UNIVERSITAT
ROVIRA I VIRGILI
DEPARTAMENT D'ECONOMIA

(creip)
CENTRE DE RECERCA EN ECONOMIA INDUSTRIAL I ECONOMIA PÚBLICA

WORKING PAPERS

Col·lecció "DOCUMENTS DE TREBALL DEL DEPARTAMENT D'ECONOMIA - CREIP"

Intra-metropolitan Agglomeration of Formal and Informal Manufacturing Activity: Evidence from Cali, Colombia

Ana I. Moreno Monroy
Gustavo A. García Cruz

Document de treball n.24- 2015

DEPARTAMENT D'ECONOMIA – CREIP
Facultat d'Economia i Empresa

UNIVERSITAT
ROVIRA I VIRGILI
DEPARTAMENT D'ECONOMIA

(creip)
CENTRE DE RECERCA EN ECONOMIA INDUSTRIAL I ECONOMIA PÚBLICA

Edita:

Departament d'Economia www.fcee.urv.es/departaments/economia/public_html/index.html Universitat Rovira i Virgili Facultat d'Economia i Empresa Avgda. de la Universitat, 1 43204 Reus Tel.: +34 977 759 811 Fax: +34 977 300 661 Email: sde@urv.cat	CREIP www.urv.cat/creip Universitat Rovira i Virgili Departament d'Economia Avgda. de la Universitat, 1 43204 Reus Tel.: +34 977 558 936 Email: creip@urv.cat

Adreçar comentaris al Departament d'Economia / CREIP

Dipòsit Legal: T – 1090 - 2015

ISSN edició en paper: 1576 - 3382
ISSN edició electrònica: 1988 - 0820

DEPARTAMENT D'ECONOMIA – CREIP
Facultat d'Economia i Empresa

Intra-metropolitan Agglomeration of Formal and Informal Manufacturing Activity: Evidence from Cali, Colombia

Ana I. Moreno-Monroy[*]

Gustavo A. García Cruz[**]

1. Introduction

The informal sector absorbs on average half of all employment in emerging and developing countries, and has proven to be quite persistent (WTO and ILO, 2009). It is composed of small private unincorporated enterprises that are not registered as mandated by national laws.[1] Owing to fast urbanization in developing countries, an important proportion of informal activity is localized in metropolitan areas. The intra-metropolitan locational patterns of informal enterprises may be different to those of comparable formal enterprises because of high levels of urban socio-economic segregation and the very nature of informal activity (Moreno-Monroy, 2012). Nevertheless, to date there is no empirical evidence on the intra-metropolitan spatial distribution of informal manufacturing activity. This empirical gap is not surprising given that comprehensive, geo-referenced data for informal enterprises is usually not available. This paper bridges this gap by comparing the agglomeration and coagglomeration patterns of formal and informal enterprises for different manufacturing industries, using a unique geo-referenced dataset including the universe of formal and informal manufacturing enterprises for a metropolitan area of an emergent economy.[2]

In the existing literature, there are no clear theoretical predictions on the locational patterns of informal enterprises in the urban space. On the one hand, the informality status *per se* may not

[*] Corresponding author. Marie Curie Research Fellow at the Economics Department and the Research Centre on Industrial and Public Economics (CREIP), Rovira i Virgili University. E-mail address: ana.i.moreno.monroy@gmail.com.

[**] Postdoctoral research fellow, CEDE and Economics Department, University of Los Andes.

[1] As defined by the 1993 15th International Conference of Labour Statisticians.

[2] In this paper we use the terms "agglomeration" and "concentration" indistinctively. Although the process of spatial concentration is sometimes referred to in the literature as "clustering", in a strict sense not all spatial concentrations are "clusters", once they are defined as "spatial concentrations of related activities based on growth-enhancing local interactions between actors and factors" (Visser, 2009, p. 180). In our empirical application, we cannot identify "clusters" because we do not have a measure of local interaction, and we do not have enough information to establish whether enterprises benefit or not from concentration. Therefore, to avoid confusion we refrain from using the terms "clusters" and "clustering".

bring any consequences in locational terms. The standard agglomeration, transaction cost and industrial district theories would predict that informal enterprises should locate close to dense final and intermediate markets located in central areas of the city, in order to benefit from supply-side and demand-side positive externalities, and minimize transaction costs. On the other hand, informal enterprises may, by preference or necessity, operate at or near home. Home-based activity arises as a result of segmented demand if informal enterprises operate in dense peripheral areas where they can cater low-income individuals. It can also arise if owners benefit from lower rental and utility costs, if accessibility to other parts of the city is difficult or costly, or if the quality of the premises is not relevant for undertaking the productive activity. These possibilities are particularly relevant for informal enterprises acting as subcontractors in disintegrated value chains.

A complete understanding of the intra-metropolitan location of informal enterprises would require not only characterizing agglomeration patterns, but also identifying the reasons why these patterns are observed. Unfortunately, we do not have enough information to discern the causes of the observed patterns, but here we take a first step by characterizing the locational patterns of comparable formal and informal enterprises. We pose two research questions: do formal and informal enterprises display different agglomeration patterns? And do formal and informal enterprises in a given industry locate in the same areas of the city? In order to provide possible answers, we use census manufacturing enterprise-level data for the metropolitan area of Cali (Colombia) for 2005. This rich dataset includes the *universe* of manufacturing enterprises geocoded at a geographical scale equivalent to the city-block level. Several criteria in the Census allow us to discriminate between formal and informal manufacturing enterprises in each industry. Furthermore, because information on the number of employees per enterprise is also provided, formal enterprises can be disentangled by size. This division is useful in distinguishing the location effects explained by size from those related to the formality status of an enterprise.

We use this information in two complementary ways. First, we calculate the degree of spatial agglomeration and coagglomeration by means of M-functions (Marcon and Puech, 2010). We prefer this method over other alternatives for measuring spatial agglomeration because, besides satisfying all the relevant criteria for a good measure of spatial agglomeration, it allows for a direct comparison between the agglomeration and coagglomeration intensity of different types of enterprises in each industry. Second, in order to fully exploit the spatial component of the data and complement the M-function results, we conduct spatial analysis on the distribution of formal and informal enterprises by means of kernel density mapping of selected industries.

Our work is connected to a recent and growing literature focusing on the causes and consequences of the spatial agglomeration and coagglomeration of formal and informal activity across geographical units in developing countries (Mukim, 2011; Sokty and Newman, 2014; Mukim, 2014). These studies contribute to our understanding of the spatial distribution of formal and informal activity, but are not informative about location within metropolitan areas, where agglomeration may be driven by other factors and carry different consequences. Previous intra-metropolitan location studies for developed and developing countries have focused solely on formal (large) manufacturing enterprises (Arauzo-Carod and Viladecans-Marsal, 2009; Chakravorty et al., 2003; 2005) or particular informal industries (Harris, 2014). Ours is the first study to provide a comprehensive account of agglomeration and coagglomeration patterns of all the informal manufacturing industries, and to offer a comparison between the locational patterns of formal and informal enterprises within a metropolitan area.

After this introduction, Section 2 proceeds with a literature review on the intra-metropolitan locational patterns of informal manufacturing enterprises. In Section 3 we describe the area of study, the manufacturing sector in Cali, and data used in the empirical analysis. In Section 4 we present the definitions of the M-functions of agglomeration and coagglomeration as well as the methodology used in the spatial analysis. In Section 5 we present the results. In Section 6 we discuss the results and conclude.

2. Theoretical predictions on the intra-metropolitan agglomeration of informal manufacturing activity

In this section we review theoretical predictions related to supply-side and demand-side drivers of agglomeration. Even though we do not test the relevance of these determinants in explaining the observed patterns, the predictions developed here can guide our interpretation of the results and set a reference for future research avenues.

We start by describing theories related to supply-side drivers of agglomeration, based on four relevant approaches identified by Visser (2009).

1. Standard agglomeration theory. This approach identifies three supply-side channels for agglomeration (Rosenthal and Strange, 2004): sharing inputs produced under increasing

returns to scale; accessing a larger pool of workers; and knowledge and technical spillovers. It can be argued that if the same supply-side mechanisms behind clustering influence the location decisions of informal enterprises, *the clustering patterns of formal and informal enterprises of similar sizes should not differ*. Empirical evidence for developed countries shows that the relatively larger tendency of small enterprises to cluster is connected to the larger sensitivity of small enterprises to final and intermediate markets accessibility (Rosenthal and Strange, 2003; Lafourcade and Mion, 2007). In a study for the informal handicraft industry in Nairobi, Harris (2014) finds that while there are agglomeration economies related to access thick intermediate, final and labor markets, their positive effects are constrained by location restrictions faced by informal enterprises, and the oversupply and hypermobility of informal labor.

2. New industrial spaces. According to Scott (1988), linkage costs rise with the quantity of goods traded between enterprises, and decline with product standardization, the stability of interactions, and the need for intermediation. Hence, large enterprises locate far away from central, congested locations, because their linkages are large in scale, standardized, stable, and more easily manageable. On the contrary, small enterprises locate near their transaction partners because their linkages are small in scale, unstandardized, unstable, and in need of personal intermediation. If the linkage costs of informal enterprises is high (as can be expected because they are small, produce mostly homogeneous products and face considerable uncertainty), the tendency to seek close proximity in central areas also apply to informal producers, if not to a higher degree. Thus, from this perspective, the prediction is that *both formal small and informal enterprises agglomerate in locations with dense networks of suppliers and consumers, while large (formal) enterprises locate outside these central areas*.
3. Flexible specialization. This approach focuses on the spatial consequences of increased disintegration, specialization and subcontracting. On a broad perspective, it is linked to an abundant literature on the informal sector and Global Value Chains (Harris, 2014; Mehrotra and Biggeri, 2005). In terms of location, informal enterprises operating on a subcontracting basis may consider proximity to input suppliers and consumers irrelevant. In this case, the prediction is that *informal enterprises prefer to locate outside central areas of the city that offer lower rent prices and costs*. This may be the case whenever

formal enterprises determine the conditions of production (Chen, 2007), for instance by directly provide materials and inputs to informal enterprises, commonly through an intermediary (Carr et al., 2000), or whenever utilities such as energy costs borne by informal producers are subsidized in peripheral areas.

4. Neo-marshallian industrial districts. Under this approach, agglomeration is also the result of vertical disintegration, specialization and subcontracting, but it involves many highly specialized small enterprises connected together in a cluster for the production of a specific good. In these industrial districts, the positive effect of agglomeration on enterprises is sustained by the balance between competition and collaboration, which is usually only possible in the presence of formal or informal cooperation agreements. Then, *formal small and informal enterprises in vertically related industries forming a cluster are found in proximity to each other*. For the case of developing countries, there are very few examples of industrial districts (the Gamarra cluster is one of them, see Visser (1999) and Visser et al., (2014)). It is more often case that observed spatial concentrations of informal enterprises do not correspond to industrial districts, and consequently may not have the same potential to generate positive agglomeration economies (Moreno-Monroy, 2012).

Demand-related factors can also drive agglomeration. Both formal and informal enterprises selling final goods and services are expected to be drawn to affluent, central areas in search for proximity consumers. However, when the segment of the population catered by informal manufacturing enterprises concentrates in peripheral areas, informal producers find it profitable to locate there. This leads to agglomerations of informal enterprises in dense low-income population residential areas. In these demand-driven agglomerations, informal enterprises may have an edge over formal enterprises because they are able to sell below production costs. Then, the observed density of informal enterprises is higher than the observed density of formal enterprises in dense lower-income areas. Thus, the prediction is that *informal enterprises locate in peripheral areas, while formal enterprises locate in central areas*.

It is worth noticing that besides this demand-side mechanism, the density of informal activity can be high in lower-income areas if informal enterprises are more likely to use the household premises for productive activities than formal enterprises. This may be because: they attribute less value to the quality of premises when compared to their formal counterparts (Sethuraman, 1997); they are unable to bid for rents vis-à-vis formal enterprises (Daniels, 2004; Harris, 2014); or because commuting is difficult or costly (Moreno-Monroy and Posada, 2014). Finally, according to the findings of Chakravorty et al. (2005) for the case of Indian metropolises, agglomerations may be simply the result factors unrelated to the demand and supply-side mechanisms, such as zoning and environmental regulations.

3. Data, study area and manufacturing sector

3.1. Data

This study uses comprehensive enterprise-level data from the Economic Census 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 (each measuring approximately 110x110 meters for Cali[3]), and there are on average of 3.7 enterprises per block, the scale is approximately equivalent to having the actual address of each enterprise. To carry out the analysis, we first obtain a unique point or plain (X-Y) coordinates for each enterprise by randomly assigning each enterprise within its block by using a standard polygon partitioning algorithm provided by the software ArcGIS. In this way, we avoid possible overlaps among enterprises.

In the analysis we only consider manufacturing enterprises. According to the Economic Census data, the *universe* of manufacturing enterprises in Cali-Yumbo is 5,130. Of this total, enterprises that could not be geographically located and those that do not operate at a fixed location were excluded, leading to a sample of 4,862 enterprises, of which 95 percent are located in Cali, and the remaining 5 percent are located in Yumbo. The information is available at the 2-, 3- and 4-digit level of the International Standard Industrial Classification of All Economic Activities (ISIC) Revision 3[4], and the analysis is conducted at the 3-digit level.[5]

Informal enterprises are defined as those enterprises not registered at the Chamber of

[3] There are 2,633 blocks in Cali. An average block is composed by 4 sides, so the length of each side is, for the case of Cali, the square root of the mean size of an average block (i.e., the square root of 12,279).
[4] For industrial codes and names, see http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=2.
[5] This level of disaggregation is used for convenience, as it makes the interpretation of results more manageable. It is worth noticing, however, that the M-functions are additive in industries, so that the M-function values in each 3-digit industry are the aggregate of the correspondent 4-digit industries.

Commerce. In Colombia, registration at 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).

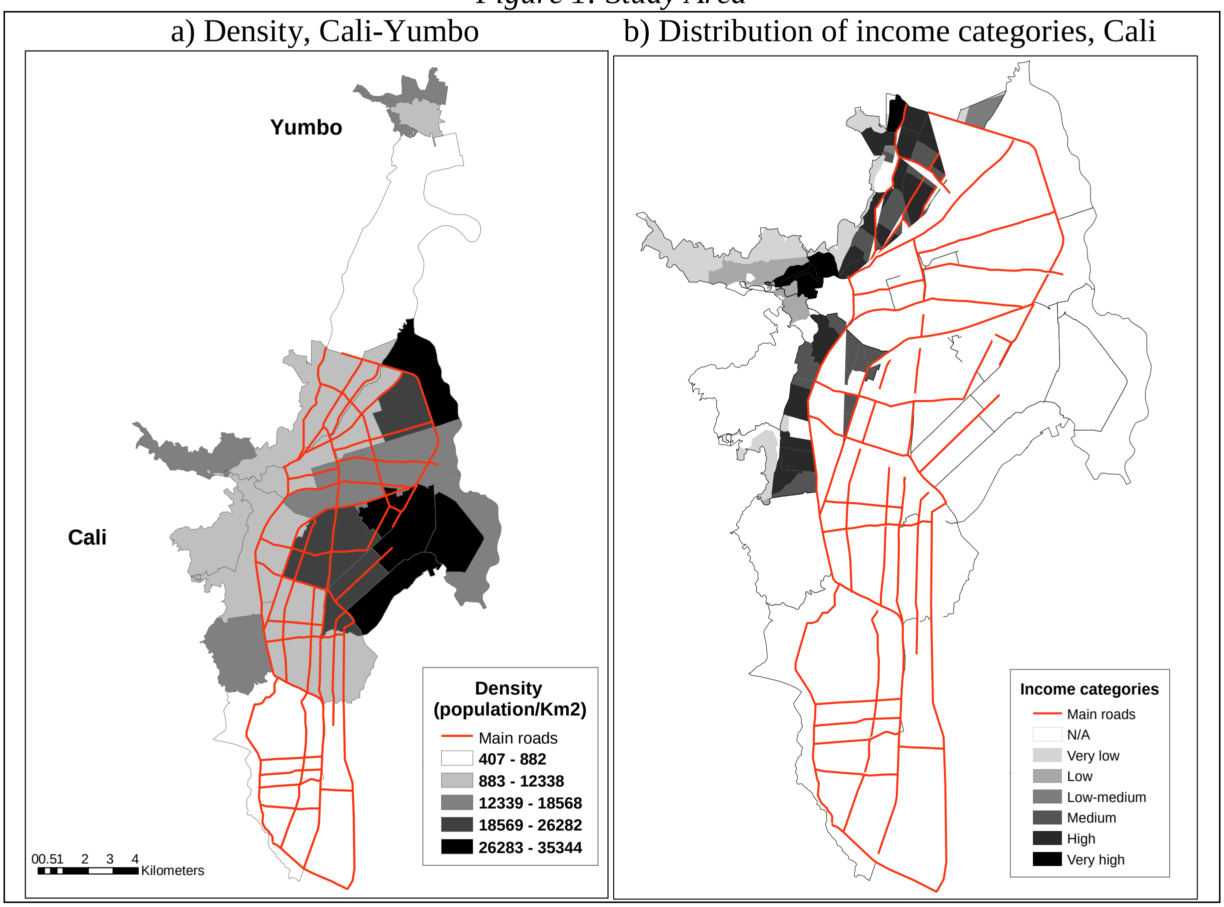
Several details regarding the quality of the data are worth mentioning. First, the census was collected door-to-door over the entire area of the study, a method that has several advantages: 1) "invisible" informal enterprises, i.e., those operating in households or shops without an external sign or banner could also be identified; 2) given that the Census covers *universe* of enterprises, sampling problems are ruled out (Cárdenas and Rozo, 2009); and 3) as the census does not rely on the prompt return of questionnaires, response rates are much higher. Second, before voluntarily answering the survey, business owners were informed that the information provided was fully confidential and that it would not carry any legal action against them. This way of proceding helps reducing the risk of false reporting and increase the owner's willingness to truthfully reveal their registration status and their compliance with other mandated regulations. In addition, in Colombia informality is not openly and widely persecuted, which also ensures that people were encourage to provide veridical information.

To carry out the comparison between formal and informal enterprises, we split formal enterprises into two categories according to the number of employees: medium-large and small. We define formal medium and large enterprises (FMLEs) as enterprises with more than 50 employees, and formal small enterprises (FSEs) as enterprises with 50 employees or less. The threshold of 50 employees is chosen based on two criteria: institutional reasons, and sample properties. Regarding the institutional criteria, according to the Colombian Micro, Small and Medium Enterprise Act 590 of 2004, enterprises with fewer than 50 employees benefit from specific incentives such as tax credits, reduced social contributions and loan interest rates. Regarding the sample properties criteria, a sample density distribution plot (not shown) revealed that most enterprises in the sample are small (see Table 1 below). Based on the density distribution, we also define a threshold of 10 workers as a robustness check to evaluate the sensitivity of the results to the formal enterprise size threshold.

3.2. Study Area: The metropolitan area of Cali

Santiago de Cali is the capital of the Valle del Cauca region (*departamento)* and third city of importance in Colombia. The metropolitan area of Cali is composed of the municipalities of Cali and Yumbo and is divided into over 14,000 *Manzanas.*[6] This last scale is comparable with the census or city-block level used in the US Census.[7] The metropolitan area stretches over 34 km (of which 24 km correspond to Cali) from north to south and 16 km from east to west. In 2005, it had a population of 2,164,098 people, and the average density was 17,217 people per sq. km. The largest and most densely populated areas of the city proper are on the center and east of the city (Figure 1). These densely populated areas are populated predominantly by low and low-middle income population, and have limited transport infrastructure. 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 and better equipped in terms of transport infrastructure.

Figure 1: Study Area

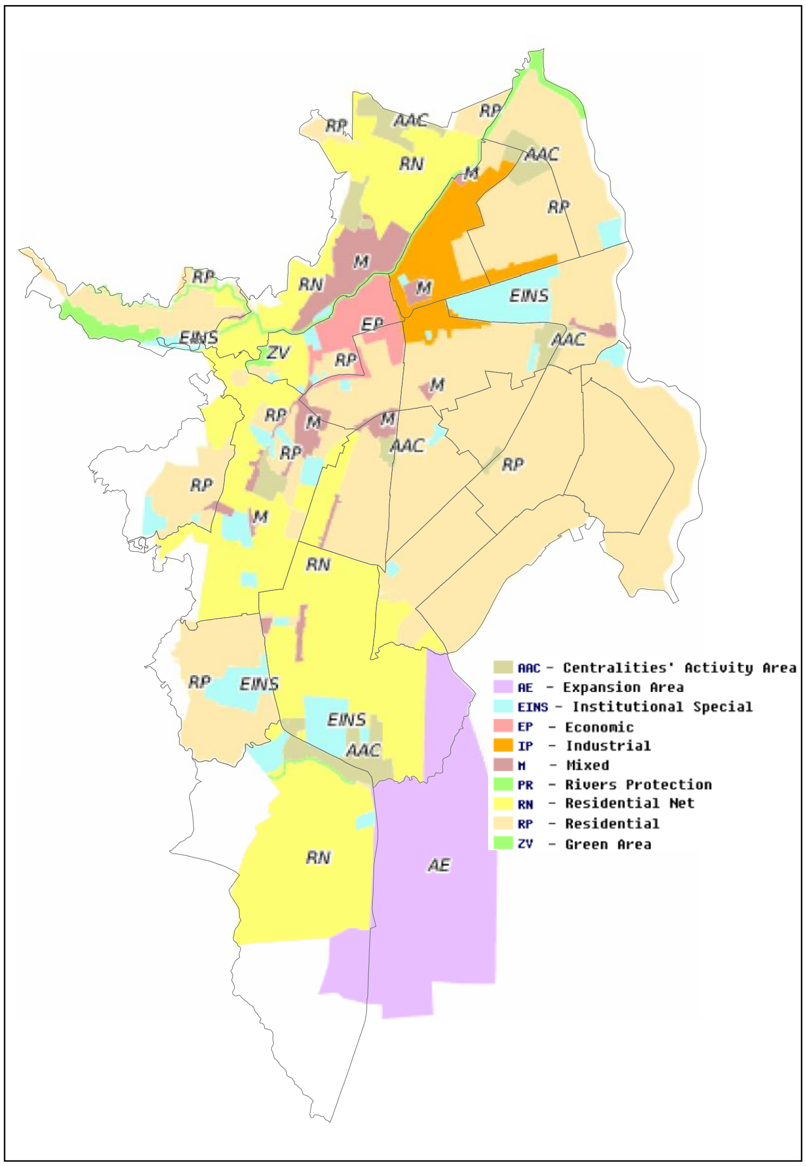


[6] Of which 1,125 (4 percent) are green areas (e.g. parks) or have no information available.
[7] For a definition see http://www.census.gov/geo/www/geo_defn.html#CensusBlock.

Source: Own elaboration based on IDESC data.
Note: Data for transportation routes and income in Yumbo is not available. Population density by Communes. Levels of income based on economic stratification in six categories (1=Very Low, 6=Very high)

In Cali, there are no specific zoning regulations for particular industrial sectors. The existing zoning regulations deal with the delimitation of residential and industrial land uses, and also environment-related zoning (e.g., river protection zones). There is an industrial zone located near the center of the city (Figure 2). A large proportion of the largest industrial enterprises are located in the corridor between Cali and Yumbo, which is also a designated industrial area.

Figure 2: Zoning



Source: Own elaboration based on IDESC data.

3.3. Manufacturing sector

The manufacturing sector is an important driver of output and employment creation in Cali-Yumbo. In 2005, it explained 16 percent of value added, 18 percent of the Gross Regional Product and 21 percent of total employment (Ministerio de Trabajo, 2011). Historically, the paper and food industries that use inputs from the fertile lands on the Cauca Valley have played a key role in the region's development. In 2005, the most important sectors in terms of employment are apparel (16 percent of total employment), other chemicals (16 percent) and other food products (7 percent). There are important differences in the scale of these activities, however. While the number of establishments in apparel explains 14 percent of the total industrial establishments, the number of establishments in other chemicals and other food products explains only 2 percent and 1 percent, respectively.

Informal manufacturing employment is highly prevalent, with 43 percent of total manufacturing workers employed in informal enterprises (Table 1). These informal enterprises operate at a very small scale. As can be seen in Table 1, roughly 99 percent of informal enterprises and 83 percent of formal enterprises can be classified as *micro* (1-10 workers). As Table 2 shows, the percentage of enterprises (45.6 percent) operating within the premises of a household is significantly larger for informal than for formal (45.6 versus 13 percent). The percentage of home-based informal activity is larger in apparel (69 percent) and spinning of textiles (68 percent), and lowest for printing (24 percent). On the other hand, the share of female labor is highest in textile and clothing-related activities, such as other textile products (73 percent) and apparel (68 percent), and lowest in carpentry, boat repairing (9 percent), and glass (11 percent). There seems to be a connection between home-based labor and female labor,[8] although this particular topic has not been addressed in the existing literature for Cali.

Table 1: Enterprise size distribution

	Formal	Informal	Total
Micro (1-10 workers)	2318 82.93%	2041 98.74%	4359 89.65%
Small (11-49 workers)	332 11.88%	24 1.16%	356 7.32%

[8] The correlation coefficient between the percentage of home-based enterprises and the share of female labor by industries is 27 percent.

Medium and Large (>50 workers)	145 5.18%	2 0.10%	147 3.02%
Total	2795 100%	2067 100%	4862 100%

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%

The industry in Cali caters mostly the local market and only around 7 percent of total revenue comes from exports.[9] The sectors with the highest percentage of export revenues are: printing (54 percent), apparel (28 percent), structural metal products (23 percent) and basic iron and steel (15 percent).

4. Methodology

4.1. The measurement of geographic agglomeration and coagglomeration

Before proceeding with the description of our measures of agglomeration and coagglomeration, it is worth clarifying two reasons why we do use cluster-based methods. First, measures based on this method are prone to suffer from the Modifiable Areal Unit Problem (MAUP), rendering the empirical results biased across geographical scales (Marcon and Puech, 2003a). Second, as noted by Duranton and Overman (2005; 2008), cluster-based measures are not able to properly capture size heterogeneity within industries.

To measure agglomeration and coagglomeration we use the M-functions, a distance-based method proposed by Marcon and Puech (2003a; 2010). This method produces a measure that fully satisfies the property of being unbiased across geographical scales, as well as the other four relevant criteria of a good measure of clustering (Duranton and Overman, 2005): 1) it is

[9] Source: DANE - Annual Manufacturing Survey 2006.

comparable across industries; 2) it controls for industrial concentration; 3) it controls for the overall aggregation pattern of industries; and 4) it is possible to test for the significance of the results. It also controls for inhomogeneous space, and consequently can account for the fact that enterprises cannot locate everywhere in the city because of the presence of parks, lakes, swamps, etc. This advantage of the M-functions is particularly relevant for intra-metropolitan studies of location, as it applies to both the extent and shape of the study area (Deurloo and de Vos, 2008).

An alternative method that satisfies these properties is the *Kd*-function used by Duranton and Overman (2005). As explained by Marcon and Puech (2012, p. 14), unlike the M-function, which relies on a probability density function (pdf) and is concerned with the question 'do externalities matter at a given distance?', the *Kd*-function relies on a cumulative distribution and is concerned with the question 'up to which distance do externalities matter?'. For the case of the intra-metropolitan location of formal and informal enterprises, both questions are relevant. Calculating both functions would indeed provide a complete picture of agglomeration and coagglomeration. Nevertheless, this approach is out of the scope of this work. Thus, in choosing between these two distance-methods, we prefer the M-function is because it is the only available distance-based method that allows for a straightforward interpretation and comparison of the value of the resultant indices (Marcon and Puech, 2003a; 2010).

4.2. M-functions agglomeration and coagglomeration

We calculate the intra and inter-industry M-functions for every 1 km between zero and 20 km at the industry level for formal enterprises, informal enterprises, FSEs and FMLEs.[10] We use the calculated plain coordinates (X-Y) for each enterprise at the city block level to measure the Euclidean distance among enterprises. In the following, we explain in more detail the definition of M-functions of agglomeration and coagglomeration.

4.2.1. Agglomeration

The M-function for intra-industrial spatial agglomeration 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. First we identify all enterprises belonging to sector *S* in the area of study. Here,

[10] We used the software Ripley v.2.8, designed by Eric Marcon and Florence Puech. Its features are now available as part of the R package dbmss.

a *sector S* refers to either an industry, a type of enterprise (formal medium and large, formal small or informal) or both (e.g., informal enterprises in the textile industry). For each of these enterprises, we draw a circle of radius *r* (e.g., 1 km). Within this distance, we count the number of employees belonging to enterprises in sector *S* (e_{iSr}). We then express the sum of this quantity over *i* as a proportion of the number of employees belonging to enterprises in all other sectors within the same circle (e_{ir}). Next we divide this ratio by sector *S'* employment weight 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. M-values *equal to one* indicate that whatever the considered distance, there are proportionally as many employees belonging to sector *S* as there are in the global area, or that there is 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 agglomeration 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*.

We calculate the statistical significance of the M-function by constructing confidence intervals for the null hypothesis of independence of enterprise locations, according to which the agglomeration pattern of enterprises in each sector is the same. We determine these intervals using Monte-Carlo methods in the following way: first, we generate a large number of simulations (1000 for the case of agglomeration and 100 for the case of coagglomeration). Next, we choose a confidence level of 5 percent so that the 95 percent confidence interval of M for each value of *r* is delimited by the outer 5 percent of the randomly generated values. There is significant relative agglomeration (dispersion) in a given sector if the corresponding M-values are larger (smaller) than one and are outside the confidence interval bands.

4.1.2. Coagglomeration

The inter-industrial version of the M-function, which has the same properties as the intra-industrial one described above, allows assessing the presence of coagglomeration. M-functions of coagglomeration for sectors S_1 and S_2 are defined as:

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

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

$M_{S_1S_2}(M_{S_2S_1})$ depicts the spatial structure of enterprises belonging to sector S_2 (S_1) that are found around sector S_1 (S_2). Thus, these 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 enterprises S_2 (S_1) located around those of sector S_1 (S_2) is larger or smaller 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 different given that the null hypothesis has to control for both S_1 and S_2 patterns. To control for the pattern of S_1, we generate the null-hypothesis point set for $M_{S_1S_2}(r)$ by keeping S_1 points unchanged and redistributing all other points onto all other locations. We follow the same process for S_2. There is significant coagglomeration whenever both values are significantly different from their respective null hypothesis (Marcon and Puech, 2003a).

4.2. Spatial analysis

The M-indicators provide useful information regarding agglomeration and coagglomeration patterns of types of enterprises and/or industries, but they are not informative of the actual spatial distribution of enterprises within the metropolitan area. As an example, while the indicators may show significant coagglomeration 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.

Based on the calculated plain (X-Y) coordinates for each enterprise, we perform the spatial analysis consist on a kernel density mapping for industries and type of enterprises selected based on the M-functions results. The Kernel estimator is given by:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right), \qquad (4)$$

where h is the chosen radius or smoothing parameter, $X_1 \ldots X_n$ are enterprises, n is the number of enterprises and K is the quadratic kernel function. We identify the q kernel density surface for each individual point with the highest value at its location. Beyond h, the density surface for a given enterprise is zero. We then obtain a continuous, smooth density surface across the entire study area by adding these individual density surfaces (Silverman, 1986). The smoothing parameter, or search radius distance, is set to 500 meters. Lastly, we map this density surface.

5. Results

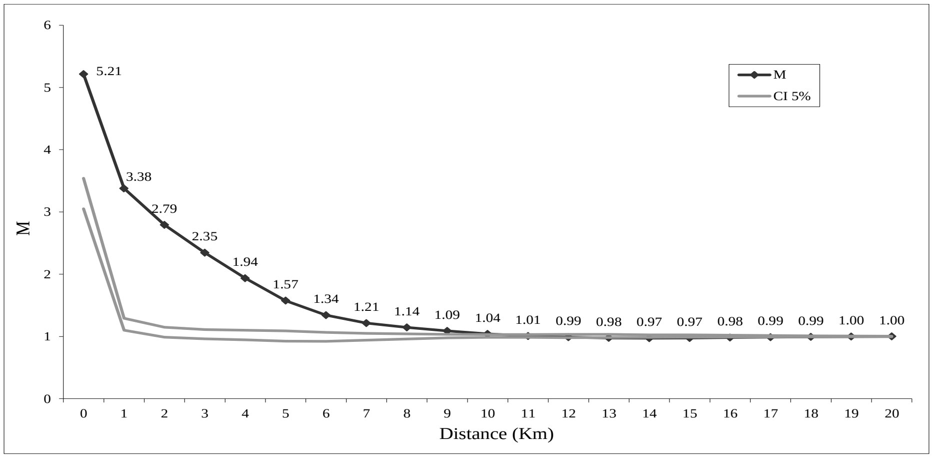
5.1. Spatial indicators of relative geographic agglomeration and coagglomeration

This section presents the results for the spatial indicators of agglomeration and coagglomeration. All of the intra and inter-industry M-functions are calculated using information for all of the 3-digit level industries in the sample (64 industries in total). However, sectors in which the number of enterprises is less than 10 are not shown, as the number of enterprises in these cases is too small to yield reliable predictions of agglomeration patterns.

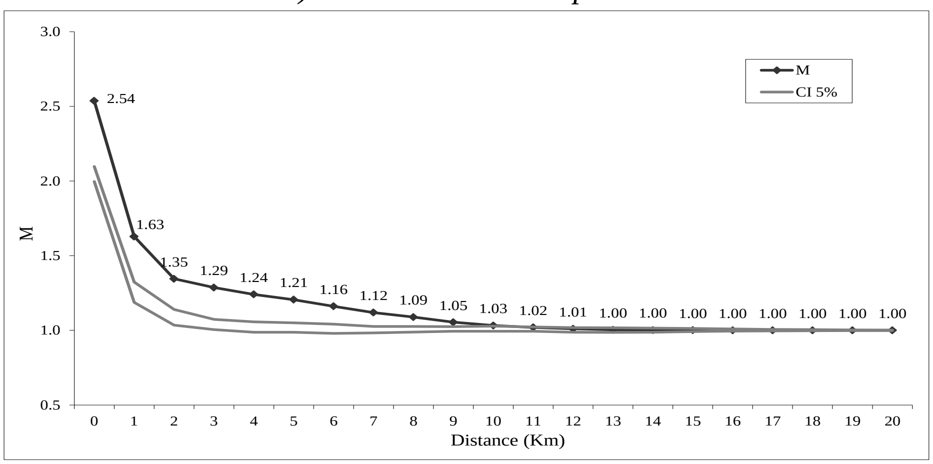
5.1.1. Agglomeration by industry

Before delving into the results by industry, it is relevant to comment on some aggregate results by enterprise type. Figure 3 shows the M-functions for the aggregate of informal, formal small and formal large enterprises. Formal small and informal enterprises show significant agglomeration at distances between 0 and 9 km. For both types of enterprises, the M-functions peak at 0-1 km and then show a continuous decay as distance increases. Finally, as expected, the degree of agglomeration for larger formal enterprises is much smaller. Interestingly, the degree of agglomeration is larger for informal enterprises than for formal enterprises of comparable size: at the peak M-value, the relative density of employees in informal enterprises in a radius of less than 1 km is over 5 times higher than that in the whole area, whereas that of formal small enterprises is only 2.5 times higher. These results indicate that, as predicted by theory, small enterprises agglomerate, and that this agglomeration takes place at fairly small scales.

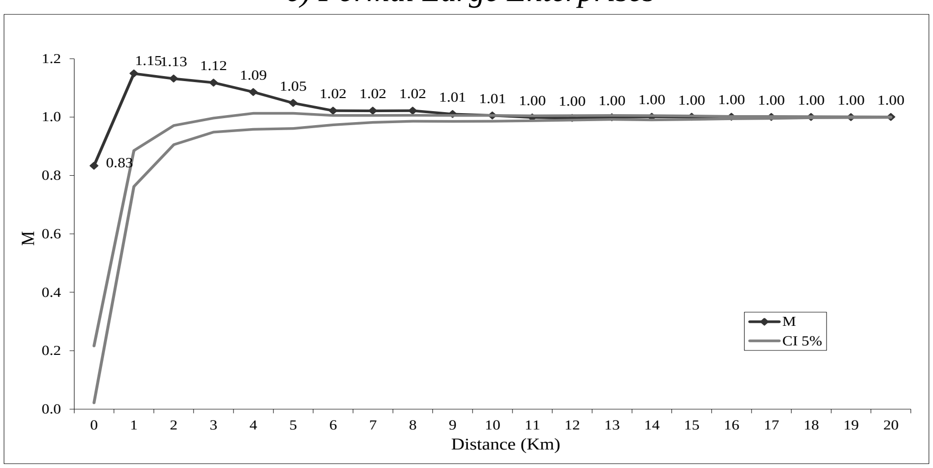
Figure 3: Intra-industry M-functions by type of enterprise
a) Informal Enterprises



b) Formal Small Enterprises



c) Formal Large Enterprises



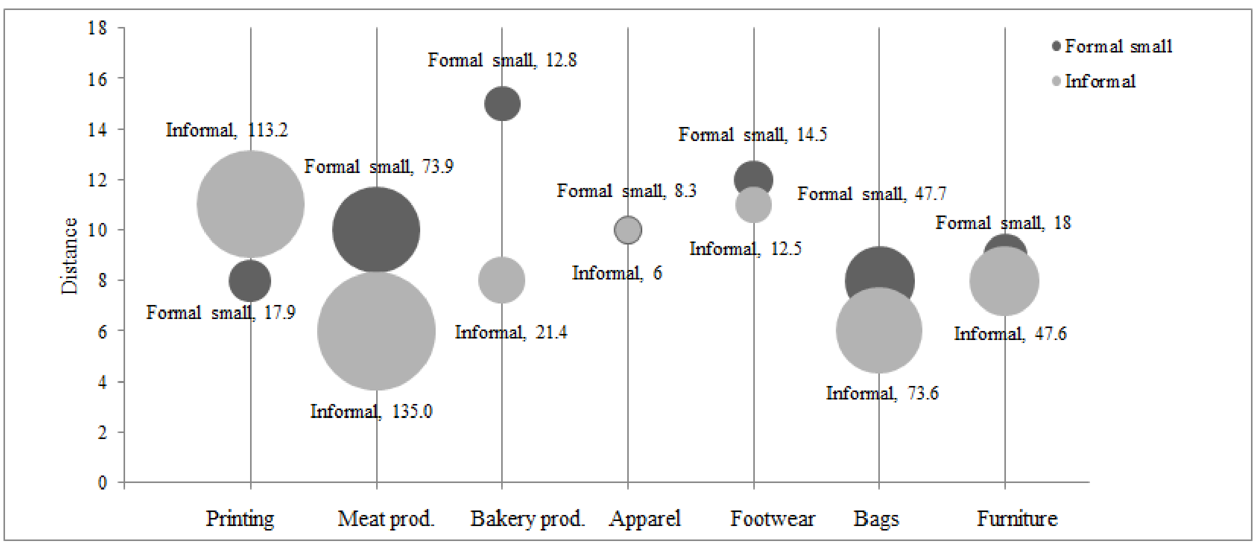
As predicted by the new industrial spaces approach, agglomeration is far more common for formal small and informal enterprises than for large formal enterprises.[11] Informal enterprises exhibit significant agglomeration in all of the 27 industries where formal small enterprise

[11] For reasons of space, we do not include the industry-level results tables, but they are available upon request.

agglomeration is significant, while the opposite is not true for three industries (beverages, non-metallic products and knitted and crocheted fabrics). The strength of agglomeration of small enterprises is indeed remarkably high in some industries. The M-peak values are larger for informal enterprises in 19 out of 27 cases, and this difference ranges from 62 to 0.08 times (other textile products and spinning of textiles, respectively). Because existing studies using the M-functions exclude small enterprises (less than 10 or 20 workers) from their samples, we do not have a benchmark against which we can compare these results. Although some of the M-values are much higher than the maximum found in other studies, they do seem plausible given the small size of informal enterprises.

Out of the 27 industries where both formal small and informal enterprises show significant agglomeration, seven are significant for a continuous range of distance of at least 1 km and display M-peaks at the same distance for formal small and informal enterprises. Figure 4 plots distance ranges and M-peak values by industry. Out of these seven industries, the strength of agglomeration is larger for informal enterprises than for small formal enterprises in five industries (printing, tanning, dressing and processing of leather, furniture, meat, mill and bakery products), while the opposite is true for two industries (apparel and footwear).

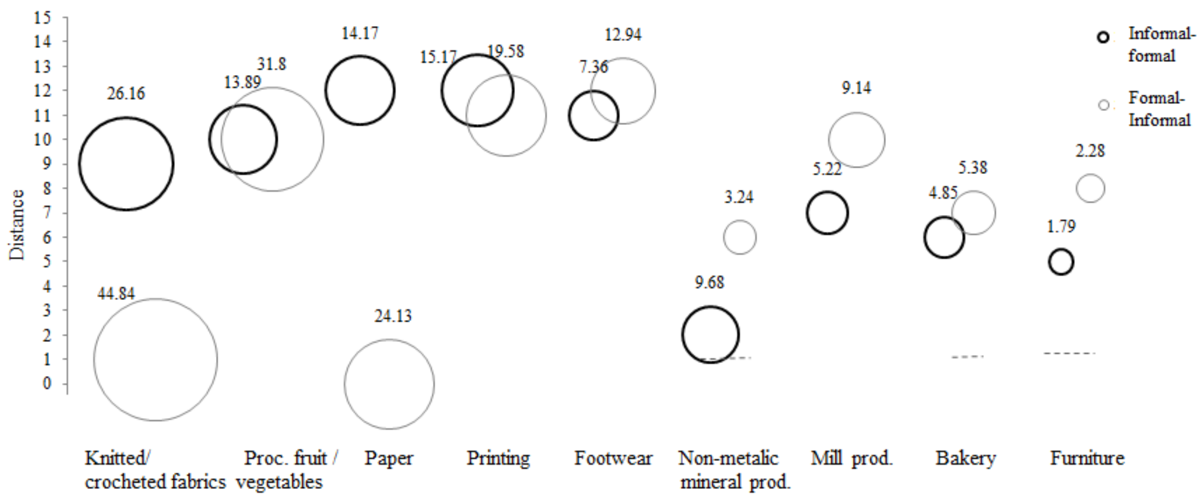
Figure 4: Intra-industry M-functions, selected industries



5.1.2. Coagglomeration of formal and informal enterprises

As expected, the coagglomeration results differ starkly by enterprise size. On the one hand, we find that only 3 percent of all possible coagglomeration pairs between either formal small and large enterprises, or informal and large enterprises, are statistically significant. On the other hand, 27 out of 30 industries display significant coagglomeration between formal small and informal enterprises. As shown in Figure 5, 9 of these 27 industries show significant coagglomeration for a continuous distance range of at least 1 km, and display M-peaks at the same distance for both types of enterprises. These results are in line with those of Mukim (2014). Using a cluster-based measure of coagglomeration of formal (of all sizes) and informal enterprises by Indian districts in 2005, she finds relatively weak coagglomeration relationships, with the highest occurring in food products, wood, apparel, fabricated metal products and other metallic products.

Figure 5: Inter-industry M-functions, selected industries



Note: Dashed line indicates the lower bound of the distance range (if absent this value is equal to zero)

We can now compare the agglomeration and coagglomeration results for formal small and informal enterprises shown in Figures 4 and 5. On the one hand, four industries (printing, bakery products, footwear and furniture) display both significant agglomeration and coagglomeration. This may be an indication that in these industries, formal small and informal enterprises co-locate in the same areas. On the other hand, for three industries (meat products, apparel and bags), we observe significant agglomeration but no significant coagglomeration. In this case, agglomeration of formal small and informal enterprises in these industries may be happening in different areas of the city. We analyze these possibilities in section 5.2.

5.1.3. Results with a different size threshold

In order to assess the sensitivity of the results to the formal enterprise size threshold, we calculate agglomeration and coagglomeration M-functions for formal micro enterprises with 10 workers or less (FMicro), formal enterprises with more than 10 workers (FSMLEs), and informal enterprises (IEs). As explained before, this threshold is chosen based on the sample size distribution.[12]

Regarding agglomeration, we start by comparing the results for informal and formal micro enterprises. We find significant informal enterprise agglomeration in all industries displaying significant formal micro enterprise agglomeration. The opposite is not true in two cases (beverages, and basic chemicals). Although there are minor changes in the rankings, informal enterprises still display larger M-peak values than formal micro enterprises in a larger proportion of industries (19 out of 28). As before, the M-peak values of agglomerations of formal micro enterprises are larger than those of informal enterprises in eight industries, although the composition of this group is somewhat changed.[13] For intra-industry functions that are significant for a continuous range of distance of at least 1 km and display M-peaks at the same distance for formal and informal enterprises, the same seven industries in Figure 4 appear, together with five more industries. Next, comparing formal micro and formal small enterprises shows that, as expected, formal micro enterprises display larger M-peak values.

The results for coagglomeration are broadly in line with those discussed in section 5.1.2. Coagglomeration in three out of the 27 industries (non-metallic mineral products, watches and clocks and recycling of non-metal waste) was no longer significant after changing the size threshold. Furthermore, a new industry displaying significant coagglomeration appeared (beverages). 14 out of 30 industries display M-peaks at the same (continuous for at least 1 km) distance for formal and informal enterprises. The composition of this group is somewhat different from that in Figure 5, but coagglomeration remains significant for most industries.

In conclusion, the agglomeration and coagglomeration results seem to be robust to the change in size threshold, which indicates that the observed differences between formal and informal enterprises are not due to differences in size.

[12] The full set of industry level results are available upon request.
[13] Specifically, bakery and watches and clocks appear while joinery and carpentry and casting of metals drop.

5.2. Spatial analysis

In this section, we show kernel density maps in order to visualize the difference between: 1) industries where both formal small and informal enterprises agglomeration and coagglomeration is significant and 2) industries where agglomeration is significant and strong but coagglomeration between formal small and informal enterprises is not significant (or significantly weaker). For the spatial analysis to be robust, the analysis is limited to industries with over 100 enterprises in each enterprise category, namely footwear, furniture and bakery products on the one hand, and apparel on the other.

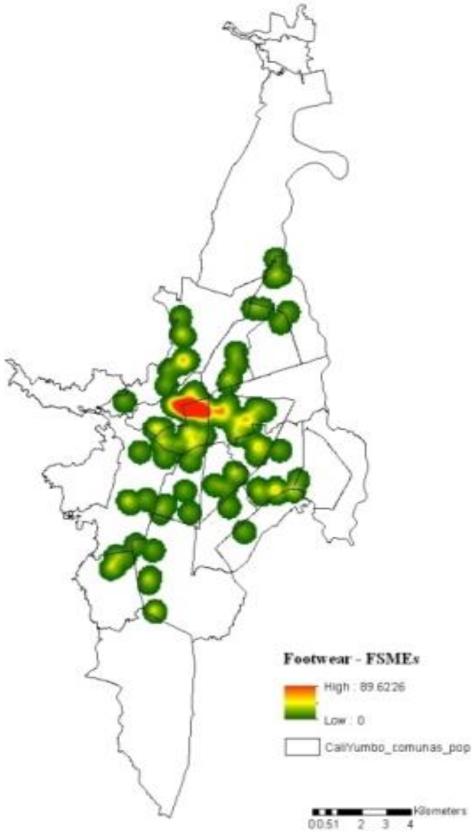
The kernel density mapping for the first group of industries shows significant overlap between the red spots indicating the highest density of enterprises for both types of enterprises (Figure 6). Interestingly, in the case of footwear, formal small and informal enterprise clusters overlap in the center of the city, where there is a dense network of consumers and suppliers. The footwear industry in Cali consists mostly of technologically-lagged, small-scale enterprises that cater mostly the local market and face strong international competition (SENA, 2004). Although we do not have sufficient data to assess the relative relevance of international orientation in shaping the productive geography of cities, it is a topic worth of further investigation.

The bakery industry displays a markedly different pattern than the footwear industry. As Figure 6 shows, in this industry clusters of formal small and informal enterprises are more dispersed in space, and coagglomeration of both types of enterprises occurs in a peripheral area characterized by a high density of lower income consumers. This points to an interesting result: coagglomeration of formal and informal enterprises is not only observed because informal enterprises locate in areas which are the domain of formal enterprises (i.e., areas with better accessibility and higher incomes), but also because formal small enterprises locate in peripheral areas. It is worth noticing that the demand for bakery products in Cali is extremely localized because of the nature of the product, meaning that the market for each enterprise is reduced to a couple of blocks (SENA, 2006b). In this case, demand-driven factors may drive the observed pattern, although we do not have sufficient information to test this hypothesis.

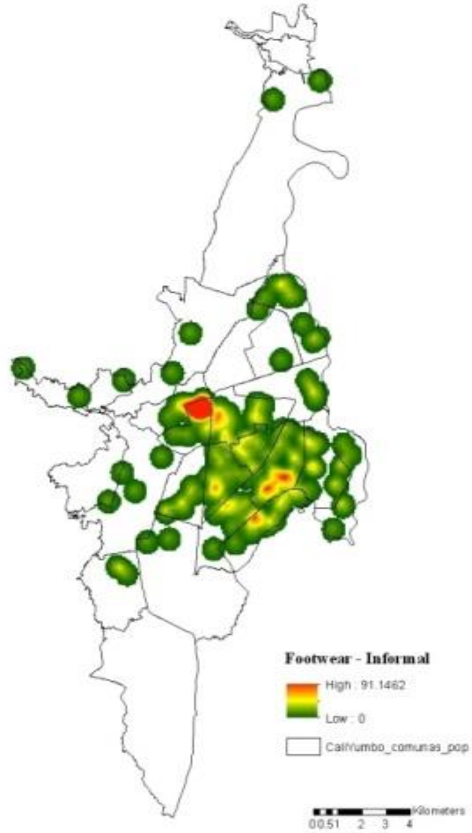
Figure 6: Kernel density maps, selected industries

a) Footwear

Formal small

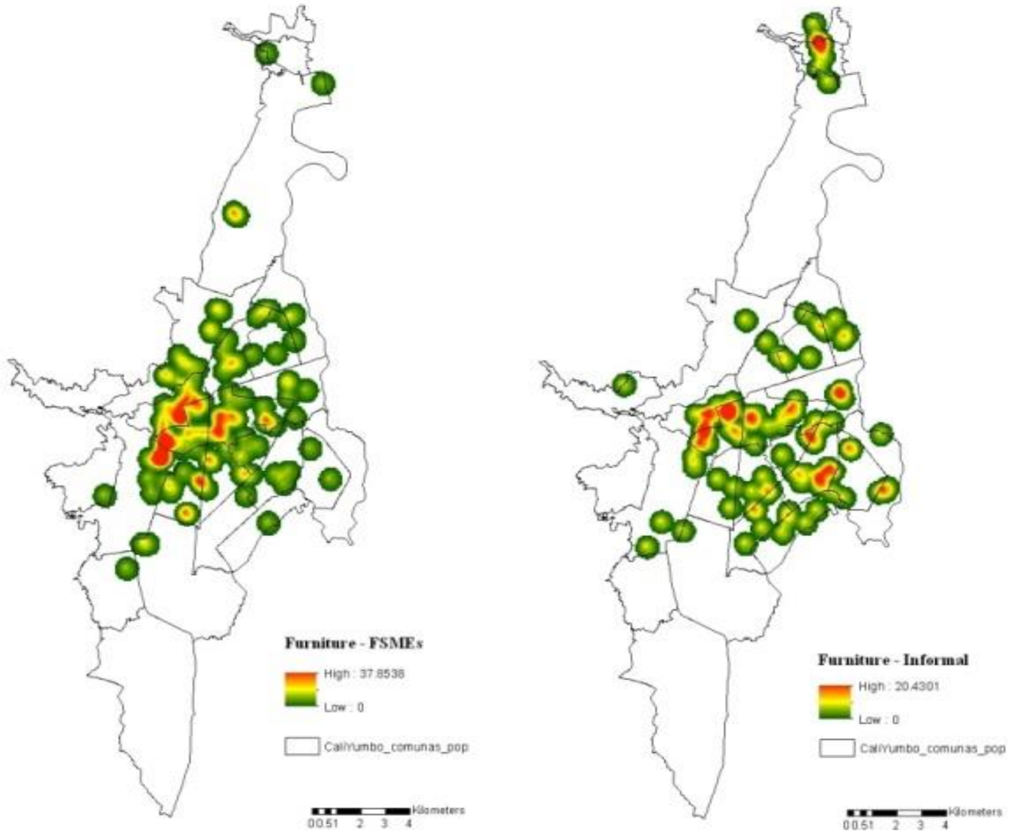


Informal



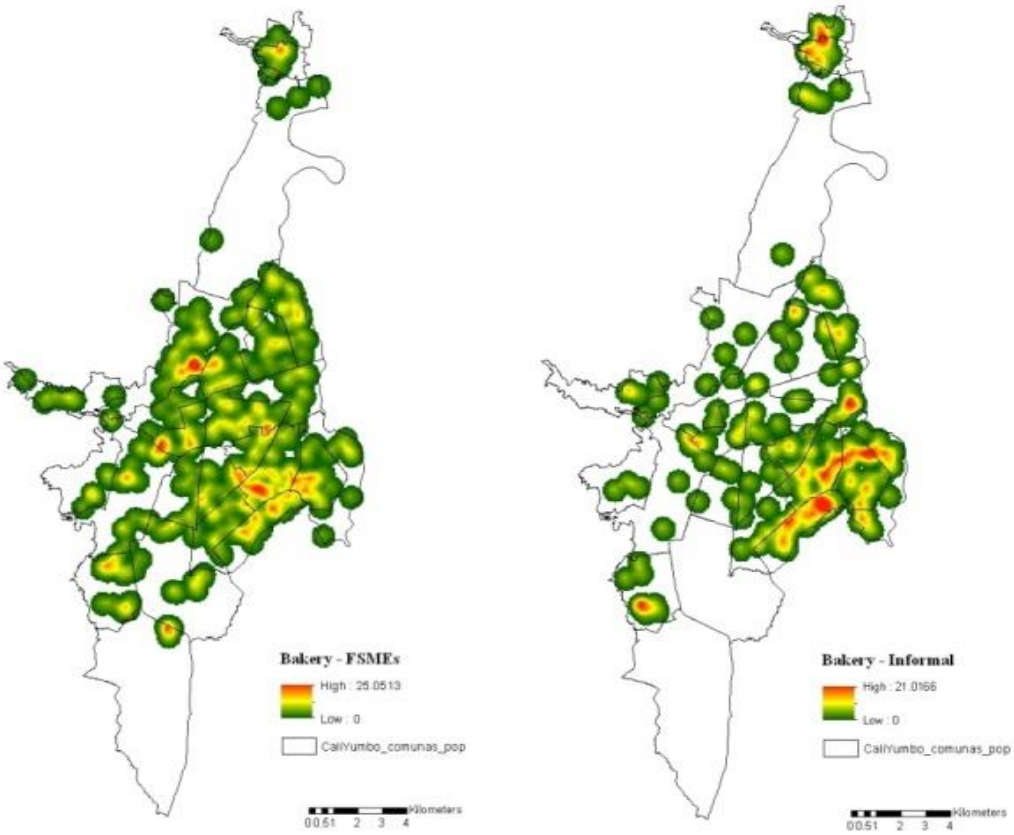
b) Furniture

Formal small Informal



c) Bakery products

Formal small Informal

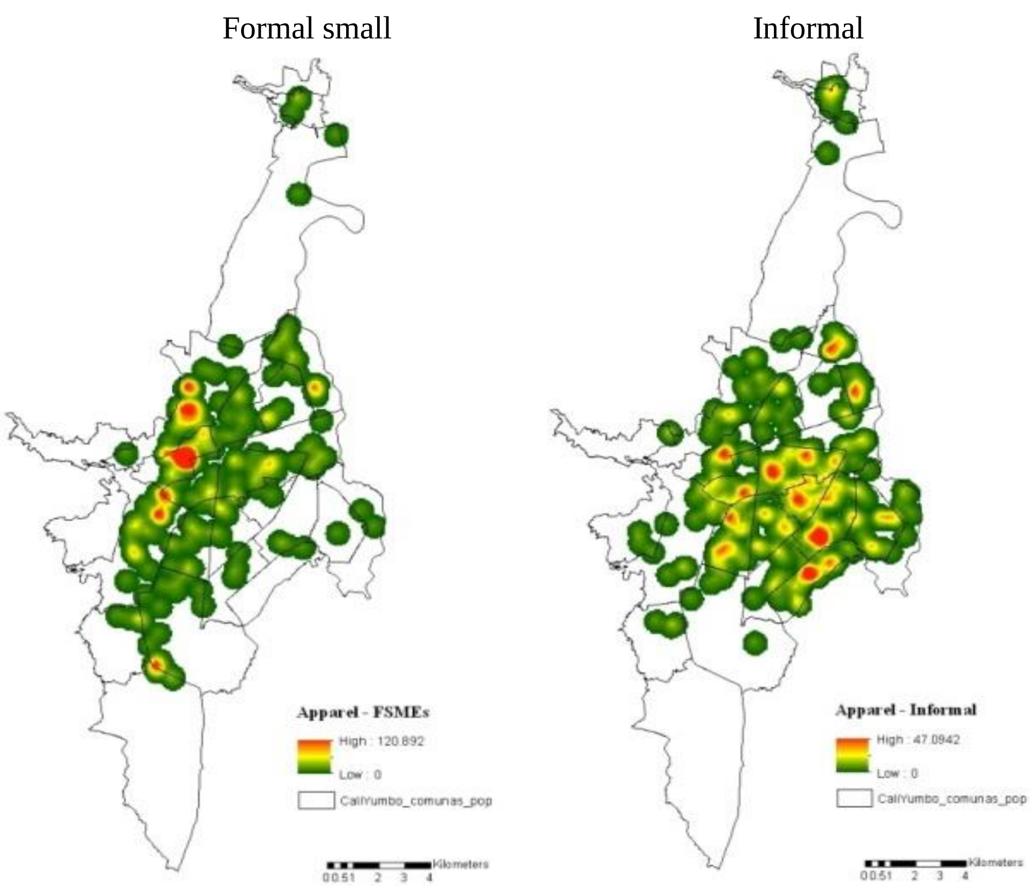


These cases can be contrasted with the case of apparel, where both formal small and informal enterprises displayed significant agglomeration but where no coagglomeration of the two types of enterprises was found.[14] As Figure 7 shows, there is clear spatial segmentation in the sense that formal small and informal enterprises locate in different parts of the city, as indicated by different areas of high density for formal small and informal enterprises. The apparel industry, an internationally-oriented industry in Cali, is characterized by vertically disintegrated value chains which often make use of home-based, cheaper (mostly female) informal labor (SENA, 2006a). In spatial terms, this may be reflected on the high density points found mostly in lower-income

[14] The inter-industry results with a different threshold indicate significant, albeit very small, coagglomeration of formal micro and informal enterprises in the apparel industry. Here reference is made to the results for formal small enterprises.

areas of the city (see Figures 1 and 2). Again, we do not have enough information to establish this with certainty, but exploring the spatial consequences of subcontracting is a research avenue worth pursuing.

Figure 7: Kernel density, Apparel formal small and informal enterprises



6. Discussion and Conclusions

Our main goal in this article was to understand the spatial distribution of informal enterprises vis-à-vis formal enterprises. Coming back to our first research question, do formal and informal enterprises display different agglomeration patterns, our results show that both types of small enterprises display a strong tendency to cluster in the same industries. As predicted by theory, this tendency is much stronger than that of larger formal enterprises. Although for the industrial sector as a whole informal enterprises display a larger degree of agglomeration than formal enterprises of similar size, in some industries, agglomeration is stronger for formal small enterprises. It would be interesting to investigate in the future whether agglomeration is driven by the same forces for both types of enterprises, and what explains the differences in agglomeration strength for different industries.

In our second research question we asked whether formal and informal enterprises coagglomerate in the same areas of the city. Our analysis showed that, in general terms, formal and informal enterprises of similar size are found in proximity to each other, whereas larger formal enterprises display location patterns that seem independent of enterprises of smaller size. We find that in some industries, formal small and informal enterprises seem to coagglomerate in the same areas of the city. However, in some other cases such as the apparel industry, agglomeration of both types of enterprises can happen in parallel in different areas of the city (for instance simultaneously in central and peripheral areas).

Our results indicate whether coagglomeration of formal and informal enterprises occurs, but are not informative about underlying mechanisms. Disentangling these mechanisms is particularly challenging (Rosenthal and Strange, 2009; Ellison et al., 2010) and in our case, not possible with the information available. In particular, it would be interesting to investigate, at an intra-metropolitan level, to what extend the observed locational patterns are driven by supply-side or demand-side mechanisms (Mukim, 2011; 2014) or by other factors (Chakravorty et al., 2005); whether informal enterprises benefit to the same extent of agglomeration externalities within cities, and under which circumstances they can become a source of negative externalities (Sokty and Newman, 2014; Duranton, 2008); what types of linkages between formal and informal producers in related industries matter for co-location (Mukim, 2014); and finally, how the rise of subcontracting and Global Value Chains affects the distribution of formal and informal production in the urban space, given an asymmetric relation between formal and informal producers (Chen, 2007; Moreno-Monroy, 2012).

Finally, the evidence found here suggests that it is also relevant to consider the sectoral and spatial heterogeneity of the urban informal sector in the design of informal sector policies. In Cali, as in most cities, past and current formalization efforts have been focused on transversal policies such as tax-break incentives, facilitation and minimization of procedures. While these efforts are important in increasing the benefits of formality (Chen, 2007), the study of agglomeration patterns and their determinants can guide the design of appropriate policies to unleash possible constrains related to locational restrictions and agglomeration diseconomies experienced in different degrees by informal industries (Harris, 2014), and minimize possible diseconomies of informal enterprise agglomeration on formal small enterprises.

References

Arauzo-Carod, J.M and Viladecans-Marsal, E. (2009) Industrial location at the intra-metropolitan level: The role of agglomeration economies. *Regional Studies* 43(4): 545-558

Cárdenas, 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): 123–142

Chakravorty, S., Koo, J. and Lall, S.V. (2003) Metropolitan industrial clusters: patterns and processes. *Policy Research Working Paper Series* 3073, The World Bank

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

Chen, M.A. (2007) Rethinking the informal economy: Linkages with the formal economy and the formal regulatory environment. *DESA Working Paper 46*, United Nations Economic and Social Affairs

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

Deurloo, M.C. and de Vos, S. (2008) Measuring segregation at the micro level: an application of the M measure to multi-ethnic residential neighbourhoods in Amsterdam. *Tijdschrift voor Economische en Sociale Geografie* 99(3): 329-347

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

Ellison, G., Glaeser, E.L. and Kerr, W.R. (2010). What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review* 100: 1195-1213

Harris, J. (2014) The messy reality of agglomeration economies in urban informality: Evidence from Nairobi’s handicraft industry. *World Development* 61:102-113

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

Mehrotra, S. and Biggeri, M. (2005) Can industrial outwork enhance homeworkers' capabilities? Evidence from clusters in South Asia. *World Development* 33(10): 1735-1757

Ministerio de Trabajo (2011) *Programa Nacional de Asistencia Técnica para el Fortalecimiento de las Políticas de Empleo, Emprendimiento y Generación de Ingresos en el ámbito Regional y Local: Diagnostico Cali*. Ministerio de Trabajo, Bogotá

Marcon, E., and Puech, F. (2003a) Evaluating the geographic concentration of industries using distance-based methods. *Journal of Economic Geography* 3(4): 409-428

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

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. (2012) A typology of distance-based measures of spatial concentration. *HAL Open Archive*, March

Moreno-Monroy, A.I. (2012) Critical commentary: Informality in space: Understanding agglomeration economies during economic development. *Urban Studies* 49 (10): 2019-2030

Moreno-Monroy, A.I. and Posada, H.M. (2012) On the effect of transport subsidies on informality rates. *CEPESP Discussion Paper*, Fundação Getulio Vargas

Mukim, M. (2011) Industry and the Urge to Cluster: A Study of the informal sector in India. *SERC Discussion Papers* 0072, Spatial Economics Research Centre, LSE

Mukim, M. (2014) Coagglomeration of formal and informal industry: evidence from India. *Journal of Economic Geography*, forthcoming

Perry, G., Maloney, W., Arias, O., Fajnzylber, P., Mason, A., and Saavedra-Chaduvi, J. (2007) *Informality: Exit and exclusion*. Washington DC, World Bank, Latin American and Caribbean Studies

Rosenthal, S. and Strange, W. (2003) Geography, industrial organization, and agglomeration. *The Review of Economic and Statistics* 85(2): 377-393

SENA (2004) *Cadena productiva del cuero. Mesa sectorial del cuero, calzado y marroquinería*. Servicio Nacional de Aprendizaje, SENA, Medellín

SENA (2006a) *Caracterización ocupacional cadena textil-confección*. Servicio Nacional de Aprendizaje, SENA, Medellín

SENA (2006b) *Caracterización ocupacional de la industria de la panificación y la repostería*. Servicio Nacional de Aprendizaje, SENA, Cali

Sethuraman, S.V. (1997) *Urban poverty and the informal sector: A critical assessment of the current strategies*. Geneva: ILO

Scott, A. (1988) *Metropolis: From the division of labour to urban form*, Berkeley, CA: The University of California Press

Silverman, B.W. (1986) *Density estimation for statistics and data analysis*. New York: Chapman and Hall

Sokty, C and Newman, C. (2014) Clustering, competition, and spillover effects: Evidence from Cambodia, WIDER Working Paper, No. 2014/065

Visser, E.J. (1999) A comparative analysis of clustered and dispersed firms in the small-scale clothing industry in Lima. *World Development* 27(9): 1553-1570

Visser, E.J. (2009) The complementary dynamic effects of clusters and networks. *Industry and Innovation* 16(2): 167-195

Visser, E.J. Tavara, J.I. and Villaran, F. (2014) Growing but not developing: Long-term effects of clustering in the Peruvian clothing industry. *Tijdschrift voor Economische en Sociale Geografie*, 106(1): 78-93

WTO and ILO (2009) *Globalization and informal jobs in developing countries: A joint study of the International Labour Office and the Secretariat of the World Trade Organization*. Geneva: WTO Publications, World Trade Organization