

Casanova, Marta R.; Orts, Vicente

**Conference Paper**

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

50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden

**Provided in Cooperation with:**

European Regional Science Association (ERSA)

*Suggested Citation:* Casanova, Marta R.; Orts, Vicente (2010) : Assessing the Tendency of Spanish Manufacturing Industries to Cluster: Co-localization and Establishment Size, 50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<http://hdl.handle.net/10419/119154>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

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

Marta R. Casanova and Vicente Orts<sup>†</sup>

**Preliminary Version.** Please do not quote without permission.

## Abstract

In this paper, we evaluate the spatial location patterns of Spanish manufacturing firms and we assess the different tendencies to cluster in each industry relative to the whole of manufacturing. To do this, we use a distance-based method (Marcon and Puech, 2003; Duranton and Overman, 2005), more concretely the Ripley's  $K$  function, which measures concentration by counting the average number of neighbours of each firm 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 envelope to improve the delimitation of our area of study, substituting the rectangular shape used by other authors and in our previous paper and thus avoiding the nuisance of empty spaces.

We apply this method to Spanish manufacturing sectors at two-digit and four-digit level, isolating like this the different behaviours of spatial distribution of each subsector caused by 'spillovers' characteristic of each activity and also preventing compensation effects due to previous aggregation. Furthermore, we examine the co-localization between horizontally-linked and vertically-linked industries to assess the importance of these spillovers across industries and, finally, we try to answer what type of establishment, depending on its size, is the driver of the Spanish industrial agglomeration.

**Keywords:** distance-based method, Ripley's  $K$  function, polygonal envelope, disaggregation, co-localization, establishment size.

**JEL classification:** C15, C40, C60, R12.

---

<sup>†</sup> Department of Economics, Universitat Jaume I.

Contact: Marta R. Casanova, Department of Economics, Universitat Jaume I, Avda. Sos Baynat s/n, 12071 Castellón, (mroig@eco.uji.es).

## Introduction

We know that the spatial distribution of economic activity presents heterogeneity, as confirmed by the literature that analyzes the characteristics of the location of such activity and its causes.<sup>1</sup>

When economic activity is distributed across space does not take into account the administrative boundaries. However many authors<sup>2</sup> insist on employing techniques that use administrative units to measure geographic distribution of activity, i.e. Gini or Ellison and Glaeser (1997) indices. In this way, the space is treated as being discrete and the analysis is restricted to just one scale. Otherwise, the treatment of the space as continuous and the use of multiple scales to analyse the distribution of activity can allow us to know and to compare the concentration intensity for every spatial scale<sup>3</sup> and can avoid the shortcomings of the administrative scale.

Results from our previous paper lead us to believe that the location patterns observed are brought about mostly by the characteristics of each sector. Besides, these patterns do not always correspond with the assumptions of the economic theory, since we obtained that the most highly concentrated sectors are both traditional and high-tech industries. Nevertheless, the aggregation in our previous analysis prevents us from knowing whether all subsectors within the same sector would follow similar location patterns. In this way, by means of the disaggregation carried out in this paper we can achieve to isolate the different behaviours of spatial distribution of each subsector caused by 'spillovers' characteristic of each activity.

Marshall (1890) emphasized the role of the spillovers, or positive externalities, and the linkages as sources of agglomeration. As well 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 the disaggregation carried out we will see if there is interaction and interdependence between the different subsectors when they locate in space. In other words, we will check if there is co-localization between horizontally-linked and vertically-linked

---

<sup>1</sup> See Fujita *et al.* (1999), Krugman (1991a, 1991b), Marshall (1890), Ottaviano and Puga (1998), Puga (1999, 2002), Venables (1995).

<sup>2</sup> For further details see Krugman (1991a), Brülhart (2001), Amiti (1997), Devereux *et al.* (2004), Rosenthal and Strange (2001), Maurel and Sédillot (1999) or Callejón (1997).

<sup>3</sup> Marcon and Puech (2003a) and Duranton and Overman (2005) also avoided the inconvenience of geographic scale in their papers.

subsectors/industries. Thereby, we can assess the importance of the spillovers across subsectors. Finally, we will try to find the location decisions of the establishments depending on its size, that is to say, what type of establishment is the driver of the Spanish industrial agglomeration, given that in the Spanish manufacturing abounds the small enterprises and the establishments are often family businesses with few employees.

Finally, we should add that in our previous paper we minimized the drawbacks associated to the use of a rectangular area when using as a benchmark the whole of manufacturing. Now, in this paper, these drawbacks disappear completely because we improve the application. It is got thanks to the use of a polygonal boundary. In this way, we incorporate an innovative technique to improve the delimitation of our area of study, by replacing the rectangular shape by a polygonal. The rectangle as the area of study is used in our previous paper and by other authors<sup>4</sup>. However, the software employed in our case 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. Besides, this incorporation provides robustness to our results.

The distance-based method we use to measure the spatial distribution of activity is Ripley's  $K$  function, as do Marcon and Puech (2003a). This measure enables us to know whether concentration exists, what its intensity is and at what distance, or spatial scale, its highest level is obtained. The continuity of the space is reached by means of the availability of the geographic coordinates (longitude and latitude) of every Spanish manufacturing firm. Like this, we locate the firms, represented by dots, accurately in space without taking administrative borders into account.

For the moment, using the locations of all manufacturing firms as our benchmark and controlling for the overall tendency of manufacturing to agglomerate, we have found that patterns of location are quite different across sectors, with dispersion in some of them and concentration in others. Such differences can be due to variations in the scope of the forces at play in each sector and this also affects the spatial scale at which clustering occurs. According to the drivers of concentration, depending on their size, our results show that both, large and small, establishments can be the drivers of aggregation; it all depends on the sector to which

---

<sup>4</sup> For instance, Marcon and Puech (2003) limited their empirical analysis to rectangular areas because they claimed that '*complexity depends on the shape of the area under study*'.

we refer. Although, we realise that large firms are often the main drivers of agglomeration in high-tech industries.

## **Data**

We use establishment level data from the Analysis System of Iberian Balances database<sup>5</sup> to carry our empirical analysis out. Our database contains Spanish manufacturing establishments at four-digit level, which are classified using the National Classification of Economic Activities<sup>6</sup>. 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 firms 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 firms with less than ten workers do not have essential information (geographical coordinates) to carry our analysis out.

In Table 1 we find additional descriptive information about each sector, as the whole number of firms of each one, their technological intensity or the proportion of establishments that each sector has, depending on their size. This table indicates that the low-tech sectors have a greater percentage of small establishments. The small enterprises in Spain represent, approximately, an average rate of 80%, considering less than 50 workers, and an average rate of 45% if we consider firms with less than 20 workers.

## **Methodology**

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.

---

<sup>5</sup> SABI

<sup>6</sup> NACE 93 - Rev. 1

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  $i^{\text{th}}$  and  $j^{\text{th}}$  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 the area of the study region; and  $w_{ij}$  is the weighting factor to correct for border effects.<sup>7</sup> The indicator function,  $I(d_{ij})$ , takes a value of 1 if the distance between the  $i^{\text{th}}$  and  $j^{\text{th}}$  firms is lower 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.

Finally, using the definition of  $\lambda$ , 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 ( $\lambda$ ).

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)*.

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. 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.

---

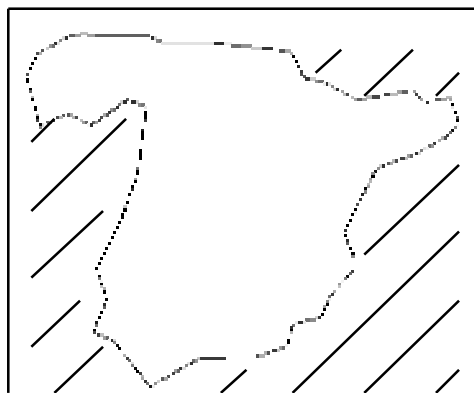
<sup>7</sup> These border-effect corrections should be incorporated to avoid artificial decreases in  $K(r)$  when  $r$  increases, because the increase in the area of the circle under consideration is not followed by the increase of firms (outside the study area there are no firms).

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.

Regarding the area of study, 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. Their reason for not using the whole country was the increasing complexity when simulating random points inside the area and when correcting the border-effects on convex shapes.<sup>8</sup> These drawbacks thus limited their empirical analysis to rectangular areas. Nevertheless, Duranton and Overman (2005) did not have this shortcoming, since border-effect corrections were not necessary in their test and this simplified their empirical analysis. In our analysis, we improve the delimitation of the area of study used by other authors by substituting the rectangular shape for a polygonal shape. In fact, it would be unproductive to analyse a rectangular area because the Spanish territory is not regular like a rectangle. Moreover, the statistical software employed, 'R',<sup>9</sup> allows border corrections to be applied adequately to any irregular polygonal shape, thereby simplifying the treatment of border effects.



<sup>8</sup> They explicitly said, "It was impossible to use the whole of France because of border-effect corrections".

<sup>9</sup> This software is downloadable from the following website: <http://www.r-project.org/>.

**Figure 5.** Polygonal shape envelope.

Here, 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 firms are found, which are represented by the oblique lines. This polygonal envelope was built by joining thirty-five points on the perimeter of the Spanish territory.

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 firms as in each of the sectors under consideration, but the location of these hypothetical firms is restricted to the sites where we can currently find firms from the whole manufacturing sector. It is 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 assess the significance of departures from randomness and to control for industrial concentration.

## **Results**

Our analysis is at an initial stage. Therefore our results are very early and our conclusions are very general. We have already introduced the 'polygonal envelope' into our research and we have begun to differentiate according the size of the firms. However, the fact of how the sectors, or the subsectors, are co-localized is still pending.

By introducing the polygonal envelope our study area is reduced, as we believe only that area where economic activity can be found. Thus, by being reduced the area of study, the value of  $K$  of each sector is also reduced, as we can see in Table 2. In other words, it is obvious that companies can not be located in sea areas, but in our previous work, when we used a rectangular area intrinsically we assumed that companies could be located anywhere within this rectangle.

### **INSERT TABLE 2 AROUND HERE**

Taking into account, now, the size of the firms, we should note that in the Spanish manufacturing abounds the small enterprises, thus we should make a detailed analysis of them. Therefore, although the theory tells us that we must consider as 'small establishments' those businesses with fewer than 50 employees, in order to analyze small businesses we have



proposed two parallel dataset and we have carried out two parallel analysis. In the first one we take into account establishments with less than 20 workers and in the second one with less than 50 workers. By this, we want to check if the location patterns vary depending on whether we regard as 'small establishment' that ones with less than 20, or 50 workers. In the following table, we can see the different  $M_{TM}$  values for the whole sector and for every subset of firms.

Sector	Significant peak ( $M_{TM}$ value)			
	all firms	<20	<50	>50
15	-0,01	-0,01	-0,01	-0,01
17	0,03	0,03	0,03	0,02
18	---	---	---	---
19	0,11	0,11	0,12	0,04
20	---			
21	0,04			
22	0,13	0,12	0,13	0,21
24	0,03	0,10	0,01	0,08
25	---			
26	-0,03	-0,02	-0,02	0,03
27	---			
28	0,07			
29	-0,03			
31	0,03	0,02	0,03	0,15
32	0,12	0,10	0,09	0,16
33	0,07			
34	---			
35	---	---	---	0,05
36	---	---	---	---
37	---			

- > Sector 19: small firms are the drivers of agglomeration
- > Sector 22: large firms are the drivers of agglomeration
- > Sector 24: large firms are the drivers of agglomeration
- > Sector 31: large firms are the drivers of agglomeration
- > Sector 32: large firms are the drivers of agglomeration

According to the drivers of concentration, depending on their size, our results show that both, large and small, establishments can be the drivers of aggregation; it all depends on the sector to which we refer. Although, we realise that large firms are often the main drivers of agglomeration in high-tech industries. On the other side, sector 19 has a different behaviour; in fact, its concentration is clearly due to small establishments. In this way, we can say that small firms are the drivers of agglomeration in this sector.

## References

- Amiti, M., (1997) Specialisation patterns in Europe. Discussion paper 363, Centre for Economic Performance, London
- 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.
- Brühlhart, 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.
- Ellison, G., Glaeser, E., (1997) Geographic concentration in U.S. manufacturing industries: a dartboard approach. *Journal of Political Economy*, 105 (5), 889-927.
- 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. and Puech, F., (2003a) Evaluating the Geographic Concentration of Industries using Distance-Based Methods. *Journal of Economic Geography*, 3 (4), 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 (5), 575-604.
- Ottaviano, G.I.P., Puga, D., (1998) Agglomeration in the Global Economy: A Survey of the 'New Economic Geography'. *The World Economy*
- 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.
- 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.

**Table 1. Additional descriptive information**

Sector	Technological intensity	Number of firms	<20		<50		>50		>250	
15	<i>L</i>	5778	2618	45,31%	4620	79,96%	1158	20,04%	220	3,81%
17	<i>L</i>	1951	944	48,39%	1620	83,03%	331	16,97%	36	1,85%
18	<i>L</i>	1712	902	52,69%	1508	88,08%	204	11,92%	19	1,11%
19	<i>L</i>	1699	894	52,62%	1565	92,11%	134	7,89%	7	0,41%
20	<i>L</i>	2349	1307	55,64%	2124	90,42%	225	9,58%	19	0,81%
21	<i>L</i>	837	295	35,24%	619	73,95%	218	26,05%	35	4,18%
22	<i>L</i>	2997	1565	52,22%	2554	85,22%	443	14,78%	58	1,94%
24	<i>H</i>	1722	566	32,87%	1162	67,48%	560	32,52%	143	8,30%
25	<i>M-L</i>	2165	875	40,42%	1690	78,06%	475	21,94%	54	2,49%
26	<i>M-L</i>	3429	1464	42,69%	2738	79,85%	691	20,15%	89	2,60%
27	<i>M-L</i>	987	397	40,22%	712	72,14%	275	27,86%	80	8,11%
28	<i>M-L</i>	8103	4297	53,03%	7104	87,67%	999	12,33%	90	1,11%
29	<i>M-H</i>	3018	1281	42,45%	2458	81,44%	560	18,56%	70	2,32%
31	<i>M-H</i>	1099	445	40,49%	809	73,61%	290	26,39%	64	5,82%
32	<i>H</i>	344	120	34,88%	240	69,77%	104	30,23%	29	8,43%
33	<i>H</i>	376	180	47,87%	302	80,32%	74	19,68%	13	3,46%
34	<i>M-H</i>	876	244	27,85%	521	59,47%	355	40,53%	134	15,30%
35	<i>M-H</i>	451	144	31,93%	313	69,40%	138	30,60%	31	6,87%
36	<i>L</i>	2927	1531	52,31%	2576	88,01%	351	11,99%	30	1,02%
37	<i>L</i>	228	125	54,82%	199	87,28%	29	12,72%	3	1,32%

**Table 2. Comparison of results ‘with’ or ‘without’ polygonal envelope**

Sectors (NACE 93 - Rev. 1)	Significant peak ( $M_{TM}$ value)	
	without polygonal envelope	with polygonal envelope
15 Food products and beverages	-0,04	-0,01
17 Textiles	0,18	0,03
18 Wearing apparel and dressing	---	---
19 Tanning and dressing of leather	0,45	0,11
20 Wood and products of wood	-0,03	---
21 Pulp, paper and paper products	0,07	0,04
22 Publishing, printing & recorded media	0,12	0,13
24 Chemical and chemical products	0,08	0,03
25 Rubber and plastic products	0,04	---
26 Other non-metallic mineral products	-0,03	-0,03
27 Basic metals	0,02	---
28 Fabricated metal products	0,01	0,07
29 Other machinery and equipment	0,04	-0,03
31 Electrical machinery	0,08	0,03
32 Radio, televisions & other appliances	0,13	0,12
33 Instruments	0,12	0,07
34 Motor vehicles and trailers	0,02	---
35 Other transport equipment	0,02	---
36 Furniture and other products	0,06	---
37 Recycling	---	---

# Appendix 1

## Establishment Size

