XV Applied Economics Meeting
A Coruña, 7th-8th June 2012

Intra-metropolitan clustering of formal and informal manufacturing activity: evidence from Cali, Colombia

Gustavo A. García*
Universitat Autònoma de Barcelona
gustavoadolfo.garcia@uab.cat
Department of Applied Economics, Campus Bellaterra,
Edificio B, 08193 Bellaterra (Cerdanyola), Spain. Phone: 935811153 – Fax: 935812292

Ana I. Moreno-Monroy
University of Groningen
A.I.Moreno.Monroy@rug.nl
Faculty of Economics and Business, Nettelbosje 2, 9747 AE Groningen,
The Netherlands. Phone: 50 363 7190 – Fax: 50 363 3716

Preliminary version (February 2012) – Please do not quote

Abstract

Informal manufacturing activity is substantial in developing countries. However, because existing studies on intra-metropolitan clustering for developing countries are based solely on data for formal enterprises, little is known regarding the clustering behavior of informal enterprises. In this article we use data from the Economic Census of the metropolitan area of Cali, which includes the universe of manufacturing formal and informal enterprises, in two complementary ways. First, we calculate the degree of spatial clustering and co-clustering by means of M-functions, which treat space as continuous. Because the resulting M-indices have a straightforward interpretation and are fully comparable across sectors, we are able to compare the geographical clustering and co-clustering of both the aggregates of Formal Large, Formal Small and Medium and Informal Enterprises, and for each of these types in selected ISIC 3-digit industries. Second, we conduct spatial analysis on the distribution of formal and informal enterprises. Given that the location of each enterprise is known at the zip-code level, we are able to visualize the spatial behavior of formal and informal enterprises. In particular, we perform kernel density estimations and present cartographic representations of the geographical distribution of formal and informal enterprises in selected clustered and co-clustered industries. We find that while formal and informal enterprises display a tendency to cluster in roughly the same industries, these clusters are not necessarily found on the same areas of the city. In fact, we observe significant co-clustering of enterprises operating in complementary industries, but this co-clustering only occurs between enterprises of the same type. Thus, while there is a common logic behind clustering of small enterprises, it seems to operate in parallel in two different parts of the city.

Key words: informal enterprises, clustering, developing country, distance-based methods, M function, spatial analysis
***JEL*:** C40, L60, R12

---

* G.A. García acknowledges financial support from the Ministerio de Ciencia e Innovación, ECO2010-20718. Any mistake or omission remains ours.

1. INTRODUCTION

The share of employment absorbed by the informal manufacturing sector in developing countries is substantial.[1] In fact, in some developing countries, informal manufacturing employment accounts for the largest share of urban manufacturing employment (WTO and ILO, 2005; Deichmann et al, 2008). However, when analyzing the spatial clustering of manufacturing activity and its potential benefits, most theories and empirical studies deal with formal (large) manufacturing enterprises and do not include information on informal (micro and small) enterprises (Wu, 1999; Chakravorty et al, 2005).

Within this context, relevant questions such as how the inter-relations of formal and informal manufacturing activities shape the urban landscape, or where is informal manufacturing activity expected to take place within metropolitan areas in developing countries remain unanswered. Providing an answer to these questions is not an easy matter because the very nature of informal activity makes it particularly hard to come by detailed data. Perhaps for this reason, to date there is no empirical studies on the spatial logic of informal manufacturing activity within cities in developing countries, let alone on the determinants of such spatial logic.

The present empirical study compares the spatial clustering patterns of formal and informal manufacturing activity within a metropolitan area of a developing country. We use census manufacturing enterprise-level data for the metropolitan area of Cali for the year 2005. Cali is the third largest city in Colombia, with over two million inhabitants. This unique dataset includes the universe of manufacturing enterprises geocoded up to a disaggregated scale equivalent to the "city block" level. Several criteria in the census allow us to identify the formality status of the enterprise, so that we are able to discriminate between "formal" or "informal" manufacturing enterprises in each 3-digit ISIC industrial sector. Furthermore, because information on the number of employees per enterprise is also provided, we are able to disentangle formal enterprises into Formal Small and Medium Enterprises (FSMEs), and Formal Large Enterprises (FLEs). This division is useful in distinguishing the location effects explained by size from those related to the formality status of an enterprise.

[1] Although there is no comprehensive cross-country data for the manufacturing sector, the numbers for the informal sector at large are telling. The share of employment absorbed by the informal sector was on average over 70% in Sub-Saharan Africa, 50% in Latin America, 47% in East Asia and North Africa and 24% in transition economies (Jütting and Laiglesia, 2009).

This paper explores two particular questions: first, do formal and informal enterprises display different clustering patterns? And second, do formal and informal enterprises locate on the same areas of the city, or is there a marked spatial segmentation between formal and informal manufacturing activity? These questions are quite relevant from a policy perspective. First, as pointed out by Duranton (2008), in order to align policies with the reality of developing countries, more evidence is needed on the potential of informal manufacturing activity to generate production externalities. Exploring the actual clustering patterns of informal enterprises is the first building block in understanding what kind of externalities may arise from the substantial presence of informal manufacturing enterprises in urban centers.

Second, policy designs should take into account the fact that the relative location of informal enterprises bears different consequences. On the one hand, if informal locate near similar formal enterprises in central areas, additional locational competition and congestion costs are generated which impact the location decisions of formal producers. On the other hand, if informal enterprises operate mostly in peripheral areas, they are cut off from accessing the best final and intermediate consumer markets available in the city. From the point of view of informal enterprises, this can lead to lower possibilities of establishing linkages with formal enterprises, and fewer chances for growth and survival.

Thus, in order to answer these questions, we use two complementary approaches. First, we calculate the degree of spatial clustering and co-clustering by means of M-functions (Marcon and Puech, 2010). This distance-based method allows us to treat space as continuous. Clustering (co-clustering) at a certain distance in a *sector* is measured as the ratio of two ratios. The numerator is the sum of the number of employees in enterprises of the same sector within a certain distance (e.g., a circle of 1Km radius) as a proportion of the sum of the number of employees in enterprises of all other sectors (in another sector) within the same distance. The denominator is the proportion of employment in the sector in total employment in the whole area (in our case the city of Cali). A *sector* can refer to either an ISIC 3-digit industry, a type of enterprise (FLEs, FSMEs or Informal Enterprises -IEs-) or both (e.g., IEs in the textile industry). This exercise is repeated for different distances to obtain continuous measures of clustering and co-clustering, or M-indices. Given that the M-indices have a straightforward interpretation and are fully comparable across sectors, we are able to compare the geographical clustering and co-clustering of both the aggregates of FLEs, FSMEs and IEs, and for each of these types of enterprises in ISIC 3-digit industries. When analyzing co-clustering, we focus our analysis on seven industries that are clustered and have a significant number of employees and enterprises.

Second, in order to complement the results obtained from the M-indices, we analyze the geographical distribution of formal and informal enterprises in different clustered and co-clustered industries. In particular, we perform a kernel density analysis on clustered industries and analyze the geographical distribution of formal and informal enterprises of selected clustered and co-clustered industries.

After this introduction, Section 2 proceeds with a literature review on the intra-metropolitan locational patterns of informal manufacturing enterprises. Section 3 describes the area of study and data used in the empirical analysis, and presents some preliminary results on the spatial distribution of formal and informal enterprises. Section 4 describes the definitions of the M-functions of clustering and co-clustering as well as the methodology used in the spatial analysis. Section 5 presents the results and Section 6 concludes.

2. THEORETICAL PREDICTIONS ON THE INTRA-METROPOLITAN CLUSTERING OF INFORMAL MANUFACTURING ACTIVITY

For the purpose of this paper, informal enterprises are defined as those enterprises producing legal goods that do not fully comply with established legal regulations.[2] A typical informal enterprise, besides being (very) small in size, faces capital restrictions and operates in highly competitive markets with very low entry costs. In the existing literature it is challenging to find studies specifically addressing the location of informal manufacturing activity (or all informal activity, for that matter) in the urban space.

A first approximation to the problem is to derive the intra-metropolitan locational patterns of informal enterprises from their linkage costs. According to Scott (1988), linkage costs rise with the quantity of goods that are traded between enterprises and decline the more standardize the goods are, the more stable interactions are, and the less need there is for intermediation.[3] In this way, large

[2] A definition specific for the case of Colombia will be introduced in the next section.
[3] For instance, an enterprise that has built a stable and trusted relationship with an input supplier will face lower linkage costs because there is an agreed degree of quality and product specifications of the good involved, setup costs need not to be incurred every time a transaction takes place and orders can be placed directly without the need of further intermediation.

enterprises are expected to locate far away from central, congested locations, given that their linkages are large in scale, standardized, stable, and more easily manageable. On the contrary, small enterprises are expected to cluster near their partners in transactions because their linkages are small in scale, unstandardized, unstable, and in need of personal intermediation.[4] Because of their characteristics, the linkage costs of informal enterprises are probably high, so that the tendency to seek proximity would also apply to informal producers, if not to a higher degree. Thus, from this perspective, both formal small and informal enterprises are expected to agglomerate in locations with dense networks of suppliers and consumers, while large (formal) enterprises are expected to locate outside these central areas.

This prediction, however, ignores the fact that informal enterprises are likely to operate in "rent free" locations (such as household premises), and outside industrial and commercial areas. There are four possible reasons for this. First, informal enterprises may have a lower intrinsic value attached to the quality of premises when compared to their formal counterparts (Sethuraman, 1997). Second, informal enterprises may be unable to bid for rents vis-à-vis formal enterprises. If location is indeed a determinant of performance, informal enterprises are marginalized from the "best" markets in the city, which renders them less competitive than their formal counterparts and effectively excludes them from competitive locations (Daniels, 2004). Third, the segment of the population catered by informal manufacturing enterprises may be concentrated precisely in peripheral areas, so that the proximity of informal producers to their consumers follows suit with the same logic of transaction costs minimization. Four, informal enterprises operating on a subcontracting basis may consider proximity to input suppliers and consumers irrelevant and thus prefer to locate in residential areas of the city with lower rent prices, and in the case of home-based enterprises, in areas where informal producers reside.[5] This may be the case because formal enterprises directly provide materials and inputs to informal enterprises (commonly through an intermediary) that transform them though a labor-intensive process and send them back to the formal enterprise (Carr et al, 2000), or

[4] Empirical evidence has confirmed that the relatively larger tendency of small enterprises to cluster compared to that of large enterprises is connected to the larger sensitivity of small enterprises to final and intermediate markets accessibility (Rosenthal and Strange, 2009; Lafourcade and Mion, 2007).

[5] In very large cities and when very close proximity is required for subcontracting to take place, formal manufacturing enterprises are also expected to relocate towards areas of the city where cheap labor is available (Scott, 1988; Holl, 2008).

because energy costs, which are paid by the informal producer, are subsidized in peripheral areas.

It can also be argued that if the same supply-side mechanisms behind clustering apply for the case of informal enterprises, and these can actually explain the location of enterprises, the clustering patterns of formal and informal enterprises of similar sizes should not differ. The literature has recognized three of such channels (Duranton and Puga, 2004): 1) sharing, or the gains derived from sharing inputs produced under increasing returns to scale; 2) matching, or the gains derived from accessing a larger pool of workers; and 3) learning, or the gains derived from knowledge spillovers. It can also be argued, however, that informal enterprises located in peripheral areas may be accessing sub-markets for intermediate inputs and labor in their own peripheral area, and may be subject to limited or no knowledge spillovers if they are surrounded mostly by similarly labor-intensive, low technology enterprises (Moreno-Monroy, forthcoming). In this case, formal and informal enterprises would not necessarily locate on the same areas of the city.

3. DESCRIPTION OF THE STUDY AREA AND DATA

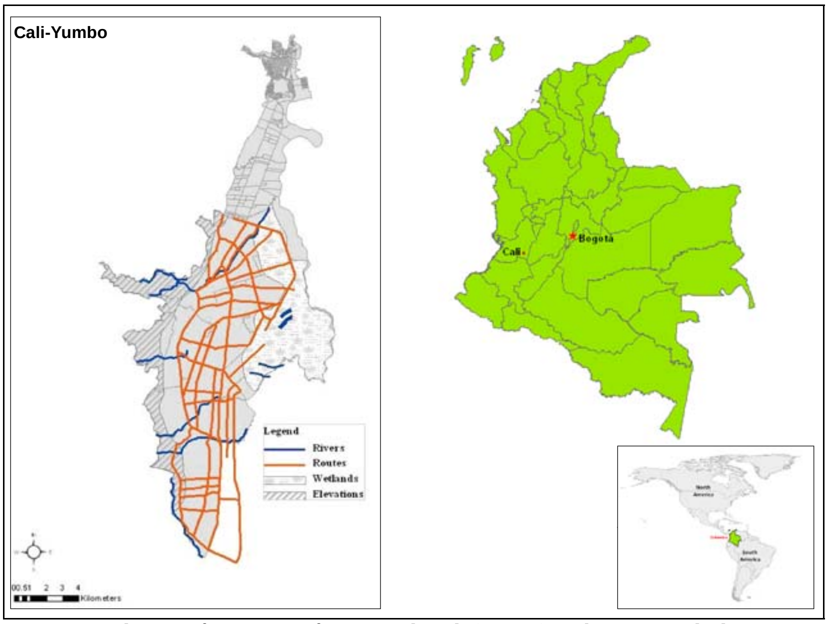
3.1. Study Area: The metropolitan area of Cali

Santiago de Cali (3°27' lat. North - 76°31' long. West), capital of the Valle del Cauca region (*departamento)* and third city in importance of Colombia, is located in the West of Colombia. The metropolitan area of Cali is composed of the municipalities of Cali and Yumbo and is divided into 26 communes, 279 neighborhoods and over 14,000 "*Manzanas*", of which 1,125 (4%) are green areas or have no information available. This last scale is comparable with the "census block" level used in the USA Census.[6] In 2005, the metropolitan area of Cali had a population of 2,164,098 people, and the average density was 17,217 people per sq. km. Cali-Yumbo stretches over 33.6km (of which 23.5km correspond to the city of Cali excluding Yumbo) from North to South and 15.7km from East to West.

Cali is a relatively flat area which registers an average elevation of 1,000 amsl, with the highest elevations in the Western part of the city and the lowest and wetlands on the East (see Figure 1). The main transportation routes, starting from the center of the city, describe a concentric and radial pattern towards the East of the city.

[6] For a definition see http://www.census.gov/geo/www/geo_defn.html#CensusBlock.

Figure 1. Study Area



Source: Spatial Data Infrastructure of Santiago de Cali (IDESC); authors' own calculations.

The largest and most densely populated areas of the city proper are on the Center and East of the city (see Figure A1 in the Appendix). These densely populated areas coincide with areas of population predominantly in low and low-middle income categories (see Figure A2 in the Appendix). On the contrary, the wealthiest areas of the city proper, located predominantly in the South and North-West of the city, are less densely populated.

<u>3.2. Data</u>

We use comprehensive enterprise-level data from the Economic Census of the city of Cali and the municipality of Yumbo, carried out by the National Department of Statistics (DANE) for the year 2005. This database contains detailed information per enterprise, including employment, economic sector, social security contributions and other legal requirements and geographical location at the block level. Because blocks are small geographical units (approximately 110x110 meters for Cali, see Table A1 in the appendix), and there are on average 3.7 enterprises per block, our information is approximately equivalent to having the actual location of each enterprise.

For our analysis we consider only manufacturing enterprises. According to the Economic Census data, the *universe* of manufacturing enterprises in Cali-Yumbo is 5,130. Of this total, we exclude those enterprises that could not be located geographically and those that do not operate on a fixed location. These exclusions led to a sample of 4,862 enterprises of which 95% are located in Cali and the remaining 5% are located in Yumbo. The information is available at an industrial disaggregation of ISIC 2, 3 and 4-digits.

Informal enterprises are defined as those enterprises not registered in the Chamber of Commerce. In Colombia, registration in the Chamber of Commerce is mandatory, as it certifies the ownership of the enterprise and its evasion can lead to penalties. Furthermore, enterprises that do not fulfill this requirement do not have access to financial credits (from formal sources) and cannot sign business contracts with public and private sector enterprises. This measure of informality is highly correlated to variables measuring other dimensions of informality, such as tax evasion, bookkeeping practices and contributions to the social security system (Cárdenas and Rozo, 2009).

Several details regarding the quality of the data are worth mentioning. First, the Economic Census was carried out by personnel who went door-to-door over the whole area of study. This way of collecting information has several advantages: 1) "invisible" informal enterprises, i.e., those operating in households or shops without an external sign or banner could also be identified; 2) given that the *universe* of enterprises in included, sampling problems are ruled out (Cardenas and Rozo, 2009) and 3) as the Census does not rely on the promptly return of formularies, responses rates are much larger. Second, at the moment of collection, business owners were made fully aware that if they declared that their business was not in compliance with all the legal regulations, they would not experience any negative legal consequences, and that the information provided was fully confidential. This, together with the fact that informality in Colombia is not openly and widely persecuted, ensures that people were encouraged to provide veridical information.[7]

As can be seen in Table 1, roughly 99% of informal enterprises and 83% of formal enterprises can be classified as "micro" (1-10 workers). We split formal enterprises into two categories: Formal Large Enterprises (FLEs), defined as those formal enterprises with more than 50 employees, and FSMEs, defined as those formal enterprises with 50 employees or less. A substantial percentage of informal

[7] It is also worth mentioning that the Census was carried out and processed the DANE, also in charge of all economic census and surveys in Colombia, including nationwide manufacturing and population censuses and household surveys.

enterprises (45.6%) operate within the premises a household (Table 2), while the large majority of formal enterprises operate in an office or plant.

Table 1. Enterprise size distribution

	Formal	Informal	Total
Micro (1-10 workers)	2318 82.93%	2041 98.74%	4359 89.65%
Small and Medium (11-49 workers)	332 11.88%	24 1.16%	356 7.32%
Large (>50 workers)	145 5.18%	2 0.10%	147 3.02%
Total	**2795** **100%**	**2067** **100%**	**4862** **100%**

Source: Economic Census Cali-Yumbo 2005; authors' own calculations.

Table 2. Activity location

	Formal	Informal	Total
Office or plant	2,423 86.69%	1,033 49.98%	3,456 71.08%
Fixed place (*Puesto fijo*)	30 1.07%	92 4.45%	122 2.51%
Household	342 12.24%	942 45.57%	1,284 26.41%
Total	**2,795** **100%**	**2,067** **100%**	**4,862** **100%**

Source: Economic Census Cali-Yumbo 2005; authors' own calculations.

3.3. Spatial distribution of formal and informal enterprises

Figure 2 displays the enterprise-based Location Quotient (LQ)[8] index for the whole sample of manufacturing enterprises (panel a), FEs (panel b) and IEs (panel c) at the city block level. This index, which measures the relative concentration of activities, is defined as:

$$LQ_i^s = \frac{c_i^s}{x^s},$$

where $c_i^s = ent_i^s / \sum_{s=1}^{s} ent$ denote sector's *s* share of enterprises in location *i*, and $x^s = \sum_{i=1}^{N} ent_i^s / \sum_{i=1}^{N} \sum_{s=1}^{s} ent_i^s$ is sector's *s* share of enterprises in the global area. If the LQ index is greater than one, the sector analyzed is relatively more concentrated in location *i* than in the global area.

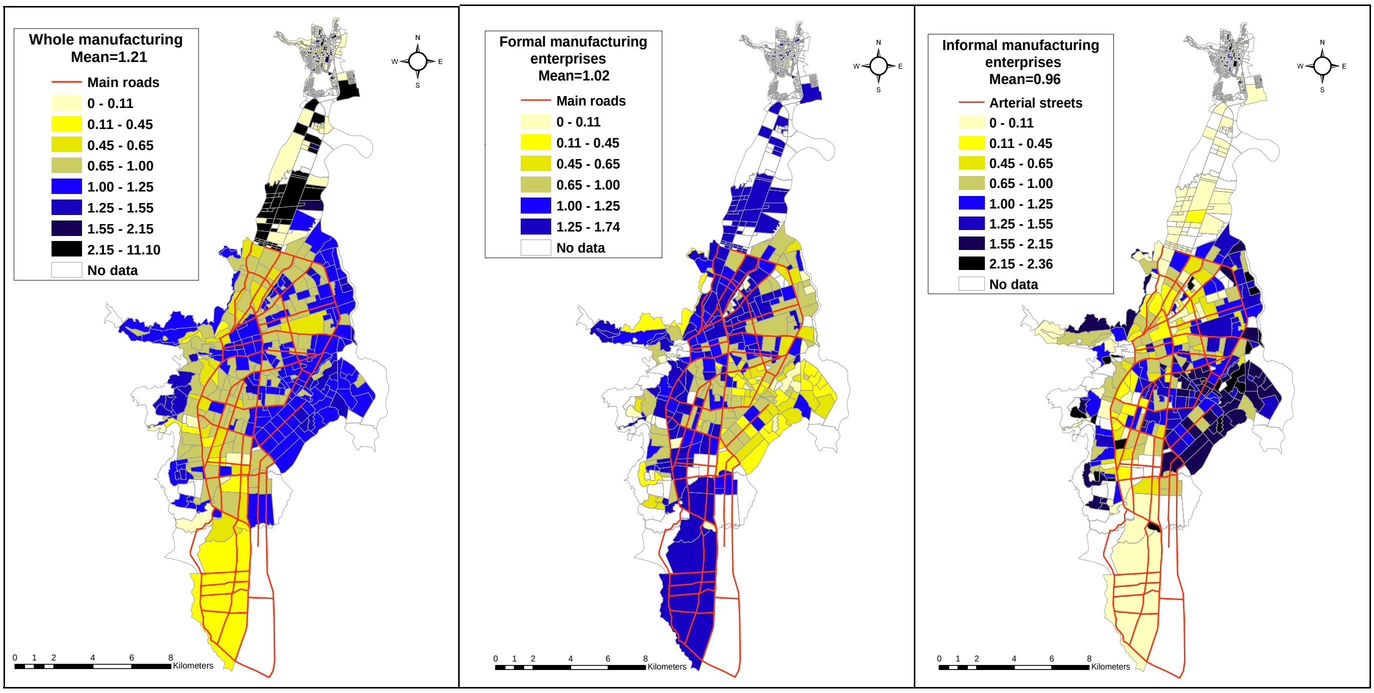
LQ indices with values larger than one are found in different areas of the city for formal and informal manufacturing activity, which in principle indicates that the

[8] Since we are interested in determining the external effects generated in the co-clustering of enterprises, we calculate the plant-based LQ index (Lafourcade y Mion, 2007).

two types of activities do not concentrate in the same areas within the city. In fact, informal manufacturing is concentrated mostly on newly urbanized areas for which income levels are lower than the average (see Figure A2 in the Appendix), while formal activity is found in wealthier and more central locations.

Figure 2. Concentration plant-based LQ index

a) All *b) Formal* *c) Informal*



Source: Economic Census Cali-Yumbo 2005; authors' own calculations.

4. METHODOLOGY

4.1. Spatial indicators of relative geographic clustering and co-clustering

In order to evaluate the degree of clustering and co-clustering of formal and informal enterprises, we calculate the intra and inter-industry M-functions of proposed by Marcon and Puech (2003; 2010), using the software Ripley v.2.8.[9]

For the implementation of this method we use available information on the plain coordinates (X-Y) for each enterprise at the city block level to measure the Euclidean distance among enterprises. We calculate M-functions for formal enterprises, informal enterprises, FSMEs (less than 50 employees), and FLEs (more than 50 employees) and/or (selected) ISIC 3-digit industries. The M-functions are calculated every 1Km between zero and 20Km, and the confidence intervals are computed at a 5% confidence level with 1000 simulations for the case of clustering

[9] This software, designed by Eric Marcon and Florence Puech, is available at http://e.marcon.free.fr/Ripley/

and 100 simulations for the case of co-clustering.[10] In the following we present in detail the definition of M-functions of clustering and co-clustering.

4.1.1. Clustering

The M-function for intra-industrial spatial clustering in a circle of radius *r* for sector *S* is:

$$M(r,S) = \sum_{i=1}^{N_S} \frac{e_{iSr}}{e_{ir}} \Big/ \sum_{i=1}^{N_S} \frac{E_S - e_i}{E - e_i} \qquad (1)$$

Where *i*=1,2….Ns is an index for enterprise and e [E] denotes [total] employment. The function works as follows. All enterprises belonging to sector *S* in the area of study are identified. For *each* of these enterprises, a circle of radius *r* (e.g., 1Km) is drawn. Within this distance, the number of employees belonging to enterprises in sector *S* is counted (e_{iSr}). The sum of this quantity over *i* is then expressed as a proportion of the number of employees belonging to enterprises in all other sectors within the same circle (e_{ir}). This ratio is then made relative to the weight of employment in sector *S* in total employment *in the whole area.*

This relative structure of the M-function allows for a direct interpretation and comparison across sectors and distances. In fact, the M-function is the only distance-based method that allows for a straightforward interpretation and comparison of the value of the resultant indices (Marcon and Puech, 2010). M-values *equal to one* indicate that whatever the considered distance, there are proportionally as many employees who belong to sector *S* as there are in the global area, indicating a completely random location of enterprises in sector *S*. M-values *larger than one* indicate that there are proportionally more employees close to enterprises in sector *S* in a radius *r* than in the global area, which corresponds to the existence of relative geographic clustering of sector *S* at distance *r*. M-values *smaller than one* indicate that there are relatively fewer employees in sector *S* within a radius *r* than in the global area, or in other words, that sector *S* is relatively dispersed at distance *r*.

The statistical significance of the M-function can be tested by calculating confidence intervals for the null hypothesis of independence of enterprise locations, according to which the clustering patterns of enterprises in each sectors is the same. Thus, a large number of independent random distributions of enterprises are generated

[10] The reason for the lower number of simulations for the inter-industry version of the M-functions is computational restrictions.

considering all possible locations and enterprise sizes. The confidence intervals are determined using Monte-Carlo techniques.[11] Then, significant relative clustering (dispersion) of a sector appears if the corresponding M-values are superior (inferior) to one and are outside the confidence interval bands.[12]

The main advantage of this method over cluster-based methods (such as the Ellison-Glaeser Indices) is that the M-indicators do not suffer from the Modifiable Areal Unit Problem (MAUP), which would render the empirical results biased across geographical scales selected by the researcher. Because the M-function considers space as continuous, it full satisfies the property of being unbiased across geographical scales.

Besides this property, the M-indicators satisfy the other four relevant criteria of a good measure of clustering: (Combes and Overman, 2004; Duranton and Overman, 2005; Marcon and Puech, 2010): 1) the measure is comparable across industries; 2) the measure controls for industrial concentration; 3) the measure controls for the overall aggregation pattern of industries; and 4) it is possible to test for the significance of the results. Another important advantage of the M-indicators is that they control for inhomogeneous space, so that the M-function already accounts for the fact that enterprises cannot locate everywhere in the city.

Furthermore, according to Rosenthal and Strange (2003), the mechanisms through which agglomeration economies operate (e.g., labor market pooling, shared inputs, and technological spillovers) are likely to attenuate with distance and may happen at very short distances. In this regard, distance-based methods are better suited for identifying clustering patterns at fine-grained and varying spatial scales, because they are informative regarding the *actual* distance at which clustering and co-clustering take place.

4.1.2. Co-clustering

Besides own-industry clustering, with the M-functions it is possible to assess the presence of co-clustering by calculating the inter-industrial version of the M-function which possesses the same properties as the intra-industrial one described above. M-functions of co-clustering for sectors S_1 and S_2 are defined as:

[11] A more detailed explanation can be found in Marcon and Puech (2010).
[12] In most of the analysis we do not consider dispersion, but the results are available upon request.

$$M_{S_1S_2}(r) = \sum_{i=1}^{N_{S_1}} \frac{e_{iS_2r}}{e_{ir}} \Big/ \sum_{i=1}^{N_{S_1}} \frac{E_{S_2}}{E-e_i} \quad (2)$$

$$M_{S_2S_1}(r) = \sum_{i=1}^{N_{S_2}} \frac{e_{iS_1r}}{e_{ir}} \Big/ \sum_{i=1}^{N_{S_{21}}} \frac{E_{S_1}}{E-e_i} \quad (3)$$

$M_{S_1S_2}$ ($M_{S_2S_1}$) depicts the spatial structure of enterprises belonging to sector S_2 *(*S_1*)* that are found around sector S_1 (S_2). Thus, these co-clustering M-functions test whether the relative density of employees of one sector located around enterprises of another sector is on average larger or smaller than that of the whole territory. The value of these equations shows whether the relative density of plants S_2 (S_1) located around those of sector S_1 (S_2) is greater (superior to one) or less (inferior to one) than that observed for the global area. The statistical significance of the inter-industries M-functions is tested using the same methodology of the intra-industry indicators described above, although the construction of the confidence intervals is slightly more complicated as the null hypothesis has to control for both S_1 and S_2 patterns (for details see Marcon and Puech, 2003).

4.2. Spatial analysis

The M-indicators provide useful information regarding clustering and co-clustering patterns of types of enterprises and/or industries. However, these indicators are not informative of the actual spatial distribution of enterprises within the metropolitan area. As an example, while the indicators may show significant co-clustering of formal and informal enterprises in a specific industry, it is not possible to establish whether it happens in the center of the city or in peripheral areas. Making this distinction is important because the first case is indicative of relatively good access of informal enterprises to central markets, while the second is indicative of formal enterprises presence in areas of predominantly informal activity.

Thus, in order to complement the results from the spatial indicators of clustering and co-clustering and fully exploit the spatial component of our data, we undertake spatial analysis of the geocoded data using ArcGIS v.10. Given that the location of each enterprise is known at the city block level, we can assign randomly each point (enterprise) within the block. In this way, we obtain unique plain (X-Y) coordinates for each enterprise. The spatial analysis is undertaken using these points on industries and type of enterprises selected based on the M-functions results. In particular, we run a kernel density analysis on data for selected clustered formal and informal industries which produces a raster showing the density of a particular type of activity over the entire area. We also analyze the geographic distribution of formal and informal enterprises in selected co-clustered industries.

5. RESULTS

5.1. Spatial indicators of relative geographic clustering and co-clustering

In this section we present the results for the spatial indicators of clustering and co-clustering by type of enterprise -IEs, FSMEs and FLEs (aggregated by industry), by industry (aggregated by type of enterprise) and by type of enterprise and industry. In what follows, we refer to "short distance" whenever the M-values are significant on a distance range of 0 to 1Km only, and "all distances" whenever the range is at least 0 to 6Km.[13]

In the analysis of some of the results we focus on seven selected industries.[14] The selection of these industries was based on three criteria: 1) significant clustering (as measured by the M-function) is observed; 2) significant number of enterprises in the industry and/or 3) significant proportion of manufacturing employment (see Table A2 in the Appendix).[15]

5.1.1. Clustering

By type of enterprise

For the case of the aggregate of IEs and FSMEs, the M-function suggests that both types of enterprises concentrate at all distances, because M-values are larger than one and outside the upper band of the confidence interval (see Figure A3 in the appendix). Interestingly, the degree of clustering is larger for informal enterprises than for formal enterprises of comparable size: at the peak M-value, the relative density of employees in IEs in a radius of less than 1km is over 5 times higher than that in the whole area, whereas that of FSMEs is only 2.5 times higher. Finally, as expected, formal and informal enterprises of smaller size have different clustering patterns than larger formal enterprises.

[13] Note that the terminology here is relative to the size of our study area.
[14] Note however that the intra and inter-industry M-functions are calculated using information for all industries.
[15] The results for all industries are available upon request.

By industry

The M-functions per industry reveal that 40 out of 67 industries are clustered at different distance ranges. As in Marcon and Puech (2003), the most important degree of concentration occurs at very small distances for all sectors but the degree of concentration widely differs from one industry to another.

Five out of this 40 industries show significant clustering at all distances, while four industries display significant yet not very prominent clustering only at very close proximity. As Table 3 shows, the Printing/Editing industry displays the highest degree of clustering of our nine selected industries, with a peak M-value of 27.11.

Table 3. Intra-industry M-functions, selected industries

Ranking	**Industry**	**Distance range (Km)**	**M-peak**	**M-peak distance (Km)**
1	Printing/Editing	0 - 6	27.11	0
2	Furniture	0 - 5	9.98	0
3	Footwear	0 - 11	9.26	0
4	Mill products	0 - 7	5.39	0
5	Bakery	0 - 7	5.16	0
6	Beverages	0	4.59	0
7	Paper	0	3.37	0
8	Plastics	0	2.66	0
9	Apparel	0	1.75	0

Source: Economic Census Cali-Yumbo 2005; authors' own calculations.

By type of enterprise and industry

The results by type of enterprise and industry, summarized in Table 4, reveal that IEs and FSMEs *alone* drive clustering in the most clustered industries (Printing/Editing, Footwear and Mill Products): while FLEs show significant clustering in some industries, especially at very short distances, the results for the aggregate of enterprises seem to be driven by formal and informal smaller establishments.

Evidently, IEs and FSMEs exhibit clustering in the same industries, but to different degrees. The strength of clustering of IEs is very high in some industries, such as Printing/Editing, where the relative density of employees within a radius of 1Km is around 113 times larger than in the whole area. Although other studies (Marcon and Puech, 2003) do no find such extremely high M-values they do seem to be plausible given the relatively small size of informal enterprises.[16] In some industries the degree of clustering of IEs is higher than that of FSMEs: for instance

[16] Most studies exclude small enterprises (<10 or 20 workers) from their samples so we do not have a benchmark to compare our results.

IEs are six times more concentrated than FSMEs in the Printing/Editing industry. However, this pattern cannot be generalized, as FSMEs clustered at all distances exhibit a higher M-value for two of the seven selected industries, Footwear and Apparel.

Table 4. Intra-industry M-functions by type of enterprise, selected industries

Type	Ranking	Industry	Distance range (Km)	M-peak	M-peak distance (Km)
Formal large	1	Bakery	0	95.93	0
	2	Furniture	0	25.25	0
	3	Apparel	0	2.46	0
	4	Plastics	0, 4-6, 8	2.08	0
Formal SMEs	1	Furniture	0-9	18	0
	2	Printing/Editing	0-8, 10-11	17.94	0
	3	Footwear	0-12	14.45	0
	4	Mill products	0-15	12.82	0
	5	Apparel	0-10	8.28	0
	6	Bakery	0-9	7.03	0
	7	Plastics	0, 5-8	5.93	0
Informal	1	Printing/Editing	0-11	113.15	0
	2	Plastics	0-9	61.14	0
	3	Furniture	0-8	47.62	0
	4	Mill products	0-8	21.36	0
	5	Footwear	0-11	12.52	0
	6	Bakery	1-8	9.12	1
	7	Apparel	0-10	6	0

Source: Economic Census Cali-Yumbo 2005; authors' own calculations.

5.1.2. Co-clustering

By type of enterprise

As suggested by the theoretical predictions, the locational patterns of larger enterprises are quite independent of those of smaller enterprises: FLEs seem to be dispersed with respect to IEs and FSMEs at distances between 0 and 11Km. However, while IEs fall into the area of random allocation with respect to FLEs, FSMEs do co-locate around FLEs. On the contrary, both formal and informal enterprises of similar size seem to seek proximity to each other: IEs co-cluster around FSMEs (and vice versa) at all distances, but the tendency of informal enterprises to "follow" formal enterprises of similar size seems to be stronger (see Figure A4 in the Appendix).

By industry

Unlike the case of clustering, significant co-clustering relationships between our seven selected industries occur at different distances ranges, and the M-peak is not necessarily found at close proximity but at different distances.[17] In fact, we only find three significant co-clustering relationships at all distances: enterprises of the Footwear industry locating around enterprises of the Furniture industry and vice versa (see Table 5), and Footwear enterprises locating around Apparel enterprises. For these cases, the mechanisms behind the observed co-clustering relationship may be related to the fact that each pair of industries uses common inputs (such as leather in the case of footwear-furniture and textiles in the case of footwear-apparel) and type of labor.

Table 5. Inter-industry concentration, selected industries (all distances)

Central industry	Around industry	Distance range (Km)	M-peak	M-peak distance (Km)
Furniture	Footwear	0-8	2.75	0
Apparel	Footwear	0-11	2.59	0
Footwear	Furniture	0-7	2.18	0

Source: Economic Census Cali-Yumbo 2005; authors' own calculations.

By type of enterprise and industry

Table 6 shows the results for significant co-clustering relationships by industry and enterprise type at all distances. Disentangling co-clustering relationships by type of enterprise reveals some interesting results. First, all of the significant co-clustering relationships at all distances occur between (formal and informal) small enterprises.[18]

Second, for some industries (Printing/Editing, Footwear, Mill Products and Bakery) the observed clustering of informal and formal enterprises of similar size belonging to the same industries (see section 5.1.1) seems to occur on the same areas of the city. This is deducted from the presence of significant co-clustering of FSMEs and IEs in the same in industry. For instance, in the Printing/Editing industry, the density of IEs around FSMEs is about 19 times larger than that of the whole area. This pattern, however, is not observed for all the industries displaying significant clustering (i.e., for the Furniture, Apparel and Plastics -see Table 4-), which indicates that for those industries FSMEs and IEs may cluster in different parts of the city. These cases will be analyzed in depth in the next section.

[17] The full results are available upon request.

[18] At short distances, some significant co-clustering relationships between large and small enterprises are observed, but for the sake of space these results will not be discussed here.

Third, IEs (and not FSMEs) explain the significant industry-level co-clustering relationship found between Footwear and Furniture enterprises at all distances. In fact, three out of the four significant relationships between industries that share common inputs and type of labor are between IEs. More importantly, whenever co-clustering in industries sharing common inputs or type of labor is observed, it happens only between enterprises of the *same type*. For instance, while Apparel IEs locate around Footwear IEs *and* Footwear FSMEs locate around Apparel FSMEs, there is no significant co-clustering of Apparel (Footwear) IEs around Footwear (Apparel) FSMEs, or vice versa.

And fourth, all of the significant inter-industry/inter-enterprise relationships seem to be in industries with no apparent link (e.g. Bakery and Apparel), suggesting that more general reasons for co-location may be at work, for instance proximity to a transportation route or large consumer market, or simply history or regulations in land use (Chakravorty et al, 2005). We will explore this case further in the next section.

Table 6. Inter-industry concentration by enterprise type, selected industries (all distances only)

Central industry	Around industry	Distance range (Km)	M-peak	M-peak distance (Km)	Common Inputs/ Type of Labor
Printing/Ed.-FSMEs	Printing/Ed.-IEs	0-10	19.47	0	Same Ind
Footwear-FSMEs	Footwear-IEs	0-11	12.93	0	Same Ind
Mill products-FSMEs	Mill products-IEs	0-10	9.14	0	Same Ind
Footwear-IEs	Footwear-FSMEs	0-11	7.36	0	Same Ind
Bakery-IEs	Bakery-FSMEs	0-6	4.85	1	Same Ind
Furniture-IEs	Footwear-IEs	0-7	12.06	0	Yes
Footwear-IEs	Furniture-IEs	0-9	7.72	0	Yes
Apparel-IEs	Footwear-IEs	0-11	7.18	0	Yes
Footwear-FSMEs	Apparel-FSMEs	0-9	3.01	0	Yes
Mill products-FSMEs	Footwear-IEs	0-13	7.74	0	No
Bakery-FSMEs	Footwear-IEs	0-11	7.12	0	No
Apparel-IEs	Bakery-FSMEs	0-9	3.92	0	No
Bakery-FSMEs	Apparel-IEs	0-10	3.81	0	No

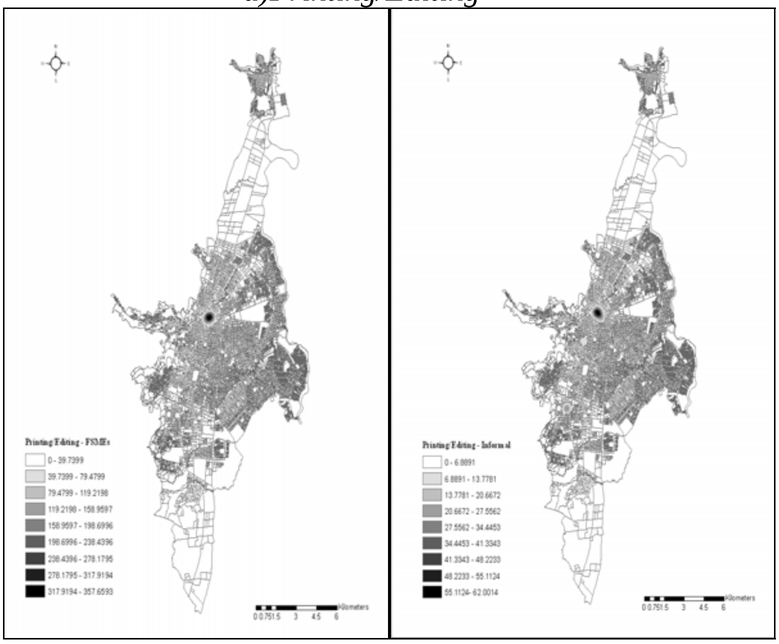
Source: Economic Census Cali-Yumbo 2005; authors' own calculations.

5.2. Spatial analysis

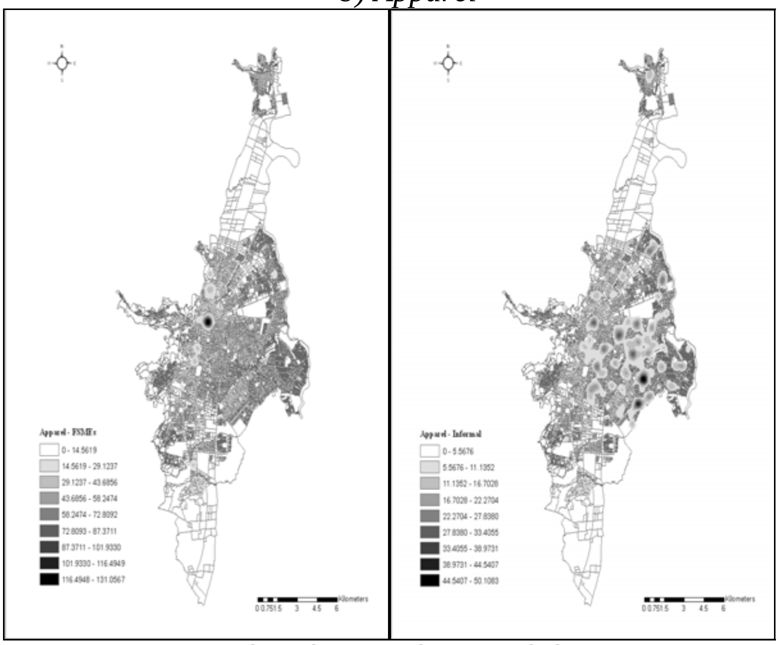
As indicated in section 5.1, IEs and FSMEs display clustering and also significant co-clustering in the Printing/Editing industry. Analyzing the kernel density output in this industry reveals that FSMEs and IEs do cluster on the same areas of the city: the spot indicating the highest density of enterprises is located approximately on the same place for both types of enterprises (see panel a) of Figure 3). In fact, the kernel density analysis of the remaining three industries for which clustering and co-clustering were significant tells a similar story (see Figure A6). These cases can be contrasted with the case of the three industries for which both FSMEs and IEs display significant clustering, but no co-clustering of the two types of enterprises is found. In particular, for the case of Apparel (see panel b) of Figure 3), there is clear spatial segmentation, in the sense that FSMEs and IEs locate in different parts of the city, as indicated by different areas of high density for FSMEs and IEs (see Figure A7).

Figure 3. Kernel density, selected industries

a)Printing/Editing



b) Apparel

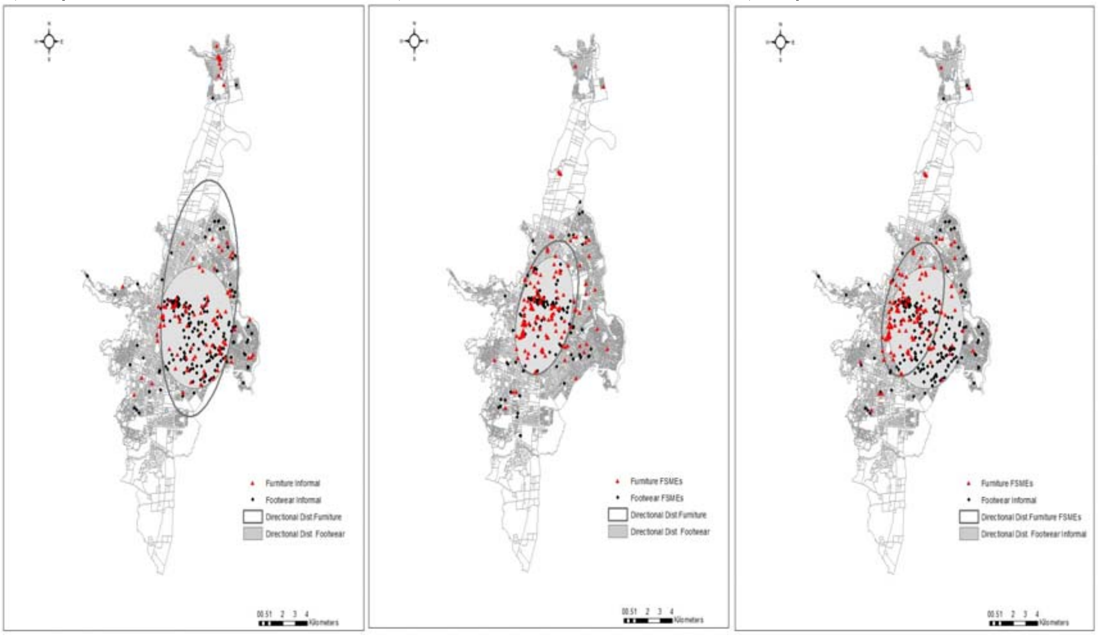


Source: Economic Census Cali-Yumbo 2005; authors' own calculations.

Figure 4 displays a panel of three maps representing the spatial distribution of enterprises of the Furniture and Footwear industries, including their respective standard deviational ellipses, which indicate the directional distribution of enterprises in space.[19] The first one shows the spatial distribution of IEs in both industries, the second one that of FSMEs and the third one that of FSMES in the footwear industry and IEs in the furniture industry. As described in the previous section, enterprises in the furniture and footwear industries display significant co-clustering, which is explained by co-clustering of IEs, and is perhaps driven by a noticeable cluster of IEs in Yumbo. Furthermore, the lack of co-clustering between FSMEs and IEs in these industries can be the result of the fact that a considerable number of FSMEs in the Footwear industry locate along the main axis of the city, as indicated by the directional distribution of FSMEs, while most IEs in the furniture industry locate on the east of Cali and in Yumbo.

Figure 4. Spatial distribution of enterprises of the Furniture and Footwear industries

a) Informal *b) FSMEs* *c) Informal and FSMEs*



Source: Economic Census Cali-Yumbo 2005; authors' own calculations.

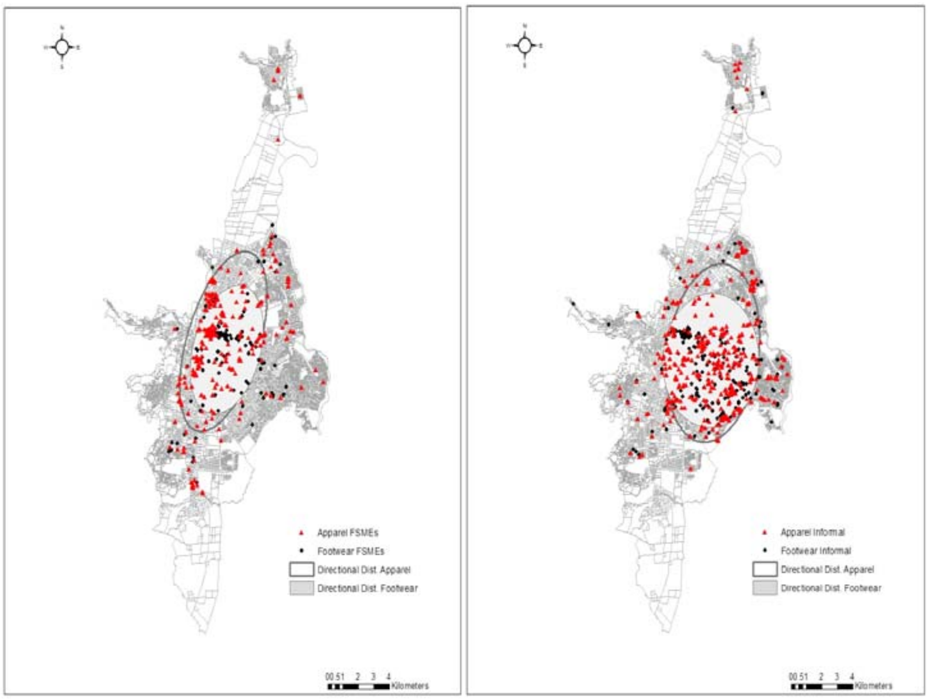
As the results of the inter-industry M-functions indicated, there is co-clustering of enterprises of the Apparel and Footwear industries, which is explained by co-

[19] Standard deviational ellipses "measures whether a distribution of features exhibits a directional trend, i.e. whether features are farther from a specified point in one direction than in another direction" (see http://resources.esri.com/help/9.3/arcgisdesktop/com/gp_toolref/spatial_statistics_tools/directional_distribution_standard_deviational_ellipse_spatial_statistics_.htm). Ellipses are calculated at one standard deviation.

clustering of Apparel IEs around Footwear FSMEs, on the one hand, and Footwear FSMEs around Apparel FMSEs on the other. Thus, while there is an evident spatial complementarity between these two industries, it is only significant for enterprises of the same type. As the standard deviational ellipses in Figure 5 illustrate, co-clustering of enterprises in these industries happens in different parts of the city for each type of enterprise. In this case, spatial segmentation between IEs and FSMEs seems to impede significant co-clustering relationships between formal and informal enterprises in industries that share common inputs and type of labor.

Figure 5. Spatial distribution of enterprises of the Apparel and Footwear industries

a)FSMEs *b) Informal*



Source: Economic Census Cali-Yumbo 2005; authors' own calculations.

Lastly, the spatial analysis reveals an interesting fact regarding the co-clustering of FSMEs and IEs. As explained earlier, the inter-industry M-functions showed three significant inter-industry/inter-enterprise co-clustering relationships, all of them between industries that do not share common inputs or type of labor. As it turns out, FSMEs in these industries have substantial presence in areas where mainly informal enterprises operate (see panels b) and c) in Figure A6, and compare to the distribution of IEs in the Apparel and Footwear industries in panel b) in Figures 5). This can be due to a larger importance of proximity to dense consumer markets for small enterprises in these industries (see Figure A1). In this case, spatial segmentation seems to be weakened by formal enterprises penetrating markets beyond central locations and operating in areas of informal manufacturing activity rather than informal enterprises locating near main transportation routes or central areas where formal activity predominates.

6. CONCLUSIONS

One of the main goals of this article was to extend our understanding of the spatial distribution of informal enterprises vis-à-vis formal enterprises. Our first research question was: do formal and informal enterprises display different clustering patterns? As the results show, both types of enterprises display a strong tendency to cluster in the same industries, and this tendency is much stronger than that of larger formal enterprises. Although on aggregate informal enterprises display a larger degree of clustering than their formal counterparts of similar size, this is not the case for each individual industry, as for some industries FSMEs display clustering to a larger degree than IEs. Thus, the same logic behind clustering seems to operate in different industries for both FSMEs and IEs but to different degrees.

Our second research question was: is there a marked spatial segmentation between formal and informal manufacturing activity within the city? Or, do formal and informal enterprises locate on the same areas? The co-clustering analysis allowed us to establish that, in general terms, while formal large enterprises display location patterns that seem independent of enterprises of smaller size, formal and informal enterprises of similar size seek each other's proximity. However, we could establish that significant clustering of both formal and informal enterprises of similar size on the same industry does not necessarily imply that this clustering takes place on the same parts of the city. As our results show, in some cases, clustering of both types of enterprises does indeed happen on the same areas of the city, but in other cases it happens simultaneously in different parts of the city. In conclusion, formal and informal enterprises of similar size belonging to the same industry may locate in different parts of the city, which is an indication of spatial segmentation, and this is perfectly compatible with significant clustering of each type of enterprise. This is an interesting result, as it points to the fact that while there is a common logic behind clustering of small enterprises, it can operate *in parallel* in two different parts of the city.

While we found some significant co-clustering of *either* formal or informal enterprises of similar size belonging to industries that share common inputs and type of labor, we did not find evidence on inter-industry/inter-enterprise co-clustering in these industries. As the spatial analysis revealed, the significant co-clustering relationships between FSMEs and IEs of unrelated industries seem to be explained by the presence of FSMEs in areas where informal activity predominates, rather than the other way around. Thus, all in all, the results give more support to the view of spatial segmentation, at least from the perspective of informal enterprises.

REFERENCES

Cardenas, M and Rozo, S (2009) Análisis cualitativo y cuantitativo de la informalidad empresarial en Colombia. *Desarrollo y Sociedad* 63: 269-296

Carr, M., Chen, M.A. and Tate J. (2000) Globalisation and Home-Based Workers. Feminist Economics 6, 3, pp. 123–142

Chakravorty, S., Koo, J. and Lall, S.V. (2005) Do localization economies matter in cluster formation? Questioning the conventional wisdom with data from Indian metropolises. *Environment and Planning A* 37: 331- 353

Combes PP and Overman H (2004) The spatial distribution of economic activities in the European Union. In JV Henderson and JF Thisse (eds) Handbook of Urban and Regional Economics. North Holland, Amsterdam: Elsevier.

Daniels PW (2004) Urban challenges: the formal and informal economies in mega-cities. Cities 21 (6):501-511

Deichmann U, Lall SV, Redding SJ and Venables AJ (2008) Industrial Location in Developing Countries. The World Bank Research Observer, 23(2): 219-46

Duranton G (2008) Cities: Engines of Growth and Prosperity for Developing Countries? Commission on Growth and Development Working Paper 12, The World Bank

Duranton, G. and Overman, H. G. (2005) Testing for Localization Using Micro- Geographic Data. Review of Economic Studies 72: 1077–1106

Holl A (2008) Production subcontracting and location. Regional Science and Urban Economics, 38(3): 299-309

Jütting J and Laiglesia JR De (2009). Is Informal Normal? Towards More and Better Jobs in Developing Countries, Paris: OECD Development Centre

Lafourcade M and Mion G (2007) Concentration, agglomeration and the size of plants. *Regional Science and Urban Economics*, 37(1): 46-68

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

Marcon E and Puech F (2003) Measures of the geographic concentration of industries: improving distance-based methods. Mimeo

Moreno-Monroy, A.I. Critical commentary: Informality in space: Understanding agglomeration economies during economic development. Accepted manuscript, *Urban Studies*, forthcoming.

Rosenthal, S. and Strange, W. (2003). "Geography, Industrial Organization, and Agglomeration", *The Review of Economic and Statistic*, 85(2): 377-393

Schneider F (2002) Size and measurement of the informal economy in 110 countries around the world. Published electronically by the World Bank at: http://rru.worldbank.org/Documents/PapersLinks/informal_economy.pdf [Accessed October 2010] Rapid Response Unit, The World Bank

Schneider, F. (2005) Shadow economies around the world: what do we really know? European Journal of Political Economy, 21: 598-642

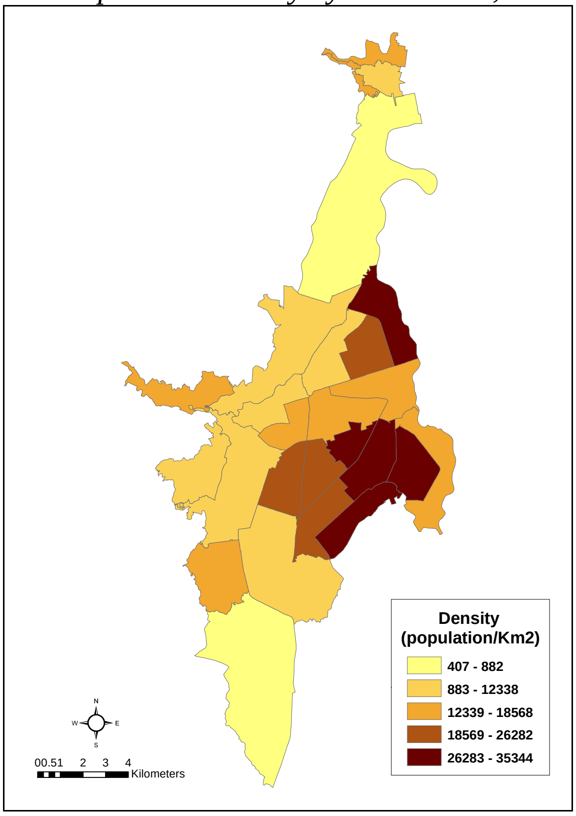
Sethuraman SV (1997) Urban Poverty and the Informal Sector: A critical. Assessment of the current strategies. ILO Geneva

Scott A (1988) *Metropolis: From the Division of Labour to Urban Form*, Berkeley, CA: The University of California Press

Wu, F. (1999) Intrametropolitan FDI enterprise location in Guangzhou, China: A Poisson and negative binomial analysis. *Annals of Regional Science* 33: 535-555

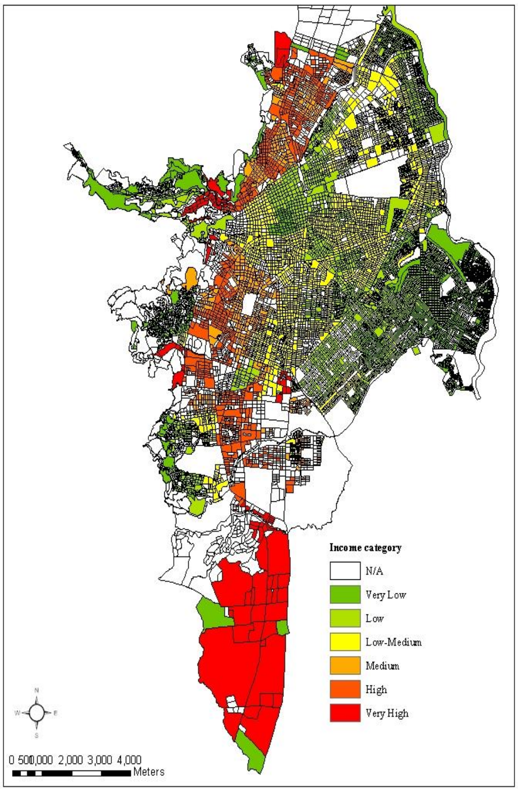
APPENDIX

Figure A1: Population density by Communes, Cali-Yumbo



Source: Population Census 2005, DANE

Figure A2. Distribution of income categories, Cali



Source: Economic Census of Cali-Yumbo 2005 and author's own calculations. Note: Data for Yumbo is not available. Levels of income based on economic stratification in six categories (1=Very Low, 6=Very high)

Table A1. Size of the blocks (Manzanas)

	Number of blocks	Size in sq. meters		
		Mean	Min	Max
Cali	2633	12279.4	807.91	3981421
Yumbo	134	81198.6	1306.1	1417601
Total	2767	15616.99	807.91	3981421

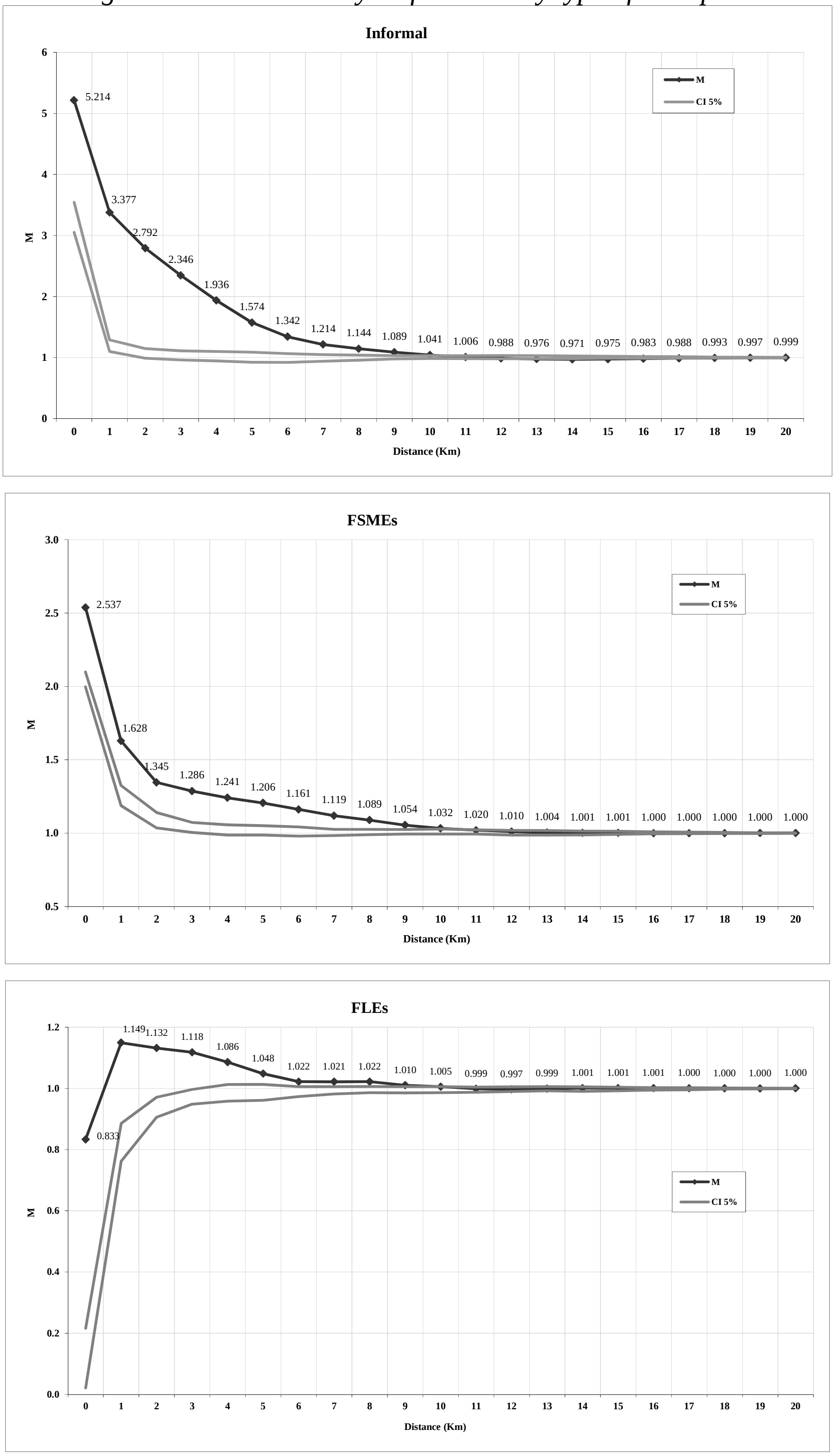
Source: Economic Census Cali-Yumbo 2005; authors' own calculations. Given that an average block is composed by 4 sides, the length of each side is, for the case of Cali, the square root of 12279 meters (110.81 meters).

Table A2. Number of enterprises and employees by selected sectors and type of enterprise

		All enterprises		**Formal SMEs**		**Formal Large**		**Informal**	
SIC	**3-digits industry**	**# enterprises**	**# employees**	**# enterprises**	**# employees**	**# enterprises**	**# employees**	**# enterprises**	**# employees**
154	**Mill products, starch and its products**	307	2387	144	854	9	1197	154	336
155	**Bakery products, pasta and its products**	578	2585	363	1659	4	386	211	540
181	**Apparel, except fur**	662	8675	270	1672	23	5365	369	1638
192	**Footwear**	354	1692	139	895	3	304	212	493
22	**Printing/Editing and similar goods**	259	1620	206	1157	4	350	49	113
252	**Plastic products**	157	2595	112	963	11	1559	34	73
361	**Furniture**	259	1495	148	732	5	495	106	268

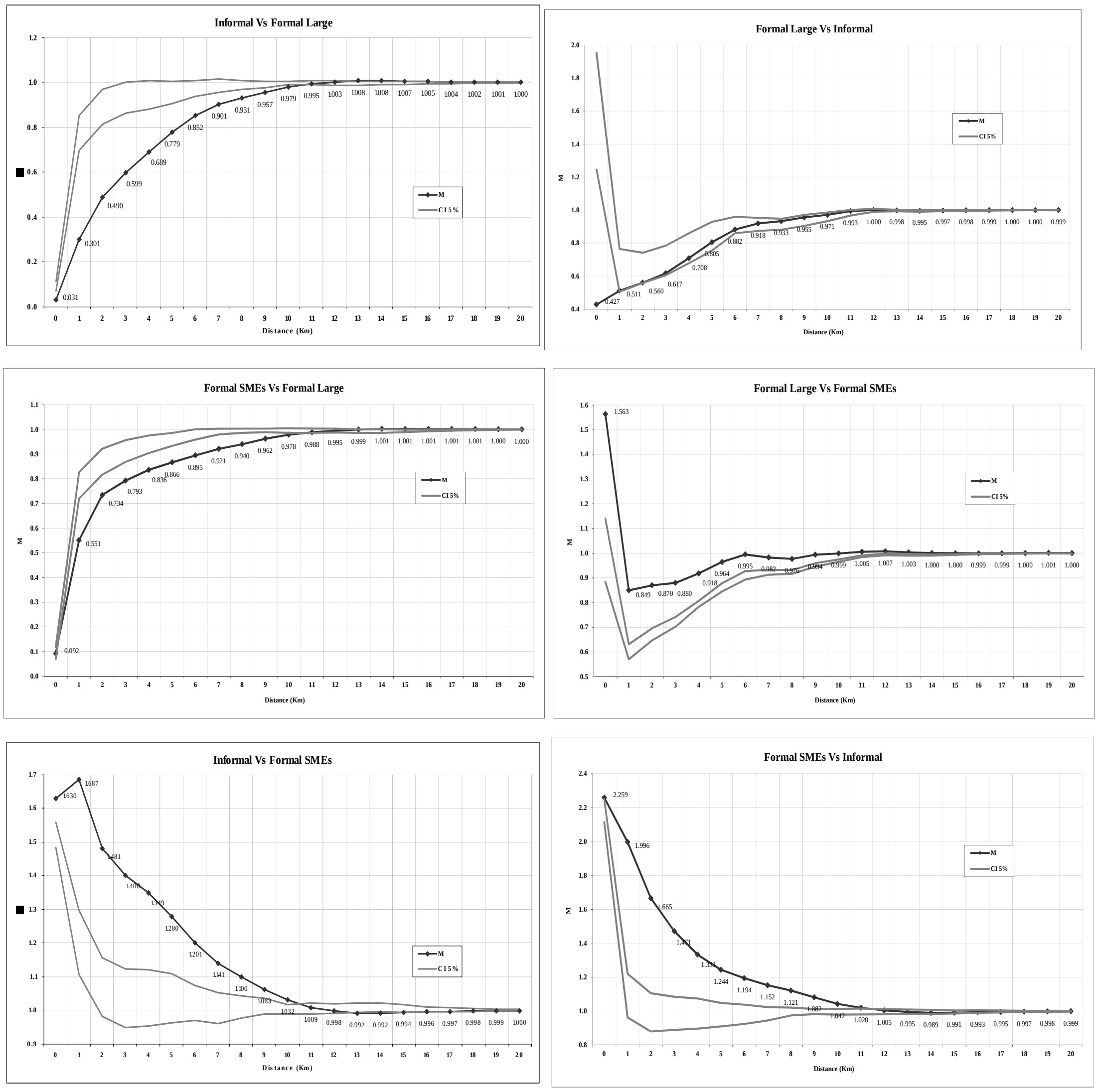
Source: Economic Census Cali-Yumbo 2005; authors' own calculations.

Figure A3. Intra-industry M-functions by type of enterprise



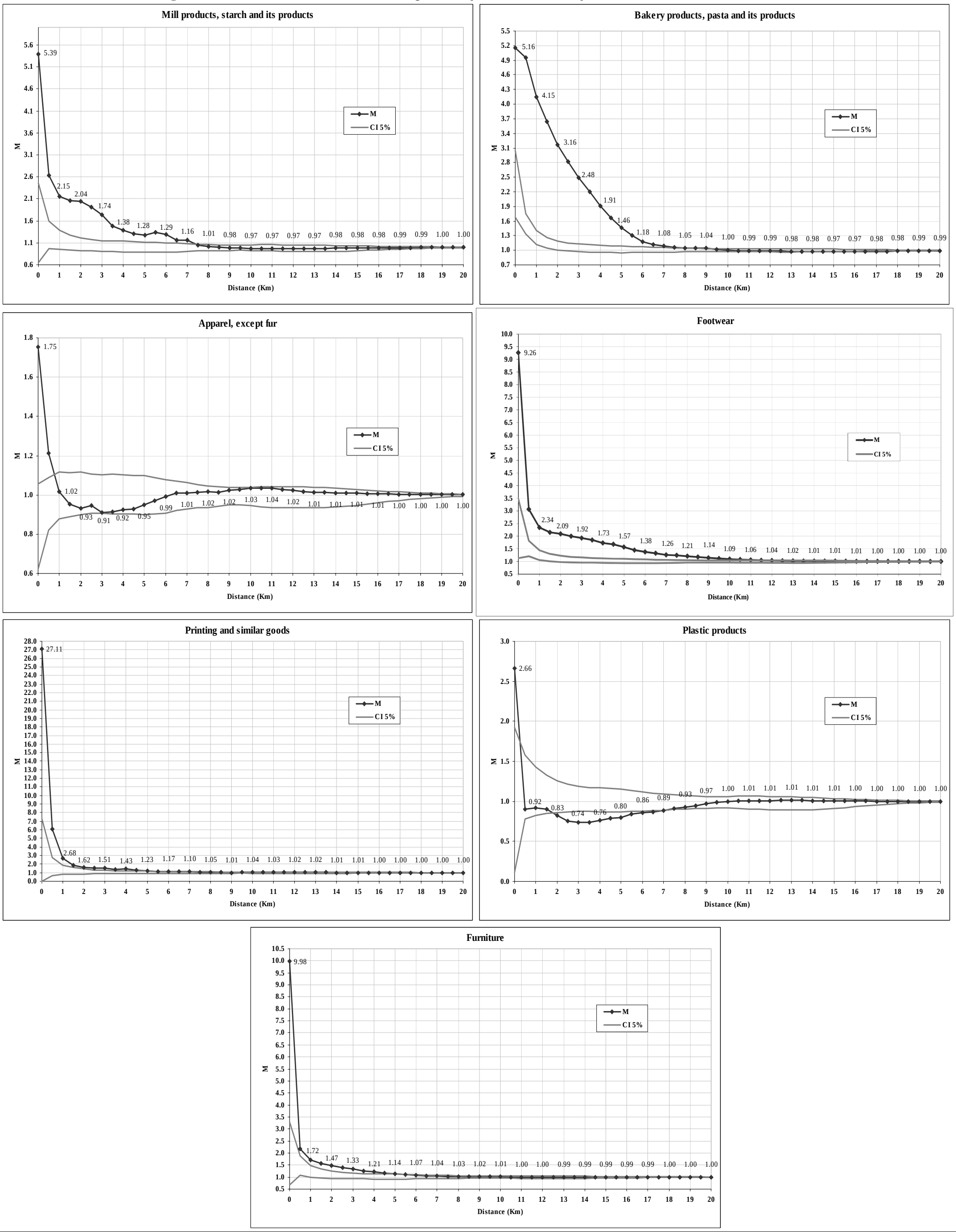
Source: Economic Census Cali-Yumbo 2005; authors’ own calculations.

Figure A4. Inter-industry M-functions by type of enterprise



Source: Economic Census Cali-Yumbo 2005; authors' own calculations.

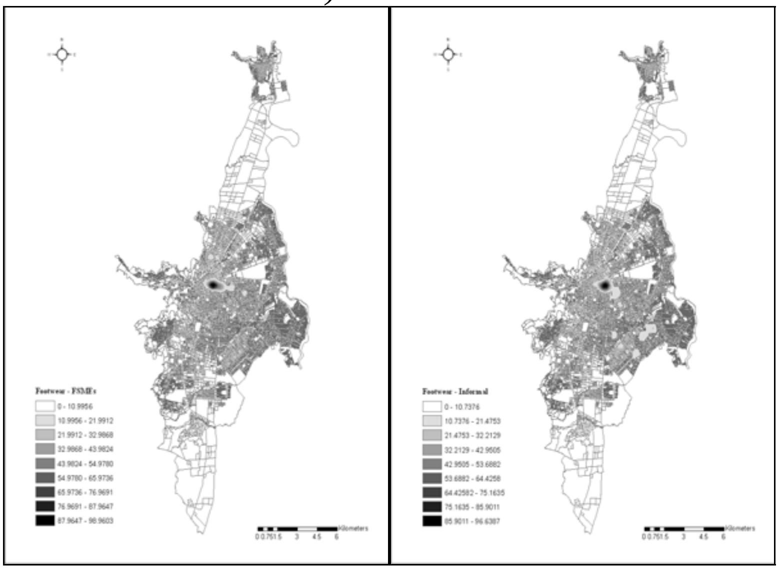
Figure A5. Intra-industry M-functions for selected industries



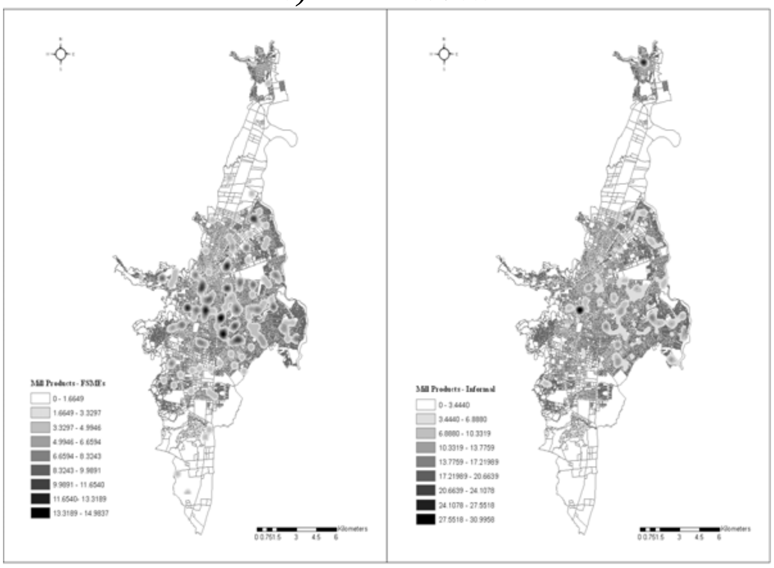
Source: Economic Census Cali-Yumbo 2005; authors' own calculations.

Figure A6. Kernel density of industries displaying significant clustering and co-clustering of FSMEs and IEs

a) Footwear



b) Mill Products



c) Bakery

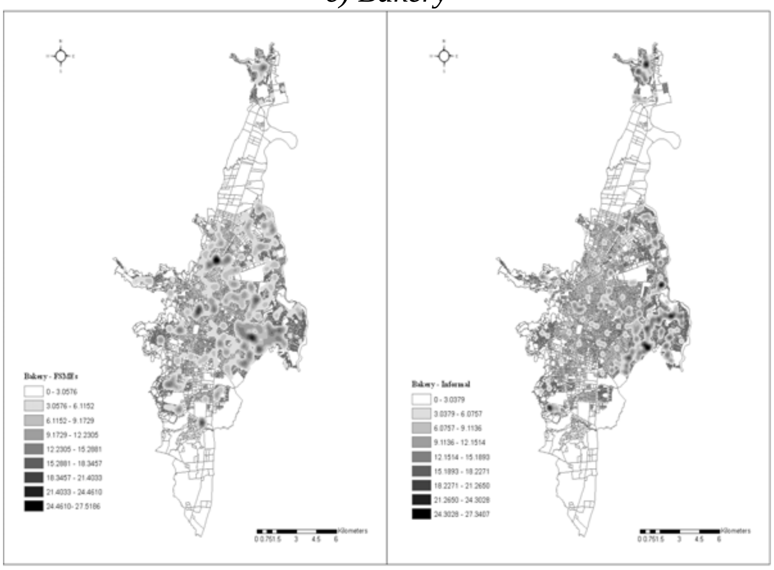
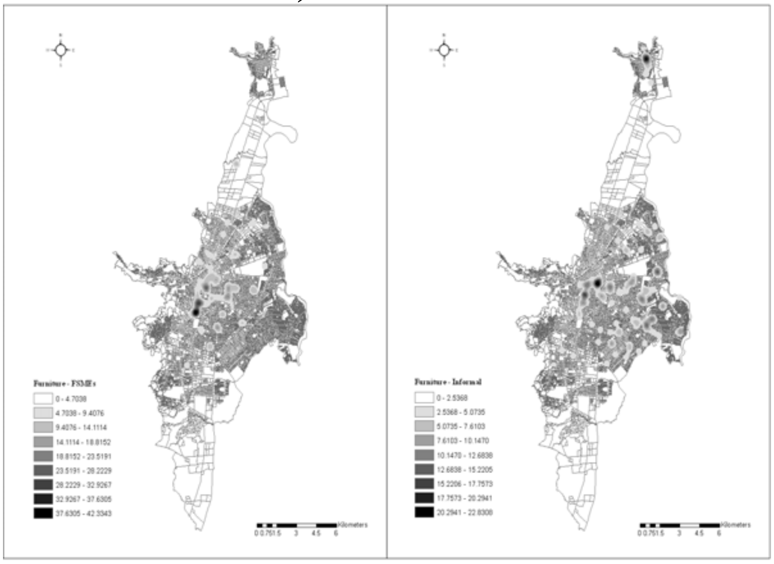


Figure A7. Kernel density of industries displaying significant clustering FSMEs and IEs but no co-clustering

a) Furniture



b)Plastics