Assessing the Tendency of Spanish Manufacturing Industries to Cluster: Colocalization and Establishment Size

Marta R. Casanova and Vicente Orts

Ivie

Los documentos de trabajo del Ivie ofrecen un avance de los resultados de las investigaciones económicas en curso, con objeto de generar un proceso de discusión previo a su remisión a las revistas científicas. Al publicar este documento de trabajo, el Ivie no asume responsabilidad sobre su contenido.

La Serie EC, coordinada por Matilde Mas, está orientada a la aplicación de distintos instrumentos de análisis al estudio de problemas económicos concretos.

Todos los documentos de trabajo están disponibles de forma gratuita en la web del Ivie http://www.ivie.es, así como las instrucciones para los autores que desean publicar en nuestras series.

Assessing the Tendency of Spanish Manufacturing Industries to Cluster: Co-localization and Establishment Size

Marta R. Casanova and Vicente Orts[*]

Abstract

In this paper, we use a distance-based method, specifically the Ripley's K function, to evaluate the spatial location patterns of Spanish manufacturing establishments and to assess the different tendencies to cluster in each sector or subsector relative to the whole of manufacturing. Specifically, we analyse the role played by the size of establishments in determining the location patterns detected in each sector, and the co-localization between horizontally- and vertically-linked industries to assess the importance of the potential spillovers across industries. We apply this methodology to Spanish manufacturing industries at the two-digit and the four-digit levels. Considering four digits of disaggregation allows us to isolate the different behaviour in the spatial distribution of each subsector as well as prevent the effects of compensation due to previous aggregation

Keywords: spatial location, distance-based method, Ripley's K function, polygonal boundary, disaggregation, co-localization, establishment size.

JEL classification: C15, C40, C60, R12.

Resumen

En este trabajo se utiliza un método basado en la distancia, específicamente la función K de Ripley, para analizar los patrones de localización espacial de los establecimientos manufactureros españoles y evaluar las diferentes tendencias a concentrarse de cada sector o subsector en relación con el conjunto de la industria manufacturera. En concreto, se analiza el papel que desempeña el tamaño de los establecimientos en la determinación de los patrones de localización detectados en cada sector. Así mismo, se evalúa la importancia de los *spillovers* potenciales entre las diferentes industrias o subsectores mediante un análisis de la tendencia a la co-localización entre empresas de industrias relacionadas horizontal y verticalmente. Aplicamos esta metodología a las industrias manufactureras españolas con un nivel de desagregación de dos y cuatro dígitos CNAE. El hecho de considerar cuatro dígitos de desagregación nos permite detectar las diferencias en el comportamiento de la distribución espacial de cada subsector, así como prevenir los efectos de compensación debidos a la agregación anterior.

Palabras clave: localización espacial, método basado en la distancia, función K de Ripley, área poligonal, desagregación, co-localización, tamaño de establecimiento.

1. Introduction

The most striking feature of the spatial distribution of economic activity is its heterogeneity. This tendency of firms and industries to become spatially localised has attracted the attention of economists since the pioneering works of Von Thünen, Marshall and Weber, to more recent contributions from the 'new economic geography' initiated by Krugman (1991).[1]

This interest has been extended, last years, to the development of empirical methods to quantify and characterise this tendency of individual firms and industries to cluster in space. The *first generation* of these measures – using the terminology employed by Duranton and Overman (2005) – was based on indicators such as *Herfindahl* or *Gini*, which did not take space into consideration.[2] The *second generation*, initiated by Ellison and Glaeser (1997), began to take space into account, but not in a proper way. This index still used administrative units to measure the spatial distribution of economic activity, treating space as being discrete.[3] Therefore, they restricted the analysis of spatial distribution to just one administrative scale, '*they transform points on a map into units in boxes*'.[4] Alternatively, the *third generation* of empirical measures of spatial localization, developed by authors from different scientific fields (economics, geography and statistics), introduced the treatment of space as being continuous. Authors like Marcon and Puech (2003), Quah and Simpson (2003), Duranton and Overman (2005) and Arbia et al. (2008), among others, were the pioneers in introducing these methods in economic geography. More recently, papers by Duranton and Overman (2008), Marcon and Puech (2010) or Albert, Casanova and Orts (2011) developed several extensions and improvements to these methodologies. These approaches use multiple scales simultaneously, are unbiased with respect to arbitrary changes in the spatial units and can allow us to know and to compare the concentration intensity for every spatial scale. Evidently, the measures included in this generation avoided the shortcomings of the administrative scale.

In this paper, we use a measure belonging to this last generation to analyse the spatial location patterns of Spanish manufacturing establishments, assessing the different tendencies to cluster in each industry or subsector relative to the whole of manufacturing. Specifically, we will use the Ripley's *K* function, a distance-based

---

[1] For further details, see Marshall (1890), Krugman (1991a, 1991b), Ottaviano and Puga (1998), Fujita *et al*. (1999), Puga (1999, 2002), or Venables (1995).

[2] See Krugman (1991a) or Amiti (1997).

[3] See Callejón (1997), Maurel and Sédillot (1999) Brülhart (2001), Rosenthal and Strange (2001), or Devereux et al. (2004).

[4] Duranton and Overman (2005), p. 1078.

method, which enables us to know whether concentration exists, its intensity, and at what distance or spatial scale its highest level is obtained.[5] This measure has also been used by other authors to analyse the spatial distribution of activity in other countries, such as the aforementioned papers of Marcon and Puech or Arbia for France and Italy, respectively.

The continuity of the space is achieved by means of the availability of the geographic coordinates (longitude and latitude) of every Spanish manufacturing establishment. Through said geographic coordinates, we locate the establishments, represented by dots, accurately in space without taking administrative borders into account. Hence, this method allows us to treat space as continuous, analysing simultaneously multiple spatial scales and avoiding the shortcomings of the administrative scale. In addition, we employ a polygonal boundary to improve the delimitation of our area of study, instead of the rectangular shape used by other authors, thus avoiding the nuisance of empty spaces.

In regards to the five requirements that any test which measures concentration should fulfil, as proposed by Duranton and Overman (2005), we should emphasise that our measure of concentration meets all of them: (1) is comparable across industries, (2) controls for the overall agglomeration tendency of manufacturing, (3) controls for industrial concentration, (4) is unbiased with respect to scale and aggregation, and (5) gives an indication of the significance of the results.

Results from previous papers lead us to believe that the manufacturing location patterns observed in Spain are mostly brought about by the characteristics of each sector.[6] Besides, these patterns do not always correspond with the assumptions of economic theory.[7] For example, in Albert, Casanova and Orts (2011) we can see that the most highly concentrated sectors in the manufacturing industry in Spain are both traditional and high-tech industries. Nevertheless, the aggregation in this analysis prevents us from knowing if all subsectors within the same sector would follow similar location patterns. In this way, through the disaggregation carried out in this paper, both by the size of establishments and by subsectors, we are searching for more detailed patterns of establishments' location in Spanish manufacturing industries. Thus, we can isolate the different behaviours of spatial distribution of each subsector caused by 'spillovers' characteristic of their respective activities.

---

[5] For further details, see Ripley (1976, 1977, and 1979).

[6] See Callejón (1997), Viladecans (2001), Alonso-Villar et al. (2003, 2004), Paluzie et al. (2004) or Albert, Casanova and Orts (2011).

[7] Audretsch and Feldman (2004) reviewed the literature related to the paper of the knowledge spillovers in the economic geography and concluded that these spillovers, as many others, matter in the formation of clusters and agglomeration.

Through this disaggregation we will see if there is interaction and interdependence between different subsectors as they locate in space. In other words, we will check if there is co-localization between horizontally- and vertically-linked subsectors. Furthermore, we will try to determine the location patterns of the establishments depending on their size; i.e. what type of establishment ('small' or 'large') is the driver of the Spanish industrial agglomeration.[8]

The remainder of the paper is organized as follows. In section 2, we present the data used in our analysis. In section 3, we introduce the methodology employed and discuss the main results of the location of Spanish industries, taking into account the size of establishments and the main subsectors at the four-digit level of aggregation. In section 4, we extend our analysis to the patterns of co-localization between vertically and horizontally linked industries, while in the last section we conclude.

2. Data

We use establishment level data, for the year 2007, from the Analysis System of Iberian Balances database[9] to carry our empirical analysis out. Our database contains Spanish manufacturing sectors at the two-digit and the four-digit level, which are classified using the National Classification of Economic Activities[10]. From every establishment we know the geographical coordinates, the industrial classification at four-digit level and the number of employees. Last, we must take into account that when we refer to NACE two-digit level we speak of '*sectors*' and when we refer to NACE four-digit level we speak of '*subsectors*'.

Our database is restricted to Spanish manufacturing establishments located only on the peninsula and not in the Canary and Balearic Islands, Ceuta or Melilla, and employing at least ten workers. This second requirement is due to the fact that most of establishments with less than ten workers do not have essential information (geographical coordinates) to carry our analysis out. After considering these requirements our database contains 43,048 establishments.

In contrast, other Spanish papers, such as Alonso-Villar et al. (2004) or Callejón (1997), analyse the geographic concentration of industry in Spain by using a dataset taken from the 'Industrial Survey of Businesses', provided by the INE (Spanish National Institute of Statistics). This survey provides data on employment according to

---

[8] We must keep in mind that in the Spanish manufacturing abounds the small enterprises and the establishments are often family businesses with few employees.

[9] SABI

[10] NACE 93 - Rev. 1

two geographical subdivisions – 17 autonomous communities (NUTS-II) and 50 provinces (NUTS-III) – with a sectorial breakdown to two and three digits and for 30 manufacturing sectors. Thus, these studies are stuck to administrative-scale data and the spatial scale chosen is a key decision that may alter the results and conclusions reached. By contrast, the concentration measure used in this paper employs a dataset that treats *space as continuous*, without taking administrative borders into account and without restricting the spatial distribution to just *one scale*, by being possible to analyse simultaneously *multiple spatial scales*.

Spanish manufacturing activities are classified into 23 sectors according to 'NACE 93 - Rev. 1': (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, 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. In our analysis, we consider twenty of these aforementioned sectors because three of them (16, 23 and 30) are too small as far as their number of establishments is concerned.

In Table 1 appears descriptive information about each one of the above-mentioned sectors. We find their technological intensity, the number of establishments and the structure of each one of them depending on the number of employees.

The European Union started to standardise the concept of small and large establishment and its current definition categorises companies with fewer than 50 employees as "small", those with fewer than 250 as "medium" and with more than 250 as "large". So, if we observe the structure of sectors taking into consideration the size of their establishments (Table 1), we find that Spain is mostly a country of small establishments and that the proportion of 'large' establishments is very low. Indeed, in most economies, smaller enterprises are much greater in number. Small and medium-sized enterprises play an important role in the economy of all countries because they can perform customized products as opposed to big companies that focus more on more standardized products. Small establishments can also serve as auxiliary to large firms. If we look again at the table, we also note that the proportion of 'small' establishment is much higher in low-tech sectors. In fact, sectors whose proportion of 'small' establishment is higher than 85% are, all of them, low-tech intensive sectors (18, 19, 20,

22, 28, 36 and 37). On the other hand, the proportion of 'medium' and 'large' establishment is much higher in high-tech sectors (24, 32, 34 and 35).

Table 1. Additional descriptive information

Sector	Technological intensity	Total of establishments	<20 employees		<50 employees		>50 employees		>250 employees	
15	L	5778	2618	45.3%	4620	80.0%	1158	20.0%	220	3.8%
17	L	1951	944	48.4%	1620	83.0%	331	17.0%	36	1.8%
18	L	1712	902	52.7%	1508	88.1%	204	11.9%	19	1.1%
19	L	1699	894	52.6%	1565	92.1%	134	7.9%	7	0.4%
20	L	2349	1307	55.6%	2124	90.4%	225	9.6%	19	0.8%
21	L	837	295	35.2%	619	74.0%	218	26.0%	35	4.2%
22	L	2997	1565	52.2%	2554	85.2%	443	14.8%	58	1.9%
24	H	1722	566	32.9%	1162	67.5%	560	32.5%	143	8.3%
25	M-L	2165	875	40.4%	1690	78.1%	475	21.9%	54	2.5%
26	M-L	3429	1464	42.7%	2738	79.8%	691	20.2%	89	2.6%
27	M-L	987	397	40.2%	712	72.1%	275	27.9%	80	8.1%
28	M-L	8103	4297	53.0%	7104	87.7%	999	12.3%	90	1.1%
29	M-H	3018	1281	42.4%	2458	81.4%	560	18.6%	70	2.3%
31	M-H	1099	445	40.5%	809	73.6%	290	26.4%	64	5.8%
32	H	344	120	34.9%	240	69.8%	104	30.2%	29	8.4%
33	H	376	180	47.9%	302	80.3%	74	19.7%	13	3.5%
34	M-H	876	244	27.9%	521	59.5%	355	40.5%	134	15.3%
35	M-H	451	144	31.9%	313	69.4%	138	30.6%	31	6.9%
36	L	2927	1531	52.3%	2576	88.0%	351	12.0%	30	1.0%
37	L	228	125	54.8%	199	87.3%	29	12.7%	3	1.3%

Regarding to treat the disaggregation and co-localization, we take into consideration and analyse seventy-seven subsectors, which have been chosen under specific selection criteria in order to have a representative sample of all the sectors. Additionally, in order to analyse the existing co-localization between pairs of these 'subsectors' we need to find those pairs with linkages. For this, the data we are going to use will be the 1996 Input-Output Table, which is published by the National Institute of Statistics (INE)[11] of Spain. This analysis provided 168 pairs of subsectors with relevant linkages.

3. Location Patterns of Spanish Manufacturing Industries

In this section we present a detailed analysis of the location of Spanish manufacturing establishments and industries, examining the location patterns of establishments according to its size, as well as the location patterns of different subgroups within an industry (subsectors). The methods that we are going to use are based on the Ripley's K function, $K(r)$. This function is a distance-based method that

---

[11] INE, (http://www.ine.es/).

measures concentration by counting the average number of neighbours each establishment has within a circle of a given radius, 'neighbours' being understood to mean all establishments situated at a distance equal to or lower than the radius (r). From here on, establishments 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 ith and jth establishments; $I(x)$ 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 the area of the study region; and w_{ij} is the weighting factor to correct for border effects.[12] The indicator function, $I(d_{ij})$, takes a value of 1 if the distance between the ith and jth establishments is lower or equal than r, or 0 otherwise, and w_{ij} will be equal to the area of the circle divided by the intersection between the area of the circle and the area of study.

In our analysis, we improve the delimitation of the area covering the study region, A, by substituting the rectangular shape used by other authors for a polygonal shape. With modern computer equipment it is feasible to consider any arbitrary window A, certainly as complicated as a map of Spain. In our case, the statistical software employed allows border corrections to be applied adequately to any irregular polygonal shape, thereby simplifying the treatment of border effects.

Here, in Figure 1, we can observe the polygonal shape that accurately delimits our territory and envelops the area of study. This allows us to avoid the nuisance of empty spaces where no establishments are found, which are represented by the oblique lines. This polygonal boundary was built by joining thirty-five points on the perimeter of the Spanish territory.

---

[12] 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).

Figure 1. Polygonal shape envelope.
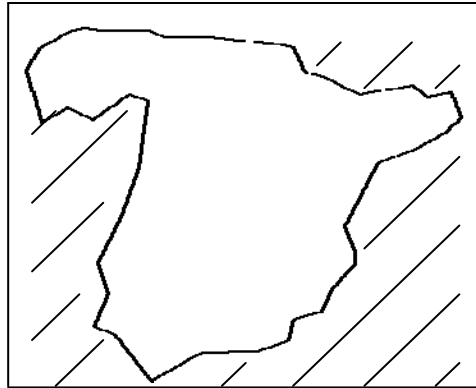
In previous works, it was usually used a rectangular area due to the increasing complexity when simulating random points inside the area and when correcting the border-effects on convex shapes. These drawbacks thus limited the empirical analysis to rectangular areas. For instance, Marcon and Puech (2003) 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.[13] In our case, the statistical software employed, 'R',[14] allows us to apply border corrections adequately in any irregular polygonal shape, thus avoiding the shortcomings associated to the use of a rectangular area as area of study. This incorporation provides robustness to our results.

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 establishments were located in the study area random and independently from each other, we would have a location pattern known as *Complete Spatial Randomness* (*CSR*).

In actual fact, considering establishments 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. Economic

---

[13] They explicitly said, "It was impossible to use the whole of France because of border-effect corrections".

[14] This software is downloadable from the following website: http://www.r-project.org/.

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 establishments to agglomerate.

Consequently, we use the whole of manufacturing as a benchmark. 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. Localization 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.

Now, to evaluate the statistical significance of departures from randomness in a robust way, we should construct a confidence interval for 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. We simulate random distributions with the same number of establishments as in each of the sectors under consideration, but the location of these hypothetical establishments is restricted to the sites where we can currently find establishments from the whole manufacturing sector. It is generated by running 1000 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 assess the significance of departures from randomness and to control for industrial concentration.

Finally, only mention that both the use of the whole of manufacturing as a benchmark and the particular way of constructing the confidence intervals, by locating the points simulated in the specifics sites where we can currently find an establishment, let us to fulfil the five aforementioned requirements of Duranton and Overman (2005).

3.1. Empirical Results

As said, through this methodology, we are going to assess in depth the location of Spanish manufacturing establishments; first, by taking into account the size of establishments and, second, by considering the disaggregation at the four-digit level.

Establishment Size

In Table 1 we have seen descriptive information about the structure of each Spanish manufacturing sector depending on the number of employees of the establishments.[15] We find that Spain's economy is mostly constituted by 'small' establishments whereas the proportion of 'large' establishments is very low; indeed, approximately 80% of the Spanish manufacturing establishments are 'small' establishments.

So, given that in the Spanish manufacturing abounds the small enterprise, we decided to make a detailed analysis of them in order to check if their location patterns vary depending on what we consider a 'small establishment'. Therefore, we propose two parallel datasets and we carry out two parallel analysis. In the first, we take into account establishments with less than twenty workers, and in the second one with less than fifty workers. From Table 1, we also concluded that the proportion of 'small establishments' is much higher in low-tech sectors, while in high-tech sectors the proportion of 'medium' and 'large' establishment is higher. Thus, since it is clear that the proportion of 'small' or 'large' establishment vary depending on the sector we observe, our following objective will be to analyse if the proportion of large or small establishments within each sector determines the pattern of location or the spatial structure of the sector as a whole; i.e. what type of establishment, depending on its size, is the driver of the industrial agglomeration in each Spanish manufacturing sector.

Table 2 contains the summary of this analysis[16] and the resulting graphs are in Appendix[17]. Column one of table 2 shows the studied sector. Column two exhibits the significant peak of the M_{TM} value of every aggregated sector, i.e. the maximum *intensity* reached by the different sectors, and the distance (*r*-value) at which this maximum intensity of the concentration is reached, i.e. the size of the cluster. The third column shows only the M_{TM} value for establishments with fewer than twenty employees, while the fourth and fifth columns present the M_{TM} value of the 'small' and 'large' establishments of each sector and the distance (*r*) at which the highest level of concentration is reached. At first glance we realise that there are not significant dissimilarities in the results with respect to this differentiation criterion to define a 'small' firm. In this way, we will consider a 'small' establishment as that one with less

---

[15] The main reason to not disaggregate this data at the four-digit level is because, due to the split of the sectors, some of the resulting point patterns had been too small as far as their number of establishments was concerned

[16] In this table we have not shown the information of those sectors which, after splitting by size of establishment, do not report significant information on the location pattern of the sector. However, in table 3, we can see the detailed information of all sectors.

[17] The dashed line in these graphs is called 'marginal M_{TM} value' by Albert, Casanova and Orts (2011) and informs us about the increase in the number of neighbours when *r* becomes higher.

than fifty employees and a 'large' establishment as that one with more than fifty employees.

Table 2. Location patterns of establishments according to its size

Sector	Significant peak (*M* value) and Distance (*r*)						
	all firms		<20	<50		>50	
15	-0.01	63 km	-0.01	-0.01	58 km	-0.01	64 km
17	0.03	20 km	0.03	0.03	20 km	0.02	40 km
19	0.11	30 km	0.11	0.12	37 km	0.04	22 km
22	0.13	82 km	0.12	0.13	82 km	0.21	96 km
24	0.03	76 km	0.1	0.01	58 km	0.08	66 km
26	-0.03	200 km	-0.02	-0.02	192 km	0.03	31 km
31	0.03	85 km	0.02	0.03	86 km	0.15	93 km
32	0.12	128 km	0.1	0.09	82 km	0.16	96 km

The general dynamics of the spatial location pattern of 'small' and 'large' establishments differs depending on the sector we refer. If we observe with detail table 2, we realise that 'small' establishments are the drivers of agglomeration in sector 19 (tanning and dressing of leather), whereas in sectors 24 (chemical and chemical products), 31 (electrical machinery) and 32 (Radio, televisions & other appliances) 'large' establishments are the drivers of agglomeration. Therefore, these results show that both 'large' as 'small' establishments may be the cause of the agglomeration on a specific sector and this will depend on the sector taken into consideration. However, we find a characteristic tendency in the high-tech sectors, by appreciating here that 'large' establishments seem to be the main drivers of agglomeration.

Disaggregation

Table 3 summarises the results obtained from computing the M_{TM} function for each 'sector' and those for the most relevant 'subsectors'. First column shows the studied sector, or subsector. In order to check which are the most and the least concentrated manufacturing industries in Spain and to compare the results after the disaggregation, we should pay attention to the second column. This column shows the significant peak of the M_{TM} value. This maximum intensity can be defined as the maximum level of concentration reached at all possible radius (*r*). Finally, third column exhibits the *distance* (*r*) at which this maximum concentration is reached. At this point, we should highlight that the spatial distribution of each sector, or subsector, comes defined by three specific features: (1) the *intensity* of the cluster, (2) the *distance* at which this highest intensity is reached, and (3) the *persistence* of its concentration in space, which combines the two previous characteristics. In fact, although the spatial

Table 3. Location patterns of Spanish manufacturing sectors and subsectors (M_{TM})

Sectors and subsectors (NACE 93 - Rev. 1)	M_{TM} value	Distance (r)
15 Food products and beverages	-0.01	63 km
1511 Production and preserving of meat	-0.01	38 km
1513 Production of meat and poultrymeat products	-0.01	56 km
17 Textiles	0.03	20 km
1725 Other textile weaving	0.09	54 km
1730 Finishing of textiles	0.04	18 km
1754 Manufacture of other textiles	0.01	18 km
18 Wearing apparel and dressing	0.04	200 km
1822 Manufacture of other outerwear	0.05	200 km
1824 Manufacture of other wearing apparel and accessories	0.15	200 km
19 Tanning and dressing of leather	0.11	30 km
1910 Tanning and dressing of leather	0.02	15 km
1920 Manufacture of luggage, handbags, saddlery and harness	0.04	24 km
1930 Manufacture of footwear	0.15	23 km
22 Publishing, printing & recorded media	0.13	82 km
2211 Publishing of books	0.18	50 km
2212 Publishing of newspapers	0.07	165 km
2213 Publishing of journals and periodicals	0.25	65 km
2222 Printing	0.05	115 km
24 Chemical and chemical products	0.03	76 km
2416 Manufacture of plastics in primary forms	0.04	63 km
2442 Manufacture of pharmaceutical preparations	0.12	49 km
2466 Manufacture of other chemical products	0.02	79 km
26 Other non-metallic mineral products	-0.03	200 km
2612 Shaping and processing of flat glass	0.02	115 km
2630 Manufacture of ceramic tiles and flags	0.26	30 km
29 Other machinery and equipment	-0.03	200km
2953 Manufacture of machinery for food, beverage and tobacco processing	-0.03	163 km
2954 Manufacture of machinery for textile, apparel and leather production	0.09	40 km
32 Radio, televisions & other appliances	0.12	128 km
3210 Manufacture of electronic valves and tubes	0.08	70 km
3220 Manufacture of television and radio transmitters	0.13	112 km
33 Instruments	0.07	81 km
3310 Manufacture of medical and surgical equipment	0.05	80 km
3320 Manufacture of instruments for measuring, testing, navigating	0.05	56 km
35 Other transport equipment	0.01	7 km
3511 Building and repairing of ships	0.03	6 km
3530 Manufacture of aircraft and spacecraft	0.06	98 km
36 Furniture and other products	-	-
3622 Manufacture of jewellery and related articles	0.07	55 km
3650 Manufacture of games and toys	0.13	29 km

distribution of some of these industries is similar to the others, the intensity and the distance at which the maximum concentration is reached may be different and this can encourage a large diversity of types of clusters. So, each industry presents its own singularity, a unique location pattern, different from the others.

Considering the concept of intensity – denoted in column 2 – we observe that the most concentrated sectors, those that reach the highest M_{TM} value, are 19, 22, 32 and 33, while the sectors with the lowest levels of concentration are 15, 26 and 29.[1] Nevertheless, the different manufacturing activities included within each sector could be very heterogeneous between them and also their location patterns. Thus, due to aggregation, when we deal with sectors at the two-digit level, the most localised and the most dispersed subsectors may compensate each other and induce inadequate conclusions about the spatial location patterns. So, we should provide a more disaggregated dataset in order to: (1) isolate the different behaviours of spatial distribution of each subsector caused by 'spillovers' characteristic of each activity; (2) prevent the compensation effects due to previous aggregation; and (3) carry out the posterior analysis of co-localization between pairs of establishments. This dataset will be generated by sectors at the four-digit level.

Through this disaggregation, we can determine if the spatial location patterns presented by the subsectors are similar to or different from those displayed by the sector itself. Regarding this, we find two different distribution behaviours. Firstly, some subsectors tend to follow similar location patterns as the aggregated sector to which they belong[2]. Secondly, other subsectors vary heavily in their spatial distribution[3] and these will be the ones discussed. Some of the most relevant cases are contained in Table 3.

Figure 2 and Figure 3 show two of the clearest examples in which some of the subsectors belonging to a particular sector present different location patterns than the sector itself.[4] In Figures 2a and 3a we can see the spatial distribution of two Spanish manufacturing sectors (22 and 26 respectively) and Figures 2c and 3c exhibit the spatial distribution of one of their corresponding subsectors (2213 and 2630). Here, each dot corresponds to an establishment and, as we can see, there are great differences in the spatial distribution of the aggregated sector and the corresponding subsector. These differences are reflected by the M_{TM} curves of Figures 2b, 2d, 3b and 3d. The M_{TM} value

---

[1] For further details about the tendency of Spanish manufacturing industries to cluster at sector level, see Albert, Casanova and Orts (2011).

[2] Some of the most remarkable examples are the subsectors belonging to the sectors 15, 18, 21, 31 and 33.

[3] Some of these subsectors are those belonging to sectors 17, 19, 22, 24, 26, 29, 35 and 36.

[4] The rest of the cases are available upon request from the authors.

gives us information about the spatial distribution of each spatial point pattern at many and different scales simultaneously, depending on the value of 'r' taken into account. Moreover, it enables one to determine the intensity of concentration or dispersion of each sector or subsector, the distance at which its maximal level is obtained, and the spatial sequence of the increases in the said intensity.

**Figure 2.** Relative location patterns of sector 22 and subsector 2213.

(a)

(b)

(c)

(d)
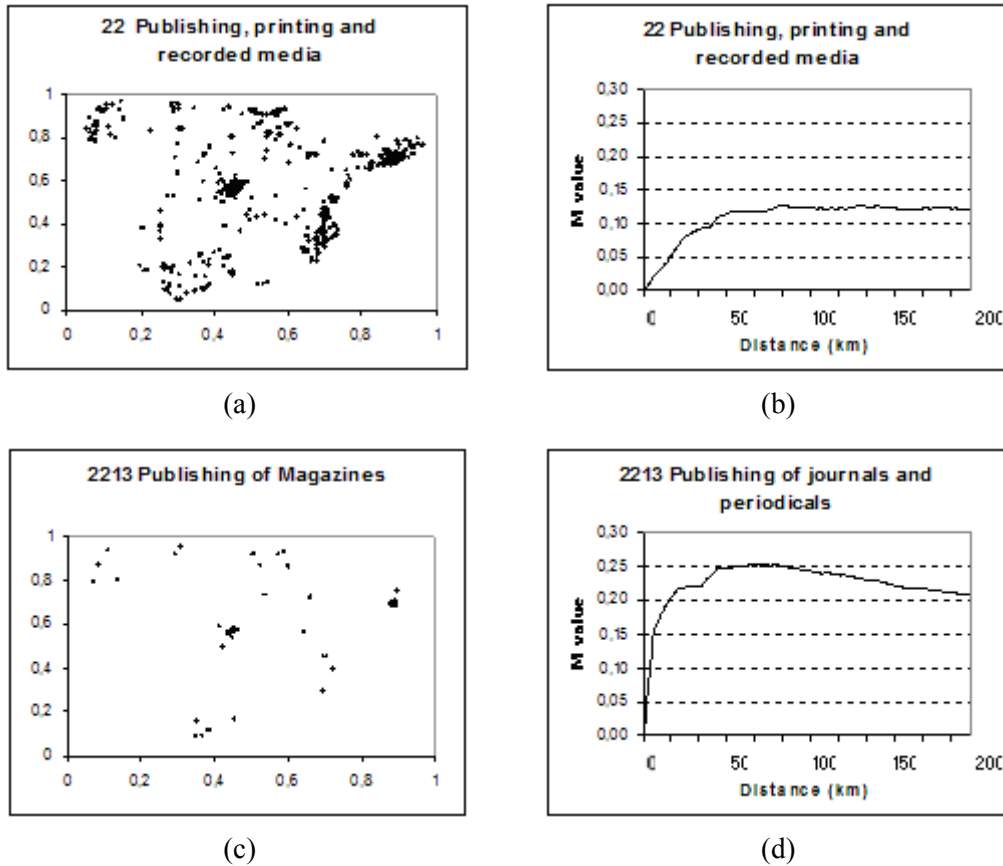
If we compare the spatial distribution of both point patterns of Figure 2a and 2c we can observe at a glance that the subsector 2213 is more concentrated in space, its clusters are more reduced and its establishments are mostly located in Barcelona and Madrid. This information appears reflected in the M_{TM} curve. In fact, the values of M_{TM} increase very fast at a very low length of the radius (r). However, there is not a sudden drop of the values because there are two very distinct and separate clusters of one another; thus, the high concentration reached at a 'small' scale descends slowly.

'Manufacture of ceramic tiles' (2630) is an industry heavily concentrated in the province of Castellón, where we can find the agglomeration of points. A particular location may specialize in a specific activity for two reasons. First, the location might have some underlying characteristic that gives a natural advantage to the activity.

Second, some type of scale economy might be reached by concentrating production at that location. This second reason would be the main cause why the Spanish ceramics is, almost entirely, located in a radius lower than 50km surrounding Castellón. In this respect, Porter (1990) argues that knowledge spillovers[5] in specialized and geographically concentrated industries stimulate growth. He gives as an example the Italian ceramics, in which hundreds of firms are located together. The same could be happening with the Spanish ceramics. If we look at the M_{TM} curve, it shows us that the increase of the M_{TM} value, and thus of the concentration, occurs at very small scales. However, unlike the previous case, this value increases very quickly and afterwards it decreases with the same speed. That is because there is a single cluster and owns the vast majority of the establishments analysed.

Figure 3. Relative location patterns of sector 26 and subsector 2630.

(a)

(b)

(c)

(d)
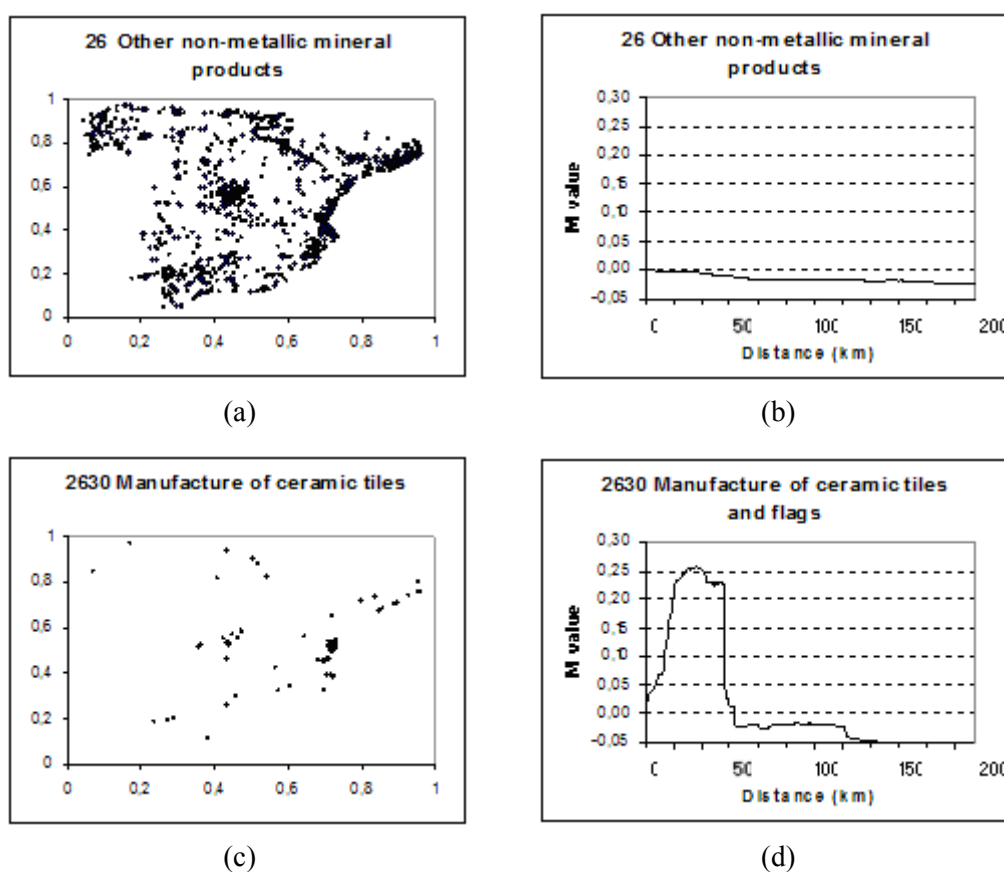
These have been two of the many examples of subsectors that are the real drivers of the location of the aggregate sectors. Some other examples are the subsectors of the industry 17, 24 or 36. These results are available upon request from the authors.

---

[5] Knowledge spillovers are one particular type of positive externalities and their importance has been emphasised, among others, by Barro (1991), Grossman and Helpman (1991) and Henderson et al. (1995).

4. Co-Localization between Spanish Manufacturing Industries

One of the reasons leading to the establishments to concentrate or to co-localise is the reduction of transport costs. Marshall (1890) highlighted that these transports costs could be for goods, labour or ideas. So, establishments can agglomerate and co-localize in order to share some of these three 'forces'. However, the close location among establishments can also be due to natural advantages. Some regions simply possess better environment for certain sectors and these can co-localize because they are attracted to the same natural advantages, even if the sectors would not have interacted through the three aforementioned 'forces'. In this way, it is difficult to ensure if two sectors are localized close to each other because they are attracted by similar characteristics of the area or by similar natural advantage, or because they have strong linkages and have deliberately decided to locate close to each other to exploit synergies.

Our measure can detect the co-localization or the spatial proximity of establishments between subsectors, but cannot detect if this proximity is due to some localized natural advantage[6] or to the deliberated decision to exploit synergies between subsectors[7]. Therefore, we cannot ensure what proportion of the co-localization found can be attributed to natural advantage and what proportion to 'economic reasons', such as linkages, spillovers, scale economies[8] or reduction of transport costs.

In this sub-section, our objective is to analyse the stylised facts to be explained about the co-localization between the Spanish manufacturing subsectors. Thus, since there are thousands of possible pairs of subsectors and it would be very tricky to analyse all of them, we need to focus only on those pairs with linkages; in other words, we must look for criteria of relationship between subsectors in order to analyse possible co-localization. To find the pairs of subsectors with linkages we will use the 1996 input-output table, as previously mentioned, and we will look for both horizontally- and vertically-linked subsectors. Once we have finished the examination of the table we obtain 168 pairs of subsectors, which present a significant percentage of forward or backward linkages and which will be analysed.[9] Moreover, we find an aspect that is striking: the percentage of sectors that present elevated linkages with the sector itself is very high. It means that the subsectors within a sector itself have many linkages

---

[6] This kind of co-localization is denominated 'fortuitous' or 'joint-localization' by Duranton and Overman (2008).

[7] This kind of co-localization is denominated 'colocation' by Duranton and Overman (2008).

[8] In this case we do not refer to the standard concept of 'scale economies' within an enterprise, but to the 'scale economies' that can be achieved when linked activities are located together.

[9] Given the large amount of information generated by this analysis, the results are available upon request from the authors.

between them or, what is the same, the subsectors tend to be self-sufficient in those subsectors that belong to the same sector, in a high percentage.

To carry out the analysis of co-localization we would need a modified K function. The previous analysis considered only the location of one event and now, in order to analyse the co-localization of the Spanish manufacturing sectors, we have to consider a multivariate spatial point pattern. For this, we use a cross-K function, $K_{ij}(r)$, where $i \neq j$ and r is the radius.

$$K_{ij}(r) = (\lambda_i \lambda_j A)^{-1} \sum_k \sum_l w(i_k, j_l) I(d_{i_k, j_l})$$

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

Here, d_{i_k, j_l} is the distance between the kth location of type i and the lth location of type j; $I(x)$ is the indicator function; $\lambda_i = N_i/A$ represents the density of points of type i and $\lambda_j = N_j/A$ represents the density of points of type j, A being the area of the study region, N_i being the total number of points of type i observed in the area of the study region and N_j the total number of points of type j; and $w(i_k, j_l)$ is the weighting factor to correct for border effects, being the fraction of the circumference of a circle centred at the kth location of process i with radius d_{i_k, j_l} that lies inside the area of study. The indicator function, $I(d_{i_k, j_l})$, takes a value of 1 if the distance between k^{th} location of type i and the lth location of type j is lower or equal than r, or 0 otherwise.

If the spatial process is stationary, corresponding pairs of cross-K functions will be equal, i.e. $K_{ij}(r) = K_{ji}(r)$. Besides, under independence between the points of type i and j, the cross-K function $K_{ij}(r)$ will be equal to πr^2 If it appears attraction between both processes at distance r, the difference will be positive, i.e. $K_{ij}(r) > \pi r^2$ and values of $K_{ij}(r)$ lower than πr^2 indicate repulsion between the processes.

The first evidence we obtain when analyse if exists co-localization between the pairs of Spanish manufacturing subsectors is that, at large spatial scales, establishments tend to locate closer to establishments in their own industry than to establishments in vertically-linked industries. Nevertheless, at small distances, the opposite occurs and some establishments decide to locate closer to other industries with which they have linkages than to establishments with which they are horizontally-linked. This result is completely contrary to that obtained by Duranton and Overman (2008) for U.K. manufacturing industries.

19

It depends on the sectors but, in general terms, establishments tend to locate closer to those establishments in one's own industry than to establishments in vertically-linked industries. Clear examples of this are the establishments in sector 17, Textiles, in sector 19, Tanning and dressing of leather, or in sector 24, Chemical and chemical products. In the first two examples, this makes sense, since the establishments of both sectors – especially sector 19 – are mostly located in a specific area and these industries are heavily concentrated. In this way, it might be likely that they deliberately locate close to each other to exploit synergies between them.

The specific case of co-localization of the establishments of sector 19 may be due to the fact that the subsectors have significant interrelations to each other – for instance Tanning and dressing of leather and Manufacture of footwear – and their final output is easy to transport.

According to vertically-linked industries, Ellison *et al.* (2010) obtained that one of the most coagglomerated industry pairs were textiles, and wearing apparel and dressing. We found that the establishments of both sectors (17 and 18) locate close to each other, but they are far of being the most co-localized sectors. Their co-localization occurs at medium distances of 'r'. Finally, only to add that some of the strongest cases of co-localization in vertically-linked industries are the pairs of establishments of sectors 21, Pulp, paper and paper products, and 36, Furniture and other products; and, the pairs of establishments of sectors 24, Chemical and chemical products and 25, Rubber and plastic products. In the first case (sectors 21 and 36), the distance at which the maximum level of co-localization is reached (r) varies depending on the subsectors analysed but, in the second case (sectors 24 and 25), the co-localization is reached at relatively small distances.

Other pairs of subsectors which their establishments co-localized are 2112-2955 and 2121-2955. Here, the subsectors of 'Manufacture of paper and paperboard' (2112) and 'Manufacture of corrugated paper and paperboard' (2121) are respectively coagglomerated with the subsector in charge of manufacturing the specific machinery for their own activities (2955). Finally, just to comment that the establishments of the pairs of subsectors 1930-1822 and 1930-1824 are shyly coagglomerated at short distances, but at large distances they show clear signs of repulsion. This might have two explanations: (1) they are deliberately located close to each other and the cluster where these establishments are is very small, or (2) their co-localization if fortuitous, i.e. they are coagglomerated because they are attracted by similar characteristics of the area or by similar natural advantage.

5. Conclusions

In this paper, we evaluate the spatial location and co-localization patterns of Spanish manufacturing firms for NACE two- and four-digit levels, and assess the different tendencies to cluster in each industry relative to the whole of manufacturing. Furthermore, we extend the analysis to explore if 'small' or 'large' establishments are the main drivers of localization in each sector. To do this, we use a distance-based method, specifically the Ripley's K function, which measures concentration by counting the average number of neighbours of each establishment within a circle of a given radius. This method allows us to treat space as continuous, analysing simultaneously multiple spatial scales and avoiding the shortcomings of the administrative scale. In addition, we employ a polygonal boundary to improve the delimitation of our area of study, substituting the rectangular shape used by other authors and thus avoiding the nuisance of empty spaces.

Our approach, through the use of the whole of manufacturing as a benchmark and the particular way of constructing the confidence intervals –by locating the points simulated in the specifics sites where we can currently find an establishment– let us to fulfil the five requirements that any test which measures concentration should fulfil, as proposed by Duranton and Overman (2005); i.e. our measure (1) is comparable across industries, (2) controls for the overall agglomeration tendency of manufacturing, (3) controls for industrial concentration, (4) is unbiased with respect to scale and aggregation, and (5) gives an indication of the significance of the results.

Our results show that 45% of Spanish manufacturing sectors are concentrated. We do not find a strong regularity in the distance at which the highest concentration of these sectors is reached. However, we find that 'Textiles' and 'Tanning and dressing of leather' sectors are concentrated at lower distances than the rest of sectors and their maximum intensity is reached at a radius of 20 and 30 km respectively. Moreover, we can add that most of the concentrated sectors – around 66% of them – are concentrated at every distance of 'r'.

When we analyse if the proportion of large or small establishments within each sector determines the pattern of location of the sector as a whole, we find that 'small' establishments are the drivers of agglomeration in 'Tanning and dressing of leather' sector and 'large' firms are the drivers of agglomeration in 'Publishing, printing & recorded media', 'Electrical machinery' and 'Radio, televisions & other appliances' sectors; hence, we realise that large firms seem to be the main drivers of agglomeration in high-tech industries.

By means of disaggregation we can observe that some subsectors tend to follow similar location patterns as the aggregated sector to which they belong, whereas other subsectors vary heavily in their spatial distribution.

Finally, when we analyse if exists co-localization between the pairs of Spanish manufacturing subsectors we find that, at large spatial scales, establishments tend to locate closer to establishments in their own industry than to establishments in vertically-linked industries. Nevertheless, at small distances, the opposite occurs and some establishments decide to locate closer to other industries with which they have linkages than to establishments with which they are horizontally-linked. Besides, in general terms, establishments tend to locate closer to those establishments in one's own industry than to establishments in vertically-linked industries.

References

Albert JM, Casanova MR, and Orts V (2011) Spatial Location Patterns of Spanish Manufacturing Firms. *Papers in Regional Science*. Forthcoming.

Alonso-Villar O, Chamorro-Rivas JM, and 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, and González-Cerdeira X (2004) Agglomeration economies in manufacturing industries: the case of Spain. *Applied Economics*, 36, 2103-2116

Amiti, M (1997) Specialisation patterns in Europe. Discussion paper 363, Centre for Economic Performance, London

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.

Audretsch, D. B. and Feldman, M. P. (2004) Knowledge spillovers and the geography of innovation. *The Handbook of Regional and Urban Economics*, IV: 2713-2739.

Brülhart, M., (2001) Evolving geographical concentration of European manufacturing industries. *Weltwirtschaftliches Archiv*, 137 (2), 215-243.

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

Devereux, M.P., Griffith, R. and Simpson, H., (2004) The geographic distribution of production activity in the UK. *Regional Science and Urban Economics*, 34, 533-564.

Duranton, G., Overman, H.G., (2005) Testing for Localization using Micro-Geographic Data. *Review of Economic Studies*, 72, 1077-1106.

Duranton, G., Overman, H.G., (2008) Exploring the detailed location patterns of U.K. manufacturing industries using microgeographic data. *Journal of Regional Science*, 48 (1), 213-243.

Ellison, G., Glaeser, E., (1997) Geographic concentration in U.S. manufacturing industries: a dartboard approach. *Journal of Political Economy*, 105 (5), 889-927.

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

Fujita, M., Krugman P. and Venables A.J., (1999) The Spatial Economy: Cities, Regions and International Trade. MIT Press, Cambridge, MA.

Krugman, P., (1991a) Geography and Trade. MIT Press, Cambridge, USA.

Krugman, P., (1991b) Increasing returns and economic geography. *Journal of Political Economy*, 99 (3), 413-499.

Marcon, E., Puech, F., (2003) Evaluating the Geographic Concentration of Industries using Distance-Based Methods. *Journal of Economic Geography*, 3 (4), 409-428.

Marcon, E., Puech, F., (2010) Measures of the geographic concentration of industries: improving distance-based methods. *Journal of Economic Geography*, 10 (5), 745-762.

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 (5), 575-604.

Ottaviano G, Puga D (1998) Agglomeration in the Global Economy: A Survey of the 'New Economic Geography'. *The World Economy* 21: 707–731

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 (2), 303-334.

Puga, D (2002) European regional policies in light of recent location theories. *Journal of Economic Geography*, 2, 373-406.

Quah D, Simpson H (2003) Spatial cluster empirics. London School of Economics Working Paper Series

Ripley, B.D., (1976) The second-order analysis of stationary point processes. *Journal of Applied Probability* 13: 255-266

Ripley, B.D., (1977) Modelling Spatial Patterns. *Journal of the Royal Statistical Society - Series B (Methodological)* 39: 172-192.

Ripley, B.D., (1979) Test of 'randomness' for spatial patterns. *Journal of the Royal Statistical Society - Series B (Methodological)* 41: 368-374.

Rosenthal, S.S., Strange, W.C., (2001) The Determinants of Agglomeration. *Journal of Urban Economics*, 50 (2), 191-229.

Venables, A.J., (1995) Economic Integration and the Location of Firms. *American Economic Review*, 85(2): 296-300.

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.

Von Thünen, J (1826) Der Isolierte Staad in Beziehung auf Landwirtschaft un NationalÄokonomie. *Hamburg: Perthes*. English translation: The Isolated state, *Oxford*, *Pergammon Press* (1966).

Weber, A. (1909) Ueber den Standort der Industrien. *Täubingen: J.C.B. Mohr*. English translation: *The Theory of the Location of Industries*, Chicago University Press(1929).
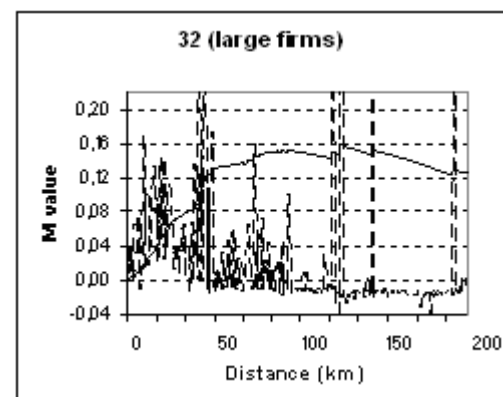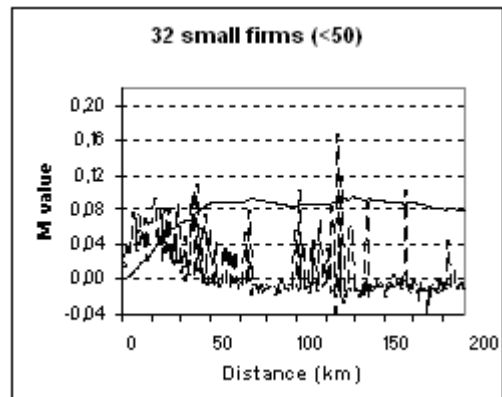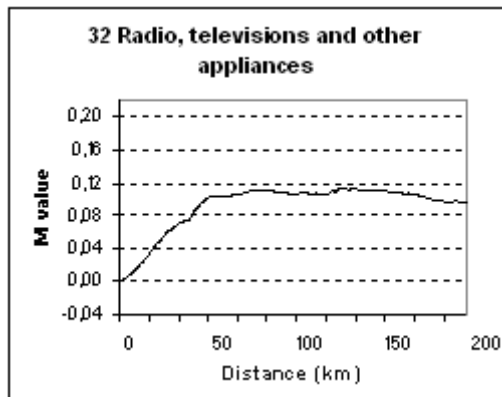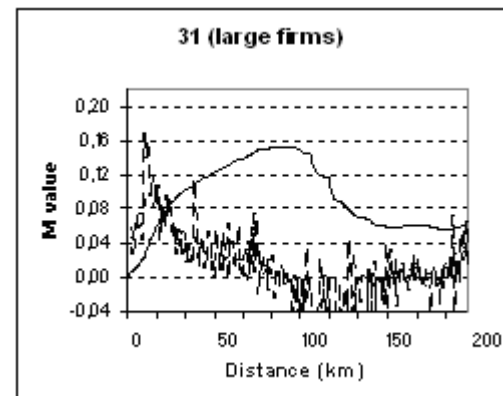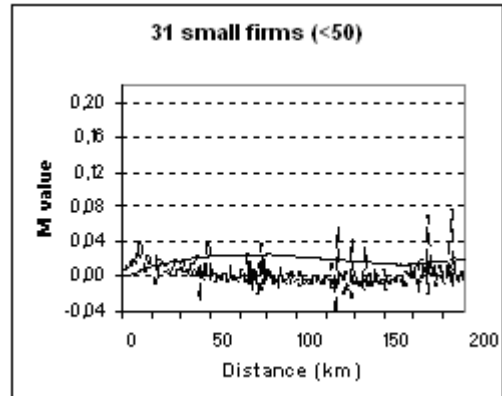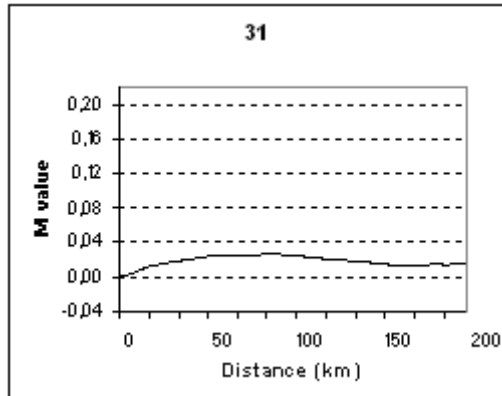
Appendix

Establishment Size

serie
ec
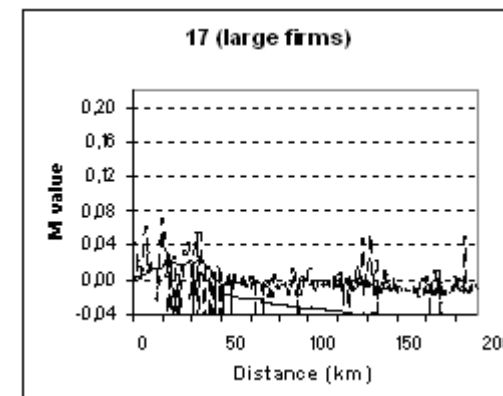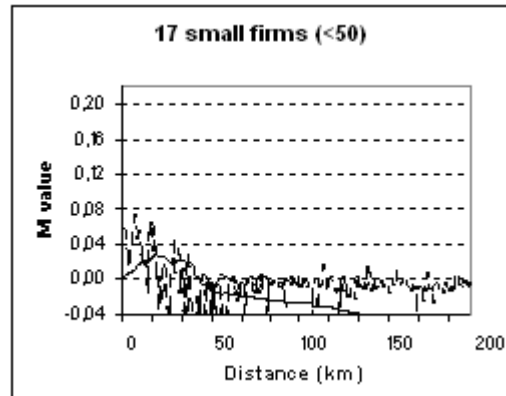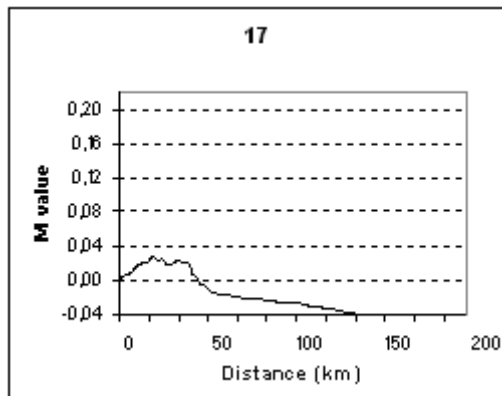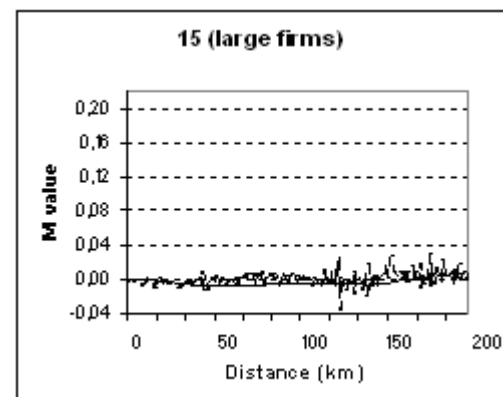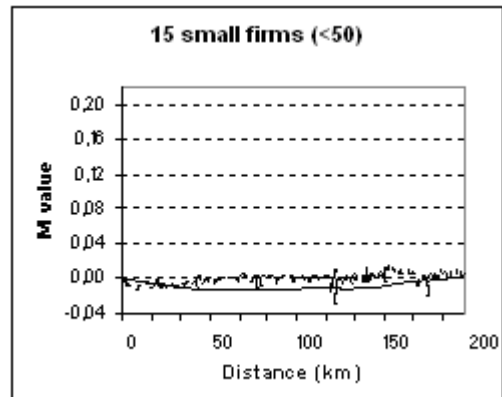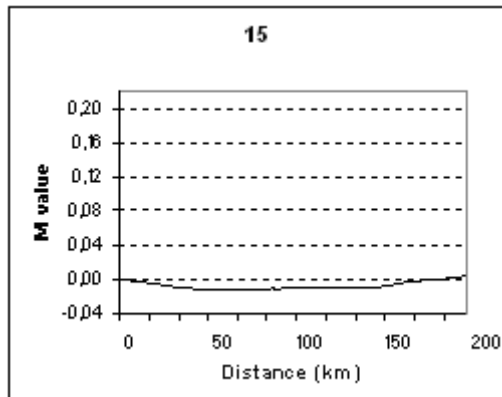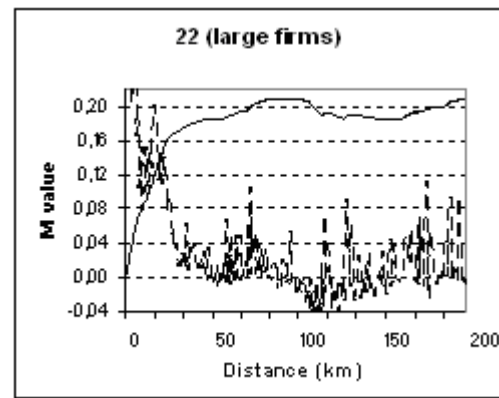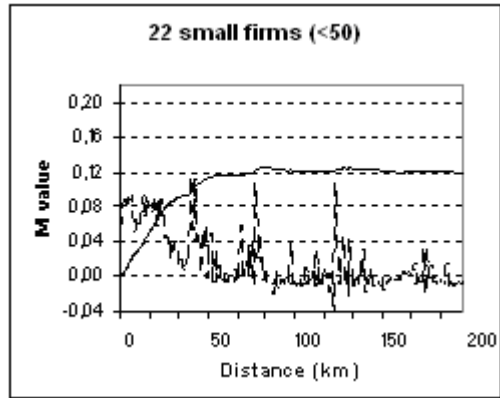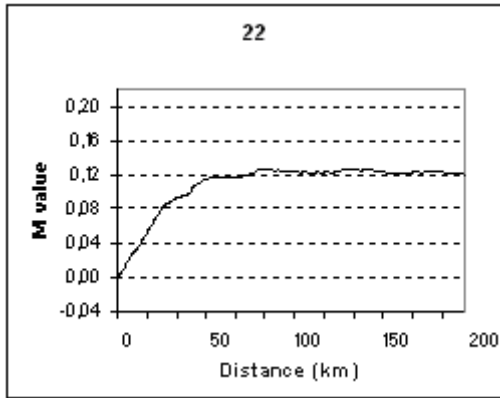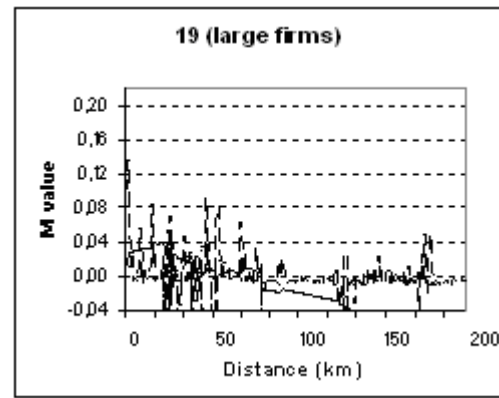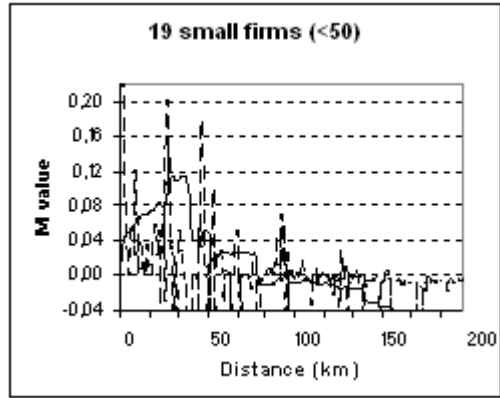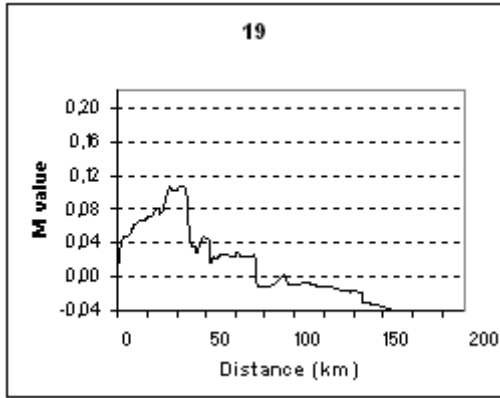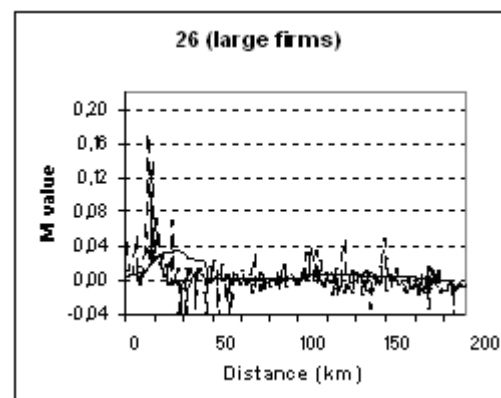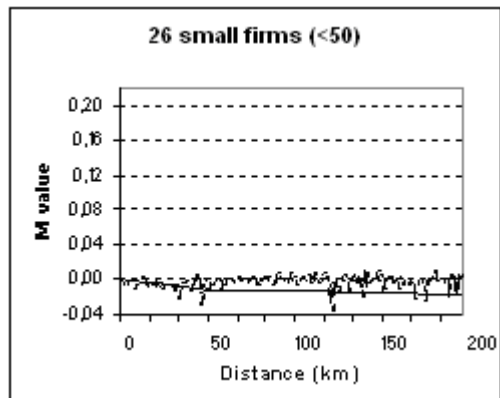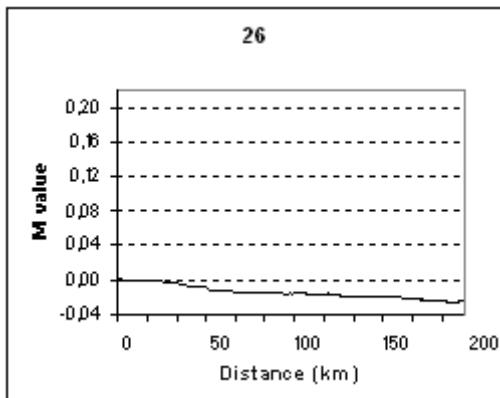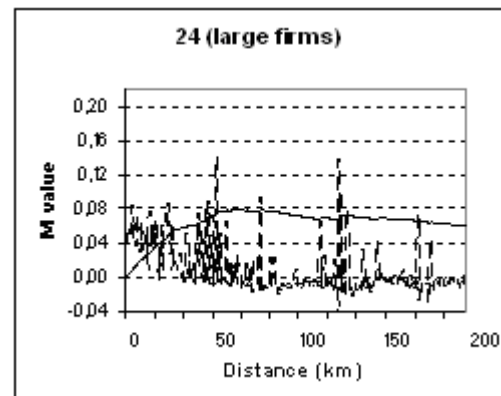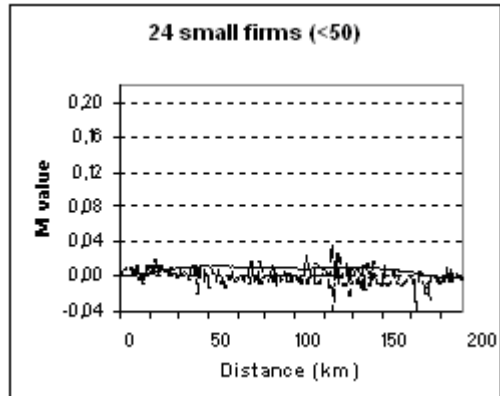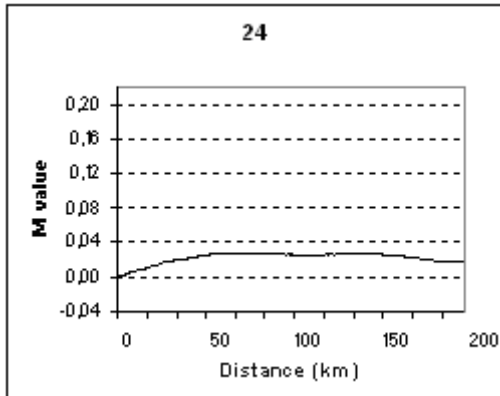
Ivie

Guardia Civil, 22 - Esc. 2, 1º
46020 Valencia - Spain
Phone: +34 963 190 050
Fax: +34 963 190 055

Website: http://www.ivie.es
E-mail: publicaciones@ivie.es