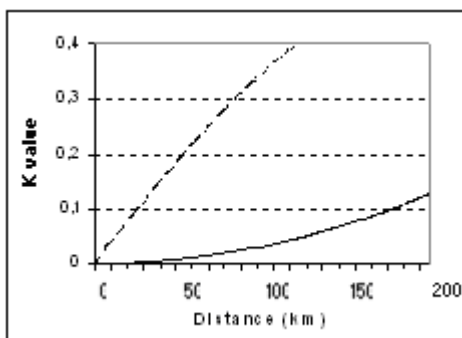
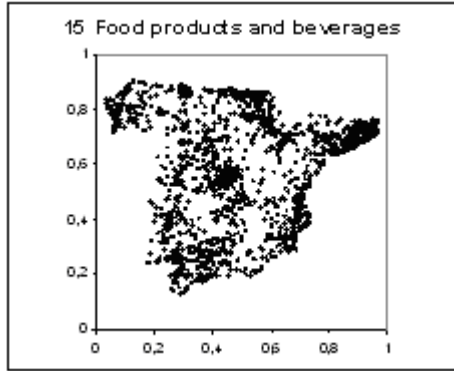
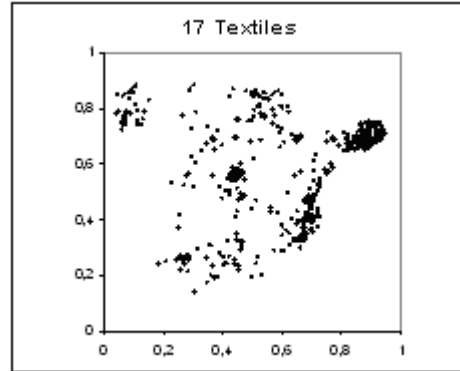


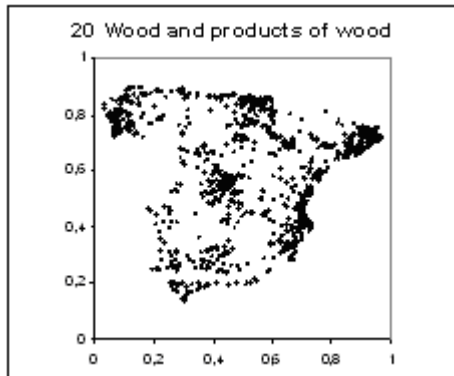
Figure 4. Ripley's K function (theoretical and empirical) and M function corresponding to the point pattern from Figure 2.



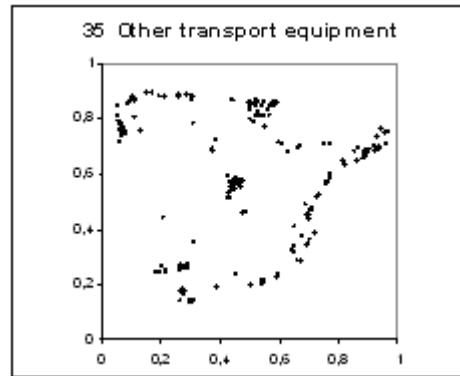
(a)



(b)

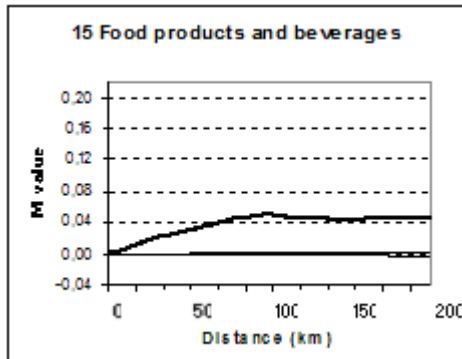


(c)

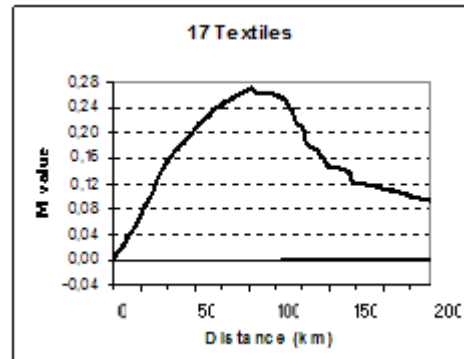


(d)

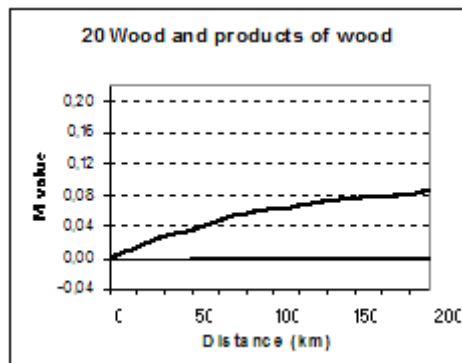
Figure 5. Maps of spatial distribution of firms.



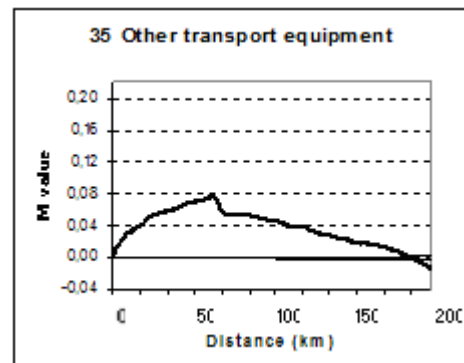
(a)



(b)

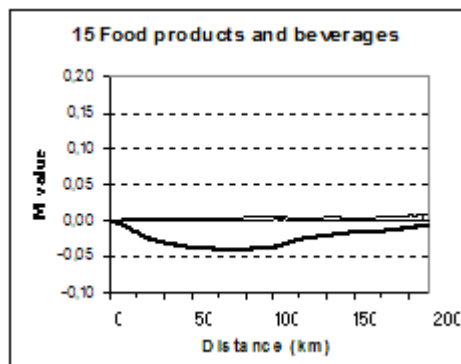


(c)

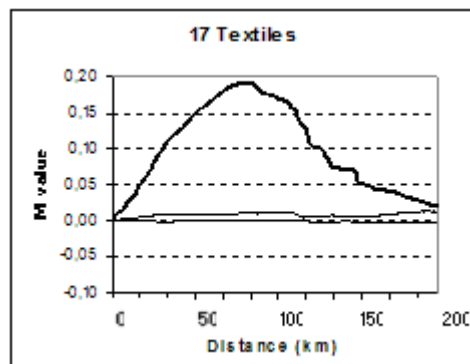


(d)

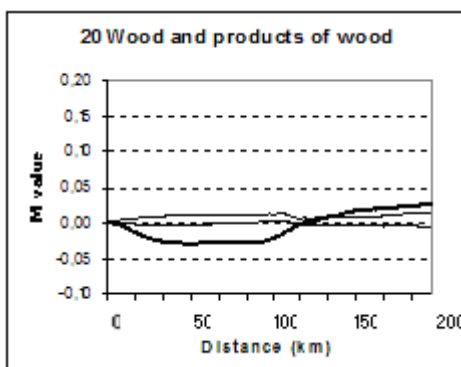
Figure 6. Spatial location patterns (M_{CSR}).



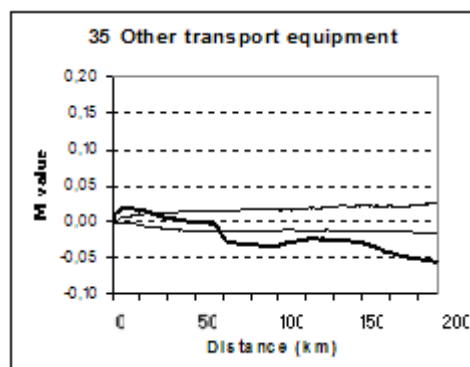
(a)



(b)

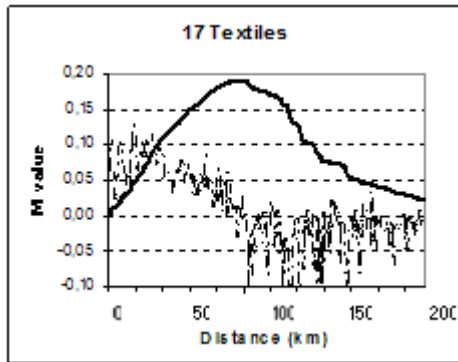


(c)

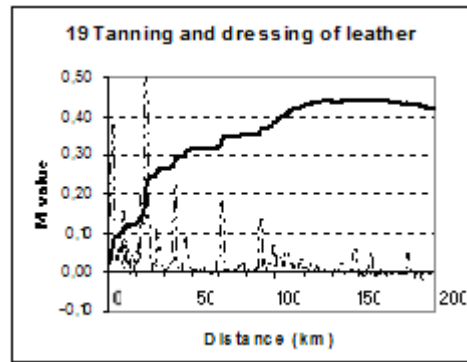


(d)

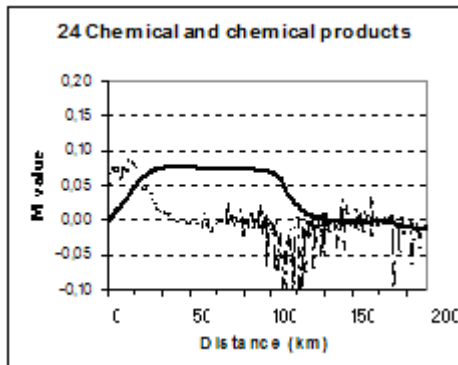
Figure 7. Spatial location patterns (M_{TM}).



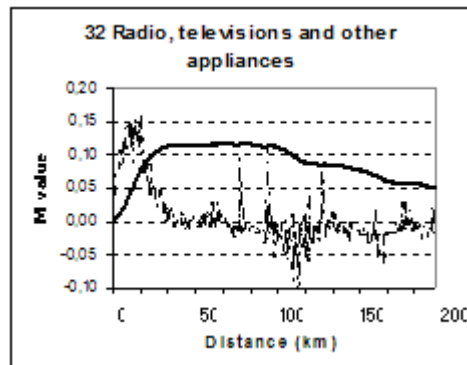
(a)



(b)



(c)



(d)

Figure 8. Spatial location patterns (M_{TM} and marginal M_{TM} value).

Table 1. Location patterns of Spanish manufacturing sectors ($M_{TM_rectangular\ bounding\ box}$)

Sectors (NACE 93 - Rev. 1)	Significant concentration	Significant dispersion	Significant peak		Type of cluster
			M_{TM} value	Distance (r)	
15 Food products and beverages	---	All distances	-0.04	80 km	2
17 Textiles	All distances	---	0.19	82 km	1
18 Wearing apparel and dressing	---	---	---	---	0
19 Tanning and dressing of leather	All distances	---	0.45	152 km	1
20 Wood and products of wood	137-200 km	0-120 km	-0.03	64 km	4
21 Pulp, paper and paper products	All distances	---	0.07	103 km	1
22 Publishing, printing & recorded media	All distances	---	0.12	28 km	1
24 Chemical and chemical products	0-128 km	---	0.08	41 km	1
25 Rubber and plastic products	0-160 km	---	0.04	90 km	1
26 Other non-metallic mineral products	145-200 km	0-123 km	-0.03	60 km	4
27 Basic metals	36-200 km	---	0.03	200 km	1
28 Fabricated metal products	---	---	---	---	0
29 Other machinery and equipment	0-169 km	---	0.04	77 km	1
31 Electrical machinery	0-131 km	---	0.07	48 km	1
32 Radio, televisions & other appliances	All distances	---	0.12	80 km	1
33 Instruments	0-140 km	---	0.11	45 km	1
34 Motor vehicles and trailers	51-100 km	---	0.02	80 km	1
35 Other transport equipment	0-25 km	57-200 km	0.02	9 km	3
36 Furniture and other products	96-200 km	25-93 km	0.05	167 km	4
37 Recycling	---	---	---	---	0

Table 2. Location patterns of Spanish manufacturing sectors ($M_{TM_polygonal\ bounding\ box}$)

Sectors (NACE 93 - Rev. 1)	Significant concentration	Significant dispersion	Significant peak	
			M_{TM} value	Distance (r)
15 Food products and beverages	---	All distances	-0.01	63 km
17 Textiles	0-37 km	49-200 km	0.03	20 km
18 Wearing apparel and dressing	116-200 km	---	0.04	200 km
19 Tanning and dressing of leather	0-82 km	86-200 km	0.11	30 km
20 Wood and products of wood	---	42-127	-0.01	70 km
21 Pulp, paper and paper products	All distances	---	0.04	95 km
22 Publishing, printing & recorded media	All distances	---	0.13	82 km
24 Chemical and chemical products	All distances	---	0.03	76 km
25 Rubber and plastic products	---	---	---	---
26 Other non-metallic mineral products	---	36-200 km	-0.03	200 km
27 Basic metals	---	---	---	---
28 Fabricated metal products	All distances	---	0.07	87 km
29 Other machinery and equipment	---	87-200 km	-0.03	200 km
31 Electrical machinery	0-159 km	---	0.03	85 km
32 Radio, televisions & other appliances	All distances	---	0.12	128 km
33 Instruments	All distances	---	0.07	81 km
34 Motor vehicles and trailers	---	---	---	---
35 Other transport equipment	0-10 km	130-200 km	0.01	7 km
36 Furniture and other products	---	---	---	---
37 Recycling	---	---	---	---

Table 3. Distance of maximum intensity of neighbouring firms

Sectors (NACE 93 - Rev. 1)	Absolute M_{TM}	Marginal $M_{TM}^{(a)}$
15 Food products and beverages	80 km	
17 Textiles	87 km	2/8/17/25 km
18 Wearing apparel and dressing	---	
19 Tanning and dressing of leather	157 km	2/9/24/26/27 km 48/81/113 km
20 Wood and products of wood	64 km	
21 Pulp, paper and paper products	103 km	17 km
22 Publishing, printing & recorded media	36 km	2-6/17 km
24 Chemical and chemical products	83 km	4-6/13/16 km
25 Rubber and plastic products	90 km	18 km
26 Other non-metallic mineral products	60 km	
27 Basic metals	63 km	
28 Fabricated metal products	66 km	
29 Other machinery and equipment	77 km	11 km
31 Electrical machinery	81 km	10 km
32 Radio, televisions & other appliances	83 km	5/7/ 9-10 km 12/16/18 km
33 Instruments	81 km	4/8-10/14/17 km
34 Motor vehicles and trailers	80 km	
35 Other transport equipment	9 km	2 km
36 Furniture and other products	193 km	
37 Recycling	---	

^(a) Those sectors in which the ‘marginal M_{TM} value’ does not appear is because it does not add relevant information.

Table 4. Location patterns of Spanish manufacturing sectors (M_M)

Sectors (NACE 93 - Rev. 1)	Significant concentration	Significant dispersion	Significant peak	
			M_M value	Distance (r)
15 Food products and beverages	0-108 km	117-200 km	0.02	73 km
17 Textiles	All distances	---	0.24	86 km
18 Wearing apparel and dressing	0-138 km	145-200 km	0.06	69 km
19 Tanning and dressing of leather	All distances	---	0.44	114 km
20 Wood and products of wood	0-141 km	153-200 km	0.03	76 km
21 Pulp, paper and paper products	All distances	---	0.12	85 km
22 Publishing, printing & recorded media	All distances	---	0.17	78 km
24 Chemical and chemical products	0-141 km	153-200 km	0.14	78 km
25 Rubber and plastic products	0-145 km	152-200 km	0.10	84 km
26 Other non-metallic mineral products	0-139 km	151-200 km	0.03	76 km
27 Basic metals	0-122 km	127-200 km	0.08	70 km
28 Fabricated metal products	0-125 km	131-200 km	0.07	72 km
29 Other machinery and equipment	0-126 km	134-200 km	0.10	74 km
31 Electrical machinery	0-137 km	144-200 km	0.13	75 km
32 Radio, televisions & other appliances	All distances	---	0.18	80 km
33 Instruments	0-160 km	171-200 km	0.17	79 km
34 Motor vehicles and trailers	0-133 km	142-200 km	0.08	80 km
35 Other transport equipment	0-103 km	112-200 km	0.06	58 km
36 Furniture and other products	All distances	---	0.06	78 km
37 Recycling	0-130 km	138-200 km	0.06	55 km

Appendix 1

Table A1. Additional descriptive information about Spanish manufacturing sectors

Sectors (NACE 93 - Rev. 1)	Number of firms	Number of employees	Technological intensity ³²
15 Food products and beverages	5,761	356,314	L
16 Tobacco products	6	1,226	L
17 Textiles	1,949	81,818	L
18 Wearing apparel and dressing	1,710	59,286	L
19 Tanning and dressing of leather	1,698	46,708	L
20 Wood and products of wood	2,340	75,844	L
21 Pulp, paper and paper products	837	56,890	L
22 Publishing, printing & recorded media	3,004	130,222	L
23 Coke, refined petroleum products	12	16,417	M-L
24 Chemical and chemical products	1,722	158,238	H
25 Rubber and plastic products	2,165	138,488	M-L
26 Other non-metallic mineral products	3,413	225,792	M-L
27 Basic metals	986	137,066	M-L
28 Fabricated metal products	8,094	267,568	M-L
29 Other machinery and equipment	3,015	161,407	M-H
30 Office machinery and computers	77	6,374	H
31 Electrical machinery	1,099	79,357	M-H
32 Radio, televisions & other appliances	344	31,593	H
33 Instruments	376	19,528	H
34 Motor vehicles and trailers	876	192,873	M-H
35 Other transport equipment	451	58,274	M-H
36 Furniture and other products	2,924	100,084	L
37 Recycling	228	8,095	L

³² This classification of sectors according to the technological intensity belongs to the National Statistics Institute, meaning H = high, M-H = medium high, M-L = medium low and L = low.

Appendix 2

The graphs situated on the left illustrate the spatial distribution of firms from each Spanish manufacturing sector. The graphs on the right show the spatial location patterns of these sectors measured by the M_{TM} function and a rectangular boundary (thick continuous line), with the addition of the marginal M_{TM} value (dashed line) and their corresponding confidence intervals (thin continuous lines).

