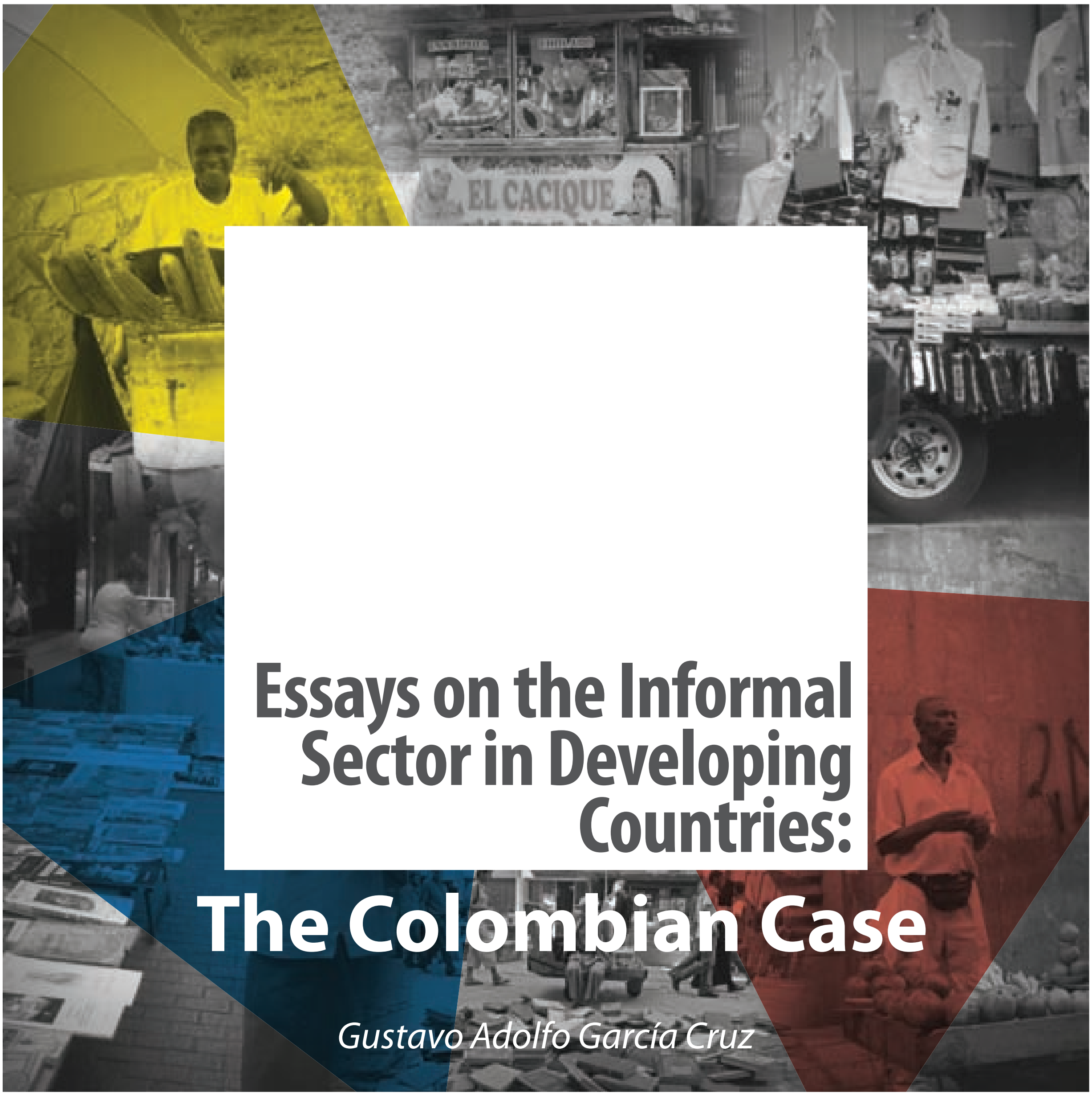
Universitat Autònoma de Barcelona



Essays on the Informal Sector in Developing Countries: The Colombian Case

Gustavo Adolfo García Cruz

Thesis in Economics

Universitat Autònoma de Barcelona
Departament d'Economia Aplicada

Essays on the Informal Sector in Developing Countries: The Colombian Case

Tesis para optar por el grado de

Doctor en Economía

Autor: Gustavo Adolfo García Cruz

Director: Dr. José Luis Roig Sabaté

Doctorado en Economía Aplicada
Departament d'Economia Aplicada
Facultat d'Economia i Empresa
Universitat Autonòma de Barcelona
2013

Agradecimientos

En primer lugar, agradezco especialmente a mi director de tesis el profesor José Luis Roig por todo su apoyo y consejos durante todo el proceso de investigación. Su colaboración, conocimientos y sobre todo su confianza depositada en mí fueron fundamentales para el desarrollo de esta tesis. También agradezco de manera especial al profesor Josep Lluís Raymond por su apoyo. Su experiencia, conocimiento y agudeza analítica hacían de cada reunión una rica clase de economía y econometría.

Gracias al Departamento de Economía Aplicada de la UAB por estos años que me alojó y apoyó. A los profesores y compañeros del doctorado con los que compartí muchas experiencias tanto dentro como fuera de las aulas.

Gracias a Ana I. Moreno por las discusiones alrededor del tema del sector informal. Su perseverancia e intensidad en el trabajo es algo admirable.

Agradezco infinitamente el apoyo de toda mi familia que desde la distancia siempre me demostraron todo su cariño y apoyo. Gracias a mi mamá, a mi papá y a mis hermanos quienes siempre me motivaron. La distancia no fue un obstáculo para enviarme su cariño, ánimo y estar junto a mí en este proceso. A ellos dedico esta tesis. Gracias totales a Erika, mi amiga incondicional, novia y esposa, gracias por tu apoyo, por estar conmigo y animarme cuando lo necesitaba. Lo mejor de todo fue haber estado juntos.

Finalmente, agradezco la financiación de la Agència de Gestio d’Ajust Universitaris i de Recerca (AGAUR) y el apoyo brindado por la Universitat Autònoma de Barcelona.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

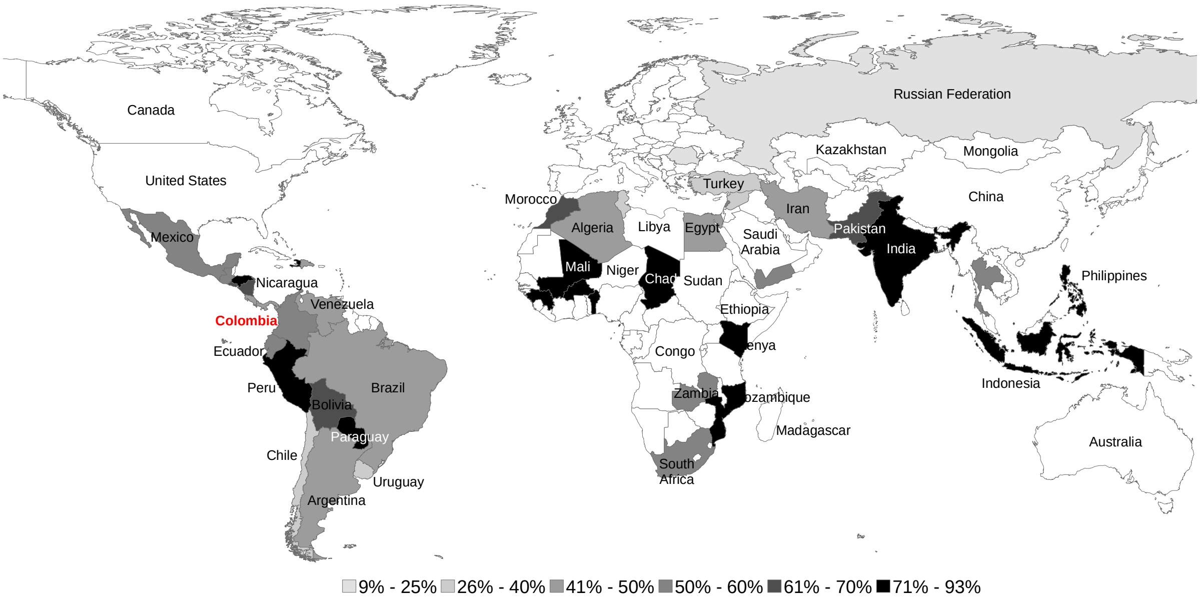
A frequently observed feature of developing countries is the scarcity of employment opportunities and the poor quality of the jobs that are available, with the result that high levels of both unemployment and informality characterize today their labor markets. While considerable attention has been devoted to unemployment, given its obvious impact on economic growth and development, an extensive body of literature has shown that informality can likewise represent a major burden to growth and social well-being (see, e.g., Perry *et al*., 2007; Jütting and De Laiglesia, 2009). As Jütting *et al*. (2008) have pointed out, the creation of good quality jobs is a key element in transforming economic growth into a process that effectively reduces inequality and poverty. The three empirical studies that make up this thesis seek to analyze and characterize this informality and, more importantly, to study the inherent heterogeneity of this phenomenon in the developing country of Colombia.[1]

In developing countries the informal sector employs a significant share of the workforce and accounts for a sizable part of economic output (see Figures 1.1 and 1.2). According to Jütting and De Laiglesia's (2009) estimates, over 55% of non-agricultural employment in developing countries is performed in activities not regulated or protected by the state (informal activities). In some countries of Sub-Saharan Africa and South Asia the proportion of informal workers can be higher than 80% (for example, Chad and India present informality rates of 95 and 83%, respectively). In Latin America the average informality rate stands at around 50%, with Honduras presenting the highest rate (76%), followed by Paraguay (70%), Peru (70%), Bolivia (69%) and Colombia (60%). As for the size of the informal economy measured as a percentage of GDP, Schneider *et al*. (2010) show that in Latin America in 2007 the shadow economy accounted for around 40% of GDP, while in Sub-Saharan Africa and South Asia it represented 39 and 30% respectively. Latin American and Sub-Saharan African

[1] More extensive research on the informal economies of other countries can be found in the ILO's Informal Economy Resource Database, available at http://www.ilo.org/dyn/dwresources/iebrowse.home?p_lang=en.

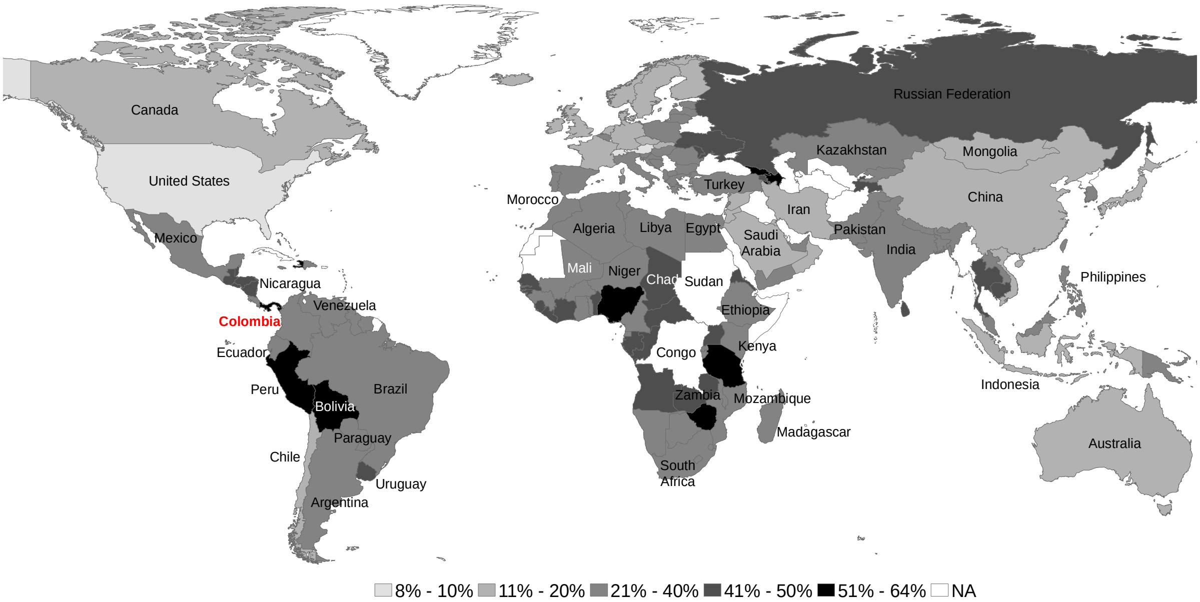
countries such as Bolivia, Peru and Zimbabwe head the rankings in this regard with percentages above 60%.

Figure 1.1. Percentage of informal employment in nonagricultural employment



Note: Data for Latin American countries correspond to the year 2010 (see ILO, 2011). Data for remaining countries correspond to the period 2000-2007 (see Jütting and De Laiglesia, 2009).

Figure 1.2. Share of informal economy in total GDP, 2007



Source: Schneider *et al*. (2010).

As can be seen, therefore, the informal sector in developing countries is large and has a crucial role to play. Much of the informal sector is associated with subsistence activities where individuals undertake all kinds of work just to ensure their own survival and that of their families. Most of the individuals who fail to obtain formal employment live in varying degrees of poverty and, because of the limited coverage and effectiveness of state social security systems, they must turn to the informal sector as their only alternative source of employment. The activities they encounter there are characterized by low wage, high levels of risk and vulnerability, and highly precarious conditions and, thus, persistence of informality means the perpetuation of poverty for a large segment of the population.

From a social perspective, high levels of informality undermine the ability of the state to protect employees (ensuring appropriate working hours and wages, among others) and to collect taxes. As such, there are not guarantees of the equality of opportunity in the labor market, or of the efficient functioning of the social security system and of social well-being in general (Jütting and De Laiglesia, 2009).

Furthermore, as mentioned above, informality can have particularly detrimental effects on a country’s economic growth, productivity and competitiveness. Informal firms are often small in scale and geographically dispersed making it easier for them to avoid detection. Additionally, they tend to be characterized by low-skilled, labor-intensive, family-operated businesses, which encounter few problems of market entry and operate under very few legal restrictions but face notable capital restrictions. All in all, these factors constitute a considerable burden on the productivity of firms and, in general on their growth. As Perry *et al.* (2007) note for seven Latin American and Caribbean countries, and the McKinsey Global Institute (2003) for Turkey, the difference in labor productivity between firms that operate informally and formally can range between 30 and 40%. Hence, combating informality is an important policy concern for governments in their efforts to achieve sustained economic growth.

1.2 Theoretical approaches on the informal sector

The informal sector encompasses a considerable variety of heterogeneous activities, ranging from traditional, casual survival-type activities, such as street vendors, shoe shiners and garbage collectors, to more modern activities, where small firms use modern sector techniques of production and compete with formal small and medium firms. It is this diversity that makes it particularly difficult to measure and to characterize the sector and which, at the same time, has led to the multiplicity of definitions and measures of informality that capture quite distinct and opposing theoretical approaches. All in all, the literature advances two main theoretical views on the informal sector: segmented labor market (or dualist) theory and the legalist or orthodox theory.

In line the first theory, informality is a residual sector in which the labor force that cannot find employment in the more productive and remunerative formal sector is absorbed. As such, the informal sector is perceived as a survival alternative for those disadvantaged or rationed out of formal employment opportunities to escape involuntary unemployment (Lewis, 1954; Dickens and Lang, 1985). Furthermore, informality is a form of market failure as it prevents individuals and firms from moving to formal activities, and in the presence of this segmentation the rewards for individuals with similar characteristics and, therefore, equal potential productivity differ between sectors. While in the formal sector there are internal markets that limit the labor supply and produce high wages, in the informal sector the ease of entry along with an abundant supply of unskilled workers lead to low wages. From this perspective, informality can be considered to be involuntary; it is not a choice but rather a last resort to escape unemployment.

From the opposing perspective, the legalist or orthodox school of thought considers the labor market to be like any another market with a competitive framework in which price flexibility and free labor mobility lead to a full employment equilibrium with equal remuneration for the same kind of work (De Soto, 1987; Saavedra and Chong, 1999; Maloney, 1999). Informality is thus seen as a dynamic sector made up of micro-entrepreneurs who based on the private cost-benefit calculation choose to be informal. As such, informal activities are considered a voluntary option taken by firms in order to avoid the bureaucratic costs associated with business start-ups and other government procedures, and by individuals who find them more profitable than formal activities or because they seek greater independence and work schedule freedom.

More recent theories recognize that the heterogeneous nature of the informal sector allows these two conflicting views to be reconciled, while stress a more complex structure of the sector. Thus, Fields (1990 and 2005) claims that the informal sector has its own internal duality that sees, in some cases, informal activities being preferred to certain formal activities (in segments he refers to as constituting the “upper tiers”), while, in others, the former offer worse conditions than the latter (the so-called “lower tiers”). In the first of these segments Fields places the workers who voluntarily seek employment in the informal sector because either their utility or their wages are higher than in the formal sector. Here they might achieve higher utility, despite lower wages, thanks to non-pecuniary advantages, such as greater flexibility in their working hours. Similarly, they might obtain higher wages thanks to greater rewards being paid for specific characteristics (e.g. entrepreneurial talent) in the informal than in the formal sector. By contrast, the segments in the lower tiers are occupied involuntarily by workers been rationed out of the formal sector.[2]

1.3 Measuring informality

As discussed, properly defining the informal sector is not an easy matter due to its very nature of not being easily observable. According to Schneider and Enste (2000) a broad definition of the informal economy includes unreported activities producing legal goods and services which would generally be taxable and reported to the tax authorities. However, the definition of informality is strongly driven by data availability and the goals of the researcher or policy maker involved.

In the literature there are at least two definitions of the informal sector: the productivity definition and the legalistic definition (Perry *et al*., 2007; Gasparini and Tornarolli, 2009; Khamis, 2012). The productivity definition focuses more on informal firms and type of job (ILO, 1991; Maloney, 2004). According to this definition informal activities are those carried out in low-productivity jobs in marginal small-scale, frequently family-based firms and using primitive technologies. Nevertheless, due to the limited information contained in surveys and the difficulties to define and measure productivity and marginal or primitive technologies, it is very complex to empirically implement this definition. For instance, Gasparini and Tornarolli (2009), trying to adjust this definition of informality to the information available in Latin American surveys, have related the informal sector to the type of job (self-employment), the type of economic unit (small), and the worker’s skills (unskilled workers). However, this kind of approximations has been criticized for focusing on small firms and not being able to capture evasion or coverage in large firms.

The legalistic definition emphasized the lack of coverage of workers by mandated labor protections or firms avoiding taxation or other legal regulations (Saavedra and Chong, 1999). The informal sector under this definition is characterized by workers non-compliant or without access to the health insurance scheme and the

---

[2] Similar arguments are made by: House (1984), Tokman (1987), Marcouiller *et al*. (1997), Ranis and Stewart (1999), Cunningham and Maloney (2001), Maloney (2004), and Perry *et al*., 2007.

pension system, and firms that not complies with the norms in terms of labor contracts, taxes, and legal registration.

This second definition also has its limitations. On the one hand, there are several dimensions to be included in the concept of labor protection and social security. Furthermore, there are differences in the extent of labor protection and social security system across countries. On the other hand, being voluntary surveys the principal sources of data, there is sensitivity to how the questions are posed, and its confidence in the respondents' willingness to truthfully reveal their labor status, tax payment or compliance of other mandated regulations.

Since the approach followed in this thesis is based on a framework for examining the factors that lead workers and firms to engage in unregistered activities, rather than analyzing the precariousness of the informal sector, we are more interested in the characteristics of individuals and firms than in the characteristics of the production, such as productivity or technology. In this sense, we follow the legalistic definition of informality. Therefore, we consider two criteria to define the informal sector: the social protection criterion to define informal employment and the criterion of registration at the Chamber of Commerce to define informality at the firm level. It is worth highlighting that this approach of single-dimensional indicator allow only describe a part of the multidimensional phenomenon of informality, providing a different aspect of reality which also has important implications for the policy makers.

1.4 Structure of thesis

The rest of this thesis is divided into three essays, each of which is self-contained and developed according to its own structure and methodological framework. However each essay is concerned with the same issue, namely, informality. More specifically, the first two essays analyze labor informality, while the third examines informality at the firm level.

From a theoretical perspective, the essays adhere to the dual informal labor markets approach proposed by Fields (1990 and 2005). As is discussed in Chapter 2 and further extended in Chapter 3, it is our contention that the two prevailing theories of the informal sector (dualist and legalist) constitute but partial explanations of the phenomenon. Even though individuals might be able to choose the type of employment they wish to perform given their personal characteristics (as argued by the legalist theory), restrictions may exist in the number of jobs actually created, particularly in the formal sector, and hence their employment options will ultimately be determined by the structure of the economy or demand conditions (as claimed by the dualist theory). In such scenario, not everyone who wants to work in the formal sector can do so ("lower tiers" or "involuntary entry" informal segment), while there are those that prefer to work in the informal sector because of the greater benefits (wages or/and non-wages) it offers compared to those available in the formal sector ("upper tiers" or "voluntary entry" informal segment).

The thesis is structured as follows. Chapter 2, *Rationing of Formal Sector Jobs and Informality* examines the determinants of the job placement process for workers in both the formal and informal sectors. The novelty in this chapter is that in the job placement process we take into account the rationing of formal jobs that exists in the economy. Despite the considerable amount of debate concerning the restrictions that exist for the creation of formal jobs in developing countries, the empirical literature has paid scant regard to this element in the job search process. To rectify this, we adopt the job queue approach first proposed by Abowd and Farber (1982) and subsequently extended by Mengistae (1999). In line with this approach, job placement in either the formal or informal sector is determined not only by individual preferences to work in one sector or the other, but also by the rationing of the formal jobs that exist. In this case, the employers' criteria for choosing workers for rationed jobs need to be taken into account. The analysis of the determinants of the job placement process is based on the estimation of an endogenous switching regression model. This model takes into consideration the dual mechanisms that exist in the job placement process: one is the workers' choice to actually join the job queue for a formal job and the other is the mechanism applied by employers in choosing workers from that job queue to fill a vacancy. This model also allows us to test the hypothesis of the existence of a job queue in the formal sector, to measure the length of this job queue for different groups of workers, and to quantify the percentage of informal workers who have been rationed out of the formal sector, i.e. those who are involuntarily informal. The research questions in this chapter are: *What factors determine whether an individual is located in the formal sector or the informal sector? Do informal workers queue for formal jobs? Is there a formal job rationing in the Colombian labor market?*

Chapter 3, *Labor Informality: Choice or Sign of Segmentation? A Quantile Regression Approach at the Regional Level* extends the previous chapter by examining the heterogeneity of the informal sector and analyzing differences in Colombia's local labor market. To this end, we begin by studying the features of the local labor markets, especially with regards to the degree of modernity characterizing the informal sector. Thus we calculate an index of modernity of the informal sector based on the type of employment generated by each economic sector (skilled or unskilled) (Ranis and Stewart, 1999). Various descriptive statistics of this index allow us to distinguish different segments within the informal sector: one associated with more traditional activities; and the other associated with more modern activities. In the second part, we decompose the wage differential between the formal and informal sectors in the entire distribution of wages using the approach proposed by Machado and Mata (2005) and the extension developed by Albrecht, Vuuren and Vroman (2009) to account for sample selection. This methodology allows us to account for a wider variety of both informal and formal employees and to determine whether different kinds of informal workers present different kinds of behavior. The specific research questions posed in this chapter are: *Is there an internal duality in the informal sector in Colombia? Why is there such a variety of workers in the informal sector? Do informal worker types differ by city? Is labor informality a choice or the result of labor market segmentation?* Empirical research on these topics is scarce and this chapter aims to fill this gap.

Chapter 4, *Agglomeration Economies and Informality in Developing Countries*, examines the clustering patterns of informal activities and their inter-relation with formal activities, as well as the effects of agglomeration economies on activities of this type. The empirical analysis in this chapter is divided in two parts. In the first, we study the spatial clustering and co-clustering patterns of formal and informal manufacturing activity within a metropolitan area of Colombia. We use census manufacturing enterprise-level data for the metropolitan area of Cali for the year 2005. Using this database we calculate the degree of spatial clustering and co-clustering by means of the M-functions proposed by Marcon and Puech (2010). Further, we analyze the geographical distribution of formal and informal enterprises in different clustered and co-clustered industries by means of kernel density mapping. In the second part, we study the effects of agglomeration economies on formal and informal productivity. In this part, we use data at the worker level for 33 local labor markets in Colombia throughout the period 2002-2006. The empirical analysis is based on the regression of individual worker wage rates (as a measure of labor productivity) on employment density (as a measure of agglomeration) conducted separately for the formal and informal wage sectors. These regressions comprise instrumental variables estimates and a bootstrap procedure to correct for the endogeneity attributable to the reverse causality between wages and agglomeration, and to avoid any bias in the statistical inferences due to the fact that we are dealing with a small number of clusters (33 municipalities). The chapter provides a systematic analysis in seeking to respond to three questions: first, *do formal and informal enterprises display different clustering patterns?* Second, *do formal and informal enterprises locate in the same areas of the city, or is there a marked spatial segmentation between formal and informal manufacturing activity?* And third *are there differences in the effects of agglomeration on productivity and wages between the formal and informal sector?*

1.5 References

Abowd, J. and Farber, H. (1982). “Job Queue and the Union Status of Workers”, *Industrial and Labor Relations Review*, 35(3): 354-367.

Albrecht, J., van Vuuren, A. and Vroman, S. (2009). “Counterfactual Distributions with Sample Selection Adjustments: Econometric Theory and an Application to the Netherlands”, *Labour Economics*, 16(4): 383-396.

Cunningham, W. and Maloney, W. (2001). “Heterogeneity among Mexico’s Microenterprises: An Application of Factor and Cluster Analysis”, *Economic Development and Cultural Change*, 50(1): 131–156.

De Soto, H. (1987). *El Otro Sendero. La Revolución Informal*, Lima, Instituto Libertad y Democracia.

Dickens, W. T. and Lang, K. (1985). “A Test of Dual Labour Market Theory”, *American Economic Review*, 4(75): 792-805.

Fields, G. (1990).” Labour Market Modelling and the Urban Informal Sector: Theory and Evidence”, in: D., Thurnham, Salome, B. and Schwarz, A. (Ed.), *The Informal Sector Revisited*. Paris, OECD.

Fields, G. (2005). “A Guide to Multisector Labor Market Models”, *Social Protection Discussion Paper Series*, No 0505, World Bank, Washington, D.C., April.

Gasparini, L and Tornarolli, L. (2009). “Labor Informality in Latin America and the Caribbean: Patterns and Trends from Household Survey Microdata”, *Desarrollo y Sociedad*, 63:13-80.

House, W.J. (1984). “Nairobi’s Informal Sector: Dynamic Entrepreneurs or Surplus Labor?”, *Economic Development and Cultural Change*, 32(2): 277-302.

ILO (1991). El Dilema del Sector no Estructurado. Memoria del Director General, CIT, 1991, Ginebra.

ILO (2011). 2011 *Labour Overview. Latin America and the Caribbean*, Lima.

Jütting, J., Parlevliet, J., and Xenogiani, T. (2008). “Informal Employment Re-loaded”, *Working Paper*, No 266, OECD Development Center, Paris, January.

Jütting, J. and De Laiglesia, J. (2009). Is Informal Normal? Towards More and Better Jobs in Developing Countries, *OECD Development Centre*, 163 pages.

Khamis, M. (2012). “A Note on Informality in the Labour Market”, *Journal of International Development*, 24: 894-908.

Lewis, W.A. (1954). “Economic Development with Unlimited Supplies of Labour”, *The Manchester School*, 22(2):139-191.

Machado, J. and Mata, J. (2005). “Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression”, *Journal of Applied Econometrics*, 20(4): 445-465.

Maloney, W. (1999), “Does Informality Imply Segmentation in Urban Labor Markets? Evidence from Sectoral Transitions in Mexico”, *The World Bank Economic Review*, 13(2):275-302.

Maloney, W. (2004). “Informality Revisited”, *World Development*, 32(7): 1159-1178.

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.

Marcouiller, D., Ruiz de Castilla, V. and Woodruff, C. (1997). “Formal Measures of the Informal Sector Wage Gap in Mexico, El Salvador, and Peru,” *Economic Development and Cultural Change*, 45(2): 367-392.

Mengistae, T. (1999). “Wage Rates and Job Queue: Does the Public Sector Overpay in Ethiopia", *Policy Research Working Paper*, No 2105. The World Bank.

McKinsey Global Institute (2003). *Turkey: Making the Productivity and Growth Breakthrough*, New York: McKinsey, 784 pages.

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.

Ranis, G., and Stewart, F. (1999). “V-goods and the Role of the Urban Informal Sector in Development”, *Economic Development and Cultural Change*, 47(2): 259-288.

Saavedra, J. and Chong, A. (1999). “Structural Reform, Institutions and Earnings: Evidence from the Formal and Informal Sectors in Urban Peru”, *Journal of Development Studies*, 35(4), 95–116.

Schneider F. and Enste, D. (2000). “Shadow Economies: Size, Causes and Consequences”, *Journal of Economic Literature*, 38: 77–114.

Schneider, F., Buehn, A. and Montenegro, C.E. (2010). “New Estimates for the Shadow Economies all Over the Word”, *International Economic Journal*, 24(4): 443-461.

Tokman, V. (1987). “El Sector Informal: Quince Años Después”, *El Trimestre Económico*, LIV 3(215): 513-536.

Chapter 2

Rationing of Formal Sector Jobs and Informality

2.1 Introduction

The neoclassical view of the labor market is based on the implicit assumption that the demand for labor is broad and diverse and that, as such, it is the characteristics of the worker that are the main determinant of employment decisions. Structural economic factors and labor demand factors are both deemed irrelevant; the only relevant factors are those associated with the conditions of supply that determine whether individuals opt for participation or inactivity, employment or unemployment and whether they work in the formal or informal sectors.

However, because this view does not consider the demand factors and, hence, the restrictions that exist in the creation of jobs (above all in the modern sector of the economy), it has significant limitations when it comes to explaining the heterogeneity present in labor markets, especially the markets of developing countries.

In this chapter we provide an alternative view of the labor market in which employment decisions are considered as being determined by factors of both supply and demand. Thus, although individuals have the opportunity to choose and to decide whether to work or not and to select the type of work they wish to perform, there are restrictions on the range of options available to them. These restrictions correspond to the limitations imposed by the demand side of the labor market, such as the few good quality jobs that are being created.

As Table 2.1 shows, the labor markets of a group of developing regions and countries present a marked heterogeneity characterized by the scarcity of employment opportunities and the poor quality of the jobs that are available. The informal market provides between 50 and 60% of jobs and is responsible for between 30 and 40% of total output, while the unemployed represent about 8% of the labor force, exceeding 10% in extreme cases such as those of North Africa and Colombia.

Given this great weight of informal economic activities, as well as the fact that the quality of employment constitutes an essential element of the quality of life and of socio-economic welfare, in addition to being a mechanism for countering social inequalities, this chapter seeks to analyze the basic characteristics and determinants of finding employment in either the formal or informal labor markets, by applying an alternative methodology that assumes the rationing of formal jobs. To date, the literature has adopted models that fail to take into account formal job rationing, so that being employed in the formal or informal sectors is deemed an exclusively individual decision dependent solely on that individual's personal characteristics and preferences. The alternative methodology proposed here is based on the so-called "job queue" approach and takes into account the criteria applied by employers in the job placement process when recruiting applicants for a limited number of jobs. The job placement process, then, can be seen to comprise two mechanisms: one is the workers' choice to actually join the job queue for a formal job and the other is the mechanism applied by employers in choosing workers from that job queue to fill a vacancy.

Table 2.1. Informality and unemployment in developing regions

Region	Informality	Unemployment	Informal GDP
Asia	70%	5.6%	35%
North Africa	47%	10.3%	36%
Latin America	54%	7.3%	41%
Argentina	53%	7.9%	25%
Brazil	51%	7.9%	40%
Chile	36%	7.8%	20%
Mexico	50%	4.9%	30%
Colombia	58%	13%	39%

Source: informality: Jütting and De Laiglesia (2009). Unemployment: ILO (2011). Informal GDP: Schneider *et al.* (2010).

This analysis of the determinants of job placement (informality vs. formality) is based on the estimation of an endogenous switching regression model. This model, in addition to taking into consideration the dual mechanism outlined above in the job finding process, includes as additional determinants the unbiased predictions of the wage premiums resulting from working in the formal sector. The estimations are conducted for the job market in Colombia which, as we have seen in Table 2.1, suffers marked quantitative and qualitative imbalances, i.e., significant levels of informality combined with a high rate of unemployment.

The rest of the chapter is organized as follows. Section 2.2 discusses the econometric model used. In section 2.3 describes the estimation procedure and discusses some statistical tests to validate the results. In section 2.4 the data are presented and the results of econometric model are discussed in section 2.5. Finally, section 2.6 concludes.

2.2 Econometric model

From a neoclassical perspective, labor market decisions are determined by the worker's characteristics, i.e., the conditions of supply. Thus, agents voluntarily choose from the options available to them: participation or inactivity, employment or unemployment, formal or informal work, etc., following a process of welfare maximization subject to a number of restrictions imposed by the individual, such as the availability of time, initial wealth, human capital endowment, among others. From this perspective, the structure of the economy and the characteristics of labor demand are not taken into account among the factors that determine employment or unemployment.

An alternative approach to the labor market might allow the workers to choose, to decide for themselves, in line with neoclassical thinking, but here the range of options is determined by the constraints imposed by the demand side. This alternative perspective is not that labor demand is broad and diverse, as is assumed in neoclassical theory, but on the contrary that there are restrictions on jobs, above all formal jobs, as the modern sector only demands the jobs that it needs (Archibald, 1977; Dickens and Lang, 1985; Lang and Dickens, 1987; Uribe and Ortiz, 2006).

Since there are restrictions on the number of jobs, the decision of wanting to work does not necessarily mean the individual will be contracted, i.e., the probability of wanting a job in a given sector is not equal to the probability of being employed in that sector. Moreover, taking into account the possibility of excess demand for a type of job ("a job queue"), finding a worker employed in a particular job depends on the job rationing rules used by employers as well as the preferences of workers for that post.

In order to model the rationing of formal jobs, this chapter follows the model of job queues proposed by Abwod and Farber (1982) and extended by Mengistae (1999). In the first of these papers, a model is developed to analyze the determinants of the union status of workers that allows the possibility of queuing for union jobs, while the second, employing Abwod and Farber's model, analyzes the existence of a job queue in the public sector and the wage differential between this and the private sector. Taking this same approach to the analysis of labor market constraints and the existence of job queues for formal jobs we find the works of Dickens and Lang (1985), Maloney (1998), Soares (2004), Co *et al*. (2005), Puentes and Contreras (2009), and Contreras *et al*. (2008).

From this perspective it is assumed that workers freely choose the sector of the labor market in which they want to work; however, this choice is constrained by the employment options available in the market. The structural constraints in the economy limit the number of formal jobs and as these are the ones that most workers prefer, given that they offer the best work conditions (stability, social security, opportunities for promotion and rewards for human capital accumulation), a queue of applicants forms for these jobs (Dickens and Lang, 1985; Lang and Dickens, 1987). The existence of a job queue for the formal sector relaxes the assumption that the probability of wanting to be employed in the formal sector is equal to the probability of obtaining a formal job. Seen from this perspective, a worker's employment in the formal sector depends on two decision processes: the worker's decision to join the queue for a formal job and the employer's decision to hire a worker who is in the queue.

Let us analyze the first of these decisions: joining the job queue for formal employment. The workers' decision to join the queue to obtain a formal job depends on the utility that such a post represents to them. A major determinant in this decision and in utility maximization is the wage premium to be gained from being employed in one sector or the other. Let W_{1i} be the hourly wage earned by individual i in the formal sector and W_{2i} the hourly wage earned in the informal sector. We also define U_{1i} as the maximum utility individual i can obtain in the formal sector and U_{2i} as the maximum utility the individual can obtain in the informal sector. Thus, it is assumed that individual i will prefer a formal job and will queue for such a job if $V_{1i}^{*} = U_{1i} - U_{2i} > 0$. Assuming that V_{1i}^{*} depends linearly on the wage premium provided by the formal sector ($\ln W_{1i} - \ln W_{2i}$) and on a set of individual traits and characteristics of his or her working environment and family (X_{1i}), the following job queue equation can be written:

$$V_{1i}^{*} = \alpha_1(\ln W_{1i} - \ln W_{2i}) + X_{1i}^{'}\beta_1 + u_{1i}\,, \tag{1}$$

where α_1 is a constant to be estimated; β_1 is a vector of coefficients to estimated; and u_1 is a random error term distributed with mean zero and variance $\sigma_{u_1}^2$.

Since V_{1i}^{*} is not directly observable and only its sign can be observed, a binary variable I_{1i} can be defined that captures an individual's willingness to obtain a formal job, as follows:

$$I_{1i} = 1 \ iff \ \ V_{1i}^{*} = \alpha_1(\ln W_{1i} - \ln W_{2i}) + X_{1i}^{'}\beta_1 + u_{1i} > 0$$

$$I_{1i} = 0 \ iff \ \ V_{1i}^{*} = \alpha_1(\ln W_{1i} - \ln W_{2i}) + X_{1i}^{'}\beta_1 + u_{1i} < 0 \tag{2}$$

Hence, I_1 is a binary variable equal to 1 if the worker joins the queue for a formal job and equal to 0 if the worker prefers not to join the queue.

The second decision, i.e., that taken by employers when selecting workers from the job queue to fill the formal vacancy, can also be modeled on the basis of a latent variable V_{2i}^*. It is assumed that the objective of the employers in the formal sector is to maximize worker productivity per unit spent on this factor. Thus, the employer will hire a given individual ($V_{2i}^* > 0$) taking into consideration characteristics of productivity (X_{2i}) and the absolute costs he expects to incur for hiring that individual ($E(\ln W_{1i})$). This can be expressed as:

$$V_{2i}^* = \alpha_2 E(\ln W_{1i} \mid V_{1i}^* > 0) + X'_{2i}\beta_2 + u_{2i} \tag{3}$$

where i represents the group of workers queuing for a job; α_2 is a constant to be estimated; $E(\ln W_{1i} \mid V_{1i}^* > 0)$ represents expected wage a worker could earn in the formal sector as has joined the queue; β_2 is a vector of coefficients to estimated; and u_2 is a random error term distributed with mean zero and variance $\sigma_{u_2}^2$.

As in the job queue equation, a binary variable I_{2i} can be defined that captures the sign of V_{2i}^* and which also captures the selection process undertaken by the employer when choosing a worker from the queue, which is:

$$I_{2i} = 1 \; iff \;\; V_{2i}^* = \alpha_2 E(\ln W_{1i} \mid V_{1i}^* > 0) + X'_{2i}\beta_2 + u_{2i} > 0$$

$$I_{2i} = 0 \; iff \;\; V_{2i}^* = \alpha_2 E(\ln W_{1i} \mid V_{1i}^* > 0) + X'_{2i}\beta_2 + u_{2i} < 0 \tag{4}$$

Then, I_2 is a dummy variable equal to 1 if the worker is chosen from the queue to fill the post available in the formal sector or equal to 0 if not selected.

Together with equations (1) - (3) the system is completed with the wage equations for the formal and informal sectors:

$$\ln W_{1i} = Z'_{1i}\gamma_1 + \upsilon_{1i} \;\; iff \;\; V_{1i}^* > 0 \text{ and } V_{2i}^* > 0 \tag{5}$$

$$\ln W_{2i} = Z'_{2i}\gamma_2 + \upsilon_{2i} \;\; \text{otherwise} \tag{6}$$

where Z_{ji} is a vector of observable individuals and labor characteristics; γ_1 and γ_2 are vector of coefficients to estimated; υ_{1i} and υ_{2i} are *iid* normal with mean zero and variances $\sigma_{\upsilon_1}^2$ and $\sigma_{\upsilon_2}^2$, respectively. Assuming the joint distribution of the error terms ($u_{1i}, u_{2i}, \upsilon_{1i}, \upsilon_{2i}$), Mengistae (1999) shows that the covariance matrix has the form:

$$\begin{bmatrix} \sigma_{u_1}^2 & \sigma_{1u_2} & \sigma_{1\upsilon_1} & \sigma_{1\upsilon_2} \\ & \sigma_{u_2}^2 & \sigma_{2\upsilon_1} & \sigma_{2\upsilon_2} \\ & & \sigma_{\upsilon_1}^2 & \sigma_{\upsilon_1\upsilon_2} \\ & & & \sigma_{\upsilon_2}^2 \end{bmatrix},$$

where, $Cov(u_{1i},\upsilon_{1i})=\sigma_{1\upsilon_1}$; $Cov(u_{1i},\upsilon_{2i})=\sigma_{1\upsilon_2}$; $Cov(u_{2i},\upsilon_{1i})=\sigma_{2\upsilon_1}$; $Cov(u_{2i},\upsilon_{2i})=\sigma_{2\upsilon_2}$; $Cov(\upsilon_{1i},\upsilon_{2i})=\sigma_{1\upsilon_1}\sigma_{\upsilon_1\upsilon_2}$; and $Cov(u_{1i},u_{2i})=\sigma_{1u_2}$.[1]

The OLS estimation of the above wage equations, regardless of the sector in which the workers choose to seek work and the employers' selection decisions, can lead to bias. $\ln W_1$ and $\ln W_2$ are censored variables, since the former is only observed for formal workers and the latter solely for informal workers, so that the disturbance terms υ_{1i} and υ_{2i} have a truncated distribution and the OLS estimates would not be consistent for the existence of a selection bias problem (Heckman, 1979; Huguet, 1996).

In order to identify the system of equations fully, we first need to deduce the wage that a worker in the queue might expect to earn in the formal sector (that is, the expression $E(\ln W_{1i} \mid V_{1i}^* > 0)$ in the equation (3)). Using equations (1) - (5) we obtain that:

$$E(\ln W_{1i} \mid V_{1i}^* > 0) = Z_{1i}'\gamma_1 + \sigma_{1\upsilon_1}\lambda_{1i} \tag{7}$$

where λ_{1i} is analogous to the inverse Mills ratio and its structure depends on whether ρ is equal to or different from zero. Its expression is developed in the following section.

The full model, therefore, comprises equations (1), (3), (5) and (6). The system is endogenous given that in the job finding equations (equations (1) and (3)) wages are included (equations (5) and (6)) as additional determinants, while the wage equations in turn are determined by the job finding decisions of the workers and the employers' selection decisions. Such models receive the name of endogenous switching regression models (Maddala y Nelson, 1975). The section that follows provides a discussion of the estimation strategy for models of this type, which depends on the degree of observability of individuals who are in the job queue and the assumption regarding the decision-making process: whether it is sequential or simultaneous.

[1] In section 2.3.1 σ_{1u_2} is refer as ρ. Mengistae (1999) assumes that this parameter is equal to cero and therefore his estimation follows a sequential structure.

2.3 Estimation procedure and the test on the existence of the job queue

In order to estimate the endogenous switching model a three-step procedure is followed. In the first step, the reduced-form bivariate probit model (which does not include the wage differential between sectors) is estimated and the residuals are taken to approximate the expectations of υ_1 and υ_2 conditional upon u_1 and u_2. These approximations, which are nothing other than the inverse Mills ratios, can, in the second step, enter the wage equation for each of the sectors and correct the selection bias. Finally, in the third step, the wage differential, estimated in the second step, is included as an additional regressor in the "in-queue" (IQ) equation. Likewise, the wage that a worker could earn in the formal sector, as estimated from the wage equations, is included in the "chosen-from-queue" (CFQ) equation.

To identify the system, a number of exclusion restrictions need to be considered for the model. On the one hand, to be able to identify the parameters of the IQ and CFQ equations, variables need to be incorporated that affect the former decision but not the latter, and vice versa. On the other hand, to identify the parameters of the wage equations, at least one variable needs to be included that determines wages but which does not affect the IQ and CFQ decisions.

A further important aspect to consider when estimating the bivariate probit model is the partial observability of the data. [2] Household surveys usually only provide information about the workers' current situation, that is, whether they are employed in the informal or formal sector, but they provide no details about the decision process that leads an individual to opt for formality or informality. This means that it is not possible to observe the different types of informal workers that might emerge from the decision process: i.e., workers that choose the informal sector and who do not queue for formal employment; or workers, who although they are in the queue, are not selected and have to opt for the informal sector. In terms of the equations described in the previous section, no information is available regarding the IQ (I_1) and CFQ (I_2) decisions. Only the product of these two decisions can be observed ($I = I_1 I_2$). If this product is equal to 1, the worker is in the formal sector and if it is equal to 0, the worker is in the informal sector.

[2] A more detail explanation on the partial observability of the data can be found in Abowd and Farber (1982), Meng and Schmidt (1985), and Tunali (1986).

To overcome this lack of information, Poirier (1980) and Abowd and Farber (1982) propose using the same dependent variable for the IQ and CFQ equations and differentiating between the characteristics that affect each of the equations. Thus the exclusion restrictions acquire considerable importance for identifying the effects.

One aspect that differentiates Poirier's method from that of Abowd-Farber's is the assumption concerning the decision-making process. The first author assumes that both decisions are taken simultaneously, while the latter assume that the decision process is sequential. According to Maddala (1983), the choice of one or other of these models depends on the purpose for which it is required. If the analysis seeks to examine the factors that influence the employers' decision to hire a specific type of worker, the simultaneous model would be the most appropriate; whereas, if it seeks to analyze the whole process of worker placement between the formal and informal sectors, taking into account the factors that affect both the preferences of workers and employers, the sequential model would be the most appropriate. The second approach, therefore, seems to be the most suitable for the case we deal with here; however, as far as possible both methods will be followed so as to observe the extent to which their respective results are consistent.

2.3.1 Estimation

As mentioned above, the estimation of the model depends on the assumption that it is made regarding the decision-making process: whether it is sequential or simultaneous. In the Abowd-Farber model it is assumed that decisions are made sequentially, which supposes that these decisions are independent. This in turn implies that the joint distribution function of the disturbance terms of the IQ and CFQ equations (u_1 and u_2, respectively) follows a bivariate normal distribution with zero means and unit variances and covariances equal to zero. Given this structure, the likelihood function with partial observability to be maximized is (Meng and Schmidt, 1985; Tunali, 1986):

$$L_1 = \prod_{I=1}\left[\Phi(X_{1i}'\beta_1)\Phi(X_{2i}'\beta_2)\right]\prod_{I=0}\left[1-\Phi(X_{1i}'\beta_1)\Phi(X_{2i}'\beta_2)\right], \tag{8}$$

where Φ is the normal cumulative distribution; X_1 and X_2 are the explanatory variables of the IQ and CFQ equations, respectively. As in the first step the reduced-form bivariate probit model is estimated, we also include as explanatory variables those variables assumed only to affect wages.

By contrast, Poirier's model assumes that the process of choosing between the formal and informal sectors is simultaneous, which implies that the correlation between the disturbance terms of the two equations is different from zero. Under this assumption therefore the likelihood function to be maximized has the following form (Meng y Schmidt (1985); Tunali (1986)):

$$L_2 = \prod_{I=1}\left[\Phi_2(X_{1i}^{'}\beta_1, X_{2i}^{'}\beta_2, \rho)\right]\prod_{I=0}\left[1-\Phi_2(X_{1i}^{'}\beta_1, X_{2i}^{'}\beta_2, \rho)\right], \tag{9}$$

where Φ_2 is the bivariate normal cumulative distribution; ρ is the covariance between u_1 and u_2, and we assume that $\sigma_{u_1}^2 = \sigma_{u_2}^2 = 1$.

Note that in the absence of a correlation between the disturbance terms of the two equations, ρ is equal to zero in equation (9) and the simultaneous model becomes a sequential model.

Continuing therefore to the second step in the estimation (as outlined at the beginning of this section), the wage equations have to be estimated next, while correcting for the bias due to the censoring of the data generated by workers choosing to work in either the formal or informal sectors. To correct for this selection bias, the procedure proposed by Maddala (1983, pp. 278-283), which is an extension of the Heckman-Lee two-stage estimation method (Heckman, 1976; Heckman, 1979; Lee, 1978), is adopted. Assuming the disturbances in the formal workers' wage equation to be normally distributed, the first moment of the truncated distribution is given by the following expression (Tunali, 1986; Mohanty, 2001):

$$E(\upsilon_{1i} \mid I_{1i} = 1, I_{2i} = 1) = \sigma_{1\upsilon_1}\lambda_{1i} + \sigma_{2\upsilon_1}\lambda_{2i}, \tag{10}$$

where λ_{1i} and λ_{2i} are analogous to the inverse Mills ratios in the univariate context. These ratio values depend on the existence or otherwise of a correlation between the IQ and CFQ decisions, that is, whether or not $Cov(u_{1i}, u_{2i}) = \rho = 0$. Assuming that the preceding expression is different from zero then (Tunali, 1986):

$$\lambda_{1i} = \frac{\phi(X_{1i}^{'}\beta_1)\Phi(X_{2i}^{'}\beta_2 - \rho X_{1i}^{'}\beta_1)/\sqrt{1-\rho^2}}{\Phi_2(X_{1i}^{'}\beta_1, X_{2i}^{'}\beta_2, \rho)} \tag{11}$$

$$\lambda_{2i} = \frac{\phi(X_{2i}^{'}\beta_2)\Phi(X_{1i}^{'}\beta_1 - \rho X_{2i}^{'}\beta_2)/\sqrt{1-\rho^2}}{\Phi_2(X_{1i}^{'}\beta_1, X_{2i}^{'}\beta_2, \rho)} \tag{12}$$

Now, assuming independence between these decisions then ρ=0, so that the above expressions can be reduced to:

$$\lambda_{1i} = \frac{\phi(X_{1i}^{'}\beta_1)}{\Phi(X_{1i}^{'}\beta_1)} \tag{13}$$

$$\lambda_{2i} = \frac{\phi(X_{2i}^{'}\beta_2)}{\Phi(X_{2i}^{'}\beta_2)} \tag{14}$$

Thus, the consistent estimation of the wage equation for workers in the formal sector is obtained by incorporating λ_{1i} and λ_{2i} as additional regressors, that is, by estimating the following equation:

$$\ln W_{1i} = Z_{1i}^{'}\gamma_1 + \sigma_{1\upsilon_1}\lambda_{1i} + \sigma_{2\upsilon_1}\lambda_{2i} + \varepsilon_{1i}\,, \tag{15}$$

where $\varepsilon_{1i} = \upsilon_{1i} - \sigma_{1\upsilon_1}\lambda_{1i} - \sigma_{2\upsilon_1}\lambda_{2i}$ and $E(\varepsilon_{1i} \mid I_i = 1) = 0$. Following the Heckman-Lee method an approximation of the expressions of λ_{1i} and λ_{2i} are the estimates from the bivariate probit, which gives consistent estimates of γ_1. That is, applying OLS to the following expression:

$$\ln W_{1i} = Z_{1i}^{'}\gamma_1 + \sigma_{1\upsilon_1}\hat{\lambda}_{1i} + \sigma_{2\upsilon_1}\hat{\lambda}_{2i} + \eta_{1i}\,, \tag{16}$$

where $\eta_{1i} = \varepsilon_{1i} + \sigma_{1\upsilon_1}(\lambda_{1i} - \hat{\lambda}_{1i}) + \sigma_{2\upsilon_1}(\lambda_{2i} - \hat{\lambda}_{2i})$.

By following the Heckman-Lee two-stage estimation method for the informal workers' wage equation ($I = 0$), problems arise in terms of the identification of different types of informal worker and, as a result, the sample selection rules are more complicated than they are for the case of formal workers. Using Poirier's model (simultaneous case), two types of informal worker can, in principle, be differentiated: those that are in the job queue but who are not chosen by the employers ($E(\upsilon_{2i} \mid I_{1i} = 1 \text{ and } \mathrm{I}_{2i} = 0)$), and those that have opted not to join the queue ($E(\upsilon_{2i} \mid I_{1i} = 0\ )$). In addition, there is another type of informal worker, in this case not very clearly defined, who while not having joined the queue may yet be selected for a formal job ($E(\upsilon_{2i} \mid I_{1i} = 0 \text{ y } I_{2i} = 1)$).[3] Given this problem of identifying types of

[3] It is possible to determine the sample selection rules and therefore to calculate the inverse Mills ratios in the case of formal workers, that is, when $I = I_1 I_2 = 1$. This product is equal to 1 only when worker is in the job queue and is chosen by the employer to fill a formal position ($I_1 = 1$ and $I_2 = 1$). However, in the case of informal workers the sample selection rules are not very clear. We observe an informal worker when $I = I_1 I_2 = 0$. Now, if we assume that the decision-making process is simultaneous, there are three possible combinations for which this last product is equal to zero, these are: 1) when a worker opt not to join the queue and he or she is not chosen by the employers ($I_1 = 0$ and $I_2 = 0$); 2) when a worker have opted to join the queue, but he or she is not chosen by the employers ($I_1 = 1$ and $I_2 = 0$); and 3) when a worker have not opted to join the queue but he or she is chosen by the employers ($I_1 = 0$ and $I_2 = 1$). This last sample

informal worker, it is not possible to either calculate the inverse Mills ratios or, therefore, correct the selection of the wage equation of the informal workers with Poirier's model.

In the case of the sequential model (Abowd-Farber's model), Mengistae (1999, pp. 11-12) proposes a solution for correcting the selection bias in the wage equation of the informal workers. The author shows that the expected wage of informal workers ($I = 0$) can be estimated as the weighted average of the wage of workers that are informal because they have not joined the job queue ($I_1 = 0$) and the expected wage of workers who are informal as they have not been chosen from the job queue despite having joined it ($I_1 = 1$ and $I_2 = 0$). That is:

$$(1-\pi)E(W_{2i}^1 \mid I_{1i} = 0) + \pi E(W_{2i}^2 \mid I_{1i} = 1 \text{ e } I_{2i} = 0)\,, \tag{17}$$

with π as the proportion of informal workers that are in the job queue and that are not chosen and (1-π) as the proportion of informal workers that are not in the queue. The above expression can be written as:

$$LnW_{2i} = Z_{2i}^{'}\gamma_2 + \sigma_{1\upsilon_2}\lambda_{3i} + \delta_1\lambda_{1i}^{*} + \delta_2\lambda_{4i} + \varepsilon_{2i}\,, \tag{18}$$

where $\lambda_{3i} = -\phi(X_{1i}^{'}\beta_1)\Big/\left[1-\Phi(X_{1i}^{'}\beta_1)\right]$; $\lambda_{4i} = -\phi(X_{2i}^{'}\beta_2)\Big/\left[1-\Phi(X_{2i}^{'}\beta_2)\right]$; $\lambda_{1i}^{*} = \lambda_{1i} - \lambda_{3i}$; $\delta_1 = \pi\sigma_{1\upsilon_2}$; $\delta_2 = \pi\sigma_{2\upsilon_2}$; $\varepsilon_{2i} = (1-\pi)\varepsilon_{2i}^1 + \pi\varepsilon_{2i}^2$; ε_{2i}^1 and ε_{2i}^2 are the disturbances in the wage equation of $I_{1i} = 0$ and, $I_{1i} = 1$ and $I_{2i} = 0$, respectively.

Consistent estimates of γ_2, $\sigma_{1\upsilon_2}$, δ_1 y δ_2 can be obtained estimating by OLS the following equation:

$$LnW_{2i} = Z_{2i}^{'}\gamma_2 + \sigma_{1\upsilon_2}\hat{\lambda}_{3i} + \delta_1\hat{\lambda}_{1i}^{*} + \delta_2\hat{\lambda}_{4i} + \eta_{2i} \tag{19}$$

where $\eta_{2i} = \varepsilon_{2i} + \sigma_{1\upsilon_2}(\lambda_{3i} - \hat{\lambda}_{3i}) + \delta_1(\lambda_{1i}^{*} - \hat{\lambda}_{1i}^{*}) + \delta_2(\lambda_{4i} - \hat{\lambda}_{4i})$ and the inverse Mills ratios can be approximate estimating the bivariate probit Abowd-Farber model.

With the Heckman-Lee two-stage procedure, consistent estimates of the parameters are obtained; however, if there is selection bias, the standard errors of these estimates will not be consistent. We had used bootstrap techniques to calculate these

selection rule is not very clear and therefore the inverse Mills ratios in the case of Poirier's model cannot be calculated.

standard errors. Notice that Tunali (1986) and Mengistae (1999) both propose methods for correcting the standard errors, but their procedure is cumbersome.

In the third step of the endogenous switching model, the IQ and CFQ equations are estimated with a structural bivariate probit model. As mentioned earlier, in the first equation the estimated difference in wages is included as an additional regressor and in the second equation the estimated wages in the formal sector are added. The structural bivariate probit can only be estimated for the Abowd-Farber's model since for Poirier's model estimates corrected by selectivity of the informal sector wages cannot be obtained.

A final aspect to take into account in estimating the endogenous switching model is that of exclusion restrictions. There are two sets of restrictions. The first ones are those that need to be imposed to distinguish between the IQ and CFQ equations. And the second ones are those that make possible the identification of the wage equations.

For the first set of restrictions it is assumed that the factors that influence the decision to join the queue for a formal job are closely related to personal characteristics, above all to the characteristics of the family and household. At the same time, the employers' hiring decisions are assumed to be more closely affected by the endowment of human capital, including such factors as education, previous employment status, the degree of commitment to work and the specific job sector. Thus, the variables excluded from the second equation but included in the first are: age, the number of years in the current job (i.e., tenure) and a number of household variables, including number of children, the occupational status of other members of the household and the educational level of the household.

Let us examine the assumptions underpinning the inclusion or exclusion of some of the aforementioned variables. The variables of age and tenure are excluded from the CFQ equation. The former is excluded to avoid problems of multicollinearity; while in the latter case, it is assumed that employers consider the experience a worker has acquired throughout his or her working life and not just that gained in one particular job (Abowd and Farber, 1982; Farber, 1983; Mengistae, 1999; Soares, 2004). On the other hand, the variables capturing a worker's position within the household and his or her marital status are included so as to measure two distinct effects. In the CFQ equation, these variables represent the worker's degree of commitment to work, to the extent that the responsibilities assumed by heads of households can be related to a greater commitment to their work; and in the IQ equation, they represent the traditional characteristics of the individual which, depending on his or her position and responsibilities within the family structure, will reflect their commitment to the taking on of a particular job.

At the same time, the inclusion of the household variables in the first equation seeks to account for the costs incurred by an individual that opts to work in either the formal or informal sectors. The number of children is a proxy for the level of household responsibility, especially as far as women are concerned. Thus, an informal job might be considered more beneficial as it is more likely to offer a flexible timetable so that the workers can distribute their time more easily between childcare and work.

As for the second set of exclusion restrictions used for identifying wage equations, it is assumed that household characteristics and previous employment status have no effect on the wage levels of formal and informal workers. Likewise, it is assumed for the identification that the specific job sector only affects wage levels, and so these variables are excluded from the structural bivariate probit model.

2.3.2 The existence and length of the formal job queue

In order to test for the existence of a formal job queue, we calculate the tests proposed by Abowd and Farber (1982) and Mengistae (1999). These tests, as well as validating the robustness of the bivariate model with respect to the univariate model, provide evidence of the barriers to obtaining a formal job. In other words, the existence of a job queue means that not all the workers who want to work in the formal sector can do so, due to the restrictions imposed by their own characteristics and the limitation on the number of formal jobs that are in fact available.

Abowd and Farber (1982) propose evaluating the existence of a job queue based on the estimates of a first-stage bivariate probit model. In the absence of a job queue the parameters of the CFQ equation are restricted to zero with the exception of the constant term. Based on these restricted and unrestricted estimates of the bivariate probit model, the null hypothesis of the non-existence of the queue based on a likelihood ratio (LR) statistic can be tested. A further test that could be applied (the opposite to that of the no-queue model) is that of the existence of a universal queue. A universal queue is equivalent to the restriction that all the parameters of the IQ equation are zero with the exception of the constant term. As in the first test, we can test the null hypothesis of the existence of a universal queue with a likelihood ratio statistic.

Mengistae (1999) proposes tests based on estimates of the conditional wage equations. The tests seek to examine hypotheses about the parameter π of equation (18), which represents the proportion of informal workers that have been rationed out of the formal sector. Testing the non-existence of a job queue involves testing the null hypothesis of π=0, while testing π=1 implies corroborating the existence of a universal queue. The first test is performed with a test of joint significance of δ_1 and δ_2 (Ho: $\delta_1 = \delta_2$=0), while in the second the restriction of $\sigma_{1\upsilon_2} = \delta_1$ is tested using an F-statistic. The limitation of this test is that it cannot be applied to Poirier's model since, as mentioned in the previous section, the wage equation of the informal workers cannot be estimated.

The length of the job queue can be calculated from the probabilities resulting from the first-stage bivariate probit estimation. Farber (1983) and Venti (1987) suggest that the length of the job queue (q) is calculated as the inverse of the average probability of being chosen from the queue given that the worker is in the queue (P(CFQ)). Likewise, the probabilities can also be deduced of joining the queue (P(IQ)) and of working in the formal sector (P(formal)). Depending on the structure of the model (sequential or simultanious), the above probabilities have different structures, namely:

	Sequential	**Simultaneous**
P(IQ)=Prob($V_{1i}^* > 0$):	$\Phi(X_{1i}^{'}\hat{\beta}_1)$	$\Phi(X_{1i}^{'}\hat{\beta}_1)$
P(CFQ)=Prob($V_{2i}^* > 0 \mid V_{1i}^* > 0$):	$\Phi(X_{2i}^{'}\hat{\beta}_2)$	$\Phi_2(X_{1i}^{'}\hat{\beta}_1, X_{2i}^{'}\hat{\beta}_2, \rho) / \Phi(X_{1i}^{'}\hat{\beta}_1)$
P(formal)=Prob($V_{1i}^* > 0$ y $V_{2i}^* > 0$):	$\Phi(X_{1i}^{'}\hat{\beta}_1)\Phi(X_{2i}^{'}\hat{\beta}_2)$	$\Phi_2(X_{1i}^{'}\hat{\beta}_1, X_{2i}^{'}\hat{\beta}_2, \rho)$
q = $1 \Big/ \frac{1}{n}\sum_{i=1}^{n}$ P(CFQ):	$1 \Big/ \frac{1}{n}\sum_{i=1}^{n}\Phi(X_{2i}^{'}\hat{\beta}_2)$	$1 \Big/ \frac{1}{n}\sum_{i=1}^{n}\Phi_2(X_{1i}^{'}\hat{\beta}_1, X_{2i}^{'}\hat{\beta}_2, \rho) / \Phi(X_{1i}^{'}\hat{\beta}_1)$

2.4 Data and descriptive evidence

The data used in this chapter come from the Great Integrated Household Survey (GIHS) for 2009, carried out by the National Administrative Statistics Department (DANE). This cross-section survey has information at micro-level on labor force, unemployment and informality of thirteen major cities with their metropolitan areas of Colombia.[4]

[4] Namely, Barranquilla, Bogotá, Bucaramanga, Cali, Cartagena, Cúcuta, Ibagué, Manizales, Medellín, Montería, Pasto, Pereira, and Villavicencio. These metropolitan areas represent 45% of total population and about 60% of urban population according to 2005 Population Census.

The sample considered in this chapter is composed of individuals between 12 and 65 years old and we further excluded agriculture workers. Our final sample is composed of 62,278 individuals.[5]

Informal workers can be defined as those workers who do not have access to the social security system to receive healthcare and a retirement pension. While this definition is not exactly in line with the one typically adopted by dualist theory, it should be noted that the approach followed here is based on a framework for examining decisions regarding the type of employment, which are mainly determined by the characteristics of the individual and the labor supply. By contrast, the definition of informality from a dualist perspective is determined by the characteristics of the production unit, and includes such factors as productivity, technology, type of employment, among others, but not those concerning the individual.[6] Following this definition of informality, we have 36,293 (58.3%) formal workers and 25,985 (41.7%) informal workers.

If we now turn to examine the characteristics of the formal and informal workers, taking into consideration their personal characteristics as well as the characteristics of the households they inhabit and the jobs they perform. Tables 2.2 and 2.3 show that formal workers are on average better educated than their informal counterparts – their being a difference of 2.4 years of schooling. Most formal workers have completed secondary education (52%) and tertiary education (36%). Similarly, most informal workers have completed secondary school (57%), but a good proportion has only primary education (26%). As far as gender is concerned, the majority of informal workers are women (51%), while the majority of formal workers are men (56%). It can additionally be noted that most informal workers are not heads of households (64%), which might indicate that the search for a secondary source of household income is conducted in the informal sector.

[5] Notice that we excluded government employees, employers and self-employed. Given this exclusion the informality rate may differ from that reported by Jütting and De Laiglesia (2009) or International Labor Organization (ILO, 2011).

[6] A more detail discussion can be found in Khamis (2012).

Table 2.2. Personal and employment characteristics

	Formal	Informal	Total sample
Education			
Less than primary	0.34%	1.65%	0.81%
Primary	11.54%	26.42%	16.86%
Secondary	51.77%	57.26%	53.73%
Tertiary	36.35%	14.67%	28.61%
Total	100%	100%	100%
Gender			
Female	44.44%	51.25%	46.87%
Male	55.56%	48.75%	53.13%
Total	100%	100%	100%
Head of household			
Non-head	56.97%	64.63%	59.70%
Head	43.03%	35.37%	40.30%
Total	100%	100%	100%
Marital status			
Unmarried	46.54%	53.86%	49.15%
Married	53.46%	46.14%	50.85%
Total	100%	100%	100%
Occupational status before the current job			
Employed	84.67%	77.33%	82.05%
Unemployed	15.33%	22.67%	17.95%
Total	100%	100%	100%
Sector			
Industry	26.32%	19.10%	23.74%
Building	4.12%	9.87%	6.18%
Commerce and hotel	22.58%	33.59%	26.51%
Transport and telecommunications	9.67%	5.60%	8.21%
Financial and education	24.42%	8.13%	18.60%
Service	12.89%	23.71%	16.75%
Total	100%	100%	100%

Note: We used personal sampling weight available in the database.

In terms of age (Table 2.3), it can be seen that there are no marked differences between formal and informal workers. By contrast, in terms of the number of years of experience acquired in the current job, differences do exist: while formal workers have six years of tenure, informal workers have just three. As for hourly wages, formal workers earn 30% more than their informal counterparts, which is indicative of the precarious and low levels of productivity that exist in the informal sector.

Table 2.3. Wages and Education

	Formal		**Informal**		**Total sample**	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Age	34.3	10.1	32.9	11.5	33.8	10.6
Education (years)	11.0	3.5	8.6	3.7	10.2	3.8
Tenure at job (years)	4.7	5.7	2.8	4.0	4.0	5.2
Education of household (other members)	5.2	2.5	4.6	2.4	5.0	2.5
Real hourly wage	3269.2	1670.7	2311.9	14384.5	2927.4	1799.5

Note: We used personal sampling weight available in the database. The wages are in Colombian pesos. We use the consumer price indexes of the biggest cities which were obtained from DANE. In Ibagué the consumer prices index is not calculated by DANE, so we used the consumer prices index of Pereira, given the similarities in population and social and cultural characteristics, as well as proximity between these cities. In December 2009 the exchange rate was 2935 Colombian pesos per euro.

In the case of the household variables (Tables 2.3 and 2.4) it can be seen that informal workers are more likely to live in households with lower levels of education. Likewise, a high proportion of informal workers live in households with school-age children (aged between 6 and 17). An equally important factor characterizing informality is the presence of other informal workers among the same household members: 60% of informal workers live in households in which there are more informal workers.

Table 2.4. Household characteristics

	Formales	**Informales**	**Total sample**
Presence of children between 0-2 years old	10.17%	11.58%	10.77%
Presence of children between 3-5 years old	11.60%	13.29%	12.37%
Presence of children between 6-10 years old	20.26%	22.54%	21.22%
Presence of children between 11-17 years old	27.12%	33.94%	29.18%
Presence of other relatives working as formal	47.83%	31.50%	41.48%
Presence of other relatives working as informal	38.97%	60.33%	48.58%

Note: We used personal sampling weight available in the database.

The distribution by economic sector (Table 2.2) shows that formal workers are concentrated primarily in the industrial (26%) and the financial and education sectors (24%), while there is a prevalence of informal workers in the retail trade and the hotel sector (34%) and in that of the personal services sector (24%). Finally, according to previous employment status records, around 23% of informal workers were unemployed before entering their current job.

The above characterization shows that informal workers are in the main unmarried women with an average age of 33, non-heads of households, and with an educational level that does not extend beyond that of secondary schooling. The households in which the informal workers are resident are characterized by a high presence of school-age children, poor levels of education (with an average of 4.6 years of schooling) and the presence of other household members working in the informal labor market. As for employment, it is found that the hourly wage of informal workers is 30% less than that of the equivalent salary paid to a formal worker and that more than half work in the commerce, hotel and personal services sectors. Informality, therefore, is a sector with precarious working conditions, characterized by an abundance of unskilled labor, concentrated in sectors with a low level of technical development.

2.5 Results

2.5.1 Univariate probit model: absence of a job queue

The probit estimation in the absence of a job queue offers a good description of the variables to be included among the determinants of the workers' choice of labor sector. The results of this estimation are shown in Table 2.5. It can be seen that as the workers' level of education improves, the probability of their working in the formal sector rises and that those who have successfully completed higher education have the greatest probability of finding work in this sector. This probability also increases with the number of years of work experience and tenure, but at a decreasing rate in the case of the first of these two variables. The coefficient of the gender variable was positive and statistic significant, which is indicative of a certain degree of discrimination in the process of deciding between the formal or informal sectors. In the case of the heads of household and civil status variables, the estimates show that being a head of household or being married means a worker is more likely to find work in the formal sector. These positive effects may reflect the fact that the greater responsibilities associated with a household are in turn associated with a greater need for a good job or are a good indication of a worker commitment to their workplace and hence their suitability to fill a formal position.

The household variables show that the presence of children, especially those in pre-school ages (3-5 years) and those that are older (11-17 years), has a negative effect on the probability of finding formal employment. By contrast, the variables associated with a higher educational level and the number of formal workers resident in the household have a positive effect on finding formal employment.

As for the workers' previous employment status, being unemployed prior to finding their current job has a negative effect on the attainment of formal employment, compared to that of having been previously employed (base category). Finally, the economic sector variable shows that finding work in sectors other than those of the commerce, hotel and building increases the probability of finding formal employment, with the exception of the personal services sector which has a negative effect on the probability of being formally employed.

Table 2.5. Estimates of univariate probit model
(Y = 1 formal, 0 informal)

	Coefficients	**z**	**Marginal effects**	**z**
Constant	-3.262***	-39.19		
Personal characteristics				
Primary	0.373***	6.52	0.137***	6.51
Secondary	1.002***	17.68	0.369***	17.65
Tertiary	1.625***	28.18	0.597***	28.15
Age	0.123***	36.71	0.045***	36.45
Age2	-0.002***	-36.16	-0.001***	-35.92
Tenure	0.061***	42.24	0.023***	43.29
Male	0.170***	14.05	0.062***	14.06
Head of household	0.234***	17.63	0.086***	17.65
Married	0.142***	11.65	0.052***	11.66
Unemployed before the current job	-0.257***	-17.03	-0.094***	-17.04
Household characteristics				
# kids 0-2	-0.051***	-3.05	-0.019***	-3.05
# kids 3-5	-0.080***	-5.18	-0.029***	-5.18
# kids 6-10	-0.068***	-6.28	-0.025***	-6.28
# kids 11-17	-0.132***	-16.69	-0.049***	-16.68
# of relatives working as formal	0.223***	24.21	0.082***	24.26
Education of household	0.020***	7.36	0.007***	7.36
Sector				
Industry	0.352***	23.93	0.130***	23.91
Transport and telecommunications	0.473***	21.88	0.174***	21.88
Financial and education	0.643***	39.15	0.237***	39.41
Services	-0.087***	-5.63	-0.032***	-5.63
N	62278			
Log L	-37307.44			
Pseudo R2	0.27			
AIC	74680.88			
BIC	74985.9			

Note: ***, **, *, denotes significance at 1%, 5% and 10%, respectively. Robust standard errors. Less than primary education and commerce, hotel and building sectors as references. We include city dummies.

2.5.2 Bivariate probit model reduced form: the existence of a queue for formal jobs

The univariate specification supposes that the process of being either formally or informally employed in fact omits the selection mechanism used by employers and, as such, the possible formal job rationing that exists. As mentioned above, a more appropriate specification for describing the process of finding work in the presence of job restrictions is the estimation of a bivariate model. The estimates of both the sequential and simultaneous bivariate probit models with partial observability are shown in Table 2.6.

Interesting patterns are observed for the IQ and CFQ equations. A higher level of education increases the likelihood of joining the queue and of being chosen for a formal job, with the most marked effect being found for those that have successfully completed higher education. With regard to the other human capital variables of age and tenure, we can see that the former has a positive and decreasing effect on the likelihood of joining the queue, while the latter has a positive effect.

Regarding the gender variable, being male increases both the likelihood of joining the job queue and of being chosen from it. In the case of the head of household variable, positive effects were recorded in both the worker's decision to seek formal employment and in the employers' selection process. As mentioned above, the head of household variable appears to be associated to a greater degree of work commitment, and so employers consider this factor positively when recruiting. Finally, it is noted that a married individual is more likely than a single person to seek formal employment, and the former is also more likely to be selected for formal employment.

As in the univariate specification, the variable of being unemployed prior to current employment is not positively related with joining the job queue. This outcome might imply that a phase of prior unemployment results in greater pressure to find employment of any kind, so that individuals would rather obtain an informal job quickly than have to go through the process of finding formal employment. Similarly, we find that workers in the commerce, hotel and building sectors are more likely to be employed informally either because they have not joined the job queue or because they have not been selected from it.

In the case of the variables that are only included in the IQ equation, the results do not differ greatly from those obtained in the univariate specification. The presence of children at both ends of the age spectrum (3-5 and 11-17 years) in the household has a negative effect, while a higher level of education in the household and the presence of other household members employed in the formal sector have a positive impact on finding work in the formal sector.

The simultaneous estimation shows that the correlation between the two equations is negative (-0.76) and statistically significant. This result indicates that finding employment in either the formal or informal sectors is more of a simultaneous process than a sequential one and that there are unobserved factors that affect the process.

Table 2.6. Estimates of bivariate probit model reduced form
(Y_j = 1 formal, 0 informal; j= IQ, CFQ)

	Abowd-Farber model (Sequential)				**Poirier model (Simultaneous)**			
	IQ		**CFQ**		**IQ**		**CFQ**	
	Coeff.	**z**	**Coeff.**	**z**	**Coeff.**	**z**	**Coeff.**	**z**
Constant	-3.529***	-38.83	0.024	0.21	-2.356***	-12.13	0.051	0.43
Personal characteristics								
Primary	0.069	0.47	0.472***	4.24	-0.198	-1.21	0.473***	4.18
Secondary	0.564***	3.89	1.067***	9.73	0.023	0.14	1.067***	9.56
Tertiary	1.141***	7.76	1.601***	14.43	0.417**	2.49	1.647***	14.62
Age	0.152***	27.7			0.131***	26.01		
Age2	-0.002***	-27.97			-0.002***	-26.11		
Tenure	0.325***	18.35			0.269***	17.42		
Male	0.183***	7.56	0.112***	4.20	0.092***	3.45	0.156***	5.60
Head of household	0.303***	11.33	0.063**	2.53	0.263***	9.42	0.042*	1.67
Married	0.097***	3.86	0.128***	5.35	0.062**	2.33	0.111***	4.56
Unemployed before the current job	-0.346***	-12.07	-0.143***	-5.80	-0.286***	-9.81	-0.107***	-4.22
Household characteristics								
# kids 0-2	-0.079***	-3.06			-0.064***	-2.85		
# kids 3-5	-0.095***	-3.97			-0.081***	-3.91		
# kids 6-10	-0.076***	-4.49			-0.065***	-4.41		
# kids 11-17	-0.165***	-13.60			-0.143***	-13.51		
# of relatives working as formal	0.302***	20.00			0.268***	19.59		
Education of household	0.032***	7.32			0.029***	7.59		
Sector								
Industry	0.235***	8.14	0.351***	11.12	0.141***	4.32	0.312***	9.31
Transport and telecommunications	0.359***	8.82	0.459***	9.68	0.224***	5.1	0.446***	9.05
Financial and education	0.564***	17.22	0.600***	18.48	0.426***	11.72	0.567***	16.09
Services	-0.043	-1.31	-0.131***	-4.13	0.067*	1.80	-0.200***	-5.99
Rho					-0.761***	-27.31		
N	62278				62278			
Log L	-36530.74				-36470.07			
AIC	73175.47				73056.14			
BIC	73702.33				73592.24			

Note: ***, **, *, denotes significance at 1%, 5% and 10%, respectively. Robust standard errors. Less than primary education and commerce, hotel and building sectors as references. We include city dummies. IQ: "in-queue" equation; CFQ: "chosen from the queue" equation.

2.5.3 The Formal and informal wage equations

The formal and informal wage equation estimates are shown in Table 2.7. It should be noted that all the estimated coefficients of the selection correction are statistically significant; moreover, the joint tests of these parameters to determine if the selection bias is significant are rejected. The main implication of the selection bias is that the rates of return to human capital, in general, and to education, in particular, are overestimated with OLS; hence, the importance of correcting it when such bias exists.

Table 2.7. Wage equations for formal and informal workers with correction for selectivity based on the bivariate probit models
(Y= Log real hourly wage)

	OLS whitout correction				Correction base on the Abowd-Farber's model				Correction base on the Poirier's model	
	Formal		Informal		Formal		Informal		Formal	
	Coeff.	t	Coeff.	t	Coeff.	t	Coeff.	t	Coeff.	t
Constant	-4.598***	-110.74	-4.941***	-100.96	-5.195***	-83.00	-3.990***	-37.76	-5.133***	-81.00
Primary	0.026	0.89	0.154***	6.14	0.163***	5.13	0.199***	6.26	0.162***	5.01
Secondary	0.257***	8.95	0.291***	11.63	0.524***	13.33	0.345***	6.01	0.520***	13.1
Tertiary	0.856***	29.45	0.739***	26.92	1.211***	26.75	0.780***	8.95	1.209***	26.29
Age	0.022***	13.03	0.034***	15.86	0.029***	15.67	0.004	1.32	0.027***	14.02
Age2	-0.0002***	-8.29	-0.0004***	-14.52	-0.000***	-11.23	0.0001	-0.55	-0.0001***	-9.79
Tenure	0.017***	38.79	0.021***	20.31	0.019***	35.96	-0.028***	-5.43	0.018***	34.38
Male	0.064***	11.65	0.085***	9.79	0.087***	14.75	0.069***	6.15	0.091***	14.75
Head of household	0.060***	10.65	0.082***	9.89	0.072***	12.54	0.065***	7.09	0.068***	12.07
Married	0.036***	6.72	0.055***	7.04	0.058***	9.96	0.068***	6.41	0.056***	9.77
Industry	0.078***	11.51	-0.010	-0.88	0.146***	15.65	0.015	0.61	0.136***	15.57
Transport and telecommunication	0.0004	0.04	-0.082***	-4.99	0.086***	6.56	-0.058*	-1.82	0.079***	6.15
Financial and education	0.251***	36.01	0.178***	11.45	0.349***	29.82	0.191***	4.87	0.341***	30.65
Services	0.098***	11.68	0.042***	4.25	0.076***	8.33	0.019	1.59	0.062***	6.28
$\sigma_{1\upsilon_1}$					0.098***	9.43			0.079***	5.63
$\sigma_{2\upsilon_1}$					0.468***	8.77			0.500***	9.79
$\sigma_{1\upsilon_2}$							-0.309***	-13.64		
δ_1							-0.168***	-10.71		
δ_2							0.273***	3.00		
N	36293		25985		36293		25985		36293	
R2 ajust	0.39		0.17		0.39		0.17		0.39	

Note: ***, **, *, denotes significance at 1%, 5% and 10%, respectively. Robust standard errors. Less than primary education and commerce, hotel and building sectors as references. All models include city dummies.

The corrected estimates obtained with the sequential model show that the human capital variables are important positive factors for determining wages both in the formal and informal sectors. Schooling, especially tertiary education, has the greatest effect. In the informal sector this positive effect might indicate that the jobs of some of these workers are not secondary jobs and, therefore, returns to education exist. Turning to personal characteristics, we find that there are gender wage differences in both sectors, that heads of household and married people enjoy a wage premium with respect to non-heads of household and to those who are single, and that in this last case the premium is higher in the informal sector.

With the simultaneous model (Poirier's model) it was only possible to correct the formal wage equation (see section 2.3) and it should be noted that there are no marked differences between this and the sequential model correction. The selection coefficients are equally significant: there being positive returns on education, while age and tenure both present positive returns, being concave in the case of age. Positive wage premiums are also found for being male, the head of a household and married.

The above estimates, as well as being used to corroborate the existence of the queue for formal jobs (next section), can be used to calculate the expected wages paid in the formal and informal sectors, which are incorporated as additional regressors in the structural bivariate probit model (section 2.5.5). In next section we analyze the existence of the job queue.

2.5.4 The existence and length of the formal job queue: results

In order to test the validity of the results obtained with the bivariate probit models and, hence, the existence of a formal job queue, we calculated the Abowd-Farber and Mengistae tests as described in section 2.3.2. Abowd and Farber (1982), based on a likelihood ratio test, examine the null hypothesis of the no-queue scenario and the existence of a universal queue using a bivariate probit model. The first hypothesis implies that the coefficients of the CFQ equation are zero with the exception of the constant, while the second restricts the coefficients of the IQ equation to zero. Employing a different test, Mengistae (1999) examines the existence or otherwise of formal job rationing by conducting tests on the π parameter that appears in the conditional wage equations. The null hypothesis of the existence of a job queue is π=0 and the null hypothesis of a universal queue is π=1. The first hypothesis is checked using a test of joint significance of $\delta_1 = \delta_2$=0 and the second is verified with an F-test of $\sigma_{1\upsilon_1} = \delta_1$. These tests are presented in Table 2.8.

Table 2.8. Test for universal queue and no queue hypothesis

	Abowd-Farber test		Mengistae test	
Ho	**Abowd-Farber model**	**Poirier model**	**Abowd-Farber model**	**Poirier model**
No queue	LR Chi2(23) = 2761.36 Prob > Chi2 = 0.000	LR Chi2(23) = 3127.55 Prob > Chi2 = 0.000	F(2, 25833) = 59.90 Prob > F = 0.000	-
Universal queue	LR Chi2(32) = 7828.04 Prob > Chi2 = 0.000	LR Chi2(32) = 7944.00 Prob > Chi2 = 0.000	F(1, 25833) = 57.77 Prob > F = 0.000	-

Note: As mentioned, Mengistae's test is not available because the informal wage equation cannot be estimated (see section 2.3.1).

The test results show that, at a significance level of 5%, both the null hypothesis of the non-existence of a job queue as well as the existence of a universal queue is rejected. This provides evidence in support of a bivariate as opposed to a univariate specification to describe the process of obtaining a formal or informal job. Furthermore, the existence of a job queue indicates a rationing of formal jobs. Hence, not everyone who wants to be employed in the formal sector has this possibility, because of the restriction on the creation of formal jobs. In line with Mengistae (1999), from equation (18), it can be deduced that the estimated proportion of informal workers who have been rationed out of the formal sector is 54% ($\pi = \hat{\delta}_1 / \hat{\sigma}_{1\upsilon_2}$). At the same time, the rejection of a universal queue indicates that not all workers seek employment in the formal sector. Thus, two types of informal worker emerge: those who benefit, or earn more, from being employed in the informal sector, and those that given their low level of qualifications opt for this sector as the only way of subsisting and escaping unemployment.

Next, the length of the job queue is calculated together with the probabilities that can be deduced from the models. Table 2.9 shows the results for the mean values of the above probabilities. It can be seen that the conditional probability of being a formal worker lies between 61 and 65% and that only individuals who have tertiary education present a probability greater than 80% of being employed in the formal sector. With regard to the length of the queue, this is calculated as being around 1.25, i.e., for each 100 workers in the formal sector, there are a further 125 who wish to obtain a formal job. This value is high in comparison to the lengths of the formal job queues in Brazil (1.19) and Chile (1.06) (Soares, 2004; Puentes and Contreras, 2009).

An important aspect of the analysis of probabilities is to see how these vary for different characteristics. Thus, it can be seen that women are less likely to seek formal employment (70%) than men are (76%), but once in the queue their probability of being selected for a formal vacancy increases by seven percentage points. In the case of education, individuals with low levels of education are more likely to work in the informal sector. By contrast, the more highly educated are more likely to seek formal employment and to receive more offers of this type of work. As for the heads of household, it can be noted that they are more likely to seek formal employment than the non-heads of household, yet the latter have similar chances of receiving formal job offers as those enjoyed by the former. As for the married/single variable, married individuals are more likely to find formal employment than single individuals, and at the same time the former are more likely to seek formal employment and to receive a greater number of offers of this type of work.

Table 2.9. Probabilities and queue length from the Abowd-Farber (A-F) and Poirier bivariates models

	Total sample		**Male**		**Female**	
	A-F	**Poirier**	**A-F**	**Poirier**	**A-F**	**Poirier**
P(IQ)	0.73	0.79	0.76	0.80	0.70	0.77
P(CFQ)	0.80	0.80	0.82	0.83	0.77	0.77
P(formal)	0.61	0.65	0.66	0.68	0.57	0.61
q	1.25	1.24	1.21	1.20	1.29	1.29

	Less than primary		**Primary**		**Secondary**		**Tertiary**	
	A-F	**Poirier**	**A-F**	**Poirier**	**A-F**	**Poirier**	**A-F**	**Poirier**
P(IQ)	0.43	0.66	0.54	0.67	0.69	0.74	0.90	0.91
P(CFQ)	0.38	0.38	0.59	0.59	0.80	0.80	0.92	0.92
P(formal)	0.17	0.25	0.33	0.41	0.56	0.60	0.83	0.84
q	2.65	2.65	1.70	1.69	1.25	1.25	1.09	1.08

	Head of household		**Non-Head of household**		**Non-married**		**Married**	
	A-F	**Poirier**	**A-F**	**Poirier**	**A-F**	**Poirier**	**A-F**	**Poirier**
P(IQ)	0.81	0.85	0.68	0.74	0.67	0.74	0.79	0.83
P(CFQ)	0.82	0.82	0.79	0.79	0.78	0.79	0.82	0.82
P(formal)	0.68	0.71	0.56	0.60	0.55	0.59	0.67	0.70
q	1.22	1.22	1.27	1.26	1.28	1.27	1.22	1.22

	Unemployed before the current job		**Employed before the current job**	
	A-F	**Poirier**	**A-F**	**Poirier**
P(IQ)	0.71	0.77	0.74	0.79
P(CFQ)	0.75	0.76	0.81	0.82
P(formal)	0.56	0.60	0.62	0.66
q	1.33	1.31	1.23	1.23

IQ: “in the queue”; CFQ: “chosen from the queue”; q: queue length.

Regarding the length of the queue, the workers with no schooling form the longest job queue: for every 100 formal employees, there are 265 employees without schooling that want a formal job. Likewise, women, the non-heads of households and those that are not married find themselves in a longer job queue for formal employment: for every female worker, non-head of household or single individual in the formal sector there are, respectively, 1.29, 1.27 and 1.28 workers that want a formal job. A period of prior unemployment also lengthens the job queue, in this case to 1.33 workers.

2.5.5 Determinants of the decision of being a formal or informal worker: structural bivariate probit model

Equation (1) suggests that an individual’s preference for formal employment depends on the wage premium offered by this sector ($\ln W_f - \ln W_{inf}$). Under this framework, a percentage change in $\ln W_f$ has the same effect as a percentage change of equal magnitude in $\ln W_{inf}$. The different valuation of wages between sectors may result from the fact that there are non-wage benefits that are directly associated with wages. Thus, an increase in formal sector wages implies access to a higher retirement pension, as well as subsidies for transport, food, recreation, housing, and education, among other benefits. It, therefore, seems reasonable to think that in the decision to join the queue for formal employment, the wages paid in the formal and informal sectors have asymmetric effects. This means that instead of including the wage differential between sectors, wages should be included separately in the IQ equation.[7]

A further important question to bear in mind are the types of wage equation estimates to be included in the structural bivariate probit. In practice there are two types: unconditional and conditional. The former are estimates of the wage equation that do not take into account the type of job performed by the workers and, therefore, are free of the effects of wage selection bias (the inverse Mills ratios). The estimates of the conditional wage equation, by contrast, are based on the relative weight of the workers' wages once the specific job sector has been determined. This wage estimate differs from that of the unconditional wage equation estimate insofar as the selection effects are due to differences in the rates of return to the unobserved characteristics of the workers and not to the levels of these characteristics (Gyourko and Tracy, 1988; Mengistae, 1999). Here, the unconditional wage equation estimates are used, since we are interested in determining whether there is a wage premium at which workers choose a formal job.

The maximum likelihood estimates of the structural and sequential probit models with partial observability are shown in Table 2.10. Note that for the reduced estimate (Table 2.6), the education variables have not been considered. In fact, they have been eliminated as they are also used as predictors of wage levels and so the possibility of the double-counting of effects is avoided. However, in Table A2.1 of the appendix these estimates are conducted with the education variables and, as can be seen, no marked differences are found.

[7] This alternative was proposed by Venti (1987).

Table 2.10. Estimates of Abowd-Farber structural bivariate probit model
(Y_j = 1 formal, 0 informal; j= IQ, CFQ)

	IQ		CFQ	
	Coeff.	**z**	**Coeff.**	**z**
Constant	4.105***	19.44	1.470***	3.51
Expected wages of the formal/informal sector				
$\ln \hat{W}_f$	1.943***	30.34	-0.159***	-3.75
$\ln \hat{W}_{inf}$	-0.537***	-16.33		
Personal characteristics				
Age	0.040***	18.43		
Age2	-0.001***	-37.89		
Tenure	-0.010***	-8.89		
Male	0.266***	16.3	1.274***	2.77
Head of household	0.120***	9.46	0.233***	6.51
Married	0.067***	4.97	0.249***	5.32
Unemployed before the current job	-0.299***	-18.30	-0.140**	-2.43
Household characteristics				
# kids 0-2	-0.050***	-3.8		
# kids 3-5	-0.096***	-7.95		
# kids 6-10	-0.081***	-9.84		
# kids 11-17	-0.134***	-19.00		
# of relatives working as formal	0.287***	22.62		
Education of household	0.018***	7.44		
N	62278			
Log L	-81739.52			
AIC	163525.03			
BIC	163753.85			

Note: ***, **, *, denotes significance at 1%, 5% and 10%, respectively. Robust standard errors. Less than primary education and commerce, hotel and building sectors as references. We include city dummies. IQ: "in the queue" equation; CFQ: "chosen from the queue" equation.

It can be seen, therefore, that formal sector wages have a marked positive effect on the probability of joining the queue for formal employment, while this same variable reduces the probability of being selected for a formal vacancy. This latter effect is attributable to the need employers have to minimize their production costs. In turn, the predicted wages in the informal sector have a negative effect on the search for a formal job, and this effect differs in magnitude to that generated by formal sector wages. The foregoing discussion validates the hypothesis of asymmetric effects in the decision to join the queue due to changes in the wages of the formal and informal sectors.

Turning to the effect of the other variables, it can be seen that there are no marked differences in the estimates of the reduced model. Being male increases the likelihood of joining the queue for a formal job and of being selected from it for a formal vacancy. Being the head of the household and being married exert a positive effect both on the probability of seeking formal employment and of being selected for a job, with a greater impact on the second of these probabilities. Finally, a prior period of unemployment has a negative effect on the whole process of finding formal employment.

In short, the estimates confirm the existence of formal job rationing. Such rationing is a reflection of the limited capacity of the formal, modern sector of the economy to generate sufficient jobs, thus resulting in a job queue and the mechanisms that employers use to fill these few vacancies. Given these two selection mechanisms that characterize the labor market, finding a better job depends to a considerable degree on the wage premium offered by the formal sector and the characteristics of the workers (especially their human capital), both of which serve to enhance the job search and to increase the number of job offers that a worker might receive.

The wage a worker can expect to receive in formal employment, along with the extent of the individual's work commitment and responsibility and the quality of his or her work history are the main factors that determine whether a worker finds employment in the formal sector. Likewise, the household variables play a key role in determining whether work is sought in the formal or informal sectors. The financial pressure on a household of having young children has a negative effect on finding formal work. The need for resources increases the opportunity cost of finding formal employment, and so individuals prefer to generate income more rapidly and with fewer legal obstacles. By contrast, a better educational environment and a greater number of formal workers in the household have a positive effect on helping individuals find formal employment.

2.6 Conclusions

Adopting an alternative methodology to those more typically used in the analysis of the factors influencing the decision to enter the formal or informal job sectors in Colombia, this chapter has presented evidence of formal job rationing in the Colombian labor market. The relative paucity of formal employment owing to the lack of development of the modern, formal sector has led to the generation of a dual selection mechanism for finding work in the formal sector. First, individuals must be prepared to join a queue for formal jobs. Joining this queue depends principally on the wage premium offered by the formal sector and the level of responsibility held by the individual in the household.

In the second stage, with the individuals now in the queue for formal employment, they must await a job offer. Such job offers depend principally on the absolute recruitment costs of the individuals, the degree of responsibility and work commitment they show and their work history. Thus, workers that manifest a high degree of responsibility and can present a good work history (or who were in employment in the last period) enjoy a greater likelihood of being selected for a formal job.

In quantitative terms this study has been able to determine that around 54% of workers employed in the informal sector owe their position to the rationing of formal jobs. This result indicates that there are two types of informal worker: those who do not have alternative other than to be employed in the informal sector, and those who are able to derive certain benefits from being employed in this sector. This latter group sees informality as a sector in which they can accumulate human capital and where they can reap returns on their education; in other words, they, unlike the former group, do not see informality as a secondary job sector. However, the characteristics of this second type of informal worker need to be examined in greater depth, since the outcomes reported here may well be biased as the preferences of workers for switching from one sector to another have not been taken into account. This limitation is inherent to the survey used.[8]

Finally, the confirmation of the existence of a dual selection mechanism operating in the labor market for finding employment in the formal or informal sectors is evidence of a process of labor segmentation that characterizes Colombia today. The scarcities of formal jobs combined with the poor quality of those that do exist are its principal characteristics.

[8] Soares (2004) for the Brazilian case can distinguish between informal workers who want to change their current job for a formal job. This information allows accounting for a greater diversity of informal workers.

2.7 References

Abowd, J. and Farber, H. (1982). "Job Queue and the Union Status of Workers", *Industrial and Labor Relations Review*, 35(3): 354-367.

Archibald, R. B. (1977). "Labor Queues and Involuntary Unemployment", *Economic Inquiry*, 15(1): 33-50.

Co, C., Gang, I. y Yun, M. (2005). "Self-Employment and Wage Earning in Hungary", *Review of Development Economics*, 9(2): 150-165.

Contreras, D., De Mello, L. and Puentes, E. (2008). "Tackling Business and Labour Informality in Chile", *OECD Economics Department Working Papers*, No 607, OECD Publishing

Dickens, W. and Lang, K. (1985). "A Test of Dual Labor Market Theory", *The American Economic Review*, 75(4): 792-805.

Farber, H. (1983). "The Determination of the Union Status of Workers", *Econometrica*, 51(5): 1417-1437.

Gyourko, J. and Tracy, J. (1988). "An Analysis of Public and Private Sector Wages Allowing for Endogenous Choices of both Goverment and Union Status", *Journal of Labor Economics*, 6: 229-253.

Heckman, J. (1976). "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simpler Estimator for such Models", Annals of Economic and Social Measurement, 5(4): 475-492.

Heckman, J. (1979). "Sample Selection Bias as a Specification Error", *Econometrica*, 47(1): 153-161.

Huguet, A. (1996). "Dualidad en el Mercado de Trabajo Español", *Revista de Economía Aplicada*, 4(11): 81-104.

ILO (2011). 2011 *Labour Overview. Latin America and the Caribbean*, Lima.

Jütting, J. and De Laiglesia, J. R. (2009). *Is Informal Normal? Towards More and Better Jobs in Developing Countries*. OECD Development Centre, 163 pages.

Khamis, M. (2012). "A Note on Informality in the Labour Market", *Journal of International Development*, 24: 894-908.

Lang, K. and Dickens, W. (1987). "Neoclassical and Sociological Perspectives on Segmented Labor Markets", *NBER Working Paper Series*, No 2127, NBER.

Lee, L. F. (1978). "Unionism and Wage Rates: A Simultaneous Equation Model with Qualitative and Limited Dependent Variables", *International Economic Review*, 19(2): 415-433.

Maddala, G. S. (1983). *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge University Press.

Maddala, G. S. and Nelson, F. (1975). "Switching Regression Models with Exogenous and Endogenous Switching", Proceedings of the American Statistical Association (Business and Economics Section), 423-426.

Maloney, W. (1998). "The Structure of Labor Markets in Developing Countries. Time Series Evidence on Competing Views", *Policy Research Working Paper*, No 1940, The World Bank.

Meng, C. and Schmidt, P. (1985). "On the Cost of Partial Observability in the Bivariate Probit Model", *International Economic Review*, 26(1): 71-85.

Mengistae, T. (1999). "Wage Rates and Job Queue: Does the Public Sector Overpay in Ethiopia", *Policy Research Working Paper*, No 2105. The World Bank.

Mohanty, M. (2001). "Testing for the Specification of the Wage Equation: Double Selection Approach or Single Selection Approach", *Applied Economics Letter*, 8(8): 525-529.

Poirier, D. J. (1980). “Partial Observability in Bivarite Probit Models”, *Journal of Econometrics*, 12(2): 209-217.

Puentes, E. and Contreras, D. (2009). “Informal Jobs and Contribution to Social Security: Evidence from a Double Selection Model”, *Serie Documentos de Trabajo*, No 307, Universidad de Chile, Santiago, Dicember.

Schneider, F., Buehn, A. and Montenegro, C.E. (2010). “New Estimates for the Shadow Economies all Over the World”, *International Economic Journal*, 24(4): 443-461.

Soares, F. (2004). “Do Informal Workers Queue for Formal Jobs in Brazil?”, *Working Paper*, No 1021, IPEA, Brasilia.

Tunali, I. (1986). “A General Structure for Models of Double-Selection and an Application to Joint Migration/Earning Process with Remigration”. In R. Ehrenberg (Ed.), *Research in Labor Economics*, Vol. 8, Part B, pp. 235-282. Greenwich. MA.: JAI Press.

Uribe, J. and Ortiz, C. (2006). *Informalidad Laboral En Colombia 1988-2000: Evolución, Teorías y Modelos*. Programa Editorial Universidad del Valle. Cali-Colombia.

Uribe, J., Ortiz, C. and García, G. (2007). “La Segmentación del Mercado Laboral Colombiano en la Década de los Noventa”, *Revista de Economía Institucional*, 9(16): 189-221.

Venti, S. F. (1987). “Wages in the Federal and Private Sectors”. In D. A. Wise (Ed.), *Public Sector Payrolls*. pp. 147-182. University of Chicago Press. Chicago.

Appendix

Table A2.1. Estimates of Abowd-Farber structural bivariate probit model
(Y_j = 1 formal, 0 informal; j= IQ, CFQ)

	IQ		CFQ	
	Coeff.	**z**	**Coeff.**	**z**
Constant	6.980***	40.21	1.810***	5.03
Expected wages of the formal/informal sector				
$\ln \hat{W}_f$	2.626***	65.71	-0.171***	-5.98
$\ln \hat{W}_{inf}$	-0.546***	-19.52		
Personal characteristics				
Primary	0.263***	4.93	0.941***	2.69
Secondary	0.209***	3.87	0.724**	2.49
Tertiary	-0.438***	-6.94	0.270	0.94
Age	0.017***	7.35		
Age2	-0.001***	-46.48		
Tenure	-0.023***	-26.51		
Male	0.120***	6.27	1.786***	6.85
Head of household	0.093***	7.85	0.247***	5.56
Married	0.064***	5.49	0.181***	3.42
Unemployed before the current job	-0.269***	-23.98	-0.094**	-1.98
Household characteristics				
# kids 0-2	-0.042***	-3.34		
# kids 3-5	-0.089***	-7.76		
# kids 6-10	-0.077***	-9.81		
# kids 11-17	-0.134***	-22.95		
# of relatives working as formal	0.261***	38.89		
Education of household	0.020***	10.33		
N	62278			
Log L	-81145.94			
AIC	162349.88			
BIC	162638.39			

Note: ***, **, *, denotes significance at 1%, 5% and 10%, respectively. Robust standard errors. Less than primary education and commerce, hotel and building sectors as references. We include city dummies. IQ: “in the queue” equation; CFQ: “chosen from the queue” equation.

Chapter 3

Labor Informality: Choice or Sign of Segmentation? A Quantile Regression Approach at the Regional Level

3.1 Introduction

One of the most notable features of developing countries is the great heterogeneity seen in their urban labor markets. It is common to observe a small productive formal sector, offering attractive labor conditions and relatively high wages, coexiting with a large informal sector which uses unskilled labor, with low earnings and productivity, and does not fully comply with established legal regulations (Dickens and Lang, 1985; Maloney, 1999 and 2004; Jütting and De Laiglesia, 2009). Nevertheless, within this large informal sector, there is a considerable variety of workers. According to the results of the previous chapter not only there is a duality between a formal and informal sector, but also a internal duality in the informal sector where coexist informal activities that are preferred to formal ones, along with other that are worse than the latter.

But why is there such diversity in the informal sector? Are there different kinds of informal workers; some who are voluntarily informal and others who end up in this sector because they do not have any alternative form of employment? Is labor informality a choice or the result of labor market segmentation?

The segmented labor markets theory considers informality as a survival alternative to escape involuntary unemployment for those who are disadvantaged or rationed out of formal employment opportunities (Dickens and Lang, 1985). The result is a dualism in terms of earnings for individuals with similar characteristics which depend on the sector in which they work. In the formal sector there are internal markets that constrain the labor supply and produce high wages, while in the informal sector there is no institutional or efficiency-wage basis that regulates wages. In addition the few entry barriers that exist and an abundant supply of unskilled workers lead to low wages. Thus wages depend on the sector in which workers are employed and not on their skills *per se* (Uribe *et al*., 2007).

In contrast, the orthodox neoclassical view of the human capital theory postulates that, like in any another market, price flexibility and free labor mobility lead to a full employment equilibrium with equal remuneration for the same kind of work (De Soto, 1987; Saavedra and Chong, 1999; Maloney, 1999). Due to this competitive market framework, being part of the informal sector may be a desirable choice for workers and firms, as it is based on the private cost-benefit calculations of belonging to the sector. Being informal can have desirable non-wage features and therefore individuals maximize their utility rather than their earnings. Alternatively, certain workers have a comparative advantage in the informal sector that they would not have in the formal sector (Gindling, 1991).

These two polarized views can be combined if the informal sector is very heterogeneous and contains elements of each scenario; namely if the informal sector has its own internal duality. Recent literature has recognized the existence of “upper” and “lower” tiers or “voluntary” and “involuntary” entry of informal employees or firms (Fields, 1990 and 2005; Cunningham and Maloney, 2001; Maloney, 2004). In such a scenario the upper-tier employees are those who are voluntarily informal because, given their specific characteristics, they expect to earn more than they would in the formal sector. In contrast, the lower-tier employees are those disadvantaged workers who see informality as a last resort.

Nevertheless, from the empirical stance this more recent view of dualism within the informal sector has not been satisfactory dealt with. For example, Magnac (1991), when testing for competitiveness or segmentation in the labor market of Colombia in the 1980s, found evidence of a competitive labor market structure. Similarly, Gindling (1991) and Pratap and Quintin (2006) found evidence of segmentation in Costa Rica and of a competitive structure in Argentina, respectively. However, in all the above studies the authors assume homogeneity of the informal sector, thus limiting their analysis.

Among the few studies that have tried to model the heterogeneous structure of the informal sector, we can list Cunningham and Maloney (2001), and Günther and Launov (2012). The former models the informal sector as a mixture of “upper-tier” and “lower-tier” enterprises, and using econometric techniques of factor and cluster analysis, allows for the segmentation of the market. However, despite finding evidence of segmentation, Cunningham and Maloney (2001) only considered informal firms, so that the alternative of being a formal firm does not exist in their model. Further, the authors do not take into account the selection bias induced by decision that individuals make about the type of employment.

The study by Günther and Launov (2012) analyzes the possible heterogeneous structure of the informal sector, estimating a finite mixture model which makes it possible to determine the number and size of segments that the informal sector might be composed of. This model uses minimal *a priori* assumptions to determine the segments and provides a new method for identifying the size of voluntary and/or involuntary employment in the informal sector. The empirical analysis uses data from the Ivory Coast at the end of the 1990s. Among their findings, the authors report that the informal sector consists of two segments: a highly-paid and a low-paid segment. They also found that 45% of informal employment is non-voluntary and is mainly located in the lower-paid informal segment, while the remaining 55% of informal employment is voluntary and is situated in the higher-paid informal segment.

In this chapter we analyze the heterogeneity of the informal sector, looking at a decomposition of the wage differential between the formal and informal sector throughout the entire distribution of wages. This methodology is conceptually similar to Günther and Launov's (2012) approach, except for the fact that it accounts for a wider variety of informal employees as well as formal ones. Our method advances beyond the studies based on the workers' mean-earnings, which are incapable of distinguishing if there are different types of behavior throughout the entire distribution of wages.

Our research focuses on the regional labor markets of Colombia. Given the geographic, demographic and social conditions, and economic dynamics, Colombia provides rich evidence from a large, heterogeneous informal sector. Furthermore, there are marked differences in the structures and dynamics of the local labor markets. In Colombia roughly six out of ten employees work in the informal sector[1], and cities such as Cúcuta or Montería have informality rates of around 75%. Others such as Medellín or Bogotá have rates of about 50% (García, 2011; Galvis, 2012).

In order to analyze the different motivations for joining the informal sector we have decomposed the formal/informal wage gap. Doing so allow us to distinguish what proportion of the wage gap is due to differences in prices related to individual characteristics and what proportion is due to characteristics which differ between the formal and informal sector. If the wage gap is mainly attributable to the former factor, it indicates that individuals in the informal sector earn less because they get lower returns

[1] According to the International Labor Organization (ILO, 2011) estimates, Colombia is the country with the fourth highest informality rate in South America after Paraguay (70.4%), Peru (70.3%) and Bolivia (69.5%).

for their skills and therefore they are part of the disadvantaged sector of a segmented market. On the other hand, if the wage gap is primarily explained by the latter factor, the labor segmentation is not as strong as in the above case and the differences in wages between sectors are due to differences in endowments. In this latter situation, being an informal worker is a choice, since these individuals can obtain non-wage benefits or earn more than they would earn in the formal sector.

To carry out the decomposition, we estimate the earnings functions for informal and formal workers using quantile regression, taking into account the possibility of self-selection into those sectors. We follow the method of Machado and Mata (2005) and the extension proposed by Albrecht, Vuuren and Vroman (2009) to account for selection, which is based on Buchinsky (1998), who uses semi-parametric methods.

Following this introduction, Section 3.2 proceeds with the description of the data. In Section 3.3 we discuss the estimation procedure. Section 3.4 describes the empirical findings, and finally conclusions are drawn in Section 3.5.

3.2 Data and descriptive evidence

The data used in this chapter come from the Great Integrated Household Survey (GIHS) for 2009, carried out by the National Administrative Statistics Department (DANE). This cross-section survey has information at micro-data level on labor force, unemployment and informality of thirteen major Colombian cities and their metropolitan areas.[2]

The sample considered in this work is composed of individuals between 12 and 65 years old, with agriculture workers also being excluded. Our final sample is composed of 62,278 individuals.[3] The main variable for the analysis is the real hourly wage, calculated as the monthly wage divided by the effective number of hours worked during that month and adjusted for the price level using the consumer price index (base year 2008) for each city as a deflator.[4]

[2] Namely, Barranquilla, Bogotá, Bucaramanga, Cali, Cartagena, Cúcuta, Ibagué, Manizales, Medellín, Montería, Pasto, Pereira, and Villavicencio. These metropolitan areas represent 45% of total population and about 60% of urban population according to 2005 Population Census.

[3] Note that we excluded government employees, employers and self-employed. Given this exclusion, the informality rate may differ from that reported by ILO.

[4] Consumer price indexes for the biggest cities in Colombia were obtained from DANE. Since each of these cities is the core of a metropolitan area, we applied the consumer prices index of the city to the whole metropolitan area. In Ibagué the consumer prices index is not calculated by DANE, so we decided to use the consumer prices index of Pereira, given the similarities in population and the social and cultural characteristics, as well as the proximity between these cities.

With regard to informality, we define informal workers as those workers who are not covered by the social security system. More precisely, informal workers are those workers who are not covered by the health insurance scheme and the pension system. Applying this condition, we have 36,293 (58.3%) formal workers and 25,985 (41.7%) informal workers. In Table 3.1 we provide some descriptive statistics for the key variables for formal and informal workers.

Table 3.1. Descriptive statistics

	Formal workers	**Informal workers**	**Total**
Real hourly wage	3269.2	2311.9	2927.4
Age (years)	34.3	32.9	33.8
Education (years)	11.0	8.6	10.2
Tenure at job (years)	4.7	2.8	4.0
Education levels			
Less than primary	0.3	1.6	0.8
Primary	24.4	49.7	33.4
Secondary	37.6	32.7	35.9
Tertiary	37.7	16.0	29.9
Male	55.6	48.7	53.1
Head of household	43.0	35.4	40.3
Married	53.5	46.1	50.8
Firm size			
1 – 10 employees	17.9	76.6	38.9
11 – 50 employees	22.3	14.1	19.3
More than 51 employees	59.8	9.3	41.8
Sample size	36,293	25,985	62,278

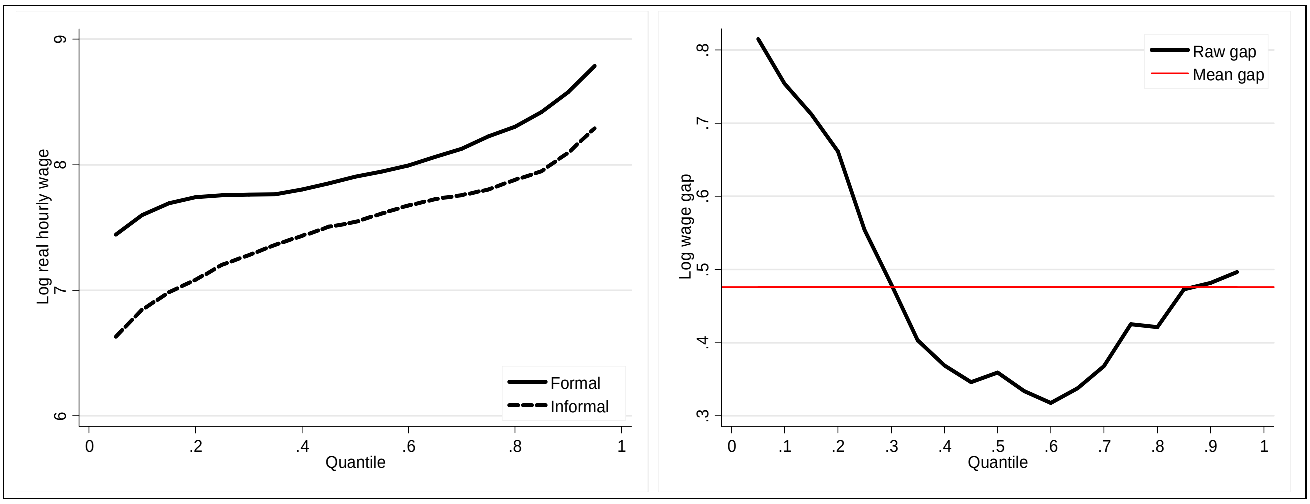
Note: We used person sampling weight available in the database. The wages are in Colombian pesos (in December 2009 the exchange rate was 2935 Colombian pesos per euro).

As can be seen from Table 3.1, the average wage among formal workers is higher than the corresponding average among informal workers: a formal worker earns on average 30% more than an informal worker. In terms of the variables that we can use to explain variation in wages, there are also some important differences between kinds of employees. Formal workers have on average a similar age to informal workers, and the years of tenure in the job are higher for formal workers than for informal workers. Turning to education, we can see that formal workers have consistently received more education than informal ones. The informal sector has a higher percentage of individuals with primary and pre-primary education (51%), while the formal sector has a much higher percentage of individuals with secondary and tertiary education (75.3%). As regards other personal characteristics, we can see that the informal workers are less likely to be men, head of the household and married than formal workers. Last, informal workers are more likely to work in firms having between 1 and 10 employees (77%), while formal workers are employed in firms with more than 51 employees (60%).

Figure 3.1 depicts the estimated kernel densities of the wages of formal and informal workers. Wage disparities between sectors are clearly visible, as wage distribution for formal workers is shifted to the right. The distribution of formal and informal sector wage and the wage gap between sectors by quantile, i.e., the difference in log wages between formal workers and informal workers at each quantile of their respective distributions, is plotted in Figure 3.2. We can see that the wage differential between sectors is positive throughout the whole wage distribution, with a large wage gap at the lower quantiles. Its size ranges between 54% at the bottom end of the distribution to 30% at the median, then increases to roughly 39% at the top end of the distribution. There are marked differences between formal and informal workers at the extremes of the distribution, which may be due to very different human capital endowments and job opportunities at these points of the earnings distribution.

Figure 3.1. Kernel density of log real hourly wage by formal and informal sector

Figure 3.2. Wage differentials between formal and informal sector over different quantiles of the wage distribution



At a city level we can see that there are also positive wage differences between sectors throughout the whole distribution, and there are different patterns between cities (see Figure A3.1 in the Appendix). Pasto, Montería and Cartagena present the largest wage gaps, with a particularly large wage gap at the lower quantiles. The common characteristic in these cities is that they present the highest levels of informality in Colombia (see Table A3.1 in the Appendix), and therefore there is an important heterogeneity of jobs and workers in the informal sector. In these cities the relative abundance of informal jobs is an important determining factor for joining the informal sector. Turning to the biggest and most developed cities, such as Bogotá, Medellín, and Cali, we can see that the wage differentials between sectors are smaller than in the former cities.

In order to simplify the presentation of the results of the empirical exercise, we define three groups of cities. In the following section we describe these groups and present some descriptive statistics of their labor markets.

3.2.1 Group of cities and their labor markets

We have divided the total sample into three sub-groups of cities corresponding to a group of central and more developed cities, and another two groups of peripheral cities which present a significant informal sector.[5]

[5] Barón (2002) and Galvis (2007) carried out studies on the economic regions in Colombia and identified these three groups of regions. The former author performed a cluster analysis on the bank deposits at department level as measure of economic performance of regions (according to Bonet and Meisel (1999) there is a correlation between GDP and bank deposits of around 0.8). While the latter author analyses the intensity of economic activity, which is measured through the bank deposits and the tax collection per capita at city level, and using spatial econometrics techniques evaluate the existence of spatial dependence in the sub-regions conformed by cities where economic activity is more concentrated.

The first group of cities (Group 1) includes Bogotá, Medellín, Cali, Bucaramanga, Manizales, Pereira and Ibagué. This group is composed of the largest industrial and most dynamic cities in Colombia, and they form the core of the country's economic activity. These cities represent 0.7% of the national territory, and according to the 2005 Population Census around 45% of the urban population is concentrated in them. In terms of economic activity the region made up of Bogotá, Cali, Medellín and Bucaramanga accounts for 70% of Colombian GDP at a department level.[6][7] Figure 3.3 shows the spatial distribution of the real GDP per capita at a department level in 2009.[8] Overall, it can be seen that, excluding the mining departments (Arauca, Casanare and Meta have the largest oil fields in the country and they accounted for 6% of Colombian GDP in 2009), the highest levels of GDP per capita are in the central region. It is also worth highlighting the fact that the ranking occupied by these cities in terms of their degree of informality has been relatively stable over time. In this respect, García (2008 and 2011) and Galvis (2012) have found that, from a regional perspective, these cities show consistently lower informality levels comparing to those cities outside this region (see Figure 3.4).

The second group of cities (Group 2) is made up of Barranquilla and Cartagena. Although these cities are among the most urbanized cities, and present an important economic dynamic (see Figure 3.3), their tourist and export vocation make them different from other cities. The country's main ports are locates in these cities and they have an important industrial cluster associated with the petrochemical-plastic sector.[9] Nevertheless, their socioeconomic and labor market indicators are unfavorable. These cities show one of the highest poverty, inequality and informality levels among the main cities of Colombia (Bonilla, 2008; Galvis, 2009). As can be seen from Figures 3.4 and 3.5, Cartagena and Barranquilla along with Montería and Cúcuta, present the highest levels of Unsatisfied Basic Needs (UBN), as well as of informality. The emphasis on

---

[6] Colombia is made up of thirty-two departments and Bogotá, the Capital District. The departments are country subdivisions similar to US states, are granted a certain degree of autonomy and each has its own capital city.

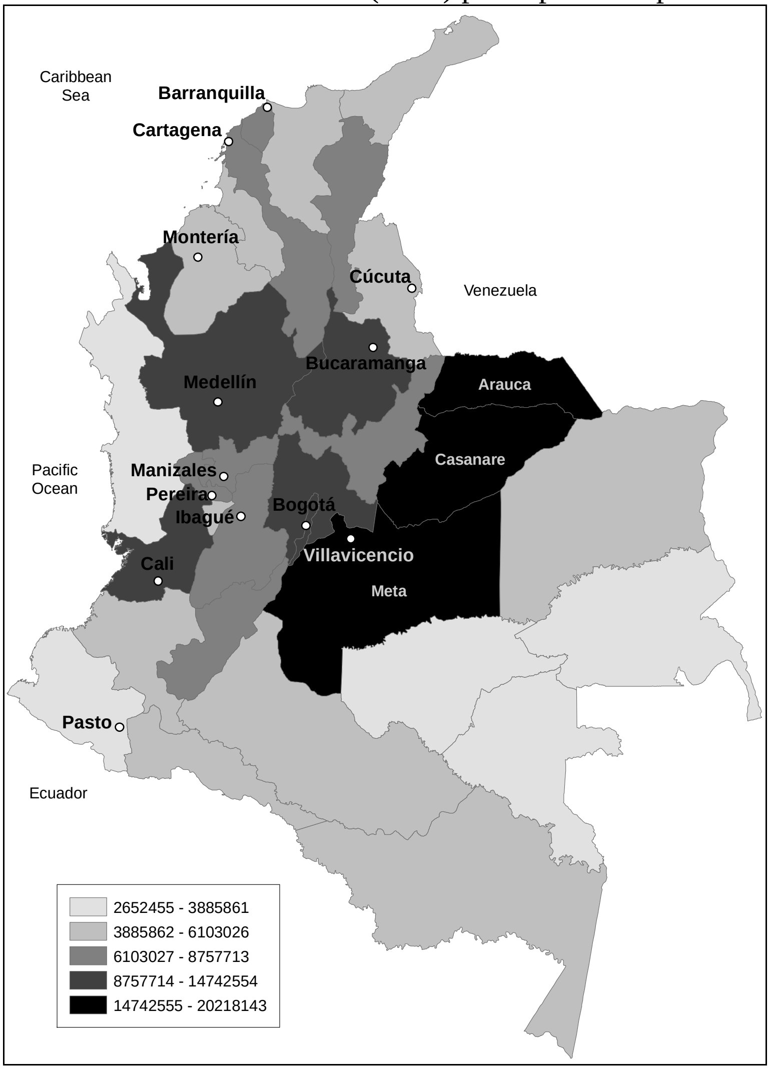
[7] Galvis (2007) reports that the region formed by these cities account for 80% of total economic activity of the country.

[8] A more relevant variable would be GDP per capita at city level, but in Colombia this data is not available.

[9] The industrial zone of *Mamonal* in Cartagena contains the second largest oil refinery in Colombia, which is integrated with the petrochemical, chemical and plastic industries. Barranquilla is highly specialized in the food and beverages, chemicals, non-metallic mineral products and basic metallurgy sectors. A more detailed economic characterization of Barranquilla and Cartagena can be found in Bonilla (2010) and Acosta (2012), respectively.

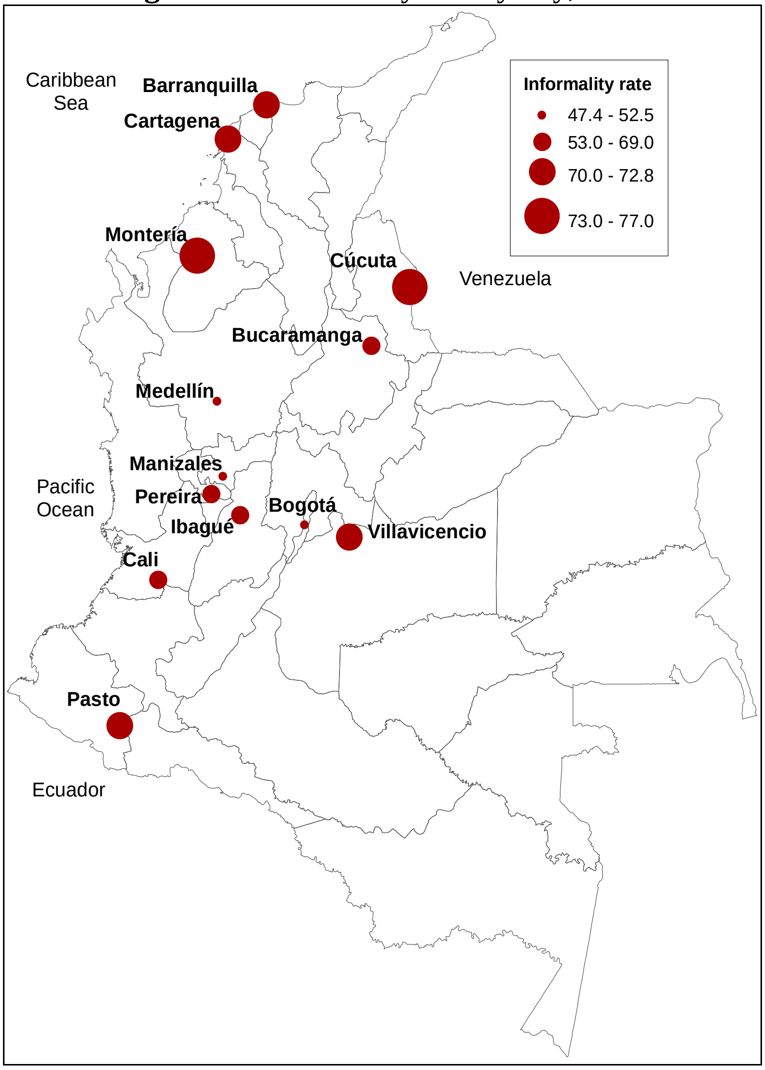
tourism in the Caribbean region and the relatively low capacity for creating jobs in the highly productive sectors, due to the fact that these are mostly made up big companies with high capital intensity and export activities, have led to a process of tertiarization of the economy. The service sector has little impact on the competitiveness of the other sectors and generates a lot of jobs but of low quality in terms of pay and working conditions (Bonet, 2005 and 2007; Bonilla, 2010; Cepeda, 2011; Acosta, 2012).

Figure 3.3. Real Gross Domestic Product (GDP) per capita at departmental level, 2009



Source: DANE - Colombian currency, constant 2005 prices.

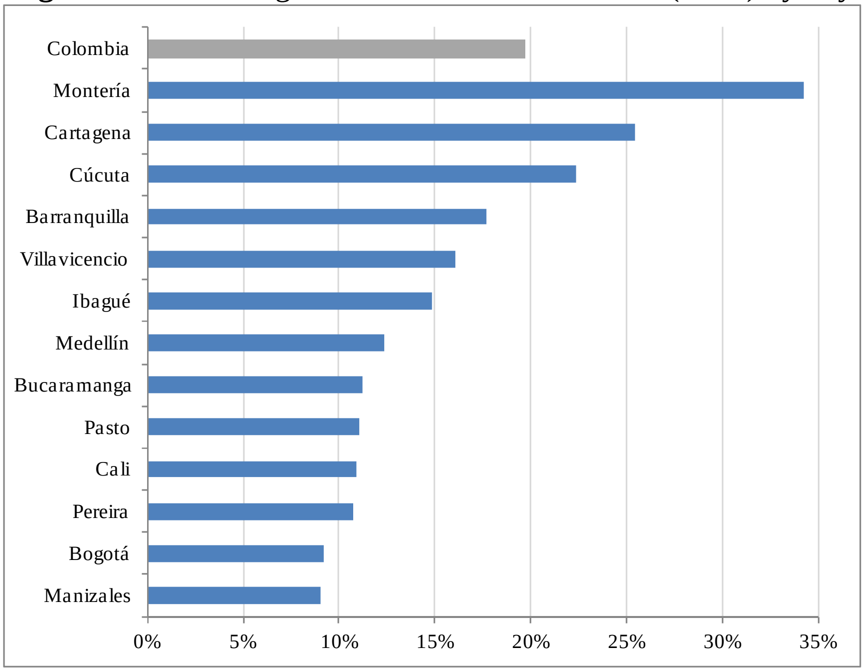
Figure 3.4. Informality rate by city, 2009



Source: Table A3.1 in the Appendix.

The third group of cities is made up Cúcuta, Montería, Pasto and Villavicencio (Group 3). These are the least developed cities, located in peripheral areas, and their activities are very much influenced by agriculture, mining and commerce (see Figure 3.3). Pasto and Cúcuta are border cities, the former sharing a border with Ecuador and the latter with Venezuela, and this is a common characteristic that can influence the type of activity and the employment generated, above all those jobs related to commerce (legal and illegal) and currency exchange (Bonet, 2007; García, 2005 and 2011). Villavicencio is the capital of the department of Meta, which currently has the largest oil-fields in the country (the department of Meta produces 47% of Colombia's oil (DANE, 2011)), and along with Montería these two are the capitals of the two main cattle farming regions of the country and therefore their economies are based mainly on these activities. Furthermore, these two regions are considered conflict zones due to the presence of paramilitary groups, guerrillas and drug trafficking activities, and this influences not only the activity economic but also the social, political and cultural make-up of the regions (Vilore de la Hoz, 2009; Sánchez *et al.*, 2012). With regard to informality, in contrast to first group of cities, this group shows the highest informality levels, with Cúcuta being the city with the highest rate (77%) (see Figure 3.4). According to García (2008 and 2011) and Galvis (2012) informality is more prevalent in less prosperous cities, which are usually located in the periphery of the country, with less resources and industrial development than cities in the center of the country.

Figure 3.5. Percentage of Unsatisfied Basic Needs (UBN) by city



Source: 2005 Population Census – DANE

Table 3.2 shows some descriptive statistics of the labor markets formed by the three groups of cities. As expected, there is a higher percentage of informal wage workers in City Groups 2 and 3 (47 and 56%, respectively) than in Group 1 (35%). We can also see that the formal workers earn more than the informal workers, and the differences are more marked in City Group 2. While the wage differences between sectors in City Group 1 is 26%, in Groups 2 and 3 the wage differences are 37 and 34%, respectively.

With regard to education, we can see that on average the difference between sectors is higher in City Group 2 than in the other two groups of cities. In terms of education levels, we can see that the group of less developed cities (Group 3) has a higher percentage of informal workers with primary and lower than primary education (56%) than the group of more developed cities (53% in Group 1 and 43% in Group 2), while these two latter groups have a higher percentage of informal workers with tertiary education (16 and 21%, respectively) than the former group (10%). There are also striking differences in terms of education levels in the formal sector between groups of cities. Interestingly, around half of the formal workers in the group of Caribbean coast cities (Group 2) have tertiary education, while in the group of more developed cities (Group 1) this percentage barely reaches 37%. The reason for these results may have to do with the higher degree of industrial specialization found in Barranquilla and Cartagena. According to Acosta (2012), these cities are among the most specialized cities in Colombia, and the industrial sectors devoted to chemicals, petrochemicals, rubber and plastic are leading such specialization. These industries are technically complex and therefore require highly skilled labor. In this regards, Arango (2011), who studied the differences in the main variables of the labor markets of the major cities of Colombia in the period from 2001 to 2011, found that Barranquilla and Cartagena (along with Bogotá) are indeed cities characterized by having the highest worker education rates in Colombia.

Table 3.2. Descriptive statistics by groups of cities

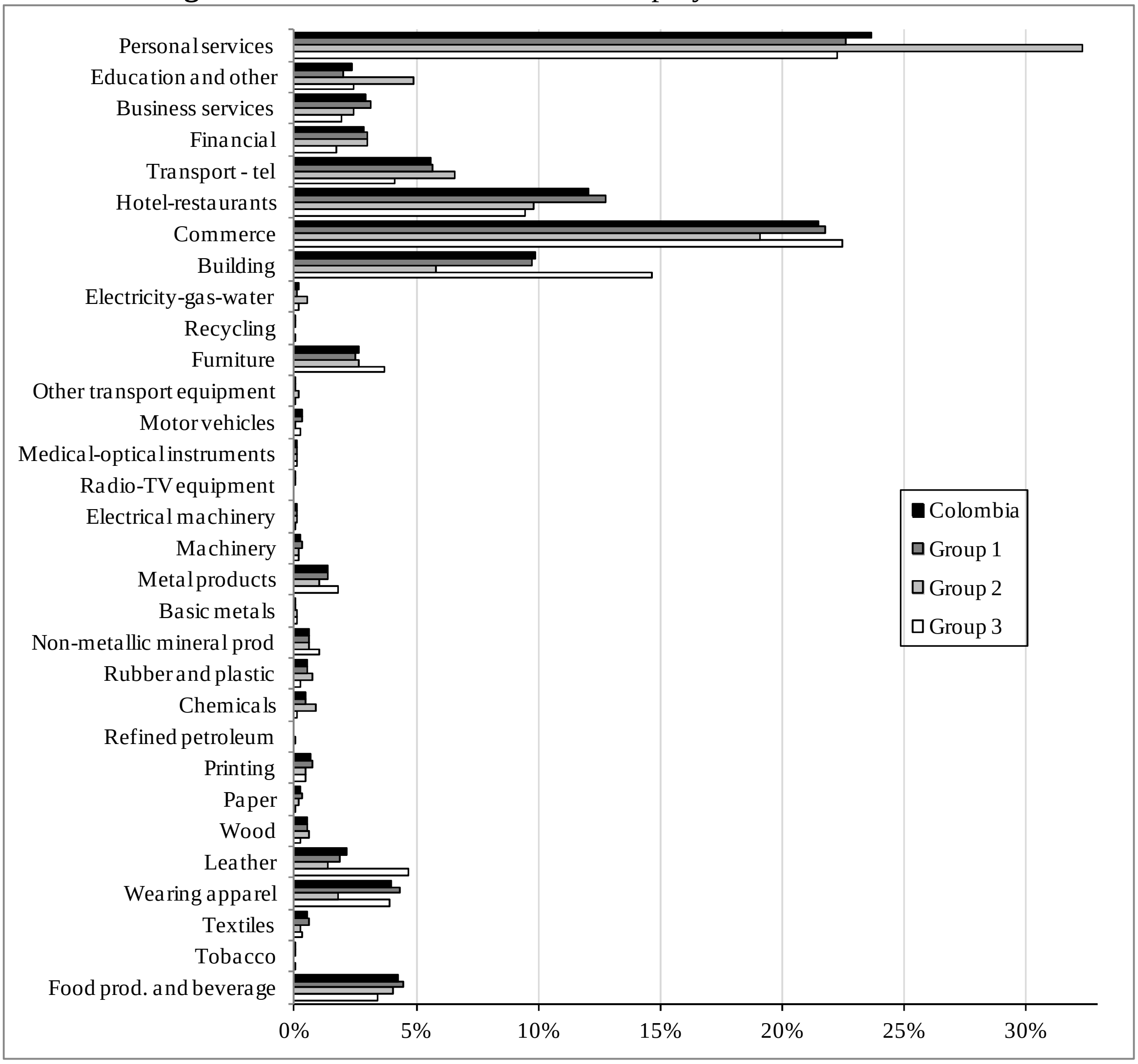
	Group 1			**Group 2**			**Group 3**		
	Formal workers	**Informal workers**	**Total**	**Formal workers**	**Informal workers**	**Total**	**Formal workers**	**Informal workers**	**Total**
Real hourly wage	3273.1	2408.5	2989.8	3240.8	2031.1	2688.0	2941.5	1946.7	2492.6
Age (years)	34.2	33.0	33.8	35.4	33.9	34.7	34.1	31.7	32.7
Education (years)	10.9	8.6	10.2	11.9	9.2	10.7	10.1	8.1	9.4
Tenure at job (years)	1.9	1.4	1.8	2.2	1.7	2.0	1.9	1.4	1.6
Education levels									
Less than primary	0.3	1.4	0.7	0.2	1.9	1.0	0.6	2.75	1.8
Primary	25.4	50.4	33.6	14.7	42.0	27.1	19.9	53.2	39.1
Secondary	37.1	32.1	35.4	39.5	35.3	37.6	45.4	33.8	38.7
Tertiary	37.2	16.1	30.3	45.6	20.8	34.3	34.1	10.3	20.4
Male	55.1	48.3	52.9	62.0	46.2	54.8	53.4	54.2	53.8
Head of household	43.0	35.8	40.7	43.6	31.0	37.9	42.4	36.5	39.0
Married	52.3	45.7	50.1	64.8	49.9	58.0	56.4	45.6	50.1
Firm size									
1 – 10 employees	18.5	77.3	37.7	9.0	67.3	35.7	20.8	81.1	55.7
11 – 50 employees	22.3	13.6	19.4	23.5	17.2	20.6	22.1	13.7	17.2
More than 51 employees	59.3	9.1	42.9	67.5	15.5	43.7	57.1	5.2	27.1
Sample size	25,368	13,723	39,091	4394	3832	8226	6531	8430	14,961

Note: We used person sampling weight available in the database. The wages are in Colombian pesos (in December 2009 the exchange rate was 2935 Colombian pesos per euro).

Another difference between the groups of cities can be found in the firm size variable. As can been seen from Table 3.2, in the City group 2 there is a substantially higher proportion of informal workers carrying out their activities in medium and large firms (around 33%) than in City Group 1 and 3 (23% and 19%, respectively). This difference is due to the presence of big firms associated with the industrial cluster of petrochemical products and export activities in the Caribbean coast cities. These activities present important productive linkages which not only benefit the formal sector but also the informal sector (DNP, 2007; Acosta, 2012).

Figure 3.6 shows the distribution of informal sector employment across 2-digit industries by group of cities. Most of the informal sector employment is in the service sector (around 80%), with personal services and commerce being the sectors where the greater share of informal employment is concentrated. Of note is the case of the group of Caribbean coast cities, where more than a third of informal workers are employed in the personal services sector. This result reflects the marked influence of tourism-related activities on the economy of this region. Within the industrial sector overall, it can been seen that informal employment is concentrated in food and beverages and wearing apparel, followed by furniture, leather and metal products. In this sector City Group 3 has a relatively higher proportion of informal employment in the leather and wearing apparel sectors, which can reveal the impact of border and cattle farming activities on the productive structure of these cities. As noted, the sectoral composition of production in the cities is an important aspect to take into consideration in explaining informality at a regional level.

Figure 3.6. Distribution of informal employment across sectors

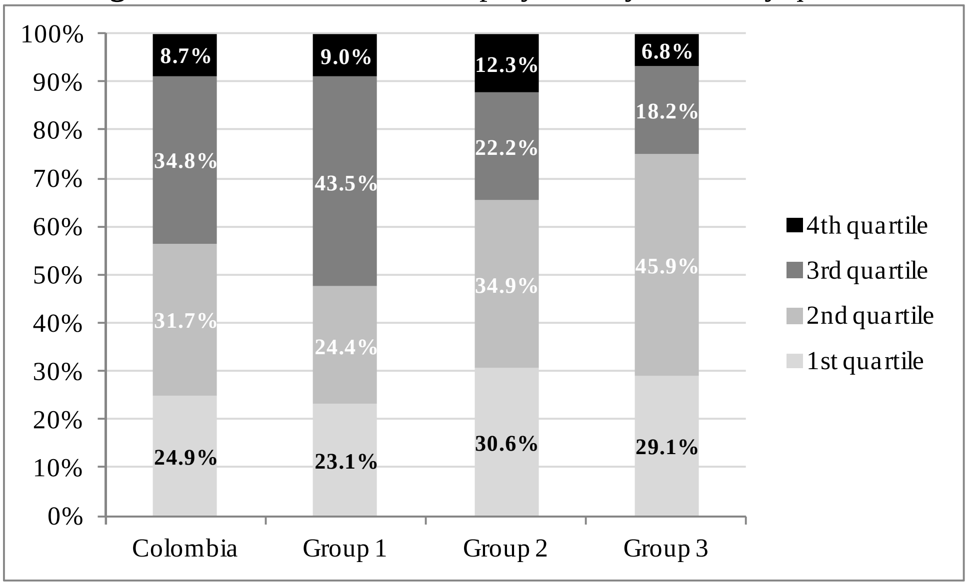


In order to measure the degree of modernity of the informal sector we calculated an index based on the type of employment generated by each economic sector, i.e. whether it is skilled or unskilled. According to Ranis and Stewart (1999) the modern informal segment is characterized by capital-intensive activities, dynamic in technology and often having skilled workers. In this sense, our index of modernity is defined as the ratio of the number of workers in skilled occupations to total employment.[10] This measurement embracers several dimensions of modernity: more modern industries have, on average, more skilled workers, are more productive (measure as wage per worker) and have a higher participation of large firms than traditional industries (see Table A3.2 in the Appendix). We calculated this index for each 2-digit industry and city.

[10] In skilled occupations we include professionals, managers and white-collar workers.

In Figure 3.7 and Table 3.3 we show the distribution of informal employment and its characteristics across modernity quartiles for the total sample and by group of cities. As shown in the figure, less than 9% of informal employment in Colombia is in sectors in the top quartile of the modernity index distribution, i.e. those where the majority of employment is highly qualified. In fact, more than half of informal employment (57%) remains in the most traditional activities (1st and 2nd quartile). Regarding the distribution of characteristics, the results show that, overall, in this modern informal segment (3rd and 4th quartile) workers are more qualified and present higher productivity levels than workers in the traditional informal segment (1st and 2nd quartile).

Figure 3.7. Informal sector employment by modernity quartile



By group of cities we can see that the more developed cities (Groups 1 and 2) have a higher degree of modernity of the informal sector than the less developed cities (Group 3). By contrast, as expected, in the less developed cities most of the informal employment is in less modern and more traditional activities (75%). The distribution of characteristics across modernity quartiles shows that informal workers in the highest quartiles in the City Group 3 are less educated and their monthly wages are around 20% lower than their informal workers counterpart in City Groups 1 and 2.

Table 3.3. Informal sector characteristics by modernity quartile

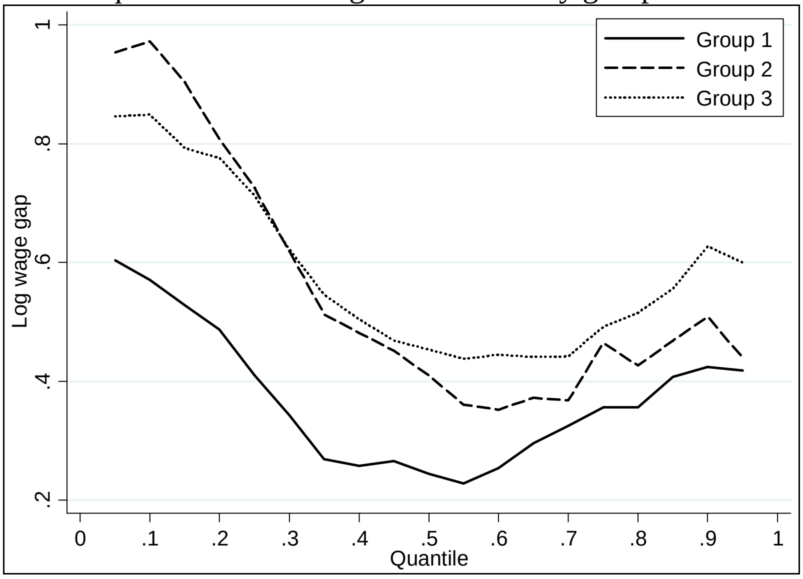
	Total sample	Group 1	Group 2	Group 3
Years of education				
Modernity quartile 1	8.4	8.4	9.3	6.5
Modernity quartile 2	8.4	8.8	9.6	6.7
Modernity quartile 3	9.4	9.1	10.1	9.1
Modernity quartile 4	11.5	11.6	11.5	9.2
Monthly wages per employee				
Modernity quartile 1	502,922	491,441	473,751	444,505
Modernity quartile 2	506,175	501,037	483,216	454,696
Modernity quartile 3	514,603	502,522	541,895	453,206
Modernity quartile 4	519,314	512,385	542,757	472,112

Note: All values are averages across informal workers within each sector and quartile, calculated using survey weights. Monthly wages are measured in Colombian currency.

Interestingly, from Figure 3.7 we can see that in City Group 2 there are a relatively high proportion of informal workers in sectors in the top quartile of the modernity index distribution (12%). This degree of modernity of the informal sector contrasts with the high participation of the more traditional informal segment: around 31% of total informal employment is in the 1st quartile. These results show the strong influence of tourism-based activities on job creation, above all of very low-skill service jobs, on the one hand, and the effect that export activities and the concentration of industries with important productive linkages have, on the other. As pointed out by Ranis and Stewart (1999) higher intermediate linkages (e.g. via subcontracting) between the formal and informal sector in the most productive and moderns sectors can lead to the expansion of the modern informal segment.

Turning now to the wage gap between sectors, in Figure 3.8 we present this at each quantile of their respective distributions by groups of cities. From the figure we can see that the wage differentials between the formal and informal sector are considerably lower in City Group 1, above all at the bottom end of the distribution. Interestingly, it can also be seen that the wage differential at the bottom of the distribution is higher in City Group 2 than in Group 3. Again this result can be due to the strong influence of personal services activities associated with tourism in the Caribbean coast cities, the majority of which are informal activities with very low qualifications and wages.

Figure 3.8. Wage differentials between the formal and informal sector over different quantiles of the wage distribution by group of cities



3.3 Estimation procedure

In order to determine which factors influence the wage gap between the formal and informal sector, taking into account the heterogeneity of workers throughout the distribution, as well as the differences that can exist between groups of cities, we made use of the quantile decomposition methodology. Quantile regression methods are particularly useful for analyzing the decomposition of the wages gap at different points of the distribution in situations where disparities are large, as is the case of a country like Colombia (Bonilla, 2008 and 2009). Furthermore, this methodology makes it possible takes into account the wage heterogeneity between group of individuals and the different impact of the determinants of wages and their gaps by type of employment at different points of the distribution (Machado and Mata, 2005). Thus, the results are more complete than those obtained by OLS.

The decomposition methods have been extensively used to analyze the gender and union wage gap, and temporal change in wages.[11] In recent years this approach has also been used to study the wage differences by race (Bucheli and Porzecanski, 2011), ethnicity (Atal *et al.*, 2009), native/immigrant (Simón *et* al., 2008; Nicodemo and Ramos, 2012) and types of workers such as private/public (Lucifora and Meurs, 2006; Bargain and Melly, 2008), full/part-time (Hardoy and Schone, 2006; Wahlberg, 2008), permanent/temporary (Bosio, 2009; Comi and Grasseni, 2009) and formal/informal (Bargain and Kwenda, 2010; Arabsheibani and Staneva, 2012).

[11] A more detail review of the literature on this methodology can be found in Fortin *et al*. (2011).

We now present a brief description of the estimation procedure of the Machado and Mata decomposition with sample selection adjustment. We follow the adaptation of the Machado-Mata procedure introduced by Albrecht *et al*. (2009) based on Buchinsky (1998), which is a non-parametric method for accounting for selection for quantile regression.

In our analysis, the potential selection bias in the estimation of wage equations may result from a self-selection of individuals into different employment types: formal or informal. There are several observable and unobservable factors which may affect whether a worker is part of the formal or informal sector. In order to correct this selection bias, we could, as a first step, follow Heckman (1979) and estimate a probit model to calculate the probabilities of workers being in the formal and informal sector. However, the methodology proposed by Buchinsky (1998) does not impose the restriction of normality and instead uses a semi-parametric method developed by Ichimura (1993), which makes no assumptions about the distribution of the residuals.

Following Buchinsky (1998), we thus let I_i be the variable that indicates the sector in which worker *i* is employed and takes the values 1 for the informal and 0 for the formal. For this binary model we have the following equation for the latent or index variable:

$$I_i^* = z_i^{'}\gamma + v_i, \tag{1}$$

where z_i is a set of observable characteristics that influence the probability that a worker *i* is employed in the informal sector; and γ is a vector of coefficients to estimate. The employment sector is determined by:

$$I_i = \begin{cases} 1 \text{ if } I_i^* > 0 \\ 0 \text{ if } I_i^* \leq 0 \end{cases} \tag{2}$$

Now, let X_{inf} and X_{for} be the stochastic vectors of characteristics for informal (*inf*) and formal (*for*) workers which have distribution functions $G_{X_{inf}}$ and $G_{X_{for}}$, respectively. The realizations of these stochastic vectors are given by x_{inf} and x_{for}. The endogenous variable that represents the log wage is Y_{inf} for the group of informal workers and Y_{for} for the group of formal workers and they have unconditional distribution functions $F_{Y_{inf}}$ and $F_{Y_{for}}$, respectively. The quantile regression can be written for each sector as:

$$Q_\theta(Y_{for} \mid X_{for} = x_{for}) = x'_{for}\beta^{for}(\theta) \tag{3}$$

and

$$Q_\theta(Y_{inf} \mid X_{inf} = x_{inf}) = x'_{inf}\beta^{inf}(\theta), \tag{4}$$

where $Q_\theta(Y \mid X = x)$ is the conditional quantile at θ^{th} quantile. The Machado-Mata procedure consists of generating a random sample of size n from a uniform distribution $U[0,1]: u_1, u_2, ..., u_n$, and calculating the conditional quantile regression for each group which yields n estimates of the quantile regression coefficients $\hat{\beta}^{inf}(u_n)$ and $\hat{\beta}^{for}(u_n)$. We then use the estimated result and a random sample of size n of the vectors of covariates x to predict simulated values of both $\hat{y}_{for} = \tilde{x}'_{for}\hat{\beta}^{for}(u)$ and the counterfactual wage distribution $\hat{y}_{inf} = \tilde{x}'_{inf}\hat{\beta}^{for}(u)$, i.e. the wage distribution of the informal sector resulting from assigning the returns of the formal sector but keeping the observed characteristics of the informal sector unaltered. These steps are repeated m times. Finally, the difference between the log wages of formal workers and the log wage given in the counterfactual distribution at the θ^{th} quantile can be decomposed as:

$$\begin{aligned} Q_\theta(Y_{for} \mid X_{for} = \tilde{x}_{for}) - Q_\theta(Y_{inf} \mid X_{inf} = \tilde{x}_{inf}) &= \underbrace{\left[Q_\theta(\tilde{x}'_{for}\hat{\beta}^{for}(u)) - Q_\theta(\tilde{x}'_{inf}\hat{\beta}^{for}(u)) \right]}_{characteristics\ effects} \\ &+ \underbrace{\left[Q_\theta(\tilde{x}'_{inf}\hat{\beta}^{for}(u)) - Q_\theta(\tilde{x}'_{inf}\hat{\beta}^{inf}(u)) \right]}_{coefficient\ effects} \end{aligned} \tag{5}$$

The first term on the right hand side of expression (5) refers to the characteristics effects. This term shows the contribution of the differences in the distribution of endowments between formal and informal workers to the wage gap at the θ^{th} quantile. The second term calculates the counterfactual value of the wage gap if the informal workers retained their observed characteristics but were paid for them like the formal workers. This term represent the coefficient effects. We use a bootstrap procedure to estimate standard errors for the reported components of the decomposition.

Since we only observe the wages of those workers who actually work in the informal or formal sector, these workers are not draw randomly from the distribution of individuals and therefore there may be a selection bias when we estimate the wage equations. Consequently, in order to correct for selection and to get unbiased estimates of β in the quantile wage equations, Buchinsky (1998) proposes to introduce an extra term in the quantile regressions, namely,

$$Q_\theta(Y_{for} \mid Z = z) = x'_{for}\beta^{for}(\theta) + h_\theta(z'\gamma) \tag{6}$$

and

$$Q_\theta(Y_{inf} \mid Z = z) = x'_{inf}\beta^{inf}(\theta) + h_\theta(z'\gamma)\,. \tag{7}$$

The vector Z includes also the set of observable characteristics that influence wages (i.e. the Xs), but for identification Z must contain at least one variable that is not included in X and should be uncorrelated with the log wage. The term $h_\theta(z'\gamma)$ plays the same role as Mill's ratio in the usual Heckman (1979) procedure, but it is quantile-specific and more general so as not to assume normality. Buchinsky (1998) suggests the following power series approximation to the term $h_\theta(z'\gamma)$

$$\hat{h}_\theta(z'\hat{\gamma}) = \sum_{k=1}^{K} (\lambda(\hat{\mu} + \hat{\sigma} z'\hat{\gamma}))^{k-1}\hat{\delta}_k(\theta)\,, \tag{8}$$

where $\lambda(\cdot)$ represents the usual inverse Mill's ratio, and $\hat{\mu}$ and $\hat{\sigma}$ are scaling parameters which are estimates of the constant and slope coefficients from the probit regression of I_i on the index $z'\hat{\gamma}$.

In order to estimate the coefficients γ in equation (1), Buchinsky (1998) proposes to use the semi-parametric least-squares (SLS) method proposed by Ichimura (1993). Since we estimate a semi-parametric sample selection model, the intercept in the wage equation is not identified. When k=1 in equation (8), $\delta_1(\theta)$ is equal to one and therefore it cannot be separately identified from the constant term in $\beta(\theta)$. To identify the constant term in the wage equation, we first remove the k=1 term from the power series expansion and estimate the resulting quantile model; and then we estimate the constant term in the wage equation without adjusting for selection by using a subsample of observations so that the probability of informal sector participation is close to one.

In summary, the extension of the Machado-Mata algorithm to adjust for selection proposed by Albrecht *et al*. (2009) is the following:

1. Estimate γ using a semi-parametric least-squares (SLS) method (Ichimura, 1993).
2. Sample u from a standard uniform distribution.
3. Compute $\hat{\beta}^{inf}(u)$ and $\hat{\beta}^{for}(u)$ using the Buchinsky technique.
4. Sample x_{inf} and x_{for} from the empirical distribution $\hat{G}_{X_{inf}}$ and $\hat{G}_{X_{for}}$, respectively.

5. Compute $\hat{y}_{for} = \tilde{x}'_{for}\hat{\beta}^{for}(u)$ and $\hat{y}_{inf} = \tilde{x}'_{inf}\hat{\beta}^{for}(u)$.
6. Repeat steps 2 – 5 *m* times. [12]
7. Compare the simulated distributions to decompose the estimated wage gap between sectors.

3.4 Results

In this section we present the results of the quantile decomposition formal/informal wage gap. The conditional quantile regression approach proposed by Machado and Mata (2005) makes it possible to decompose the difference between the formal and informal workers log wage distributions and identifying how much of the wage gap estimated at different quantiles of the wage distribution can be attributed to differences in characteristics and how much can be attributed to differences in returns to those characteristics.

3.4.1 SLS estimation and the quantile regression models

As mentioned in Section 3.3, in the first step we estimated the semi-parametric least squares (SLS) model for the probability of being informal, and in the second step we estimate the quantile regression models for the wage equation including the power series expansion to deal with selection. In both the probability and the quantile regression models we included variables for education levels, gender, and dummies for size of firm, industry and occupation. In order to identify the probability models we included variables for the presence of children between 0 and 12 years old at home, the presence of other relatives working as formal workers, the average number of years of education of members of the household as a measurement of the educational environment of the household, whether the individual is head of the household and the marital status. Table 3.4 shows the results for the probit and SLS probability models for the total sample and by group of cities.

In order to test whether in effect the probability of being informal relies on the normality assumption for the residuals, we performed a Hausman test. As pointed out by Buchinsky (1998), the SLS estimate is consistent and independent of the distribution of the residuals, while the probit estimate is efficient under normally distributed residuals, and therefore a Hausman type test can be performed. Test statistics for

[12] Our estimations are based on *m*=1000.

Hausman's test, reported at the bottom of Table 3.4, clearly indicate that for the total sample and by groups of cities the null hypothesis of normal errors is rejected at the 5% significance level. Therefore we use the estimates from the SLS models in the quantile regression models.

The results presented in Table 3.4 indicate that, overall, younger, less educated, females, non-head of household and non-married individuals are more likely to work in the informal sector. These higher probabilities of individuals in less important positions in the family may indicate that the secondary incomes of the households are earned in informality.

Table 3.4. Estimates of the informal employment models
(y = 1 informal; 0 formal)

	Total sample			Group 1		Group 2		Group 3	
	Probit	Probit[a]	SLS	Probit	SLS	Probit	SLS	Probit	SLS
Constant	2.658***	2.474***	2.474	2.652***	2.652	2.707***	2.707	2.988***	2.988
	(66.61)	(51.13)	(·)	(49.12)	(·)	(22.31)	(·)	(35.02)	(·)
Age	-0.019***	-0.018***	-0.018	-0.019***	-0.019	-0.016***	-0.016	-0.023***	-0.023
	(-26.39)	(-20.52)	(·)	(-19.16)	(·)	(-8.24)	(·)	(-14.73)	(·)
Education levels									
Primary	-0.151***	-0.132***	-0.142***	-0.113	-0.119***	-0.180***	-0.018	-0.253***	-0.267***
	(-7.02)	(-5.15)	(-4.82)	(-4.07)	(-3.29)	(-2.62)	(-0.23)	(-5.30)	(-4.80)
Secondary	-0.498***	-0.452***	-0.480***	-0.458***	-0.534***	-0.614***	-0.753***	-0.688***	-0.680***
	(-25.10)	(-19.10)	(-16.56)	(-17.42)	(-12.69)	(-10.02)	(-7.58)	(-16.16)	(-11.74)
Tertiary	-0.766***	-0.700***	-1.028***	-0.722***	-1.074***	-0.948***	-1.451***	-1.046***	-1.136***
	(-22.29)	(-17.05)	(-16.84)	(-14.73)	(-11.27)	(-10.83)	(-8.27)	(-14.40)	(-10.38)
Male	-0.122***	-0.119***	-0.179***	-0.148***	-0.197***	-0.144***	-0.351***	-0.117***	-0.167***
	(-7.82)	(-6.42)	(-8.45)	(-7.11)	(-6.45)	(-3.44)	(-6.08)	(-3.43)	(-4.19)
Head of household	-0.162***	-0.057***	-0.140***	-0.165***	-0.244***	0.114**	-0.310***	0.155**	-0.160***
	(-9.94)	(-2.90)	(-6.54)	(-7.61)	(-6.46)	(-2.54)	(-4.31)	(-4.44)	(-3.48)
Married	-0.084***	-0.094***	-0.145***	-0.098***	-0.105***	-0.090**	-0.238***	-0.174**	-0.191***
	(-5.69)	(-5.34)	(-6.60)	(-4.93)	(-3.49)	(-2.27)	(-4.35)	(-5.53)	(-4.64)
Presence of children at home	0.021	0.043**	0.115***	-0.018	-0.009	0.029	0.087**	0.045	0.085**
	(1.42)	(2.44)	(6.08)	(-0.91)	(0.34)	(0.73)	(2.04)	(1.46)	(2.41)
Other relatives working as formal	-0.361***	-0.272***	-0.293***	-0.215***	-0.287***	-0.476***	-0.734***	-0.325***	-0.380***
	(23.84)	(16.80)	(12.77)	(10.70)	(8.42)	(-11.50)	(-7.97)	(-9.43)	(-7.83)
Education of household	-0.014***	-0.022***	-0.036***	-0.009**	-0.021***	-0.018**	-0.001	0.007	-0.014**
	(-4.96)	(-6.72)	(-8.68)	(-2.45)	(-3.58)	(-2.27)	(-0.08)	(1.17)	(-2.06)
Size of firm									
11 - 50 employees	-0.982***	-0.995***	-1.083***	-0.967***	-1.260***	-1.173***	-1.580***	-1.044***	-1.502***
	(-57.82)	(-49.05)	(-22.16)	(-42.01)	(-14.48)	(-24.28)	(-8.67)	(-29.33)	(-12.72)
More than 51 employees	-1.617***	-1.608***	-1.934***	-1.552***	-2.180***	-1.778***	-2.790***	-1.870***	-2.501***
	(-98.03)	(-81.80)	(-23.36)	(-69.15)	(-14.44)	(-38.86)	(-8.90)	(-51.30)	(-13.13)
Observations	62,278	43,595		39,091		8226		14,961	
Hausman test		216.1		198.6		384.7		207.4	
p-value		[0.000]		[0.000]		[0.000]		[0.000]	

Note: ***, **, *, denotes significance at 1%, 5% and 10%, respectively. ( ) z-statistics. The constant and the coefficient on variable age in the SLS models were normalized, they are equal to their values in the probit models, so that the probit and SLS models are comparable. All models include industry dummies and occupation dummies. Less than primary school and 1-10 employees are the excluded categories in education and size of firm variables, respectively.

[a] Given computational restrictions on the total sample we take a sample randomly selecting 70% of the observation in each metropolitan area. The resulting sample is 43,595 observations.

Turning to the household characteristics variables, the findings show that having a child at home has a positive impact on the propensity to work in the informal sector but this variable is not significant in more developed cities. At the same time, the presence at home of other relatives working in the formal sector has a negative impact on the probability of being informal and this effect is greater in the group of Caribbean coast cities. In addition, households with a higher education level imply a negative effect on the likelihood of being an informal worker being of particular importance in City Groups 1 and 3. As noted, family environment has a significant effect on the decision to be informal.

Last, the size of firm variables are significant and show that, as the size of firm increases, the probability of being part of the informal sector decreases and this effect is higher in City Groups 2 and 3 than in Group 1.

As described above, in the second step we used the estimates from the SLS to calculate the power series expansion and introduce this term in the quantile regression models to correct for selectivity. To calculate this correction term we included two terms of orthogonal polynomials in the series expansion.[13] At the same time, to implement the identification of the constant term in the wage equations, we used a subsample of workers with a high probability of being informal, namely, those who are younger or older, less educated (less than primary education), with presence of children at home and other relatives working in the informal sector. In Tables A3.3 to A3.6 in the Appendix we present results for corrected quantile regressions for the 5th, 10th, 25th, 50th, 75th, 90th and 95th quantiles.

It can be seen from data in Tables A3.3 to A3.6 that in the City Groups 2 and 3, as well as in the total sample, most of the selection terms are statistically significant, while in more developed cities (Group 1) not all such terms are significant. These results indicate the presence of sample selection bias for individuals across the whole wage distribution in Groups 2 and 3, but not in Group 1. Given these results, we used the estimations of wage equation for City Group 1 without correcting for selectivity in the decomposition. Table 3.5 summarizes the results for corrected and uncorrected quantile regressions at three representative quantiles. The results obtained from OLS and other quantiles for City Group 1 are shown in Table A3.7 in the Appendix.

[13] In fact we tested by including a third term of polynomials in the series expansion, but the estimates presented severe multicollinearity problems. This problem was also mentioned by Buchinsky (1998).

From Table 3.5 we can see that in City Group 2 informal workers receive higher returns to education than formal workers, above all at high quantiles. Similar results, but this time at the median and lower quantiles of the distribution, are found in City Group 3. With regards to other basic human capital variables, such as experience and job tenure, the results show that more experience has a positive and decreasing impact on wages and this effect is particularly higher at low quantiles in the informal sector and is similar in magnitude among groups of cities. An extra year of tenure in a job has a positive impact on wages, and this is relatively constant across the distribution in the formal sector independently of the group of cities. Meanwhile, in the informal sector an extra year of tenure also has a positive effect but this decreases across the distribution.

Regarding the gender variable, the results reveal that there is a strong discrimination against women in the informal sector. This characteristic is more marked in the less developed cities (Group 3) and at low quantiles of the distribution: a woman’s expected earnings at the 10^{th} percentile is approximately 15% lower than a man’s. Meanwhile, in City Groups 1 and 3 similar results are found, but at high quantiles: the difference in wage between a female and a male informal worker is around 11%.

Table 3.5. Quantile regressions by group of cities

(y = Log real hourly wage)

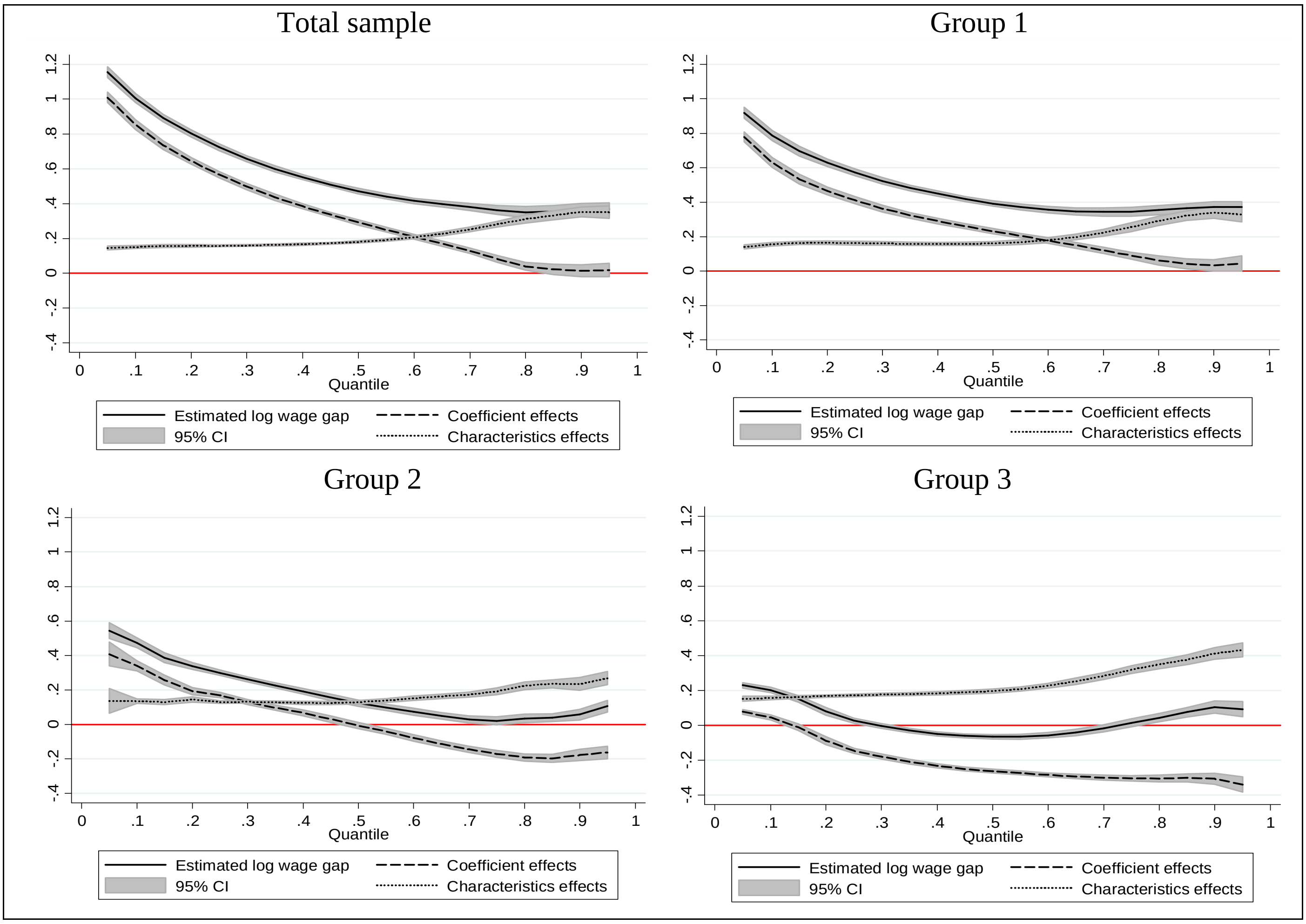
	Group 1						Group 2						Group 3					
	Formal			Informal			Formal			Informal			Formal			Informal		
	10%	50%	90%	10%	50%	90%	10%	50%	90%	10%	50%	90%	10%	50%	90%	10%	50%	90%
Constant	7.245***	7.525***	7.662***	6.702***	7.287***	7.719***	7.475***	7.617***	7.844***	6.437***	7.030***	7.440***	7.429***	7.577***	7.573***	6.429***	7.004***	7.451***
	(412.1)	(709.99)	(262.73)	(203.86)	(475.94)	(250.72)	(94.29)	(133.63)	(64.44)	(125.83)	(205.84)	(150.60)	(116.43)	(238.93)	(94.60)	(176.96)	(342.06)	(193.48)
λ							0.111**	0.070**	0.172**	0.201***	0.073***	0.101*	0.196***	0.055***	-0.031**	-0.231***	-0.180***	-0.212***
							(2.54)	(2.36)	(2.55)	(3.18)	(2.86)	(1.65)	(6.64)	(3.63)	(-2.07)	(-6.52)	(-9.02)	(-5.91)
Education levels																		
Primary	0.067***	0.038***	0.091***	0.071***	0.089***	0.091***	0.115***	0.047**	0.084*	0.065*	0.055**	0.115***	0.067**	0.069***	0.126***	0.126***	0.105***	0.097***
	(5.30)	(4.84)	(4.20)	(3.38)	(8.75)	(4.50)	(3.82)	(2.08)	(1.65)	(1.93)	(2.37)	(3.64)	(2.05)	(4.17)	(3.08)	(5.38)	(7.72)	(3.95)
Secondary	0.164***	0.119***	0.296***	0.179***	0.196***	0.216***	0.162***	0.100***	0.212***	0.155***	0.208***	0.238***	0.166***	0.139***	0.300***	0.243***	0.246***	0.278***
	(13.21)	(15.72)	(14.20)	(7.82)	(18.35)	(10.27)	(5.44)	(4.34)	(4.12)	(4.09)	(8.00)	(6.78)	(5.03)	(8.80)	(7.57)	(9.01)	(16.69)	(10.56)
Tertiary	0.406***	0.536***	0.666***	0.388***	0.547***	0.765***	0.259***	0.430***	0.530***	0.410***	0.528***	0.659***	0.344***	0.503***	0.691***	0.766***	0.691***	0.695***
	(22.41)	(51.48)	(22.74)	(7.38)	(21.49)	(15.11)	(6.38)	(14.84)	(8.14)	(5.90)	(10.17)	(9.57)	(7.95)	(24.50)	(13.28)	(12.69)	(19.42)	(10.88)
Experience	0.002***	0.004***	0.006***	0.012***	0.010***	0.005**	-0.002	0.002	0.007**	0.013***	0.009***	0.004	-0.003	0.004***	0.008***	0.015***	0.014***	0.016***
	(2.37)	(7.42)	(3.58)	(6.64)	(10.42)	(2.55)	(-1.24)	(1.45)	(2.22)	(4.43)	(4.12)	(1.38)	(-1.23)	(4.12)	(2.95)	(6.72)	(11.19)	(6.86)
Experience^2	-0.0001**	-0.0001***	-0.0001**	-0.0002***	-0.0002***	-0.00005	0.00004	-0.00004	-0.0001	-0.0002***	-0.0001***	-0.0001	-0.00006	-0.0001***	-0.0001*	-0.0002***	-0.0002***	-0.0002***
	(-2.50)	(-6.66)	(-1.96)	(-6.46)	(-8.25)	(-1.29)	(0.99)	(-1.14)	(-1.27)	(-3.36)	(-2.88)	(-0.54)	(-1.20)	(-3.07)	(-1.76)	(-5.19)	(-8.49)	(-5.15)
Tenure	0.011***	0.011***	0.019***	0.045***	0.019***	0.026***	0.007**	0.015***	0.016***	0.028***	0.025***	0.023***	0.005	0.003**	0.010**	0.023***	0.018***	0.021***
	(7.04)	(11.92)	(7.64)	(10.25)	(8.55)	(5.86)	(2.53)	(7.13)	(3.27)	(5.95)	(7.05)	(5.13)	(1.36)	(1.99)	(2.44)	(4.66)	(5.97)	(4.36)
Tenure^2	-0.0002***	0.00001	-0.0002**	-0.002***	-0.001***	-0.001***	-0.0001	-0.0002***	-0.0001	-0.001***	-0.0006***	-0.0005***	-0.00005	-0.0002***	0.0001	-0.0006***	-0.0004***	-0.0005**
	(-3.36)	(0.29)	(-1.98)	(-10.61)	(-5.15)	(-4.02)	(-1.33)	(-3.11)	(-0.92)	(-4.70)	(-4.30)	(-2.82)	(-0.34)	(-3.05)	(0.35)	(-2.85)	(-3.15)	(-2.47)
Male	0.016***	0.052***	0.112***	0.083***	0.094***	0.105***	-0.023	0.020**	0.051**	0.040	0.104***	0.116***	-0.029*	0.018**	0.065***	0.150***	0.137***	0.143***
	(2.24)	(12.16)	(9.96)	(4.72)	(10.79)	(6.21)	(-1.55)	(1.97)	(2.32)	(1.43)	(5.23)	(4.01)	(-1.87)	(2.31)	(3.23)	(6.98)	(11.01)	(6.24)
Size of firm																		
11 – 50 employees	0.100***	0.059***	0.072***	0.205***	0.130***	0.134***	0.111***	0.064***	0.034	0.275***	0.185***	0.142***	-0.007	0.015	0.076*	0.314***	0.196***	0.238***
	(10.14)	(9.52)	(4.31)	(9.70)	(12.07)	(6.31)	(3.39)	(2.62)	(0.67)	(6.45)	(6.06)	(3.30)	(-0.22)	(0.98)	(1.94)	(10.61)	(11.68)	(8.25)
More than 51 employees	0.151***	0.106***	0.151***	0.189***	0.125***	0.221***	0.077*	0.046	-0.001	0.292***	0.213***	0.235***	-0.010	0.056***	0.210***	0.529***	0.481***	0.575***
	(17.45)	(19.75)	(10.46)	(7.17)	(9.47)	(8.69)	(1.78)	(1.44)	(-0.31)	(4.16)	(3.97)	(3.18)	(-0.29)	(3.16)	(4.63)	(8.90)	(16.10)	(11.26)
Observations		18,018			8304			4394			3832			6531			8430	

Note: ***, **, *, denotes significance at 1%, 5% and 10%, respectively. ( ) t statistics. Experience is calculated as (age-year of education-6). All models include industry dummies and occupation dummies. Up to primary school and 1-10 employees are the excluded categories in education and size of firm variables, respectively.

3.4.2 Decomposition results

In this section we present the results of the decomposition. Figure 3.9 plots the wage gap that remains after we take into account the difference in the returns of observed characteristics between sectors and correcting for selection for the total sample and by group of cities.

Figure 3.9. Quantile decomposition of the wage gap between the formal sector and informal sector



Source: Table A3.8 and A3.9 in the Appendix

As can be seen from Figure 3.9, for the total sample a significant positive wage gap across the whole distribution remains, with a large gap at the bottom of the distribution. A characteristic of the wage gap is that this decreases throughout the distribution, which would be consistent with greater freedom of choice between formal and informal sector working as workers move up the distribution. At the top end, informal sector workers may to some extent accept lower earnings in order to avoid having to contribute to social protection systems which are perceived to be ineffective.[14]

Regarding the contribution of each set of factors (coefficients and characteristics), we can see that at the bottom of the distribution much of the wage gap is due to informal workers being paid less for their remunerated characteristics than those in the formal sector. The coefficient effects fall over all the distribution, while the characteristics effects rise, particularly toward the upper end of the distribution where they largely exceed the coefficient effects. These results indicate that low-paid informal workers earn less because not only are they less skilled, but they also get lower returns to such skills, whereas high-paid informal workers earn less because formal workers have much better skills.

It is possible to distinguish at least two groups of informal workers who are very different in their position relative to the formal sector. On the one hand, informal workers at the bottom of the distribution represent the disadvantaged segment, which is due to the fact that, even with equal characteristics to those of formal workers, these informal workers obtain lower rates of return to their characteristics. In Fields's (1990) formulation this segment refers to the "easy-entry informal sector" and consists of free-entry, low wage employment, undesirable relative to the formal sector employment and containing a large amount of residual and underemployed labor.

On the other hand, informal workers at the top of the distribution refer to workers who exhibit a higher wage and lower wage differential than those at the bottom of the distribution and their rates of return to characteristics are very similar to those in the formal sector. This group of informal workers corresponds to the advantaged segment, entry into which requires certain characteristics such as a sizeable accumulation of financial and/or human capital. Although these informal workers earn less than their counterparts among formal workers, they find informal activities more

[14] Using data from the 2008 Quality of Life Survey, we calculated that 25% of salaried workers in Colombia reported that they were not covered by any health insurance plan because it was very costly and 20% of salaried workers reported excessive red tape as an obstacle to being covered. In this survey there are questions about the quality and use of the health system which do not exist in the GIHS.

profitable than formal activities. In this regards, Maloney (1999) argues that highly-paid informal workers have specific characteristics or abilities which may imply a non-wage advantage compared to potential earnings in the formal sector. Additionally, he also claims that the high administrative costs of social security combined with the low quality of the services may discourage some workers from getting a job in the formal sector. In this case, informality can be seen as a deliberate choice in order to avoid such administrative costs which are perceived to have a low value given their cost.

At the groups of cities level we can see different patterns in the wage gap and their determinants. The pattern in City Group 1 is similar to the total sample in that the wage gap is positive over the whole distribution, the extent of the coefficient effect is higher at the bottom and median of the distribution, and at the top the characteristics effect explains most of the wage gap. In City Groups 2 and 3 the wage gap between sectors is smaller over the whole distribution; indeed the trend for this gap is towards zero at the top of the distribution in City Group 2 and is negative between the 30th and 70th quantile of the distribution in City Group 3. This lower wage gap may suggest that in cities where informal activities are the main source of income, the informal sector is no longer considered the poor and marginal sector. This result is in line with Marcouiller *et al*. (1997), and Arabsheibani and Staneva (2012) who find a wage premium associated with work in the informal sector in Mexico and Tajikistan, respectively. These authors claim that the scarcity of regulations, the low lvel of enforcement of labor laws and higher tolerance for informal activities can mean higher wage benefits associated with working in the informal sector.

With regard to the contribution of the coefficient and characteristics effects on the wage differential, we can see that in the group of Caribbean coast cities at low quantiles the former effect is positive and makes an important contribution to the wage gap, while at the top half of the distribution the extent of the characteristics effect is higher than the negative coefficient effect.

These results suggest that at the lower quantiles, levels of human capital and other remunerable characteristics are lower in the informal sector than in the formal one, but more importantly the rates of return to those characteristics are lower in the former sector than in the latter. These more disadvantaged individuals in this segment of informal workers could be consistent with the hypothesis of segmentation for these workers. As mentioned, this part of the distribution contains the traditional or more disadvantaged segment of the informal sector, which in the case of this group of cities represents around 66% of the total informal employment (see Figure 3.7). It seems fair to think that at these points in the wage distributions there is no room for these workers in the formal sector, and informality is a last resort strategy escape unemployment.

On the other hand, at the top end of the distribution informal workers obtain higher returns to their characteristics than formal workers and the wage gap is almost zero, indeed this is only 2% at the 75th quantile, although it then increases to 10% at the 95th quantile. These better conditions in this part of distribution are determined by the presence of informal workers with certain skills who, despite having job opportunities in the formal sector as a result of such skills, prefer the combination of monetary rewards and higher flexibility in terms of working hours, specific work relationships, responsibility, etc. in their informal jobs (Fields, 1990). As discussed in Section 3.2, in the Caribbean coast cities a third of informal workers are in sectors where most of the employment is highly qualified, and furthermore they present the highest education and productivity levels compared to the other two groups of cities. Therefore, the low wage gap can be easily compensated for by the cost saving and non-pecuniary aspects associated with being unregistered and hence there will be incentives for choosing informality voluntarily as a form of employment.

Last, in City Group 3, we can see that only at the extremes of the distribution is there a positive wage gap, and this is primarily explained by the characteristic effects, whereas at the median of the distribution there is an informal employment wage premium which is mainly explained by the negative coefficient effect. The high rationing of formal jobs and the relative abundance of informal workers, in particular of those with very low qualification levels (see Figure 3.7 and Table 3.2), are an important determinant of the conditions in the labor markets in this group of cities. In this situation only formal workers who have a significant advantage in characteristics and/or rates of return are superior to informal workers with regard to wages, whereas at the median of the distribution, where there is an important concentration of informal workers, the benefits of being formal are undermined. From the point of view of informal workers, the higher relative disadvantage at the bottom of the distribution implies that workers in this group have no alternative other than to be employed in the informal sector, while at high quantiles the lower relative wages can be compensated with the benefit of a greater flexibility which may allow them to enjoy their work more.

3.5 Conclusions

In this chapter we have examined the heterogeneity of the informal sector at a regional level in Colombia by analyzing the decomposition of the wage gap between the formal and informal sector. We have used the quantile regression decomposition method and corrected by selectivity using semi-parametric methods. This econometric model has allowed us to analyze individuals across the entire distribution of wages and determine if the informal sector has its own internal duality.

Our results show that there is a marked heterogeneity in the informal sector in Colombia. We find that in general there are two distinct segments of workers in the informal sector who have a different motivation for working in this way. On the one hand, there is a lower-paid informal segment in which informal workers are particularly disadvantaged with respect to formal workers, not only in terms of characteristics but also in terms of rates of return to those characteristics. These individuals are rationed out of the formal labor market and informality is seen as the only alternative form of employment. On the other hand, there is a higher-paid informal segment that represents a competitive part into which the wage gap between the two sectors is much narrower than at the bottom and individuals receive similar returns to their characteristics than those in the formal sector. In this segment, informal workers, despite earning less than formal workers, will prefer informality because the benefits of becoming a formal worker may not be attractive. These results suggest that, just as formal and informal activities co-exist, voluntary and involuntary informal employment co-exists. Informality may be a choice as well as being the result of labor market segmentation. Certainly, these are two concurrent scenarios for the same phenomenon.

We also find that there are striking differences in labor market characteristics between groups of cities, in particular regarding the kind of informal employment that exists. The results show that the largest share of informal employment is in the most traditional activities, i.e. where much of the employment is low skilled. In the less developed cities (Group 3) this segment represents about 75% of the total informal employment, whereas in the more developed cities (Group 1) it represents around 47%. With regard to the modern informal segment, the results show that whereas in the group of Caribbean coast cities (Group 2) this segment represents 34% of the total informal employment, in the group of more developed cities it is 52%. These results might suggest that in the more developed cities the informal sector may be associated with more modern activities through intermediate linkages which would lead to an expansion of the modern informal segment.

Turning to the wage differential, once the difference in the returns of observed characteristics between formal and informal sector has been taken into account, the results show that the wage gap over the whole distribution is much narrower in City Groups 2 and 3 than in Group 1. This result supports the idea that the relatively higher abundance of informal activities can come to undermine the ability of the state to provide employee protection and therefore there can be higher benefits associated with working in the informal sector.

With regard to decomposition we found that the wage gap at the very bottom of the distribution is mainly explained by the differential in returns to characteristics of individuals, in particularly in City Groups 1 and 2. In this segment levels of human capital and other remunerable characteristics are very low and given greater importance of the differential in rates of return to characteristics between sectors on wage gap, there is a marked segmentation effect. This result indicates that, at these points of the distribution, the informal sector represents the disadvantaged sector where workers end up as a last resort option for obtaining a paid job.

At the upper half of the distribution the characteristics effect dominates the coefficient effect and the wage gap is positive. These findings suggest that choosing to be an informal worker at these points of the distribution can be in part due to the fact that highly-paid informal workers may to some extent accept lower wages in order to avoid the administrative cost of social security, which is perceived to be costly and ineffective, or because they seek a job with greater flexibility in terms of responsibilities or work schedule. For example, the results showed that the sizes of the estimated wage differential in City Groups 2 and 3 are 10% and 9% at the 95th quantile respectively, which can be easily compensated for by the non-wage benefits associated with working in the informal sector.

3.6 References

Acosta, J. (2012). “Cartagena, entre el Progreso Industrial y el Rezago Social” *Documentos de Trabajo sobre Economía Regional*, No. 178, Banco de la República, Cartagena, December.

Albrecht, J., van Vuuren, A. and Vroman, S. (2009). “Counterfactual Distributions with Sample Selection Adjustments: Econometric Theory and an Application to the Netherlands”, *Labour Economics*, 16(4): 383-396.

Arabsheibani, G. and Staneva, A. (2012). “Is there an Informal Employment wage Premium? Evidence from Tajikistan”, *IZA Discussion Paper*, No. 6727, IZA, Bonn, July.

Arango, L. (2011). “Mercados de Trabajo de Colombia: Suma de Partes Heterogéneas”, *Borradores de Economía*, No 671, Banco de la República, Bogotá, September.

Atal , J. P., Ñopo, H. and Winder, N. (2009), “New Century, old Disparities : Gender and Ethnic Wage Gaps in Latin America”, *IDB Working Papers Series*, No 109, Inter-Amercan Development Bank, Washington, January.

Bargain, O. and Kwenda, P. (2010). “Is Informality Bad? Evidence from Brazil, Mexico and South Africa”, *IZA Discussion Paper*, No. 4711, IZA, Bonn, January.

Bargain, O. and Melly, B. (2008). “Public Sector Pay Gap in France: New Evidence Using Panel Data”, *IZA Discussion Paper*, No. 3427, IZA, Bonn, April.

Barón, J. (2002). “Las Regiones Económicas de Colombia: Un Análisis de Clusters” *Documentos de Trabajo sobre Economía Regional*, No. 23, Banco de la República, Cartagena, January.

Bonet, J. (2005). “Desindustrialización y Terciarización Espuria en el Departamento del Atlántico, 1990-2005”, *Documentos de Trabajo sobre Economía Regional*, No. 60, Banco de la República, Cartagena, July.

Bonet, J. (2007). “La Terciarización de las Estructuras Económicas Regionales en Colombia”, *Revista de Economía del Rosario*, 10(1): 1-19.

Bonet, J. and Meisel, A. (1999) “La Convergencia Regional en Colombia: Una Visión de largo plazo, 1926-1995”, *Coyuntura económica*, 29(1): 15-51.

Bonilla, L. (2008). “Diferencias Regionales en la Distribución del Ingreso en Colombia”, *Documentos de Trabajo sobre Economía Regional*, No. 108, Banco de la República, Cartagena, December.

Bonilla, L. (2009). “Determinantes de las Diferencias Regionales en la Distribución del Ingreso en Colombia, un Ejercicio de Microdescomposición”, *Ensayos sobre Política Económica*, 27(59): 46-82.

Bonilla, L. (2010). “El Sector Industrial de Barranquilla en el siglo XXI: ¿Cambian Finalmente las Tendencias?, *Documentos de Trabajo sobre Economía Regional*, No. 136, Banco de la República, Cartagena, December.

Bosio G., (2009) “Temporary Employment and Wage Gap with Permanent Jobs: Evidence from Quantile Regression”, *MPRA Paper*, No 16055, University of Muenchen, July.

Bucheli, M. and Porzecanski, R. (2011), “Racial Inequality in the Uruguayan Labor Market: An Analysis of Wage Differentials between Afrodescendants and Whites”, *Latin American Politics and Society*, 53(2): 113-150.

Buchinsky, M. (1998). “The Dynamics of Changes in the Female Wage Distribution in the USA: A Quantile Regression Approach”, *Journal of Applied Econometrics*, 13(1): 1-30.

Cepeda, L. (2011). “Los Sures de Barranquilla: La Distribución Espacial de la Pobreza”, *Documentos de Trabajo sobre Economía Regional*, No. 142, Banco de la República, Cartagena, April.

Chernozhukov, V., Fernández-Val, I. and Melly, B. (2012). "Inference on Counterfactual Distribution", *cemmap Working paper,* No CWP05/12, cemmap, Institute for Fiscal Studies, February.

Comi, S. and Grasseni, M. (2009). "Are Temporary Workers Discriminated Against? Evidence from Europe", *ChilD Working Papers*, No 17/2009, ChilD, July.

Cunningham, W. (2001), "Breadwinner Versus Caregiver: Labour Force Participation and Sectoral Choice over the Mexican Business Cycle", in E. Katz and M. Correia (Ed.), *The Economics of Gender in Mexico: World, Family, State and the Market*, World Bank, Washington, D.C.

Cunningham, W. and Maloney, W. (2001). "Heterogeneity among Mexico's Microenterprises: An Application of Factor and Cluster Analysis", *Economic Development and Cultural Change*, 50(1): 131–156.

DANE (2011). Informe de Coyuntura Económica Regional, Meta 2011.

De Soto, H. (1987). *El Otro Sendero*. *La Revolución Informal*, Lima, Instituto Libertad y Democracia.

Dickens, W. T. and Lang, K. (1985). "A Test of Dual Labour Market Theory", *American Economic Review*, 4(75): 792-805.

DNP (2007). Cadena petroquímica-plásticos, cauchos, pinturas, tintas y fibras. Documento sectorial.

Fields, G. (1990)." Labour Market Modelling and the Urban Informal Sector: Theory and Evidence", in: D., Thurnham, Salome, B. and Schwarz, A. (Ed.), *The Informal Sector Revisited*. Paris, OECD.

Fields, G. (2005). "A Guide to Multisector Labor Market Models", *Social Protection Discussion Paper Series*, No 0505, World Bank, Washington, D.C., April.

Fortin, N., Lemieux, T. and Firpo, S. (2011). "Decomposition Methods in Economics ", In O. Ashenfelter and D. Card (ed.) *Handbook of Labor Economics*, Amsterdam: North-Holland, 4(4): 1-102

Galvis, L. (2007). "La Topografía Económica de Colombia". In: J. Bonet (Ed.), *Geografía Económica y Análisis Espacial en Colombia*. Bogotá: Banco de la República.

Galvis, L. (2009). "Geografía Económica del Caribe Continental", *Documentos de Trabajo sobre Economía Regional*, No. 119, Banco de la República, Cartagena, December.

Galvis, L. (2012). "Informalidad Laboral en las Áreas Urbanas de Colombia", *Coyuntura Económica*, 42(1):15-51.

García, G. A. (2005). "El Componente Local de la Informalidad Laboral para las Diez Principales Áreas Metropolitanas de Colombia, 1988-2000", *Desarrollo y Sociedad*, 56: 113-146.

García, G. (2011). "Determinantes Macro y Efectos Locales de la Informalidad Laboral en Colombia", *Sociedad y Economía*, 21: 69-98.

Gindling, T. (1991). "Labor Market Segmentation and the Determination of Wages in the Public, Private-formal and Informal Sectors in San-Jose, Costa-Rica", *Economic Development and Cultural Change*, 39(3): 585-603.

Gunther, I. and Launov, A (2012). "Informal Employment in Developing Countries. Opportunity or Last Resort?", *Journal of Development Economics*, 97: 88-98.

Hardoy, I. and Schone, P. (2006). "The Part-Time Wage Gap in Norway: How Large is It Really?", *British Journal of Industrial Relations*, 44(2): 263-282.

Heckman, J. (1979). "Sample Selection Bias as a Specification Error", *Econometrica*, 47(1): 153-161.

House, W. (1984). “Nairobi's Informal Sector Dynamic Entrepreneurs or Surplus Labor?, *Economic Development and Cultural Change*, 32(2): 277-302.

Ichimura, H. (1993). “Semiparametric Least Squares (SLS) and Weighted SLS estimation of Single Index Models”, *Journal of Econometrics*, 58: 71–120.

ILO (2011). 2011 *Labour Overview. Latin America and the Caribbean*, Lima.

Jütting, J., Parlevliet, J., and Xenogiani, T. (2008). “Informal Employment Re-loaded”, *Working Paper*, No 266, OECD Development Center, Paris, January.

Jütting, J. and De Laiglesia, J. (2009). Is Informal Normal? Towards More and Better Jobs in Developing Countries, *OECD Development Centre*, 163 pages.

Lucifora, C. and Meurs, D. (2006). “The Public Sector Pay Gap in France, Great Britain and Italy”, *Review of Income and Wealth*, 52(1): 43-59.

Machado, J. and Mata, J. (2005). “Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression”, *Journal of Applied Econometrics*, 20(4): 445-465.

Magnac, T. (1991). “Segmented or Competitive Labor Markets”, *Econometrica*, 59(1): 165-187.

Maloney, W. (1999), “Does Informality Imply Segmentation in Urban Labor Markets? Evidence from Sectoral Transitions in Mexico”, *The World Bank Economic Review*, 13(2):275-302.

Maloney, W. (2004). “Informality Revisited”, *World Development*, 32(7): 1159-1178.

Marcouiller, D., Ruiz de Castilla, V. and Woodruff, C. (1997). “Formal Measures of the Informal-sector Wage Gap in Mexico, el Salvador, and Peru”, *Economic Development and Cultural Change*, 45(2): 367-392.

Marjit, S. (2003). “Economic Reform and Informal Wage – a General Equilibrium Analysis”, *Journal of Development Economics*, 72: 371-378.

Melly, B. (2007), “Estimation of Counterfactual Distributions Using Quantile Regression”, forthcoming. University of St. Gallen

Moreno-Monroy, A., Pieters, J. and Erumban, A. (2012). “Subcontracting and the Zize and Composition of the Informal Sector: Evidence from Indian Manufacturing”, *IZA Discussion Paper*, No. 6785, IZA, Bonn, August.

Nicodemo, C and Ramos, R. (2012). “Wage Differentials Between Native and Immigrant Women in Spain: Accounting for Differences in Support”, *International Journal of Manpower*, 33(1): 118-136.

Pratap, S. and Quintin, E. (2006). “Are Labor Markets Segmented in Developing Countries? A Semiparametric Approach”, *European Economic Review* 50(7):1817-1841.

Ranis, G., and Stewart, F. (1999). “V-goods and the Role of the Urban Informal Sector in Development”, *Economic Development and Cultural Change*, 47(2): 259-288.

Saavedra, J. and Chong, A. (1999). “Structural Reform, Institutions and Earnings: Evidence from the Formal and Informal Sectors in Urban Peru”, *Journal of Development Studies*, 35(4), 95–116.

Sánchez, A., Días, A., Peláez, A., Castelblanco, O., Tautiva, J., González, C. and Ángel, L. (2012). “Evolución Geográfica del Homicidio en Colombia”, *Documentos de Trabajo Sobre Economía Regional*, No 169, Banco de la República, Cartagena, June.

Sánchez, F. and España, I. (2012). “Urbanización, Desarrollo Económico y Pobreza en el Sistema de Ciudades Colombianas 1951-2005”, *Documentos CEDE*, No 13, Universidad de los Andes, Bogotá, July.

Simón, H., Sanromá, E. and Ramos, R. (2008). “Labour segregation and immigrant and native-born wage distributions in Spain: an analysis using matched employer–employee data”, *Spanish Economic Review*, 10(2): 135-168.

Uribe, J., Ortiz, C. and García, G. (2007). “La Segmentacion del Mercado Laboral Colombiano en la Década de los Noventa”, *Revista de Economía Institucional*, 9(16): 189-221.
Viloria de la Hoz, J. (2009). “Geografía Económica de la Orinoquia”, *Documentos de Trabajo Sobre Economía Regional*, No 113, Banco de la República, Cartagena, June.
Wahlberg, R. (2008). “Part-Time Penalty in Sweden: Evidence from Quantile Regression”, *Working Papers in Economics*, No 315, Universidad of Gothenburg, Göteborg, September.

Appendix

Table A3.1. Informality rate by metropolitan area

Colombia	58.28
Medellín	47.38
Bogotá	52.13
Manizales	52.50
Pereira	58.15
Cali	65.66
Bucaramanga	66.46
Ibagué	69.03
Barranquilla	70.76
Villavicencio	71.53
Cartagena	72.64
Pasto	72.75
Montería	75.55
Cúcuta	76.93

Note: we included government employees, employers and self-employees to calculate the informality rate.

Figure A3.1. Wage differentials between formal and informal sector over different quantiles of the wage distribution by metropolitan area

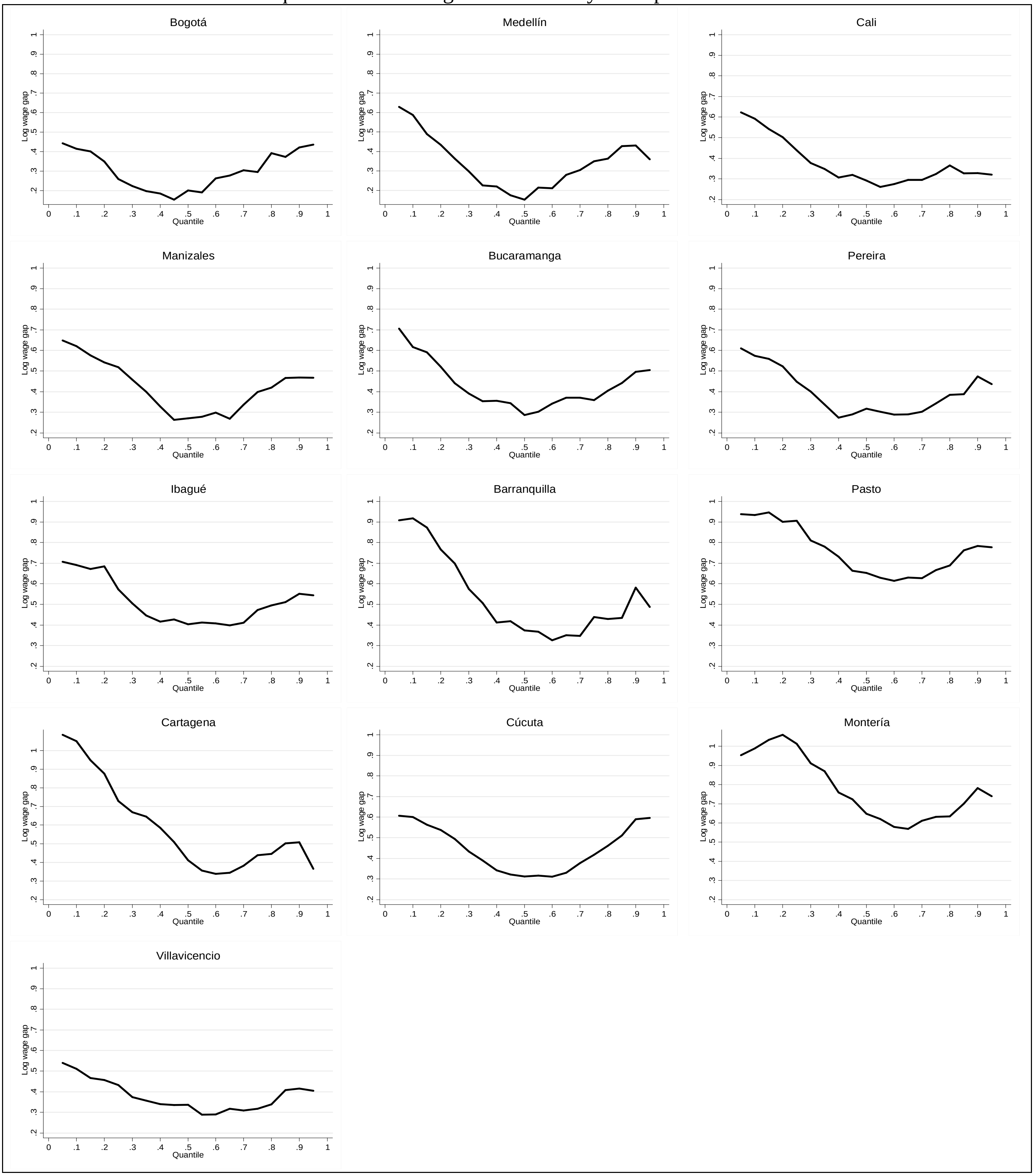


Table A3.2. Sector characteristics by modernity quartile

	Total sample	Group 1	Group 2	Group 3
Size firm (> 50 employees)				
Modernity quartile 1	41.14%	40.90%	51.65%	21.26%
Modernity quartile 2	48.00%	47.92%	56.89%	26.25%
Modernity quartile 3	58.94%	58.50%	63.42%	45.19%
Modernity quartile 4	61.23%	61.73%	69.79%	61.39%
Years of education				
Modernity quartile 1	9.3	9.3	9.9	7.0
Modernity quartile 2	9.8	9.6	10.5	8.5
Modernity quartile 3	10.6	10.8	11.4	10.3
Modernity quartile 4	11.9	11.9	12.0	11.6
Monthly wages per employee				
Modernity quartile 1	613,144	620,117	566,055	510,694
Modernity quartile 2	635,087	666,367	605,707	510,975
Modernity quartile 3	700,139	696,530	665,480	590,461
Modernity quartile 4	706,167	705,173	673,680	691,979

Note: All values are averages across all workers within each sector and quartile, calculated using survey weights. Monthly wages are measured in Colombian currency.

Table A3.3. Quantile regressions for the total sample with corrections for selectivity (y = Log real hourly wage)

	Formal workers								**Informal workers**							
	OLS	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**OLS**	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**
Constant	7.540***	7.271***	7.453***	7.572***	7.556***	7.594***	7.689***	7.829***	7.117***	6.316***	6.525***	6.834***	7.169***	7.429***	7.619***	7.782***
	(410.75)	(182.36)	(282.55)	(515.1)	(385.98)	(301.87)	(190.59)	(153.84)	(506.18)	(184.07)	(244.24)	(350.91)	(467.5)	(463.19)	(293.46)	(223.75)
λ	0.065***	0.164***	0.177***	0.087***	0.024***	0.036***	0.026	0.040	-0.046***	-0.014	-0.049***	-0.035**	-0.024*	-0.033***	-0.073***	-0.101***
	(6.17)	(7.46)	(11.91)	(10.18)	(2.16)	(2.47)	(1.07)	(1.25)	(-3.89)	(-0.51)	(-2.23)	(-2.15)	(-1.87)	(-2.51)	(-3.47)	(-3.67)
Education levels																
Lower secondary education	0.063***	0.073***	0.059***	0.052***	0.041***	0.066***	0.087***	0.085***	0.095***	0.098***	0.083***	0.086***	0.096***	0.106***	0.095***	0.114***
	(7.20)	(3.90)	(4.81)	(7.36)	(4.46)	(5.52)	(4.56)	(3.59)	(9.98)	(4.27)	(4.67)	(6.56)	(9.29)	(10.03)	(5.67)	(5.15)
Higher secondary education	0.172***	0.153***	0.141***	0.100***	0.116***	0.195***	0.274***	0.324***	0.228***	0.194***	0.215***	0.227***	0.210***	0.201***	0.239***	0.272***
	(19.74)	(7.81)	(11.19)	(14.20)	(12.51)	(16.50)	(14.62)	(14.03)	(22.18)	(7.44)	(10.71)	(15.73)	(18.75)	(17.56)	(13.07)	(11.28)
Bachelor/Master	0.493***	0.289***	0.320***	0.396***	0.497***	0.589***	0.647***	0.670***	0.582***	0.487***	0.508***	0.530***	0.551***	0.624***	0.630***	0.640***
	(42.22)	(10.87)	(18.20)	(41.14)	(39.85)	(36.93)	(25.25)	(21.04)	(24.49)	(8.49)	(11.10)	(15.96)	(21.29)	(23.57)	(14.72)	(11.39)
Experience	0.005***	0.003**	0.002***	0.002***	0.004***	0.005***	0.006***	0.005***	0.010***	0.013***	0.015***	0.013***	0.010***	0.009***	0.009***	0.008***
	(7.15)	(2.34)	(2.35)	(3.93)	(5.74)	(5.98)	(4.48)	(3.01)	(11.84)	(6.43)	(8.72)	(11.07)	(10.80)	(8.83)	(5.38)	(3.72)
Experience^2	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001	-0.0001***	-0.0002***	-0.0002***	-0.0002***	-0.0001***	-0.0001***	-0.0001***	-0.0001
	(-6.10)	(-3.77)	(-3.63)	(-5.20)	(-5.29)	(-4.20)	(-2.45)	(-1.06)	(-8.00)	(-4.95)	(-6.61)	(-8.53)	(-7.68)	(-5.83)	(-3.03)	(-1.62)
Tenure	0.013***	0.015***	0.010***	0.008***	0.012***	0.013***	0.016***	0.016***	0.023***	0.035***	0.031***	0.028***	0.020***	0.017***	0.020***	0.015***
	(12.62)	(6.98)	(6.79)	(9.63)	(11.29)	(9.61)	(7.47)	(5.97)	(12.05)	(7.82)	(9.09)	(10.62)	(9.64)	(8.29)	(5.96)	(3.38)
Tenure^2	-0.0001***	-0.0004***	-0.0002***	-0.0001**	-0.0001**	0.00001	-0.0001	-0.0001	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.0004***	-0.001***	-0.0003
	(-2.95)	(-4.33)	(-2.77)	(-2.31)	(-2.11)	(-0.42)	(-1.25)	(-1.34)	(-8.24)	(-6.69)	(-7.41)	(-7.49)	(-5.99)	(-5.00)	(-3.65)	(-1.55)
Male	0.044***	-0.001	-0.001	0.011***	0.041***	0.077***	0.094***	0.094***	0.103***	0.091***	0.091***	0.092***	0.094***	0.103***	0.130***	0.120***
	(9.60)	(-0.14)	(-0.13)	(3.08)	(8.39)	(12.55)	(9.75)	(7.74)	(12.51)	(4.47)	(5.80)	(8.25)	(10.49)	(11.07)	(8.87)	(6.30)
Size of firm																
11 – 50 employees	0.055***	0.055***	0.012	0.033***	0.049***	0.059***	0.064***	0.058***	0.207***	0.293***	0.272***	0.210***	0.162***	0.142***	0.175***	0.217***
	(6.37)	(3.10)	(0.99)	(4.82)	(5.31)	(5.08)	(3.43)	(2.44)	(18.43)	(11.05)	(12.84)	(13.49)	(13.18)	(11.15)	(8.7)	(8.30)
More than 51 employees	0.103***	0.066***	0.020	0.049***	0.091***	0.127***	0.150***	0.137***	0.312***	0.386***	0.392***	0.295***	0.248***	0.243***	0.342***	0.433***
	(10.30)	(3.17)	(1.38)	(6.09)	(8.59)	(9.34)	(6.81)	(4.84)	(17.58)	(8.92)	(11.50)	(11.88)	(12.83)	(12.36)	(11.06)	(10.66)
Observations	25,392								18,203							

Note: ***, **, *, denotes significance at 1%, 5% and 10%, respectively. ( ) t statistics. Experience is calculated as (age-year of education-6). All models include industry dummies and occupation dummies. Less than primary school and 1-10 employees are the excluded categories in education and size of firm variables, respectively.

Table A3.4. Quantile regressions for City Group 1 with corrections for selectivity
(y = Log real hourly wage)

	Formal workers								Informal workers							
	OLS	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**OLS**	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**
Constant	7.497***	7.239***	7.429***	7.575***	7.534***	7.529***	7.628***	7.812***	7.255***	6.532***	6.696***	6.984***	7.287***	7.512***	7.723***	7.886***
	(431.42)	(207.56)	(313.61)	(600.03)	(519.89)	(311.57)	(180.33)	(141.75)	(431.26)	(121.72)	(205.26)	(350.75)	(472.1)	(452.85)	(253.31)	(184.71)
λ	0.035***	0.146***	0.165***	0.083***	0.009	0.0002	-0.031	0.017	-0.006	0.028	0.022	0.021	-0.011	-0.024	-0.029	-0.037
	(3.10)	(6.53)	(10.93)	(10.14)	(0.95)	(0.01)	(-1.09)	(0.45)	(-0.54)	(0.81)	(1.07)	(1.55)	(-1.04)	(-1.13)	(-1.36)	(-1.26)
Education levels																
Lower secondary education	0.068***	0.065***	0.059***	0.039***	0.037***	0.073***	0.093***	0.096***	0.076***	0.004***	0.07***	0.083***	0.090***	0.084***	0.093***	0.086***
	(7.87)	(3.96)	(5.20)	(6.27)	(5.15)	(6.12)	(4.46)	(3.63)	(6.77)	(0.11)	(3.35)	(6.30)	(8.74)	(7.79)	(4.64)	(3.06)
Higher secondary education	0.184***	0.147***	0.132***	0.087***	0.117***	0.213***	0.303***	0.355***	0.198***	0.116***	0.174***	0.195***	0.200***	0.197***	0.220***	0.254***
	(21.30)	(8.55)	(11.40)	(13.82)	(16.27)	(17.85)	(14.63)	(13.61)	(16.36)	(3.03)	(7.56)	(13.47)	(17.97)	(16.89)	(10.23)	(8.55)
Bachelor/Master	0.534***	0.323***	0.356***	0.435***	0.533***	0.632***	0.676***	0.685***	0.512***	0.349***	0.389***	0.413***	0.550***	0.565***	0.768***	0.810***
	(44.90)	(13.42)	(21.12)	(49.19)	(53.78)	(38.12)	(23.17)	(18.30)	(17.84)	(4.07)	(7.30)	(12.45)	(20.91)	(20.15)	(14.52)	(11.27)
Experience	0.004***	0.004***	0.002**	0.002***	0.004***	0.005***	0.006***	0.005***	0.009***	0.012***	0.012***	0.012***	0.010***	0.007***	0.005***	0.004*
	(6.69)	(3.05)	(2.18)	(4.18)	(8.10)	(5.90)	(3.65)	(2.23)	(8.85)	(3.89)	(6.30)	(10.22)	(10.42)	(7.47)	(2.93)	(1.68)
Experience^2	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001**	-0.00002	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0001***	-0.0001	-0.00003
	(-5.39)	(-4.42)	(-3.39)	(-5.65)	(-7.38)	(-3.73)	(-1.89)	(-0.46)	(-7.23)	(-4.11)	(-6.25)	(-9.17)	(-8.24)	(-5.49)	(-1.55)	(-0.58)
Tenure	0.014***	0.014***	0.010***	0.008***	0.011***	0.016***	0.019***	0.017***	0.025***	0.049***	0.045***	0.029***	0.019***	0.021***	0.025***	0.017***
	(13.69)	(6.74)	(7.36)	(10.34)	(13.15)	(11.08)	(7.92)	(5.34)	(10.15)	(6.78)	(10.31)	(10.38)	(8.33)	(9.16)	(5.63)	(2.64)
Tenure^2	-0.0001***	-0.0002***	-0.0002***	-0.0001*	0.0001	-0.0001	-0.0002**	-0.0002	-0.001***	-0.002***	-0.002***	-0.001***	-0.0005***	-0.001***	-0.001***	-0.0004
	(-3.38)	(-2.85)	(-3.32)	(-1.77)	(0.33)	(-1.51)	(-2.08)	(-1.52)	(-7.03)	(-7.31)	(-10.62)	(-8.20)	(-4.96)	(-5.65)	(-3.76)	(-1.15)
Male	0.057***	0.010	0.004	0.014***	0.052***	0.097***	0.114***	0.098***	0.098***	0.098***	0.083***	0.101***	0.096***	0.102***	0.104***	0.100***
	(12.01)	(1.15)	(0.62)	(4.21)	(13.17)	(15.04)	(10.34)	(6.89)	(10.19)	(3.29)	(4.71)	(9.16)	(10.87)	(10.97)	(6.22)	(4.26)
Size of firm																
11 – 50 employees	0.068***	0.057***	0.026***	0.041***	0.055***	0.080***	0.085***	0.059***	0.165***	0.205***	0.197***	0.143***	0.134***	0.125***	0.145***	0.178***
	(7.99)	(3.38)	(2.21)	(6.54)	(7.67)	(6.76)	(4.18)	(2.19)	(12.54)	(5.45)	(8.48)	(9.53)	(11.05)	(9.66)	(6.09)	(5.40)
More than 51 employees	0.124***	0.089***	0.048***	0.060***	0.101***	0.156***	0.169***	0.132***	0.169***	0.121***	0.163***	0.112***	0.138***	0.192***	0.259***	0.324***
	(13.05)	(4.74)	(3.75)	(8.68)	(12.72)	(11.98)	(7.46)	(4.41)	(8.66)	(2.08)	(4.71)	(5.01)	(7.74)	(10.32)	(7.50)	(6.66)
Observations				25,368								13,723				

Note: ***, **, *, denotes significance at 1%, 5% and 10%, respectively. ( ) t statistics. Experience is calculated as (age-year of education-6). All models include industry dummies and occupation dummies. Less than primary school and 1-10 employees are the excluded categories in education and size of firm variables, respectively.

Table A3.5. Quantile regressions for City Group 2 with corrections for selectivity
(y = Log real hourly wage)

	Formal workers								Informal workers							
	OLS	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**OLS**	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**
Constant	7.641***	7.369***	7.475***	7.592***	7.617***	7.819***	7.844***	7.925***	6.969***	6.275***	6.437***	6.724***	7.030***	7.262***	7.440***	7.480***
	(128.28)	(58.68)	(94.29)	(556.54)	(133.63)	(92.39)	(64.44)	(53.88)	(248.69)	(114.92)	(125.83)	(180.51)	(205.84)	(255.79)	(150.60)	(96.46)
λ	0.121***	0.081	0.111**	0.028***	0.070**	0.168***	0.172**	0.196**	0.125***	0.327***	0.201***	0.188***	0.073***	0.053	0.101*	0.075
	(3.88)	(1.19)	(2.54)	(3.93)	(2.36)	(3.76)	(2.55)	(2.44)	(3.29)	(4.60)	(3.18)	(3.86)	(2.86)	(1.44)	(1.65)	(0.83)
Education levels																
Primary	0.072***	0.126***	0.115***	0.074***	0.047**	0.070**	0.084*	0.059	0.076***	0.077**	0.065*	0.036	0.055**	0.099***	0.115***	0.102**
	(3.04)	(2.71)	(3.82)	(13.58)	(2.08)	(2.04)	(1.65)	(1.05)	(3.95)	(2.12)	(1.93)	(1.40)	(2.37)	(5.22)	(3.64)	(2.19)
Secondary	0.142***	0.125***	0.162***	0.086***	0.100***	0.156***	0.212***	0.217***	0.210***	0.125***	0.155***	0.143***	0.208***	0.221***	0.238***	0.280***
	(5.90)	(2.70)	(5.44)	(15.71)	(4.34)	(4.47)	(4.12)	(3.65)	(9.81)	(2.98)	(4.09)	(4.95)	(8.00)	(10.55)	(6.78)	(5.35)
Tertiary	0.420***	0.212***	0.259***	0.345***	0.430***	0.510***	0.530***	0.544***	0.524***	0.408***	0.410***	0.407***	0.528***	0.603***	0.659***	0.632***
	(13.85)	(3.43)	(6.38)	(49.01)	(14.84)	(11.65)	(8.14)	(7.23)	(12.28)	(5.62)	(5.90)	(7.30)	(10.17)	(14.66)	(9.57)	(6.03)
Experience	0.004**	0.006*	-0.002	0.0004	0.002	0.004**	0.007**	0.012***	0.008***	0.008***	0.013***	0.010***	0.009***	0.005***	0.004	0.008*
	(2.44)	(1.79)	(-1.24)	(1.21)	(1.45)	(1.97)	(2.22)	(3.49)	(4.74)	(2.49)	(4.43)	(4.52)	(4.12)	(3.08)	(1.38)	(1.73)
Experience^2	-0.0001**	0.00003	0.00004	-0.00001	-0.00004	-0.0001	-0.0001	-0.0002**	-0.0001***	-0.0001*	-0.0002***	-0.0002***	-0.0001***	-0.0001*	-0.0001	-0.0001
	(-2.06)	(0.47)	(0.99)	(-1.35)	(-1.14)	(-1.43)	(-1.27)	(-1.99)	(-3.10)	(-1.85)	(-3.36)	(-4.02)	(-2.88)	(-1.65)	(-0.54)	(-1.05)
Tenure	0.013***	0.011***	0.007**	0.001***	0.015***	0.018***	0.016***	0.012**	0.027***	0.033***	0.028***	0.030***	0.025***	0.026***	0.023***	0.023***
	(5.62)	(2.80)	(2.53)	(2.67)	(7.13)	(5.44)	(3.27)	(2.20)	(9.37)	(7.18)	(5.95)	(8.03)	(7.05)	(9.66)	(5.13)	(3.60)
Tenure^2	-0.0002**	-0.0001	-0.0001	-0.0001***	-0.0002***	-0.0003**	-0.0001	-0.0001	-0.0007***	-0.001***	-0.001***	-0.001***	-0.0006***	-0.0007***	-0.0005***	-0.0005**
	(-1.96)	(-0.93)	(-1.33)	(-4.18)	(-3.11)	(-2.24)	(-0.92)	(-0.41)	(-5.97)	(-6.32)	(-4.70)	(-5.31)	(-4.30)	(-6.74)	(-2.82)	(-2.29)
Male	0.018*	-0.023	-0.023	0.001	0.020**	0.016	0.051**	0.047*	0.096***	0.001	0.040	0.082***	0.104***	0.117***	0.116***	0.167***
	(1.64)	(-1.03)	(-1.55)	(0.58)	(1.97)	(1.05)	(2.32)	(1.85)	(5.87)	(0.05)	(1.43)	(3.92)	(5.23)	(7.06)	(4.01)	(3.74)
Size of firm																
11 – 50 employees	0.062**	0.166***	0.111***	0.080***	0.064***	-0.009	0.034	0.052	0.190***	0.196***	0.275***	0.232***	0.185***	0.134***	0.142***	0.169**
	(2.41)	(3.23)	(3.39)	(13.76)	(2.62)	(-0.25)	(0.67)	(0.86)	(7.55)	(4.29)	(6.45)	(7.11)	(6.06)	(5.34)	(3.30)	(2.55)
More than 51 employees	0.039	0.139**	0.077*	0.076**	0.046	-0.039	-0.001	0.007	0.215***	0.157**	0.292***	0.187***	0.213***	0.231***	0.235***	0.249**
	(1.16)	(2.00)	(1.78)	(10.12)	(1.44)	(-0.85)	(-0.31)	(0.09)	(4.89)	(1.97)	(4.16)	(3.47)	(3.97)	(5.22)	(3.18)	(2.27)
Observations	4394								3832							

Note: ***, **, *, denotes significance at 1%, 5% and 10%, respectively. ( ) t statistics. Experience is calculated as (age-year of education-6). All models include industry dummies and occupation dummies. Less than primary school and 1-10 employees are the excluded categories in education and size of firm variables, respectively.

Table A3.6. Quantile regressions for City Group 3 with corrections for selectivity
(y = Log real hourly wage)

	Formal workers								Informal workers							
	OLS	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**OLS**	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**
Constant	7.507***	7.264***	7.429***	7.564***	7.577***	7.508***	7.573***	7.808***	6.965***	6.230***	6.429***	6.728***	7.004***	7.232***	7.451***	7.560***
	(203.25)	(100.59)	(116.43)	(203.96)	(238.93)	(123.73)	(94.60)	(81.75)	(341.80)	(144.39)	(176.96)	(237.28)	(342.06)	(270.28)	(193.48)	(167.78)
λ	0.058***	0.173***	0.196***	0.113***	0.055***	-0.020	-0.031**	0.006	-0.197***	-0.209***	-0.231***	-0.213***	-0.180***	-0.186***	-0.212***	-0.227***
	(3.28)	(5.22)	(6.64)	(6.24)	(3.63)	(-0.70)	(-2.07)	(0.13)	(-9.96)	(-5.02)	(-6.52)	(-7.68)	(-9.02)	(-7.27)	(-5.91)	(-5.52)
Education levels																
Primary	0.087***	0.046	0.067**	0.057***	0.069***	0.091***	0.126***	0.082*	0.112***	0.146***	0.126***	0.110***	0.105***	0.111***	0.097***	0.132**
	(4.63)	(1.24)	(2.05)	(2.99)	(4.17)	(2.97)	(3.08)	(1.68)	(8.29)	(5.16)	(5.38)	(5.90)	(7.72)	(6.30)	(3.95)	(4.60)
Secondary	0.190***	0.128***	0.166***	0.121***	0.139***	0.223***	0.300***	0.262***	0.264***	0.257***	0.243***	0.246***	0.246***	0.268***	0.278***	0.314***
	(10.32)	(3.35)	(5.03)	(6.43)	(8.80)	(7.35)	(7.57)	(5.52)	(18.03)	(8.09)	(9.01)	(11.79)	(16.69)	(14.32)	(10.56)	(10.48)
Tertiary	0.501***	0.309***	0.344***	0.345***	0.503***	0.594***	0.691***	0.674***	0.720***	0.736***	0.766***	0.688***	0.691***	0.760***	0.695***	0.623***
	(20.99)	(6.16)	(7.95)	(14.00)	(24.50)	(15.19)	(13.28)	(10.87)	(20.30)	(10.06)	(12.69)	(13.89)	(19.42)	(16.85)	(10.88)	(8.02)
Experience	0.006***	0.004	-0.003	0.002*	0.004***	0.006***	0.008***	0.009***	0.014***	0.014***	0.015***	0.015***	0.014***	0.014***	0.016***	0.016***
	(4.43)	(1.61)	(-1.23)	(1.71)	(4.12)	(2.75)	(2.95)	(2.61)	(11.25)	(5.15)	(6.72)	(8.44)	(11.19)	(8.51)	(6.86)	(6.15)
Experience^2	-0.0001***	-0.0001**	-0.00006	-0.00004	-0.0001***	-0.0001	-0.0001*	-0.0001	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(-3.24)	(-2.18)	(-1.20)	(-1.53)	(-3.07)	(-1.45)	(-1.76)	(-1.51)	(-8.37)	(-3.67)	(-5.19)	(-6.54)	(-8.49)	(-6.39)	(-5.15)	(-4.17)
Tenure	0.007***	0.012***	0.005	0.006***	0.003**	0.007**	0.010**	0.010**	0.022***	0.027***	0.023***	0.023***	0.018***	0.020***	0.021***	0.025***
	(3.56)	(2.91)	(1.36)	(2.95)	(1.99)	(2.07)	(2.44)	(2.09)	(7.39)	(4.54)	(4.66)	(6.00)	(5.97)	(5.47)	(4.36)	(4.65)
Tenure^2	-0.00001	-0.0005***	-0.00005	-0.00005	-0.0002***	0.0001	0.0001	0.0001	-0.0006***	-0.0006***	-0.0006***	-0.0006***	-0.0004***	-0.0005***	-0.0005**	-0.0006***
	(0.22)	(-3.05)	(-0.34)	(-0.61)	(-3.05)	(0.95)	(0.35)	(0.33)	(-4.39)	(-3.34)	(-2.85)	(-4.21)	(-3.15)	(-3.12)	(-2.47)	(-3.06)
Male	0.018**	-0.013	-0.029*	-0.010	0.018**	0.047***	0.065***	0.079***	0.149***	0.142***	0.150***	0.140***	0.137***	0.163***	0.143***	0.135***
	(2.01)	(-0.75)	(-1.87)	(-1.11)	(2.31)	(3.11)	(3.23)	(3.34)	(12.04)	(5.66)	(6.98)	(8.20)	(11.01)	(9.98)	(6.24)	(4.92)
Size of firm																
11 – 50 employees	0.038**	0.055	-0.007	-0.006	0.015	0.068**	0.076*	0.033	0.245***	0.368***	0.314***	0.245***	0.196***	0.203***	0.238***	0.246***
	(2.12)	(1.62)	(-0.22)	(-0.31)	(0.98)	(2.35)	(1.94)	(0.73)	(14.64)	(10.50)	(10.61)	(10.46)	(11.68)	(9.59)	(8.25)	(7.01)
More than 51 employees	0.104***	0.069*	-0.010	0.014	0.056***	0.177***	0.210***	0.169***	0.523***	0.523***	0.529***	0.491***	0.481***	0.502***	0.575***	0.650***
	(5.02)	(1.73)	(-0.29)	(0.65)	(3.16)	(5.25)	(4.63)	(3.20)	(17.62)	(7.60)	(8.90)	(11.04)	(16.10)	(13.75)	(11.26)	(10.85)
Observations				6531									8430			

Note: ***, **, *, denotes significance at 1%, 5% and 10%, respectively. ( ) t statistics. Experience is calculated as (age-year of education-6). All models include industry dummies and occupation dummies. Less than primary school and 1-10 employees are the excluded categories in education and size of firm variables, respectively.

Table A3.7. Quantile regressions for City Group 1 without corrections for selectivity
(y = Log real hourly wage)

	Formal workers								Informal workers							
	OLS	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**OLS**	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**
Constant	7.457***	7.075***	7.245***	7.479***	7.525***	7.528***	7.662***	7.791***	7.255***	6.547***	6.702***	6.981***	7.287***	7.507***	7.719***	7.883***
	(645.22)	(317.65)	(412.1)	(799.3)	(709.99)	(461.05)	(262.73)	(220.65)	(432.07)	(115.50)	(203.86)	(332.00)	(475.94)	(460.77)	(250.72)	(179.44)
Education levels																
Lower secondary education	0.071***	0.077***	0.067***	0.052***	0.038***	0.073***	0.091***	0.097***	0.076***	0.003	0.071***	0.087***	0.089***	0.084***	0.091***	0.085***
	(8.25)	(4.77)	(5.30)	(7.51)	(4.84)	(6.17)	(4.20)	(3.66)	(6.75)	(0.09)	(3.38)	(6.24)	(8.75)	(7.94)	(4.50)	(2.98)
Higher secondary education	0.191***	0.177***	0.164***	0.109***	0.119***	0.213***	0.296***	0.359***	0.196***	0.118***	0.179***	0.200***	0.196***	0.193***	0.216***	0.245***
	(23.18)	(10.93)	(13.21)	(16.39)	(15.72)	(18.57)	(14.20)	(14.32)	(16.77)	(2.97)	(7.82)	(13.46)	(18.35)	(17.40)	(10.27)	(8.21)
Bachelor/Master	0.545***	0.378***	0.406***	0.466***	0.536***	0.632***	0.666***	0.690***	0.509***	0.368***	0.388***	0.416***	0.547***	0.548***	0.765***	0.794***
	(47.94)	(16.49)	(22.41)	(49.54)	(51.48)	(39.83)	(22.74)	(19.14)	(18.19)	(4.15)	(7.38)	(12.11)	(21.49)	(20.51)	(15.11)	(11.19)
Experience	0.005***	0.004***	0.002***	0.002***	0.004***	0.005***	0.006***	0.005**	0.009***	0.012***	0.012***	0.013***	0.010***	0.007***	0.005**	0.003
	(6.94)	(3.48)	(2.37)	(4.31)	(7.42)	(5.94)	(3.58)	(2.26)	(8.87)	(3.73)	(6.64)	(10.04)	(10.42)	(7.42)	(2.55)	(1.28)
Experience^2	-0.0001***	-0.0001***	-0.0001**	-0.0001***	-0.0001***	-0.0001***	-0.0001**	-0.0001	-0.0002***	-0.0003***	-0.0002***	-0.0002***	-0.0002***	-0.0001***	-0.00005	-0.00001
	(-5.32)	(-4.31)	(-2.50)	(-4.89)	(-6.66)	(-3.76)	(-1.96)	(-0.40)	(-7.22)	(-3.98)	(-6.46)	(-8.85)	(-8.25)	(-5.50)	(-1.29)	(-0.16)
Tenure	0.014***	0.014***	0.011***	0.008***	0.011***	0.016***	0.019***	0.017***	0.025***	0.050***	0.045***	0.029***	0.019***	0.022***	0.026***	0.016**
	(13.75)	(7.45)	(7.04)	(9.56)	(11.92)	(11.10)	(7.64)	(5.41)	(10.15)	(6.52)	(10.25)	(10.07)	(8.55)	(9.42)	(5.86)	(2.47)
Tenure^2	-0.0001***	-0.0003***	-0.0002***	-0.0001*	0.00001	-0.0001	-0.0002**	-0.0002	-0.001***	-0.002***	-0.002***	-0.001***	-0.001***	-0.001***	-0.001***	-0.0003
	(-3.48)	(-3.64)	(-3.36)	(-1.82)	(0.29)	(-1.51)	(-1.98)	(-1.58)	(-7.03)	(-7.02)	(-10.61)	(-8.03)	(-5.15)	(-5.65)	(-4.02)	(-1.05)
Male	0.059***	0.021**	0.016***	0.019***	0.052***	0.097***	0.112***	0.100***	0.097***	0.094***	0.083***	0.103***	0.094***	0.099***	0.105***	0.097***
	(12.61)	(2.45)	(2.24)	(5.17)	(12.16)	(15.28)	(9.96)	(7.10)	(10.18)	(3.01)	(4.72)	(8.85)	(10.79)	(10.91)	(6.21)	(4.01)
Size of firm																
11 – 50 employees	0.085***	0.127***	0.100***	0.078***	0.059***	0.080***	0.072***	0.068***	0.162***	0.216***	0.205***	0.153***	0.130***	0.115***	0.134***	0.155***
	(12.63)	(10.30)	(10.14)	(14.51)	(9.52)	(8.52)	(4.31)	(3.25)	(13.76)	(5.87)	(9.70)	(10.64)	(12.07)	(10.26)	(6.31)	(5.21)
More than 51 employees	0.147***	0.189***	0.151***	0.113***	0.106***	0.156***	0.151***	0.145***	0.162***	0.157***	0.189***	0.131***	0.125***	0.170***	0.221***	0.277***
	(25.03)	(17.39)	(17.45)	(24.11)	(19.75)	(19.33)	(10.46)	(8.10)	(11.14)	(3.39)	(7.17)	(7.38)	(9.47)	(12.53)	(8.69)	(7.73)
Observations	25,368								13,723							

Note: ***, **, *, denotes significance at 1%, 5% and 10%, respectively. ( ) t statistics. Experience is calculated as (age-year of education-6). All models include industry dummies and occupation dummies. Less than primary school and 1-10 employees are the excluded categories in education and size of firm variables, respectively.

Table A3.8. Decomposition results for the total sample

Q	Raw log wage gap	Estimated log wage gap	Characteristics	Coefficient
0.05	0.812	1.155	0.145	1.010
		(0.016)	(0.006)	(0.015)
0.10	0.756	1.005	0.151	0.854
		(0.013)	(0.005)	(0.014)
0.15	0.716	0.892	0.157	0.735
		(0.011)	(0.004)	(0.013)
0.20	0.67	0.803	0.158	0.645
		(0.01)	(0.004)	(0.010)
0.25	0.555	0.725	0.158	0.567
		(0.010)	(0.003)	(0.010)
0.30	0.479	0.657	0.160	0.498
		(0.009)	(0.003)	(0.009)
0.35	0.404	0.600	0.163	0.437
		(0.009)	(0.003)	(0.009)
0.40	0.37	0.552	0.167	0.385
		(0.008)	(0.003)	(0.007)
0.45	0.346	0.509	0.172	0.337
		(0.007)	(0.003)	(0.007)
0.50	0.359	0.472	0.180	0.292
		(0.008)	(0.004)	(0.008)
0.55	0.335	0.441	0.192	0.250
		(0.008)	(0.004)	(0.007)
0.60	0.317	0.416	0.207	0.209
		(0.008)	(0.005)	(0.007)
0.65	0.338	0.397	0.227	0.171
		(0.008)	(0.006)	(0.008)
0.70	0.368	0.381	0.252	0.129
		(0.010)	(0.007)	(0.008)
0.75	0.427	0.364	0.282	0.081
		(0.013)	(0.009)	(0.010)
0.80	0.421	0.351	0.311	0.040
		(0.016)	(0.012)	(0.011)
0.85	0.472	0.354	0.332	0.022
		(0.018)	(0.013)	(0.015)
0.90	0.492	0.367	0.353	0.014
		(0.017)	(0.015)	(0.017)
0.95	0.506	0.369	0.351	0.018
		(0.017)	(0.018)	(0.020)

Note: ( ) Bootstrap standard errors based on 1000 repetitions.

Table A3.9. Decomposition results by group of cities

	Group 1				Group 2				Group 3			
Q	Raw log wage gap	Estimated log wage gap	Characteristics	Coefficient	Raw log wage gap	Estimated log wage gap	Characteristics	Coefficient	Raw log wage gap	Estimated log wage gap	Characteristics	Coefficient
0.05	0.604	0.919	0.140	0.779	0.953	0.545	0.137	0.408	0.846	0.230	0.152	0.078
		(0.017)	(0.006)	(0.015)		(0.023)	(0.036)	(0.035)		(0.007)	(0.007)	(0.007)
0.10	0.570	0.786	0.157	0.63	0.973	0.474	0.135	0.339	0.849	0.202	0.157	0.045
		(0.015)	(0.005)	(0.014)		(0.015)	(0.007)	(0.015)		(0.008)	(0.005)	(0.008)
0.15	0.528	0.696	0.164	0.533	0.903	0.387	0.130	0.257	0.792	0.153	0.163	-0.010
		(0.015)	(0.005)	(0.015)		(0.015)	(0.008)	(0.015)		(0.010)	(0.004)	(0.010)
0.20	0.487	0.630	0.165	0.465	0.807	0.338	0.145	0.193	0.776	0.081	0.169	-0.088
		(0.011)	(0.005)	(0.012)		(0.010)	(0.008)	(0.011)		(0.012)	(0.004)	(0.012)
0.25	0.409	0.574	0.162	0.411	0.725	0.299	0.129	0.170	0.713	0.027	0.173	-0.146
		(0.010)	(0.006)	(0.011)		(0.008)	(0.004)	(0.008)		(0.008)	(0.004)	(0.007)
0.30	0.343	0.524	0.161	0.363	0.618	0.262	0.131	0.130	0.622	-0.003	0.177	-0.180
		(0.010)	(0.005)	(0.01)		(0.007)	(0.003)	(0.007)		(0.007)	(0.004)	(0.007)
0.35	0.269	0.482	0.158	0.324	0.512	0.226	0.129	0.097	0.546	-0.030	0.179	-0.209
		(0.008)	(0.005)	(0.008)		(0.007)	(0.004)	(0.007)		(0.007)	(0.004)	(0.007)
0.40	0.258	0.449	0.158	0.291	0.481	0.192	0.125	0.067	0.505	-0.049	0.184	-0.233
		(0.008)	(0.005)	(0.008)		(0.008)	(0.005)	(0.009)		(0.006)	(0.004)	(0.006)
0.45	0.265	0.418	0.159	0.26	0.451	0.156	0.124	0.032	0.469	-0.061	0.190	-0.250
		(0.008)	(0.005)	(0.008)		(0.009)	(0.005)	(0.009)		(0.006)	(0.005)	(0.006)
0.50	0.244	0.392	0.160	0.232	0.410	0.124	0.13	-0.006	0.453	-0.066	0.197	-0.263
		(0.008)	(0.006)	(0.008)		(0.009)	(0.005)	(0.009)		(0.007)	(0.006)	(0.006)
0.55	0.228	0.372	0.168	0.204	0.360	0.099	0.139	-0.041	0.438	-0.066	0.208	-0.273
		(0.009)	(0.007)	(0.007)		(0.010)	(0.006)	(0.009)		(0.007)	(0.006)	(0.006)
0.60	0.253	0.357	0.180	0.177	0.352	0.074	0.152	-0.078	0.445	-0.057	0.228	-0.285
		(0.010)	(0.007)	(0.008)		(0.010)	(0.006)	(0.010)		(0.008)	(0.007)	(0.006)
0.65	0.296	0.347	0.197	0.15	0.372	0.050	0.163	-0.114	0.441	-0.040	0.253	-0.294
		(0.011)	(0.008)	(0.009)		(0.010)	(0.006)	(0.009)		(0.010)	(0.009)	(0.007)
0.70	0.325	0.343	0.223	0.12	0.368	0.029	0.173	-0.145	0.442	-0.017	0.283	-0.300
		(0.012)	(0.010)	(0.009)		(0.010)	(0.007)	(0.009)		(0.010)	(0.010)	(0.007)
0.75	0.357	0.345	0.255	0.09	0.465	0.022	0.192	-0.170	0.491	0.015	0.319	-0.304
		(0.013)	(0.013)	(0.01)		(0.012)	(0.01)	(0.01)		(0.011)	(0.011)	(0.008)
0.80	0.356	0.354	0.292	0.062	0.427	0.034	0.225	-0.191	0.515	0.044	0.350	-0.306
		(0.014)	(0.014)	(0.013)		(0.013)	(0.011)	(0.011)		(0.012)	(0.013)	(0.010)
0.85	0.407	0.364	0.322	0.042	0.468	0.039	0.236	-0.197	0.556	0.076	0.377	-0.301
		(0.014)	(0.014)	(0.015)		(0.012)	(0.012)	(0.012)		(0.014)	(0.015)	(0.012)
0.90	0.425	0.373	0.340	0.033	0.509	0.057	0.234	-0.177	0.627	0.105	0.412	-0.308
		(0.016)	(0.017)	(0.017)		(0.016)	(0.019)	(0.016)		(0.018)	(0.017)	(0.016)
0.95	0.418	0.373	0.329	0.044	0.440	0.107	0.269	-0.162	0.600	0.093	0.433	-0.340
		(0.016)	(0.022)	(0.023)		(0.018)	(0.02)	(0.019)		(0.022)	(0.021)	(0.022)

Note: ( ) Bootstrap standard errors based on 1000 repetitions.

Chapter 4

Agglomeration Economies and Informality in Developing Countries*

4.1 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, 2009; Deichmann *et al.*, 2008). Nonetheless, when analyzing the spatial clustering of manufacturing activity and its potential benefits, most theories and empirical studies on intra-metropolitan location for developing countries only consider formal (large) manufacturing enterprises and do not include informal (micro and small) enterprises (Wu, 1999; Chakravorty *et al.*, 2003; 2005).

Within this context, relevant questions such as how the inter-relationships between formal and informal activities shape the urban landscape, where informal activity is expected to take place within metropolitan areas in developing countries, or which, the formal or informal sector has achieved greater benefits from the spillovers associated with agglomeration economies, remain unanswered. Providing an answer to these questions is not an easy matter, however, because the very nature of informal activity makes it particularly difficult to retrieve detailed data. Perhaps for this reason, there have been no empirical studies to date on the spatial logic of informal manufacturing activity within cities in developing countries (Moreno-Monroy, 2012), let alone on the determinants of such spatial logic.

The empirical analysis in this chapter is divided into two parts. In the first, we compare the spatial clustering or agglomeration patterns of formal and informal manufacturing activity within a metropolitan area of a developing country. In this part 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

---

* Section 2 and 3 of this chapter are based on a paper jointly written with Ana I. Moreno-Monroy, which is under review in a journal listed in the ISI-JCR.

[1] The informal sector was defined by the 15th International Conference of Labor Statisticians (ICLS) in 1993 as "private unincorporated enterprises that are unregistered or small in terms of the number of employed persons. An enterprise is unincorporated if it is not constituted as a separate legal entity independently of its owner(s) and does not maintain a complete set of accounts."

inhabitants. This uniquely rich dataset includes the full range of manufacturing enterprises geocoded up to a disaggregated scale equivalent to the city block level. Several criteria in the Census make it possible to discriminate between formal and informal manufacturing enterprises by industries. For the purpose of this chapter, informal enterprises are defined as those enterprises producing legal goods that do not fully comply with established legal regulations. More precisely, informal enterprises are defined as those enterprises not registered in the Chamber of Commerce. Furthermore, since information on the number of employees per enterprise is also provided, formal enterprises can be disentangled into formal small and formal medium and large enterprises. This division is useful in distinguishing the location effects explained by size from those related to the formality status of an enterprise.

In order to complement the above analysis, the second part examines the effects of agglomerations economies on formal and informal productivity. Although the census manufacturing data for Cali contains rich information on the spatial location of enterprises, it does not contain any measurement of workers' educational qualifications, and therefore any estimation of productivity may be biased. In order to overcome this shortfall, data at worker level from the Continuous Household Survey (CHS) for the period 2002-2006 is used for this purpose. This survey includes information on the labor force, unemployment, informality, and other socioeconomic and socio-demographic characteristics of thirty-three municipalities in Colombia. It is thus possible to analyze thirty-three territorial labor markets through the period 2002-2006.

This chapter explores three particular questions: firstly, do formal and informal enterprises display different clustering patterns? Secondly, do formal and informal enterprises locate in the same areas of the city, or is there a marked spatial segmentation between formal and informal manufacturing activity? And thirdly, are there differences in the effects of agglomeration on productivity and wages between the formal and informal sector? These questions are quite relevant from a policy perspective. 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 for generating production externalities. The identification of these locational effects requires a characterization of the actual clustering patterns of informal enterprises, and an analysis of the difference between the clustering patterns of formal and informal enterprises.

Thus, in order to answer these questions, we use two complementary approaches. First, the degree of spatial clustering (intra-industrial) and co-clustering (inter-industrial) is calculated by means of M-functions (Marcon and Puech, 2010). This method is preferred over other alternatives for measuring spatial clustering because, besides satisfying all the relevant criteria for a good measure of spatial clustering, it enables a direct comparison between the clustering and co-clustering intensity of formal medium and large, formal small and informal enterprises in different industries. Furthermore, in order to exploit the spatial component of the data fully and complement the M-function results, the geographical distribution of formal and informal enterprises in different clustered and co-clustered industries is analyzed. In particular, the spatial analysis of the distribution of formal and informal enterprises is conducted by means of kernel density mapping and geographical distribution analyses of formal and informal enterprises in selected clustered and co-clustered industries.

In order to complement the above results, in the second approach the effect of agglomeration economies on formal and informal productivity is investigated, and an analysis performed of which sector, formal or informal, achieved greater benefits from the diversity of activities and those spillovers associate with urbanization economies. In this part individual worker wages as a measurement of labor productivity are regressed on employment density as a measurement of urban agglomeration, separately for the formal and informal wage sector, controlling by several socioeconomic, socio-demographic and regional characteristics.

After this introduction, Section 4.2 proceeds with a literature review. Section 4.3 presents analysis of intra-metropolitan clustering of formal and informal manufacturing activity from Cali. Section 4.4 discusses the results of econometric estimation of wage equation, and Section 4.5 concludes.

4.2 Literature review

The review of the literature is divided into two parts. In the first part, we provide a brief outline of the literature related to agglomeration economies in developing countries, a more detailed literature review can be found in Scott (2002), Henderson (2005), Overman and Venables (2005), Quigley (2009) and Duranton (2009). In the second part, we discuss the theoretical predictions regarding the intra-metropolitan clustering of informal manufacturing activity.

4.2.1 Agglomeration economies in developing countries

Developing countries are characterized by certain structural conditions such as economic and political instability, high rates of unemployment and underemployment, shallowness of markets, and low industrial and infrastructure development. These conditions can affect the extent of external economies associated with agglomeration economies. Several studies have found quantitative evidence of both localization and urban effects. In the seminal work by Henderson (1986), which analyzes the role of localization and urbanization economies in productivity in the metropolitan areas of Brazil, it was found that localization economies play an important part in this regard, while urbanization economies are present, but only weakly so. Its results show that if employment in any sector in any region were to double, productivity measured by value-added would increase by 11%. A comparable result is found by Lee and Zang (1998) in their study of manufacturing industry in South Korea. The authors found that doubling the employment of a given sector and region is associated with an increase in value-added per worker as a measurement of productivity of 7.9%. From Indian cities, the studies by Mills and Becker (1986), Becker *et al*. (1992), and Shukla (1996) show that equally significant increases in productivity are generated by urbanization. In a more recent analysis of city growth in Brazil, Da Mata *et al*. (2007) found that the urban elasticity, measuring urbanization economies as market potential, is 11%.

Nonetheless, these evidences have an important bias, since most of the findings are concerned with the formal sector and do not take into account the informal sector. As pointed out by Duranton (2009), between the formal and informal sector there are intense linkages, which suggests that agglomeration effects are generated within both sectors, with benefits that accrue to both. According to Overman and Venables (2005) there are two possibilities in which the existence of an informal sector can affect the benefits of agglomeration economies. On the one hand, the existence of an informal sector can drive up urban costs and crowd out the formal sector. On the other hand, the informal sector also contributes to agglomeration economies. In this sense, the informal sector is made up of small enterprises producing on a small scale, which establishes important networks that contribute to the formation of clusters. Furthermore, as in the formal sector, the informal sector can achieve benefits from the productivity effects associated with the concentration of the activity and employment.

4.2.2 Theoretical predictions on the intra-metropolitan clustering of informal manufacturing activity

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 standardized the goods are, the more stable interactions are and the less need there is for intermediation.[2] Hence, large 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 transaction partners because their linkages are small in scale, unstandardized, unstable, and in need of personal intermediation.[3] Because of their characteristics, the linkage costs of informal enterprises are probably high, so that the tendency to seek close 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 why informal enterprises can be expected to locate in different areas than formal enterprises. 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

[2] 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 product involved, setup costs need not be incurred every time a transaction takes place, and orders can be placed directly without the need of further intermediation.

[3] 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, 2003; Lafourcade and Mion, 2007).

population catered by informal manufacturing enterprises may be concentrated precisely in peripheral areas, so that the proximity of informal producers to their consumers follows the same logic of transaction costs minimization. Fourth, informal enterprises operating on a subcontracting basis may consider proximity to input suppliers and consumers irrelevant and, as a consequence, prefer to locate in residential areas of the city with lower rent prices. In the case of home-based enterprises, preference would be in areas where informal producers reside.[4] This may be the case whenever formal enterprises directly provide materials and inputs to informal enterprises, commonly through an intermediary (Carr *et al.*, 2000), or whenever 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 actually influence the location decisions of enterprises (Arauzo-Carod and Viladecans-Marsal, 2009), the clustering patterns of formal and informal enterprises of similar sizes should not differ. The literature has recognized three such channels (Duranton and Puga, 2004): first, sharing inputs produced under increasing returns to scale; second, accessing a larger pool of workers; and third, knowledge and technical spillovers. It can also be argued, however, that informal enterprises located in peripheral areas may be accessing sub-optimal markets for intermediate inputs and labor in their own 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, 2012). In this case, formal and informal enterprises would not necessarily locate in the same areas of the city.

4.3 Intra-metropolitan clustering of formal and informal manufacturing activity

4.3.1 Description of the study area and data

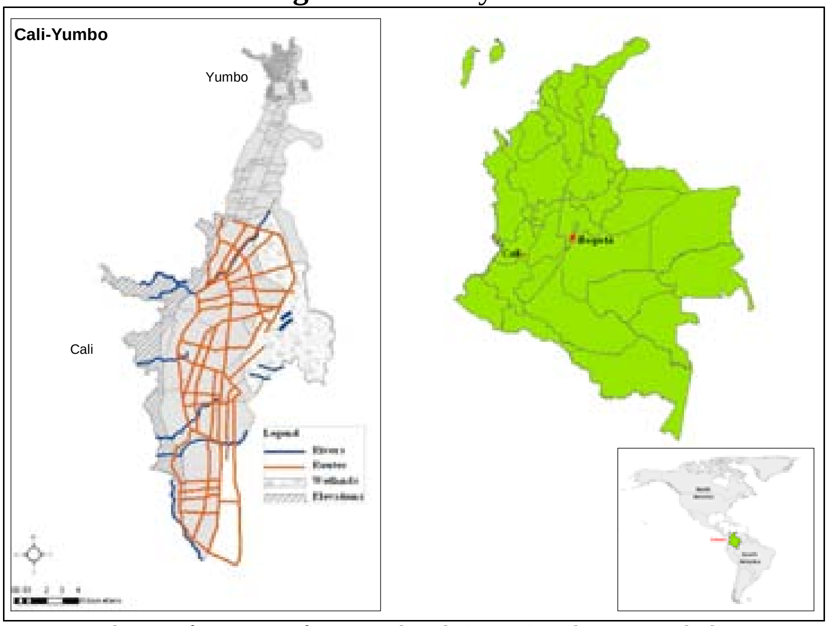
Santiago de Cali (3°27' lat. North - 76°31' long. West), capital of the Valle del Cauca region (*departamento)* and third city in importance of Colombia, is located in the west of Colombia. The metropolitan area of Cali is composed of the municipalities of Cali and Yumbo and is divided into 26 communes, 279 neighborhoods and over 14,000 “*Manzanas*”, of which 1,125 (4%) are green areas or have no information available.

[4] 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).

This last scale is comparable with the "census block" level used in the USA Census.[5] 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 4.1). The main transportation routes, starting from the center of the city, describe a concentric and radial pattern towards the east of the city.

Figure 4.1. 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 A4.1 in the Appendix). These densely populated areas coincide with areas of population predominantly in low and low-middle income categories (see Figure A4.2 in the Appendix). On the contrary, the wealthiest areas of the city proper, located predominantly in the south and south-west of the city, are less densely populated.

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

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

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, see Table A4.1 in the Appendix), and there are on average 3.7 enterprises per block, the scale is approximately equivalent to having the actual address of each enterprise.

In the analysis, only manufacturing enterprises are considered. 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% are located in Cali, and the remaining 5% 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, and the analysis is conducted at the 3-digit level (see Table A4.2 in the Appendix for codes and names).[6]

Informal enterprises are defined as those enterprises not registered in the Chamber of 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 conducted by personnel who went door-to-door over the entire area of study. This manner 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 Census covers *universe* of enterprises, sampling problems are ruled out (Cardenas and Rozo, 2009) and 3) as the Census does not rely on the prompt 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

[6] 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 (Alonso-Villar and del Rio, 2013).

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]

Formal enterprises are split into two categories: formal medium and large enterprises, defined as those formal enterprises with more than 50 employees, and formal small enterprises, defined as those formal enterprises with 50 employees or less.

As can be seen in Table 4.1, roughly 99% of informal enterprises and 83% of formal enterprises can be classified as “micro” (1-10 workers). A substantial percentage of informal enterprises (45.6%) operate within the premises of a household (Table 4.2), while the large majority of formal enterprises operate in an office or plant.

Table 4.1. 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%**

Table 4.2. 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%**

4.3.2 Spatial distribution of formal and informal enterprises

Figure 4.2 displays the enterprise-based Location Quotient (LQ)[8] index for the whole sample of manufacturing enterprises (panel a), formal enterprises (panel b) and 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.

[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 and Mion, 2007).

enterprises (panel c) at the neighborhood level. This index, which measures the relative concentration of activities, is defined as:

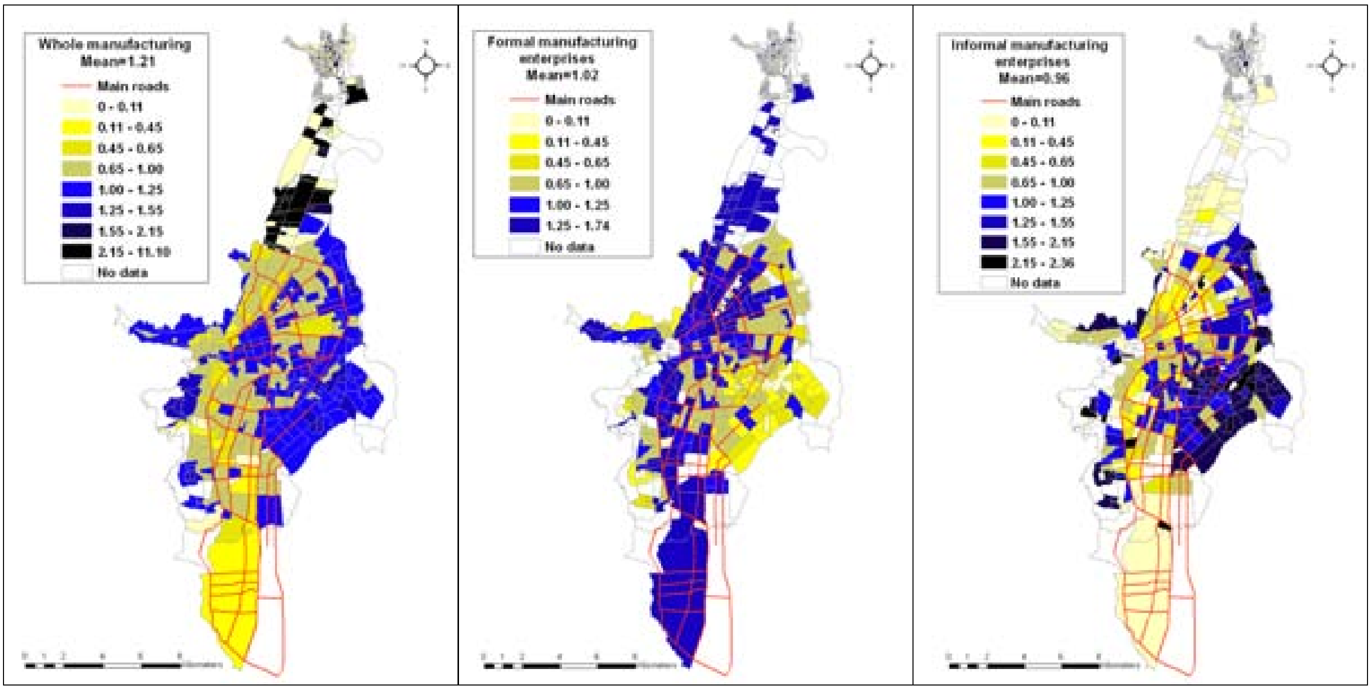
$$LQ_i^s = \frac{c_i^s}{x^s},$$

where $c_i^s = ent_i^s \Big/ \sum_{i=1}^{N} ent_i$ denote sector's s share of enterprises in location i, and $x^s = \sum_{s=1}^{S} ent_i \Big/ \sum_{i=1}^{N}\sum_{s=1}^{S} ent_i^s$ 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 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 A4.2 in the Appendix), while formal activity is found in wealthier and more central locations.

Figure 4.2. Concentration plant-based LQ index

a) All b) Formal c) Informal



4.3.3 Spatial indicators of relative geographic clustering and co-clustering

Currently, there are two alternative ways of measuring clustering and co-clustering: cluster-based methods (e.g., Ellison-Glaeser (EG) or LQ indices) and distance-based methods. For the purpose of the empirical application developed in this chapter, the cluster-based method option is dismissed for two reasons. First, identifying the spatial scale at which clustering occurs using cluster-based methods would involve calculating the indices for different spatial scales and comparing the results at these different scales. This strategy is problematic because it is likely to suffer from the scale problem resulting from 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), there are marked differences in the nature of the location patterns of small and large enterprises, and cluster-based measures are not able to properly capture size heterogeneity within industries. This makes the analysis of the differences in the nature of locational choice of different types of enterprises cumbersome.

In contrast with cluster-based methods, distance-based methods, of which the *Kd*-function of Duranton and Overman (2005) and the M-functions of Marcon and Puech (2010) are examples, fully satisfy the property of being unbiased across geographical scales. Besides this property, these distance-based methods satisfy the other four relevant criteria of a good measure of clustering (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 distance-based methods is that they control for inhomogeneous space (and consequently can account for the fact that enterprises cannot locate everywhere in the city).

The key difference between the *Kd*-function and the M-function is that the first one relies on a probability density function (pdf) and the second one on a cumulative distribution. Thus, "if the question is 'up to which distance do externalities matter?', then a cumulative function is more appropriate [e.g., *Kd*], while a pdf will answer 'do externalities matter at a given distance?' better" (Marcon and Puech, 2012, p. 14). 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 clustering and co-clustering. Nevertheless, this approach is out of the scope of this work. Thus, in choosing between these two distance-methods, the M-functions are preferred because they have an important advantage over the *Kd*-function: the M-function 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). Thus, the M-functions serve well the purpose of this chapter: *comparing* the intensity of intra-metropolitan clustering and co-clustering of formal and informal enterprises in different industries.

4.3.3.1 M-functions clustering and co-clustering

In order to evaluate the degree of clustering and co-clustering of formal and informal enterprises, the intra and inter-industry M-functions proposed by Marcon and Puech (2010) are calculated using the software Ripley v.2.8.[9] For the implementation of this method, available information on the plain coordinates (X-Y) for each enterprise at the city block level are used to measure the Euclidean distance among enterprises. M-functions are calculated for formal enterprises, informal enterprises, formal small enterprises (less than 50 employees), and formal medium and large enterprises (more than 50 employees) and/or (selected) ISIC 3-digit industries. The M-functions are calculated every 1 km between zero and 20 km. In the following, the definition of M-functions of clustering and co-clustering are explained in detail.

4.3.3.1.1 Clustering

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

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

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. Here, a *sector S* refers to either an ISIC 3-digit industry, a type of enterprise (formal medium and large, formal small or informal enterprises) or both (e.g., informal enterprises in the textile industry). For each of these enterprises, a circle of radius r (e.g., 1 km) 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

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

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. 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 sector is the same. The confidence intervals are determined using Monte-Carlo methods in the following way: first, a large number of simulations (1000 for the case of clustering and 100 for the case of co-clustering[10]) are generated. Next, a confidence level of 5%, is chosen, so that the 95% confidence interval of M for each value of r is delimited by the outer 5% of the randomly generated values. Then, significant relative clustering (dispersion) of a sector appears if the corresponding M-values are larger (smaller) than one and are outside the confidence interval bands.

4.3.3.1.2 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:

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

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

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

$M_{S_1S_2}$ ($M_{S_2S_1}$) 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 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 S_1 pattern, the null-hypothesis point set for $M_{S_1S_2}(r)$ is generated by keeping S_1 points unchanged and redistributing all other points onto all other locations. To control for the S_2 pattern, the same process applied to $M_{S_2S_1}(r)$. There is a significant co-clustering relationship whenever both values are significantly different from their respective null hypothesis (Marcon and Puech, 2003a).

4.3.4 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 relevant for the purposes of this chapter 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.

In the spatial analysis, given that the location of each enterprise is known at the city block level, each point (enterprise) is randomly assigned within its block. In this way, unique plain (X-Y) coordinates for each enterprise are obtained. The spatial analysis is undertaken using these points on industries and type of enterprises selected based on the M-functions results. In particular, data for selected clustered formal and informal industries is analyzed through a kernel density mapping. The Kernel estimator is given by:

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

where h is the chosen radius or smoothing parameter, $X_1 \ldots X_n$ are enterprises, n is the enterprises and K is the quadratic kernel function. A kernel density surface is defined for each individual point with the highest value at its location diminishing to zero with distance away from the point. Beyond h, the density surface for a given point is zero. A continuous, smooth density surface across the entire study area is obtained by adding these individual density surfaces (Silverman, 1986). The smoothing parameter, or search radius distance, is set to 500 meters.

The geographic distribution of formal and informal enterprises in selected co-clustered industries is also analyzed via cartographic representations of the spatial distribution of enterprises and visualization of standard deviational ellipses which measure whether a point distribution exhibits a directional trend. These ellipses are calculated at one standard deviation.[11]

4.3.5 Results of M-functions

This section presents the results for the spatial indicators of clustering and co-clustering by type of enterprise – informal, formal small and formal medium and large enterprises (aggregated by industry), by industry and by type of enterprise and industry. 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, see Table A4.2 in the Appendix).

4.3.5.1 Clustering

By type of enterprise

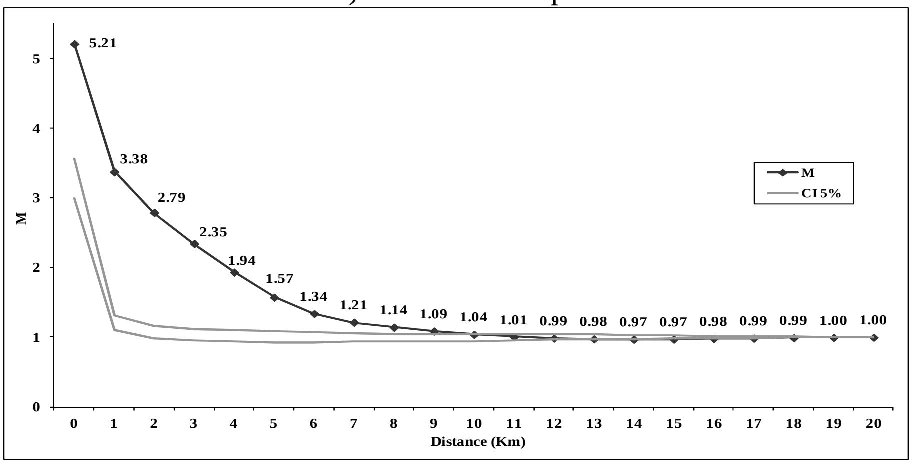
Figure 4.3 shows the M-functions for the aggregate of informal, formal small and formal medium and large enterprises. Formal small and informal enterprises show significant concentration at distances between 0 and 9 km, as can be seen from the fact

[11] Standard deviational ellipses "measure 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).

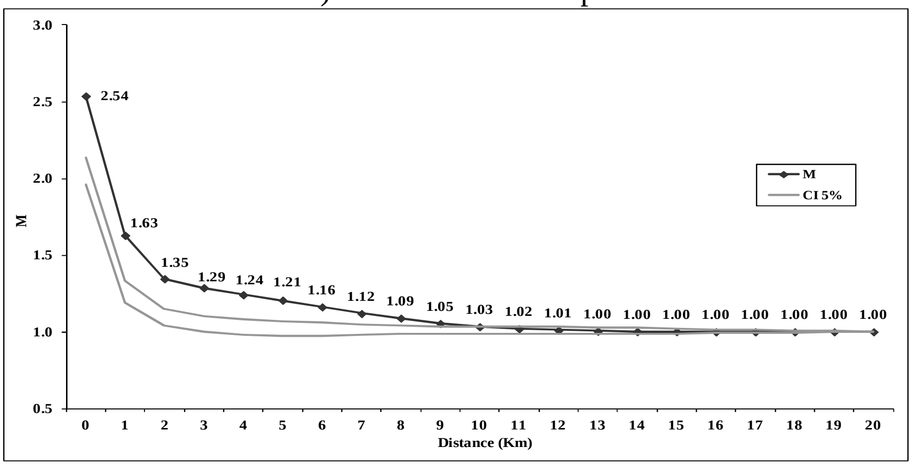
that M-values at these distances are larger than one and outside the upper band of the confidence interval. For both types of enterprises, the M-functions peaks at 0-1 km and then shows a continuous decay as distance increases. Interestingly, the degree of clustering is larger for informal enterprises than for formal enterprises of comparable size: at the peak M-value, the relative density of employees in informal enterprises within 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. Finally, as expected, formal and informal enterprises of a smaller size have different clustering patterns than larger formal enterprises: the degree of clustering, although significant between 1 and 8 km, is much smaller, with a peak value of 1.15.

Figure 4.3. Intra-industry M-functions by type of enterprise

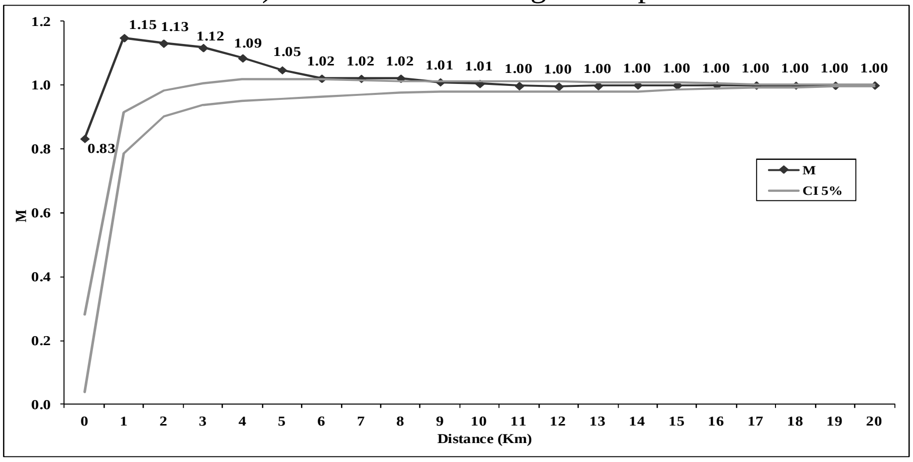
a) Informal enterprises



b) Formal small enterprises



c) Formal medium-large Enterprises



By industry

The M-functions per industry reveal that 40 out of 64 industries are clustered at different distance ranges (see Table 4.3). As in Marcon and Puech (2003b), the most important degree of concentration occurs at very small distances for all sectors, but the degree of concentration differs widely from one industry to another. Out of these 40 industries, 11 show significant M-values at least at continuous distance ranges of 0-5 km.

By type of enterprise and industry

Table 4.4 summarizes the results for the intra-industry M-functions by type of enterprise. The first thing to note is the small number of significant results for large enterprises (9 out of 64) compared to formal small and informal enterprises (38 out of 64 and 40 out of 64, respectively). As predicted by theory, the significant clustering for formal large enterprises occurs mostly at very close proximity, between 0 and 1 km.

Evidently, formal small and informal enterprises exhibit clustering in roughly the same industries but to different degrees and at different distance ranges. Formal small enterprises exhibit significant clustering in 33 of the 40 industries where informal enterprises clustering is significant (except industries 159, 172, 175, 191, 241, 269 and 371), while the opposite is not true for five industries (201, 292, 322, 323 and 351). Informal enterprises present higher strenghth of clustering than formal small; in fact the M-peak values are larger for informal enterprises in 28 out of 33 cases. The strength of clustering of small enterprises is indeed remarkably high in some industries. Although other studies (Marcon and Puech, 2003b) do no find such extremely high M-values, they do seem plausible given the small size of informal enterprises.[12] However, these extreme results may be influenced by the small number of observations by type of enterprise and industry in some industries (see Table A4.3 in the Appendix).

Table 4.3. Intra-industry M-functions

Ranking	Industry	# enterprises	# employees	Distance range (Km)	M-peak	M-peak distance (Km)
1	202	11	44	0 - 10	306.32	0
2	371	14	42	0, 3	288.02	0
3	261	14	24	0 - 10	149.70	0
4	201	26	97	0, 3 - 7	108.61	0
5	272	9	17	0, 2	48.47	0
6	372	36	169	0, 5	46.71	0
7	293	26	817	0 - 12	33.34	0
8	333	44	281	0 - 11	30.51	0
9	271	50	202	0, 3 - 8	29.91	0
10	22	259	1620	0 - 6	27.11	0
11	173	77	314	0, 5 - 6	20.43	0
12	292	38	543	0	16.20	0
13	193	120	789	0	13.12	0
14	203	205	625	0 - 9	12.87	0
15	151	49	622	0	11.83	0
16	152	124	1278	0	10.35	0
17	361	259	1495	0 - 5	9.98	0
18	289	188	903	0, 2 - 9	9.61	0
19	192	354	1692	0 - 11	9.26	0
20	209	158	610	0 - 7	8.41	0
21	174	81	562	0 - 1, 8	7.33	0
22	154	307	2387	0 - 7	5.39	0
23	155	578	2585	0 - 7	5.16	0
24	159	64	1299	0	4.59	0
25	369	152	1205	0	4.01	0
26	210	122	2725	0	3.37	0
27	322	8	63	3 - 12	2.85	3
28	252	157	2595	0	2.66	0
29	251	33	885	4 - 7	2.46	4
30	181	662	8675	0	1.75	0
31	175	22	106	3 - 6	1.72	5
32	281	51	336	3 - 6	1.56	4
33	359	13	1230	0	1.56	0
34	204	32	132	4 - 6	1.52	5
35	153	23	172	6 - 9	1.32	6
36	269	46	350	5	1.24	5
37	311	4	61	13	1.22	13
38	158	59	3867	0	1.20	0
39	242	100	8650	4, 6 - 10	1.14	4
40	291	14	437	9	1.14	9

[12] Since most studies exclude small enterprises (<10 or 20 workers) from their samples, there is no existing benchmark against which these results can be compared.

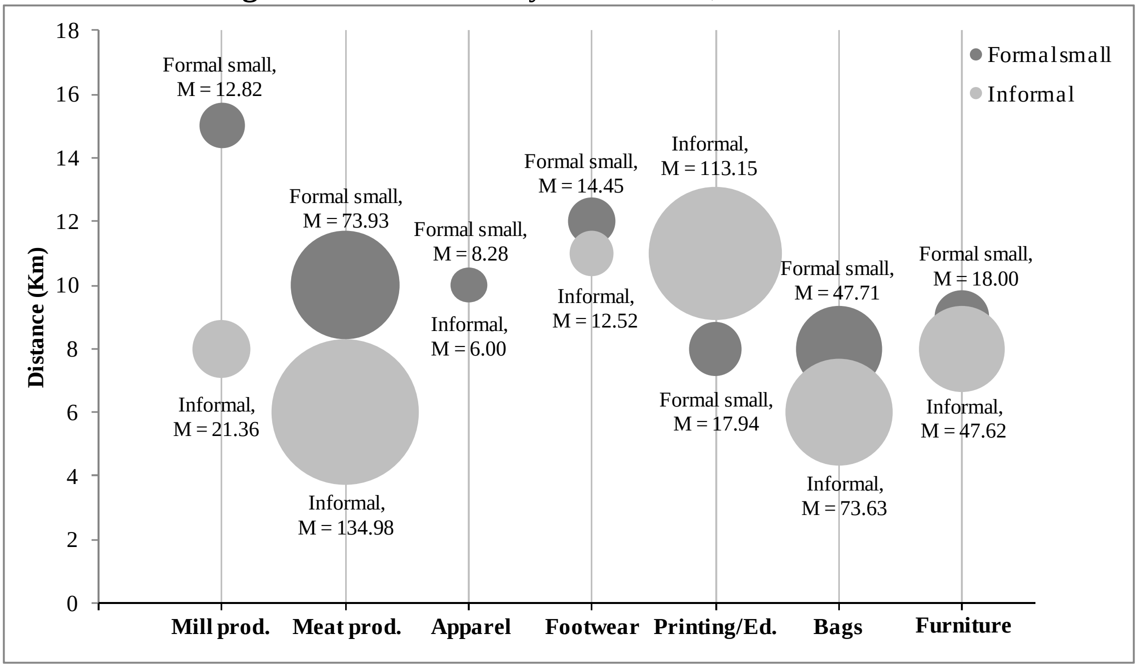
Table 4.4. Intra-industry M-functions by type of enterprise

	FMLEs				FSEs				IEs			
Ranking	Industry	Distance range (Km)	M-peak	M-peak distance (Km)	Industry	Distance range (Km)	M-peak	M-peak distance (Km)	Industry	Distance range (Km)	M-peak	M-peak distance (Km)
1	155	0	95.93	0	202	0-3, 5-9	1036.35	0	271	0	431.61	0
2	293	0-15, 18-19	82.79	0	201	0, 3-8	132.64	0	159	0-8	137.15	0
3	361	0	25.25	0	272	0	112.01	0	151	0-6	134.98	0
4	210	0-3	9.00	0	359	0	101.68	0	22	0-11	113.15	0
5	181	0	2.46	0	151	0-10	73.93	0	372	0, 7	109.32	0
6	252	0, 4-6, 8	2.08	0	157	0	61.73	0	174	0-7	83.72	0
7	242	2	1.71	2	333	0-12	52.16	0	193	0-6	73.63	0
8	159	9-11, 13	1.34	9	372	0-1	51.05	0	152	0-10	71.38	0
9	209	10	1.25	10	193	0-8	47.71	0	191	1-2	66.34	1
10					171	0-8	47.15	0	252	0-9	61.14	0
11					152	0, 4-8	43.57	0	333	0	55.79	0
12					292	0-1, 6-7	30.94	0	269	0-1, 3-6	52.72	0
13					203	0-7	30.00	0	361	0-8	47.62	0
14					271	0, 3-6	26.6	0	289	0-10	36.82	0
15					158	0, 3-4	22.93	0	158	0-6, 8	33.40	0
16					361	0-9	18.00	0	154	0-8	21.36	0
17					22	0-8, 10-11	17.94	0	175	0	20.81	0
18					289	0, 2-7	14.87	0	209	0-10	18.48	0
19					192	0-12	14.45	0	272	2-3, 7, 9-11	16.73	2
20					210	0, 2-7	13.89	0	300	1, 8, 9	14.69	1
21					154	0-15	12.82	0	204	1-6	13.94	1
22					369	0, 2-5	9.23	0	192	0-11	12.52	0
23					181	0-10	8.28	0	157	2-3	12.03	2
24					155	0-9	7.03	0	261	1-10	11.18	2
25					273	1-2, 6-7	6.93	1	359	2-5, 7-10	9.23	3
26					252	0, 5-8	5.93	0	155	0-8	9.12	1
27					242	0	5.62	0	281	1-10	7.92	1
28					322	2-12	3.84	2	153	1-6, 8-9	7.37	1
29					261	3-9	2.69	3	202	3, 5, 8-9	6.19	3
30					300	3-4	2.01	3	181	0-10	6.00	0
31					204	4-5	1.62	4	273	1-9	5.54	1
32					323	5-7	1.56	5	203	1-9	5.35	1
33					153	5	1.52	5	173	1-2	4.61	1
34					281	4-8	1.51	4	171	1	4.14	1
35					173	4-7	1.44	4	371	3-6	4.08	3
36					174	4, 7	1.36	4	172	4-6	3.25	4
37					209	3-5	1.31	3	369	1-8	3.18	1
38					351	12	1.14	12	210	1-10	2.91	1
39									242	3-10	2.89	2
40									241	3-4, 7	2.83	3

Out of the 33 industries where both formal small and informal enterprises show significant clustering relationships and excluding those industries in which the number of enterprises is less than 30, seven are significant for a continuous range of distance of at least 0 to 6 km and display M-peaks at the same distance for formal small and informal enterprises. Figure 4.4 plots distance ranges and M-peak values by industry.[13] Out of these seven industries, the strength of clustering is larger for informal enterprises than for small formal enterprises in five industries (printing, bags, furniture, meat and mill products), while the opposite is true for two industries (apparel and footwear).

[13] All M-peak values occur at 0 km.

Figure 4.4. Intra-industry M-functions, selected industries



Source: Table 4.4.

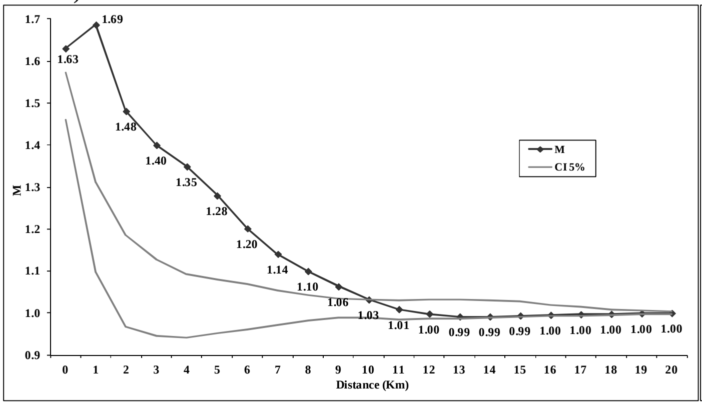
4.3.5.2 Co-clustering

By type of enterprise

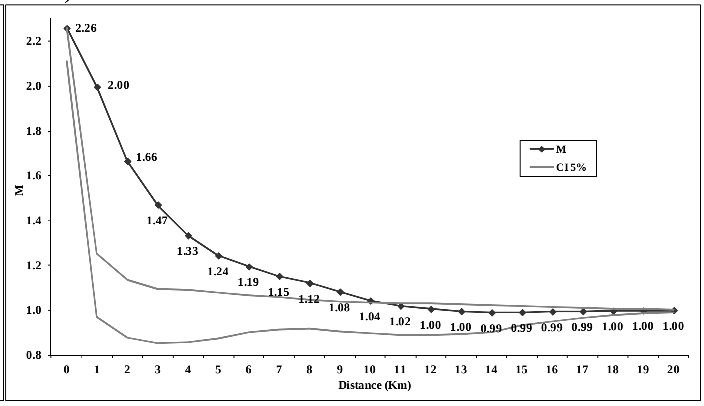
Figure 4.5 shows the inter-industry M-functions by type of enterprise. As suggested by the theoretical predictions, the locational patterns of larger enterprises are quite independent from those of smaller enterprises: formal medium and large enterprises seem to be dispersed with respect to informal and formal small enterprises at distances between 0 and 11 km. However, while informal enterprises fall into the area of random allocation with respect to formal medium-large enterprises, formal small enterprises do co-locate around formal medium and large enterprises. On the contrary, both formal and informal enterprises of similar size seem to seek proximity to each other: informal enterprises co-cluster around formal small enterprises (and vice versa) between 0 and 9 km, but the tendency of informal enterprises to “follow” formal enterprises of similar size seems to be stronger.

Figure 4.5. Inter-industry M-functions by type of enterprise

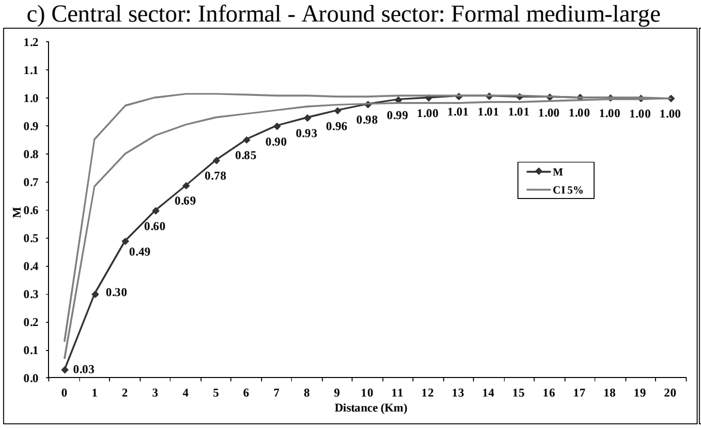
a) Central sector: Informal - Around sector: Formal small



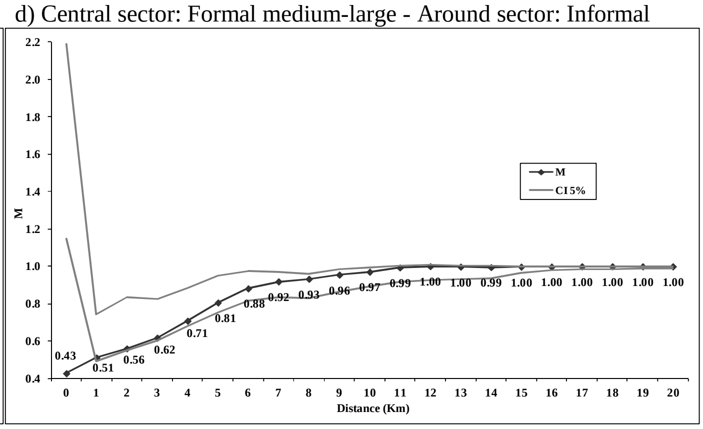
b) Central sector: Formal small - Around sector: Informal



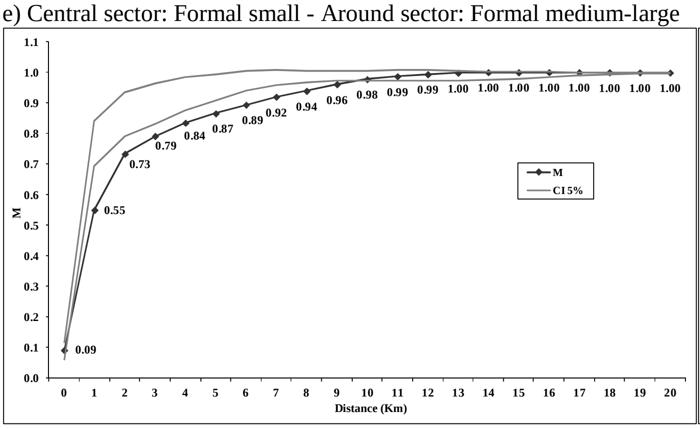
c) Central sector: Informal - Around sector: Formal medium-large



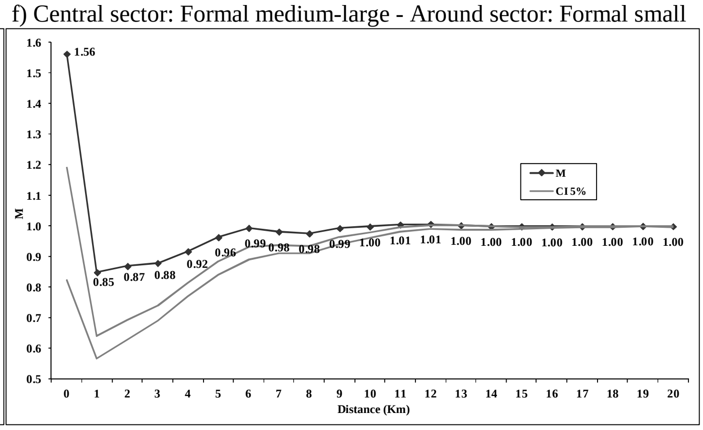
d) Central sector: Formal medium-large - Around sector: Informal



e) Central sector: Formal small - Around sector: Formal medium-large



f) Central sector: Formal medium-large - Around sector: Formal small



By industry

With regard to the co-clustering phenomenon by industry, we find 2410 pairs of sectors significantly agglomerated at different distance ranges. Out of the 2410 pairs of sectors, there are 16 significant co-clustering relationships at different distances ranges, i.e. where the enterprises belonging to a sector are around enterprises of another sector and vice versa, and display the maximum M-peaks at the same distance. In turn, out of these 16, only four significant co-clustering relationships remain once we exclude those industries in which the number of enterprises is less than 30. The most important co-clustering relationship is found for enterprises of the other wood products industry locating around enterprises of the wood for buildings industry and vice versa, and bags enterprises locating around footwear enterprises (see Table 4.5). 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 cases of footwear-furniture and bags-footwear, and wood in the case of other wood products-wood for building), type of labor and/or more general reasons (such as proximity to large consumer markets in the cases of apparel-watches).

Table 4.5. Inter-industry concentration, selected industries

Central industry	Around industry	Distance range (Km)	M-peak	M-peak distance (Km)	Input-output linkage[a]	Correlation type of labor[b]
Other wood products	Wood for buildings	0-9	6.63	0	NA	0.975
Wood for buildings	Other wood products	0-8	5.83	0		
Bags	Footwear	0-10	3.85	0	NA	0.975
Footwear	Bags	0-6	2.85	0		
Apparel	Watches and clocks	0-11	3.80	0	-0.070	0.807
Watches and clocks	Apparel	0-9	1.64	0		
Furniture	Footwear	0-8	2.75	0	0.115	0.989
Footwear	Furniture	0-7	2.18	0		

[a] Input-output relationship are measured as correlation between Input-Output coefficients. We use the Input-Output Accounts published by DANE for 2005 for Colombia valued in constant prices.
[b] The correlation for type of labor were built based on data from the Continuous Household Survey carried out by DANE for the metropolitan area of Cali, 2005 on nine occupational categories.

By type of enterprise and industry

Given that the resulting number of results by enterprise type and industry is indeed very large (up to 18,336), for expositional purposes part of the descriptive analysis of results of inter-industry clustering, which have been calculated using the full sample, is limited to seven industries. The selection of these industries was based on two criteria: 1) significant clustering (as measured by the M-function) is observed; 2) there are a significant number of enterprises and employees in the industry after the sample is split by type of enterprise (see Table A4.3 in the Appendix). Applying these conditions, the seven industries are: mill products (154), apparel (181), footwear (192), bags (193), printing/editing (22), plastic products (252) and furniture (361). These industries give a representative view of the manufacturing sector in Cali, as they account for around 40% of the total employment and 56% of enterprises. These seven industries also include considerable variation in terms of the type of labor they use, their land use and their subcontracting intensity, providing enough variation to explore industry and type of enterprise variation in inter-industry clustering patterns in a descriptive way.

The aggregate results show that, out of the 18,336 total possible cases of co-clustering relationships, 1069 are significantly co-clustering at different distance ranges and display the maximum M-peaks at the same distance. With regard to our seven selected industries, Table 4.6 shows the results for significant co-clustering relationships for a continuous range of distance of at least 0 to 6 km. Disentangling co-clustering relationships by type of enterprise reveals some interesting results. Firstly, all of the significant co-clustering relationships occur between (formal and informal) small enterprises.[14]

Table 4.6. Inter-industry concentration by enterprise type, selected industries

Central industry	Around industry	Distance range (Km)	M-peak	M-peak distance (Km)	Input-output linkage	Correlation type of labor
Printing/Ed.-Formal small	Printing/Ed.-Informal	0-11	19.58	0	Same Industry	
Printing/Ed.-Informal	Printing/Ed.- Formal small	0-12	15.17	0	Same Industry	
Footwear-Formal small	Footwear-Informal	0-12	12.94	0	Same Industry	
Footwear-Informal	Footwear-Formal small	0-11	7.36	0	Same Industry	
Mill products-Formal small	Mill products-Informal	0-10	9.14	0	Same Industry	
Mill products-Informal	Mill products- Formal small	0-7	5.22	0	Same Industry	
Footwear-Formal small	Bags-Formal small	0-6	16.81	0	NA	0.989
Bags-Formal small	Footwear-Formal small	0-12	8.39	0		
Furniture-Informal	Footwear-Informal	0-7	12.06	0	0.115	0.975
Footwear-Informal	Furniture-Informal	0-9	7.72	0		
Apparel-Formal small	Footwear-Formal small	0-11	3.12	0	-0.004	0.991
Footwear-Formal small	Apparel-Formal small	0-10	3.02	0		
Apparel-Formal small	Bags-Formal small	0-7	7.22	0	-0.004	0.998
Bags-Formal small	Apparel-Formal small	0-11	5.39	0		

Secondly, for some industries (printing/editing, footwear and mill products) the observed clustering of informal and formal enterprises of similar size belonging to the same industries (see Section 4.3.5.1) seems to occur in the same areas of the city. This is deducted from the presence of significant co-clustering of formal small enterprises and informal enterprises in the same industry. The most important co-agglomeration is found for informal enterprises in the Printing/Editing industry which are established around formal small enterprises from the same industry: the density of informal enterprises around formal small enterprises is about 19 times larger than that of the whole area. This pattern, is not however observed for all our seven selected industries

[14] 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.

displaying significant clustering (i.e. for the case of furniture, apparel, plastic products and bags industries), which indicates that for those industries formal small and informal enterprises may cluster in different parts of the city. These cases will be analyzed in depth in Section 4.3.6.

Thirdly, informal, and not formal small enterprises, account for the significant industry-level co-clustering relationship found between footwear and furniture. In contrast, formal small enterprises account for the co-location between bags and footwear. In fact the other two significant co-clustering relationships between industries that share a common type of labor (apparel-footwear and apparel-bags) are between formal small enterprises.

We can see that the relationship of location between informal and formal small, or vice versa, accounts for industry-level co-clustering in some sectors. In order to corroborate this relationship on a more general level, we performed a regression analysis where we related the inter-industry concentration levels with the inter-industry concentration levels by enterprise type. We also included some proxies of the Marshallian forces that lead enterprises to locate near each other. Some statistics of these variables are show in Table A4.4 in the Appendix.

To evaluate the proximity to customers and suppliers we used the Input-Output Accounts in 2005 for the whole economy valued in constant prices published by DANE. The input-output tables have information at a 3-digits level. Using this information we calculated Input-Output coefficients which show the relationship production technique between input and output. Hence we defined $linkageIO_{S_1S_2}$ as an input-output linkage of sector S_1 with sector S_2, and $linkageIO_{S_2S_1}$ as an input-output linkage that goes from S_2 to S_1. To construct a single proxy for the connection in inputs-outputs between a pair of industries, we defined a unidirectional version of Input-Output coefficients by $linkageIO=\max\{linkageIO_{S_1S_2}, linkageIO_{S_2S_1}\}$ (Ellison *et al.*, 2010).

With regard to the extent to which sectors use similar types of labor, which measures the importance of labor market pooling generated from the co-agglomeration, we calculated the level of industrial employment by occupations and analyzed the similarity of employment between sectors. We used the information from the Continuous Household Survey (CHS) carried out by DANE for the metropolitan area of Cali for the year 2005. This database provides industry level employment in 99 occupations. Since our study area is a metropolitan area we find that in most sectors there will of course be very few employees, even zero, in many occupations. In this regard we decided to group this variable into nine categories according to qualification level and type of occupation (professional and similar, sales, manual and services).[15] Based on this information we estimated the share of industry S_1's (S_2's) employment in occupation *o*, i.e. $Share_{S_1o}$ and $Share_{S_2o}$ respectively. In order to measure the similarity of employments between industry S_1 and S_2, we calculated the correlation between $Share_{S_1o}$ and $Share_{S_2o}$.

As mentioned above, the values of the dependent variable for each sector *S* is that obtained from the function $M_{S_1S_2}$ and $M_{S_2S_1}$ as explained in the Section 4.3.3.1.2 for a specific radius. Turning to the independent variables of the inter-industry concentration levels by enterprise type, we included by sector the possible combinations between formal medium and large, formal small and informal enterprises. In order to construct a single measure for the inter-industry and inter-industry/inter-enterprise agglomeration across an industry pair, we calculated a unidirectional version of this index in a manner analogous to our construction of *linkageIO*.[16]

An overview of the M-functions (inter-industry and inter-industry/inter-enterprise) shows that the most important degree of co-location is observed at very small radius values. Consequently, it seems reasonable to analyze the determinants of co-agglomeration at very low distance. We therefore take the M-functions for two radii: less than one and five kilometers. We estimated the proposed model for each of these radii.

In Table 4.7 we report the OLS estimations for each of the radii. In order to make it easier to interpret the results, we normalize all variables so that they have a standard deviation of one. Beginning with the proxies of the Marshallian forces, we find that the coefficients of input-output linkage and labor correlation are positive and statistically significant, although only at less than one kilometer. This may indicate that only at a very short distance the reduction in the costs due to input-output linkage and

[15] We have the following occupational categories: 1) professionals; 2) managers and owners of enterprise 3) qualified white-collar workers; 4) clerk; 5) sales employees; 6) qualified manual workers; 7) non-qualified manual workers; and 8) qualified service workers; and 9) non-qualified service workers. We also tested using different groups and the results obtained are qualitatively similar.

[16] We have six combinations of inter-enterprise co-agglomeration: Formal medium-large–Formal medium-large; Formal small–Formal small; Informal–Informal; Formal medium-large–Formal small; Formal medium-large–Informal; and Formal small–Informal. In this case we are interested in co-agglomeration between types of enterprises regardless of the type of enterprise reference and those locate around (central industry and around industry). Note that we only include sectors which have a certain degree of inter-industry concentration level in each of six combinations of enterprise type.

the access to a concentrated labor market have important positive effects on industrial co-agglomeration. These findings are broadly in line with those found by Ellison *et al.* (2010).

The estimates of the inter-industry concentration levels by enterprise type show that co-agglomeration of formal small and informal enterprises, and between them, are the main determinants of inter-industry co-agglomeration at a very short distance. These results suggest that there are more general factors for co-location of this type of enterprises that affect inter-industry co-agglomeration. For instance, it may be due to formal small enterprises seeking to locate in areas of dense consumer markets, where informal manufacturing activity prevails; or that the informal enterprises locating close to formal activity are seeking to capture consumers who have not been absorbed by the formal small sector.

Table 4.7. Results of the OLS estimates
(dependent variable: inter-industry concentration)

	<1Km	<5Km
Co-agglomeration between:		
Formal medium-large – Formal medium-large	0.0098	-0.0021
	(0.94)	(-0.26)
Formal small – Formal small	0.1650***	0.4238***
	(11.07)	(6.71)
Informal – Informal	0.0793***	0.0458
	(5.54)	(0.39)
Formal medium-large – Formal small	0.0221***	0.3644***
	(3.73)	(7.43)
Formal small – Informal	0.1074***	0.2571***
	(8.99)	(2.60)
Formal medium-large – Informal	0.0083	0.1059***
	(1.35)	(2.90)
LinkageIO	0.0511*	0.0038
	(1.81)	(0.65)
Labor correlation	0.1054**	0.0026
	(2.90)	(0.36)
Constant	-0.1569***	-0.0205***
	(-4.28)	(-2.59)
N	253	253
R2	0.5673	0.3806

Note: ***, **, *, denotes significance at 1%, 5% and 10%, respectively. ( ) t statistics.

We also find that only when formal medium and large enterprises are co-agglomerated with small enterprises (formal or informal) do they have a positive and significant effect on industrial co-agglomeration, although the co-agglomeration with informal enterprises is only statistically significant when we move to a radius of less than five kilometers. These results show the location effect that exists between formal medium and large and small enterprises, on the one hand, and on the other the limited effect that the concentration of formal medium and large enterprises has, given that there are very few enterprises of this type. Formal small enterprises are located very close to formal medium and large enterprises, and their concentration generates dense consumer markets that can contribute to industrial co-agglomeration. For their part, informal enterprises are far from formal medium and large enterprises, so that when we move to higher radii there are more possibilities of accessing a greater number of consumers (large consumer markets to formal medium and large enterprises and more central markets to informal enterprises) or transportations routes that can help industrial concentration.

4.3.6 Results of spatial analysis

As indicated in Section 4.3.5, informal and formal small enterprises display clustering and also significant co-clustering in the printing/editing, footwear and mill products industries. Analyzing the kernel density output in the printing/editing industry reveals that formal small and informal enterprises do cluster in the same areas of the city: there is significant overlap between the red spots indicating the highest density of enterprises for both types of enterprises (see panel a) of Figure 4.6). The kernel density analysis of the remaining two industries for which clustering and co-clustering were significant tells a similar story (see Figure A4.3 in the Appendix). These cases can be contrasted with those of the four industries for which both formal small and informal enterprises displayed significant clustering but no co-clustering of the two types of enterprises was found (apparel, bags, plastic products and furniture industries). In particular, for the case of apparel (see panel b) of Figure 4.6), 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 these two types of enterprises (see Figure A4.4 in the Appendix for the cases of furniture, plastic and bags).[17] Thus the degree of spatial segmentation seems to be industry-specific and may be explained by factors such as intensity of subcontracting in the industry. The apparel industry is characterized by vertically disintegrated value chains which often make use of home-based, cheaper informal labor. The specialization of informal enterprises as subcontractors in this

[17] Arguably the presence of spatial segmentation is not as clear-cut for these industries as it is for the case of apparel. A future refinement to this analysis would be to incorporate a precise measure of spatial segmentation of economic activities based, for instance, on urban segregation measures such as dissimilarity, exposure and isolation (see Feitosa *et al.*, 2007).

industry, as opposed to competitors in final markets, may explain the different spatial pattern observed for this industry.

Figure 4.6. Kernel density mapping, selected industries

a) Printing/Editing **b) Apparel**

Formal small Informal Formal small Informal

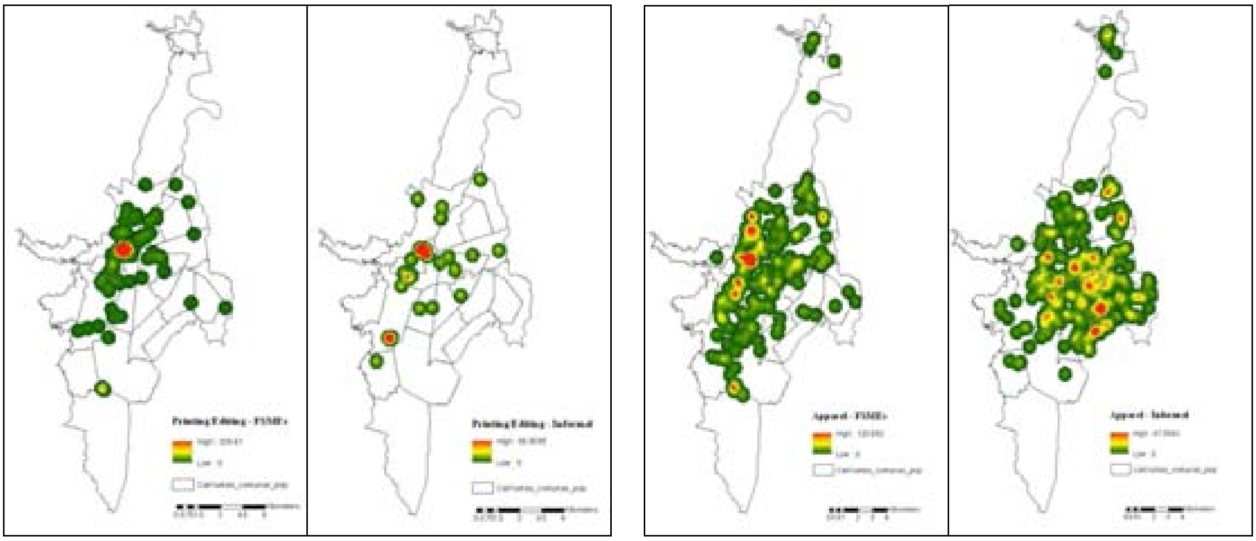
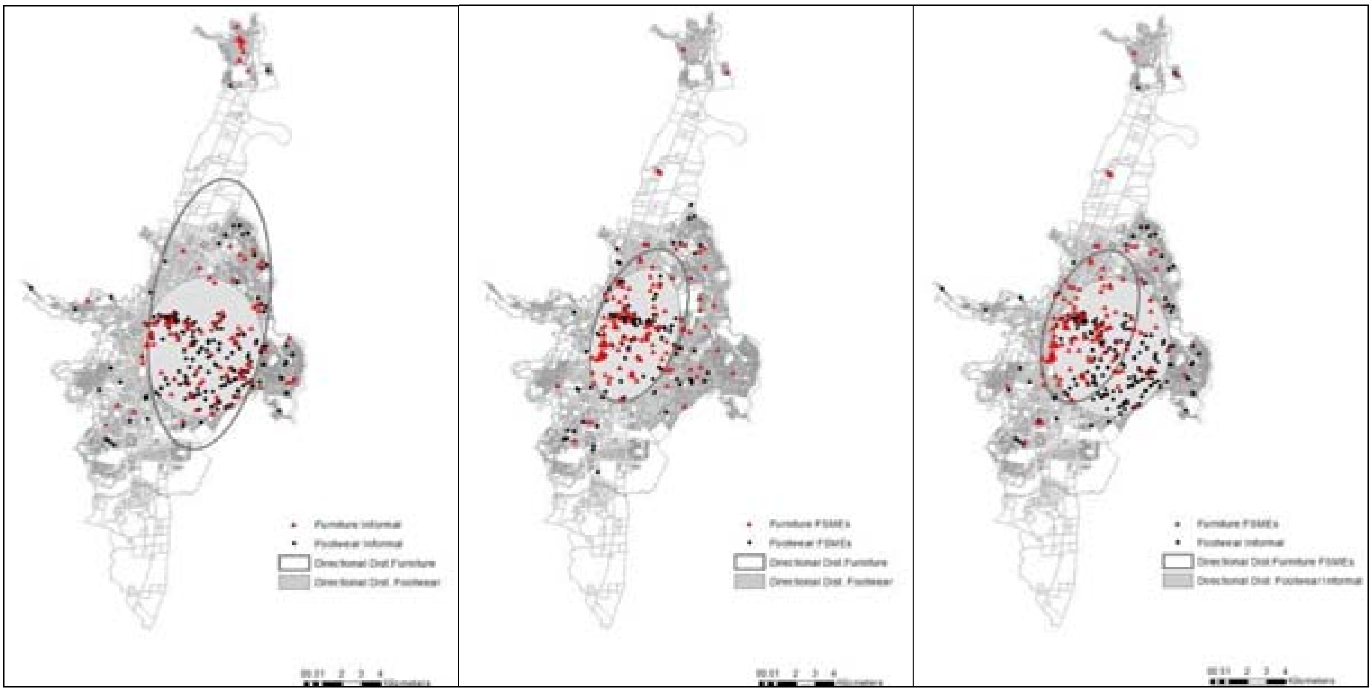


Figure 4.7 displays a panel of three maps. The first one show the spatial distribution of informal enterprises in the footwear and furniture industries, the second one that of formal small enterprises and the third one that of formal small enterprises in the footwear industry and informal enterprises 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 informal enterprises, and is perhaps driven by a noticeable cluster of such enterprises in the north or the municipality of Yumbo. Furthermore, the lack of co-clustering between formal and informal enterprises in these industries can be explained by the fact that a considerable number of formal small footwear enterprises locate along the main axis of the city, as indicated by the directional distribution of formal small enterprises, while most informal enterprises in the furniture industry locate to the east of Cali and in Yumbo.

Figure 4.7. Spatial distribution of furniture and footwear enterprises

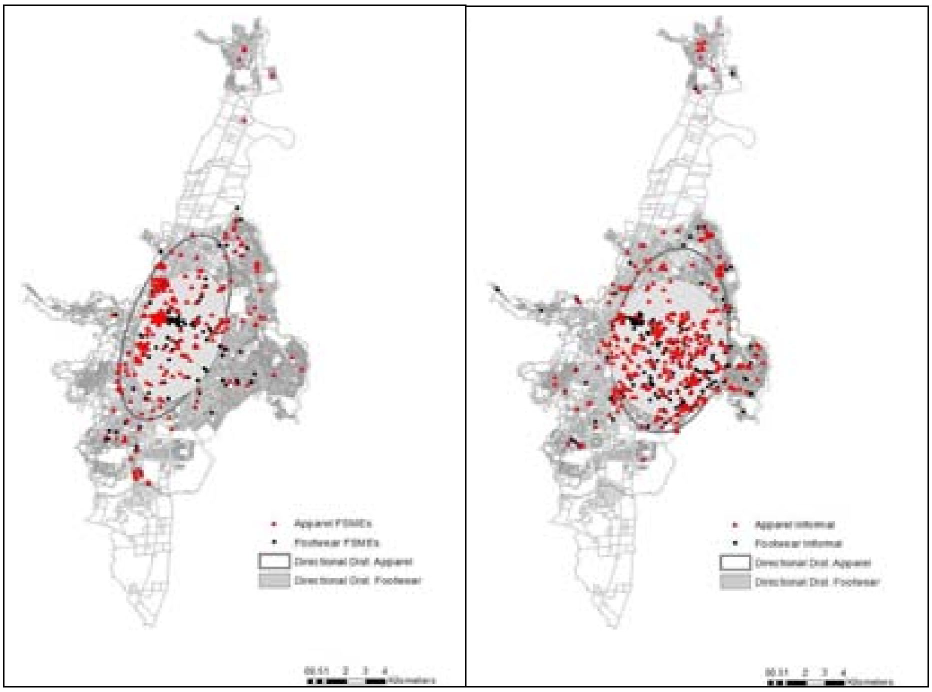
Informal | Formal small | Informal and formal small



Lastly, as the results of the inter-industry M-functions indicated, there is co-clustering of enterprises in the apparel and footwear industries, which is explained by co-clustering of formal small apparel around formal small footwear enterprises. Thus, while there is an evident spatial complementarity between these two industries, it is only significant for enterprises of the same type, in particular for formal small enterprises. As the standard deviational ellipses in Figure 4.8 illustrate, the spatial distribution of formal small and informal enterprises in these industries occurs in different parts of the city for each type of enterprise. In this case, spatial segmentation between informal and formal small enterprises seems to impede significant co-clustering relationships between formal and informal enterprises in these related industries.

Figure 4.8. Spatial distribution of enterprises of the Apparel and Footwear industries

Formal small | Informal



4.4 Agglomeration and formal and informal wages

In this section we complement the results of the previous section by attempting to clarify the effects of agglomeration economies on formal and informal productivity, and if there are such effects, which sector achieves greater benefits. We begin by regressing individual worker wage rate as a measurement of labor productivity on employment density separately for the formal and informal wage sectors. As in the previous section, we define informal workers as those workers employed in enterprises not registered with the Chamber of Commerce. The wage equation employed for the estimate has the following structure:

$$\ln w_{i(t)} = \alpha_0 + \beta \ln density_{mt} + X_{i(t)}\varphi + \rho Y_{i(t)} + \delta_t + \sigma_{si(t)} + \eta_{ri(t)} + \varepsilon_{i(t)} \tag{4}$$

where *i* identifies the worker, *m* represents the municipalities, *s* refers to the economic sector (two-digit level), *t* specifies the time period, and *r* identifies the region. The "*i(t)*" subscripts indicate that the observations are an independent cross-sectional series where N individuals are only available in each period. The dependent variable is the logarithm of the real hourly wage which has been adjusted for the price level using the consumer price index of Colombia for 1998 as a deflator.

Our measurement of urban agglomeration is the logarithm of the employment density of the municipalities, ln*density*$_{i(t)}$, which is defined as the number of people employed per square kilometer (the person sampling weight available in the databases is used to ensure a representative sample in each the municipalities). The basic idea behind this variable is that a high employment density is a potential source of increasing returns resulting from stronger knowledge and technological spillovers in areas of dense economic activity. We use the municipality as the spatial unit of analysis, and although this is not an ideal unit, it is the best available approximation of a self-contained labor market in Colombia. The municipalities are areas where a high proportion of people who live (work) in the area also work (live). As Dominicis *et al.* (2007) argues, if there is evidence of a concentration of residential activities, of work activities as well as of those social relationships that are created within it, this area can be considered as a self-contained labor market or a Local Labor System.[18]

The vector $X_{i(t)}$ contains the variables that measure a standard set of demographic attributes such as the worker's level education, experience and its square, and gender. We also included variables for the size of the firm in which the workers work ($Y_{i(t)}$), and we included sets of dummy variables with different number of workers.

In addition, in our model we included sets of dummy variables to control for several sources of heterogeneity that can lead to an omitted variable bias and inconsistency of the model parameter estimates. In order to capture macro level changes in wage rates that are common to all individuals, we include time dummies, δ_t. Similarly, to control for economic sector heterogeneity we added a set of dummy variables at the 2-digit industry level. We also incorporated regional fixed effects, $\eta_{ri(t)}$. Regarding the latter variable we included dummy variables for the region instead of metropolitan areas to avoid problems of multicollinearity. There are metropolitan areas with a single municipality (for example Bogotá, Cartagena and Pasto) with which there may be multicollinearity between variables measure only at municipality level (*density* and *diversity*) and the metropolitan area dummies.

The data used in this part come from the Continuous Household Survey (CHS), a micro database level cross-section survey on labor force, unemployment and informality in Colombia carried out by the National Department of Statistics (DANE).

[18] As Openshaw and Taylor (1979) have pointed out, the municipalities or metropolitan areas are much more related to the concept of local labor markets than the usual administrative areas, so they are a good option for overcoming the Modifiable Areal Unit Problem (MAUP)

The CHS is statistically representative at the level of the 33 municipalities which form the 13 main metropolitan areas in Colombia.[19] We analyzed the period between 2002 and 2006. Databases for other years are not comparable because several methodological changes were carried out by DANE in 2000 and 2007.[20]

One aspect to be considered in the estimation is the endogeneity bias caused by reverse causality between wages and agglomeration. Wages can increase due to higher employment density, but it is also possible that higher wages may attract more workers and firms to a given area. In order to avoid the endogeneity bias we implemented instrumental variable (IV) techniques. In the literature long-lagged values of endogenous variables have been widely used as instruments since Ciccone and Hall's (1996) pioneering work. The basic idea behind these instrument variables is that deep time lags of urban density can to some extent explain the distribution of present densities, but does not explain the distribution of current urban productivity levels. Unfortunately, in Colombia there is no data for population in the nineteenth century at a municipal level, thus long-lagged values of urban density cannot be obtained. Hence we prefer to follow an alternative strategy, which will be explained in the next section.

4.4.1 Description of the instruments

An alternative instrumentation strategy is to use the factors that have affected the process of urban development in Colombia which can be related to the rural-urban exoduses caused by internal armed conflict (Aprile-Gniset and Mosquera, 1978; Aprile-Gniset, 1992).[21] For more than four decades, Colombia has been involved in an armed conflict between the political parties (Conservative and Liberal), and this, plus growing violence from leftist guerrillas (such as the FARC and ELN), and the establishment of paramilitary groups and drug trafficking, have been the main reasons for violence in

[19] 1) *Medellín metropolitan area (MA)*: Medellín, Barbosa, Bello, Caldas, Copacabana, Envigado, Girardota, Itagüí, La Estrella, and Sabaneta. 2) *Barranquilla MA*: Barranquilla and Soledad. 3) *Cartagena MA*: Cartagena. 4) *Manizales MA*: Manizales and Villamaría. 5) *Montería MA*: Montería. 6) *Villavicencio MA*: Villavicencio. 7) *Pasto MA*: Pasto. 8) *Cúcuta MA*: Cúcuta, El Zulia, Los Patios, and Villa del Rosario. 9) *Pereira MA*: Pereira, Dos Quebradas, and La Virginia. 10) *Bucaramanga MA*: Bucaramanga, Floridablanca, Girón, and Piedecuesta. 11) *Ibagué MA*: Ibagué. 12) *Cali MA*: Cali and Yumbo. 13) Bogotá D.C. According to the 2005 Population census, about 60% of the urban population was located in these 33 cities and their metropolitan areas. As regards the regions have been built according to the Regional Council of Economic and Social Planning (CORPES) sub-regions: *Central region*: Bogotá, Manizales MA, Pereira MA, Ibague, Medellin MA; *Eastern region*: Cúcuta MA, Bucaramanga MA, Villavicencio; *Western region*: Cali MA, Pasto; *Atlantic region*: Barranquilla MA, Cartagena, Monteria.

[20] The 2001 survey is excluded from the analysis because of lack of information about the informal sector.

[21] A more detailed explanation of conflict and emigration in Colombia can be found in Sánchez (2008).

Colombia. The civilian population has been the main victim of the conflict. The worst affected by the confrontation tend to be displaced civilians, killed or expropriated by the organized criminal groups. According to government records, the number of internally displaced people caused by some type of violence associated with internal armed conflict or violation of human rights is 3.6 million[22] (about 850,000 households), affecting 69% of the municipalities in Colombia (773 out of 1119). Further, between 1980 and 2010 at least 6 million hectares were taken from their owners by using some type of violence associated with the armed conflict. The land taken represents about 15% of agriculture land in Colombia (CSPPDF, 2010 and 2011; IDMC, 2011; CODHES, 2011).

The phenomenon of forced displacement has not only affected the main cities but also the entire urban system. As can be seen in Table 4.8, about 17% of the forced migration is located in the three main cities of Colombia: Bogotá, Medellín and Cali. According to a relative index the first two municipalities have a 4 percent higher number of displaced people than the national average. Turning to the displacement rate we can see that 6 out of the 13 main cities are above the national rate: Cartagena, Barranquilla and Ibagué have more than 15,000 displaced people per 100,000 inhabitants.

Hence the hypothesis is that a large part of the rural-urban migration in Colombia has not been driven by current urban productivity levels, but rather due to a forced displacement caused by the violence. In this regard, forced displacement can to some extent explain of population settlements in Colombia. Thus we used the relative index of displaced people that arrived in each of the municipalities in each year between 2002 and 2006 as an instrumental variable. This index facilitates comparison with the national level and allows us to account for changes in the phenomenon over time (see Table 4.8).

[22] According to the non-government observer Consultancy for Human Rights and Displacement (CODHES) the total number of internally displaced people is as high as 5.2 million, and accounts for over 11% of the country’s population. The Internal Displacement Monitoring Centre IDMC (2011) reveals that Colombia is the country with the largest number of internally displaced people and is the third with the highest proportion of its population internally displaced.

Table 4.8. Forced Displacement by municipality of arrive in Colombia, 1998 – 2011[a]
(13 main municipalities)

	Number of displaced people	%	Displacement Rate (per 100,000 inhabitants)	Relative Index[b]
Bogotá	313,315	8.59	13,414.9	4.54
Medellín	204,805	5.61	17,217.7	4.12
Cali	82,343	2.26	1105.0	3.21
Villavicencio	73,355	2.01	8049.1	3.09
Cartagena	68,073	1.87	18,759.0	3.01
Barranquilla	57,973	1.59	18,163.8	2.85
Cúcuta	56,763	1.56	13,564.1	2.83
Ibagué	53,379	1.46	15,519.9	2.77
Bucaramanga	42,304	1.16	7009.3	2.54
Pasto	38,403	1.05	9945.7	2.44
Montería	34,599	0.95	6674.6	2.34
Pereira	30,990	0.85	6173.8	2.23
Manizales	10,005	0.27	448.0	1.10
Colombia	**3,649,294**	**100**	**10,461.4**	

Source: Department for Social Prosperity, Colombia.
[a] Official statistics on forced displacement per municipality is only available since 1998.
[b] By relative we mean that variable is expressed (in logarithm) as ratio to the national average.

4.4.2 Instrument relevance and exogeneity

In order to yield unbiased estimates in the estimation of the effect of employment density on wage by using instrumental variables, our instrument must satisfy two conditions for it to be valid. These conditions are: relevance and exogeneity. We shall begin by discussing the first condition.

The first condition demands that our instrument be correlated with the contemporaneous employment density: $Cov(\ln density,\ Forced\ displacement) \neq 0$. As argued above, forced displacement due to the violence is a phenomenon that has affected a significant percentage of the population. The displaced population, mainly a rural population, arrives in the cities in very poor conditions, which forcing these people to participate in the local labor markets and generating a pressure in the labor supply. However this effect is not immediate, since the displaced people exhibit a number of adverse socio-demographic factors[23] that make access to the labor market difficult, and

[23] The households of displaced people are composed of 4.9 people on average; 47% of displaced people are female heads of households; there is a high presence of children and adolescents: 56% of people under 20 years of age; 20% of displaced heads of households are illiterate, and their average number of years in education is 4.6 (CSPPDF, 2010).

where they succeeding in finding a job, this will probably be informal (the rate of informality is as high as 96% for displaced workers (CSPPDF, 2010)).

In order to show some correlation between our instrument and employment density we calculated the partial correlation and the R-squared of the corresponding univariate regression. The results show that the correlation is 0.44 and the R-squared is 0.20. Although these values are not very high, it shows that our instrument has an appreciable explanatory power. For instance, Combes *et al*. (2010) used the geological characteristics as instruments and found R-squareds ranging between 1% and 13%. Nevertheless, as they themselves say, the relevance of an instrument depends on the partial correlation of the instrumental variables and the endogenous regressor. We show the results of the partial correlations and other instrument diagnostic tests in the next section.

The second condition requires that our instrument be uncorrelated with the error term: $Cov(\varepsilon, \textit{Forced displacement})=0$. As mentioned, forced displacement is a result of internal armed conflict and does not depend on the local characteristics of the place of arrival, it is an exogenous shock: families migrate to save their lives and not to improve economic conditions, they migrate to one particluar city and not to another because of its proximity and not because of the economic conditions in that city (Calderón *et al.* 2011). Thus forced displacement could affect productivity only indirectly, as a result of its effect on the supply of labor, i.e. on the spatial distribution of population, but not directly.

4.4.3 Estimation procedure

We estimated the wage equations taking into account for within-municipality dependence, which mainly affects the standard errors of regression parameter estimates. A common correction is to calculate cluster-robust standard errors which permit within-cluster error correlation. However, this correction presumes that the number of cluster is large. As pointed out by Rogers (1993) and Cameron *et al*., (2008), when there are few (5-50) clusters, standard asymptotic tests can over-reject the null hypothesis considerably because the cluster-robust standard errors will be downwardly biased.

Recent work has highlighted the use of different techniques to improve inference when there is within-cluster dependence. One such technique is the use of the bootstrap. In OLS estimations Cameron *et al*. (2008) show that a variant of the wild bootstrap with cluster-based sampling performs well in a variety of cases, and bootstrap techniques dominate any cluster standard error estimation technique. In the case of linear regression models estimated by instrumental variables, Davidson and MacKinnon (2010) propose a wild bootstrap procedure which is applied to *t* tests, the Aderson-Rubin test of weak-instrument-robust inference (Anderson and Rubin, 1949) and the Kleibergen test (Kleibergen, 2002).

In this work we use the wild restricted efficient residual bootstrap (WRE bootstrap) algorithm proposed by Cameron *et al*. (2008), and Davidson and MacKinnon (2010). We use this algorithm to calculate the *P* values of a *t* test of regression parameter estimates, and instrument diagnostic tests such as the Aderson-Rubin test of weak-instrument-robust inference and the Kleibergen-Paap test of under-identification (Kleibergen and Paap, 2006). We briefly outline how the WRE bootstrap works in the case of linear regression models estimated by IV.[24]

As usual the IV model has the following structure of two equations, in matrix notation:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{W}\boldsymbol{\gamma} + \mathbf{u}_1 \tag{5}$$

$$\mathbf{X} = \mathbf{Z}\boldsymbol{\pi}_z + \mathbf{W}\boldsymbol{\pi}_w + \mathbf{u}_2, \tag{6}$$

where $\mathbf{y}$ is an $n\times 1$ vector of n observations on an endogenous variable, $\mathbf{X}$ is an $n\times k_x$ matrix of observations on endogenous variables, $\mathbf{W}$ is an $n\times k_w$ matrix of observations on included instruments, and $\mathbf{Z}$ is an $n\times k_z$ matrix of excluded instruments. Equation (5) is a structural equation, and equation (6) is a reduced-form equation. So, the WRE bootstrap works in the following way:

1. Under the null hypothesis $H_0\colon \beta = 0$, run the restrict structural equation and obtain the restricted OLS estimator $\hat{\boldsymbol{\gamma}}^R$ and the associated restricted OLS residuals $\hat{\mathbf{u}}_1^R$. Also run the reduced-form equation and obtain the OLS estimator $\hat{\boldsymbol{\pi}}_z$, $\hat{\boldsymbol{\pi}}_w$ and the associated residuals $\hat{\mathbf{u}}_2$.
2. Form a sample of G clusters $\left\{(\hat{\mathbf{y}}_1^*, \mathbf{W}_1),\ldots,(\hat{\mathbf{y}}_G^*, \mathbf{W}_G)\right\}$ and $\left\{(\hat{\mathbf{X}}_1^*, \mathbf{Z}_1, \mathbf{W}_1),\ldots,(\hat{\mathbf{X}}_G^*, \mathbf{Z}_G, \mathbf{W}_G)\right\}$ by the following method: for each cluster g = 1,…,G, form either $\hat{\mathbf{u}}_{1g}^{R*} = \omega_1\hat{\mathbf{u}}_{1g}^R$ ($\hat{\mathbf{u}}_{2g}^* = \omega_2\hat{\mathbf{u}}_{2g}$) with probability 0.5 or $\hat{\mathbf{u}}_{1g}^{R*} = -\omega_1\hat{\mathbf{u}}_{1g}^R$ ($\hat{\mathbf{u}}_{2g}^* = -\omega_2\hat{\mathbf{u}}_{2g}$) with probability 0.5, where $\omega_1 = \left(n/(n-k_w)\right)^{1/2}$ and

[24] The algorithm of the WRE bootstrap for linear regression models estimated by OLS can be found in Cameron *et al*. (2008).

$\omega_2 = \left(n/(n-l)\right)^{1/2}$ with $l = k_w + k_z$. Then form $\hat{\mathbf{y}}_g^* = \hat{\mathbf{u}}_{1g}^{R*}$ and $\hat{\mathbf{X}}_g^* = \mathbf{Z}_g \hat{\boldsymbol{\pi}}_z + \mathbf{W}_g \hat{\boldsymbol{\pi}}_w + \hat{\mathbf{u}}_{2g}^*$, $g = 1,\ldots G$.[25]

3. Estimate the IV model using the sample of G cluster creates above and compute t statistic $\tau_b^* = \hat{\beta}_b^* / s(\hat{\beta}_b^*)$, where $s(\hat{\beta}_b^*)$ is the standard error of $\hat{\beta}_b^*$.

4. Repeat steps (2) and (3) B times.

5. Calcule the simulated P value of the original statistic $\hat{\tau}$:

$$\hat{p}^*(\hat{\tau}) = \frac{1}{B}\sum_{b=1}^{B} I\left(\left|\tau_b^*\right| > \left|\hat{\tau}\right|\right),$$

and reject the null hypothesis whenever $\hat{p}^*(\hat{\tau}) < \alpha$.

Our estimations are based on B=20,999. See Davidson and MacKinnon (2000) for more on how to choose B appropriately. Turning to P value of Aderson-Rubin test and Kleibergen-Paap test, we followed the procedure proposed by Davidson and MacKinnon (2010).

4.4.4 Estimation results

In this section we present the results of the estimation of wage equation for the full sample, formal and informal sector. We tested whether the formal equation is significantly different from the informal equation. Using the Chow test, the null hypothesis was rejected at either a 1% or 5% level of significance, with an F statistic of 100.66.

Table 4.9 presents the main results of the wage equations for the full sample, and separated into formal and informal sectors. The coefficients on the demographic attributes (e.g. education, experience, etc.) were consistent with estimates in the labor literature, however they are not discussed or included further, given our focus on agglomeration (see Table A4.5 in the Appendix to full estimation). We began by discussing the instrument diagnostic test reported at the bottom of the table. Since we have a single instrument variable, the tests will mainly help us to determine whether or not our instrument is relevant or, where appropriate, whether it is a weak instrument.

[25] A more detailed explanation can be found in Davidson and MacKinnon (2010).

Table 4.9. Urbanization elasticity
dependent variable: log of real hourly wage
(t-statistics in parentheses)

	Full sample		Formal		Informal	
	OLS	IV	OLS	IV	OLS	IV
Log of employment density	0.0489	0.0564	0.0877	0.1150	0.0412	0.0430
t cluster-robust	(3.85)	(2.45)	(4.81)	(3.30)	(3.30)	(1.81)
Asymptotic *P* values	0.0005	0.0143	0.0000	0.0010	0.0024	0.0701
WRE bootstrap *P* values	0.0055	0.0445	0.0008	0.0319	0.0170	0.0958
Observations	49209	49209	8903	8903	40306	40306
R2	0.32	0.32	0.35	0.35	0.26	0.26
Root MSE	0.8315	0.8312	0.8508	0.8491	0.8141	0.8136
Time, industry and region dummies	yes	yes	yes	yes	yes	yes
Results for first-stage regressions						
Shea partial R2		0.1433		0.1232		0.1491
Under-identification test						
Kleibergen-Paap rk F stat.		8.612		10.242		7.890
Asymptotic *P* values		0.0033		0.0014		0.0045
WRE bootstrap *P* values		0.0344		0.0412		0.0347
Weak-instrument-robust inference						
Anderson-Rubin (AR) F stat.		60.930		33.715		34.206
Asymptotic *P* values		0.0000		0.0000		0.0000
WRE bootstrap *P* values		0.0000		0.0000		0.0000

Standard errors are cluster-robust. Clustering is on municipalities.
Notes: First-stage regression, under-identification and weak identification test statistics cluster-robust.

The first stage regressions results indicate that the instrument for employment density has appreciable explanatory power. The Shea partial R-squared score a value of 0.14 for the first stage regression of employment density for the IV full-sample model, while the Shea partial R-squared for the IV formal and informal model is 0.12 and 0.15, respectively.

With regard to the relevance of the instrument, we performed the Kleibergen-Paap test of under-identification which tests whether the model is identified, where identification requires that the excluded instrument be correlated with the endogenous regressor. When the instrument is uncorrelated with the endogenous regressor, the matrix of reduced-form coefficients is not of full rank and the model will be unidentified. Since we allow intra-group correlation, the relevant test in this case is the Kleibergen and Paap (2006) rank statistics test. In our special case of a single endogenous regressor, this test reduces to a familiar F statistic of significance on the coefficient of the excluded instrument from the OLS estimator of the single reduced-form equation (Baum *et al*., 2007). If we fail to reject the null hypothesis that the matrix of reduced-form coefficients is under-identified, it means that the instrument variable bias of the parameter estimates will be increased. The values presented in Table 4.9

show that the tests reject the null hypothesis of under-identification at a 5 percent level of significance, implying that the instrument is relevant.

However, a rejection result for the null hypothesis in the Kleibergen-Paap test should be treated with caution because weak instrument problems may still be present. Weak identification arises when the instrument is correlated with the endogenous regressor, but only weakly. As pointed out by Murray (2006) and Stock and Yogo (2005) when the instruments are poorly correlated with the endogenous regressors, the estimates from the instrumental variable model will be biased. In this case, and allowing intra-group correlation, the relevant test is the Anderson-Rubin F statistic. This test involved testing the significance of the excluded instrument in the structural equation, i.e. the F statistic for $\boldsymbol{\pi}_2 = 0$ in the regression $\mathbf{y} = \mathbf{W}\boldsymbol{\gamma}_2 + \mathbf{Z}\boldsymbol{\pi}_2 + \mathbf{v}$, which results in the substitution the reduced-form expression for the endogenous regressor **X** in the main equation for the model (Baum *et al*., 2007; Davidson and MacKinnon, 2010). The results in Table 4.9 reveal that the hypothesis is rejected, suggesting that our instrument is not weak.

Consider now the estimates of the impact of agglomeration on wages. In the OLS model for the full sample (the first column), doubling the employment density of a given municipality is associated with an increase in wages of 5.0%. This estimate is highly significant either with the asymptotic *P* value or the WRE bootstrap *P* value. In the IV model (column two) the corresponding estimate is 5.6% and it is significant at 5% with both kinds of *P* values, but less significant with the WRE bootstrap *P* value. These estimates are consistent with the previous evidence for developing countries shown in Section 4.2.1. In case of developed countries, Ciccone (2002) estimates an elasticity of 4.5% in countries such as Germany, Italy, France, Spain and the UK, and for the United States Ciccone and Hall (1996) estimate an elasticity of 5.0%.

The full sample regression assumes that the wage effect associated with agglomeration is the same for both formal and informal workers. However, there are reasons to believe that this might not be true. Informal workers are characterized by having a limited education and working in very small enterprises (Jütting and Laiglesia, 2009; Perry *et al*., 2007) which might have a greater potential for benefiting from the knowledge spillovers and input-output linkages associated with agglomeration (Rosenthal and Strange, 2008). Nevertheless, low education levels can also imply less ability to absorb the knowledge, and the activities of small enterprises are more geared towards small local markets than towards generating input-output linkages, as we saw in Section 4.3.5. On the other hand, formal workers are workers with more education than informal workers and might have a greater ability for learning from nearby human capital. Furthermore, formal workers work in medium-large enterprises (see Section 4.3.5) and this kind of enterprises can achieve greater benefits of labor market pooling and input sharing associated with agglomeration. Hence, it is not clear whether formal or informal workers would benefit more or less from agglomeration.

Because of this ambiguity of the effects of agglomeration on wages of formal versus informal workers, we estimated the model for formal and informal workers. The results in Table 4.9 show that, regardless of the method of estimation, estimates of the effects of agglomeration are larger for formal workers than for informal workers. Based on the IV models, the urban elasticity is 11.5% for formal workers and 4.3% for individuals working in the informal sector, but the latter of these IV estimates is only marginally significant in the case of the WRE bootstrap *P* value (the *P* value is 0.096). These results imply that a doubling of occupation employment density would serve to widen the difference in wages between formal and informal workers by roughly 7.2%.

To summarize, we have presented evidence that wages vary with agglomeration, with the strongest pattern being for formal workers. They earn higher wages in cities with a higher density. With regard to informal workers, their wages also vary positively with regard to density, but weakly, and there is an important difference in returns from agglomeration with respect to formal workers.

4.5. Conclusions

One of the main goals of this chapter was to extend the understanding of the agglomeration effects in the presence of an informal sector, in particular how the informal sector contributes in agglomeration and the benefits achieved by both the formal and informal sector. The first research question was: do formal and informal enterprises display different clustering patterns? As the results show, both types of small enterprises display a strong tendency to cluster in the same industries, and 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 clustering than their formal counterparts of similar size, this is not the case for each individual industry: for some industries, clustering is stronger for formal small than for informal enterprises. Thus, the same logic behind clustering seems to apply in different industries for both formal small and informal enterprises but to different degrees.

The 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 in the same areas? From the co-clustering analysis, it can be concluded that, in general terms, while formal medium and large enterprises display location patterns that seem independent of enterprises of smaller size, formal and informal enterprises of a similar size seek each other’s proximity. However, significant clustering of both formal and informal enterprises of a similar size on the same industry does not necessarily imply that this clustering takes place in the same parts of the city. As the results show, in some cases clustering of both types of enterprises can happen in both the same or different areas 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.

Finally, the third research question was: are there differences in the effects of agglomeration on productivity and wages between the formal and informal sector? As the results show, there are greater returns to the agglomeration of the activities and all those spillovers associated with agglomeration economies in the formal sector than the informal sector. Factors such as constraints to physical and human capital, small-scale production, location in the periphery and other factors, in the informal sector can limit its access to external economies associated with agglomeration.

4.6 References

Alonso-Villar, O. and del Rio, C. (2013) “Concentration of Economic Activity: An Analytical Framework”, *Regional Studies*, 47(5): 756-772.

Anderson, T.W. and Rubin, H. (1949) “Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations”. *Annals of Mathematical Statistics*, 20:46–63.

Aprile-Gniset, J. (1992) *La Ciudad Colombiana*. *Siglo XIX y siglo XX*. Bogotá: Biblioteca Banco Popular.

Aprile-Gniset, J. and Mosquera, G. (1978) “Dos Ensayos sobre la Ciudad Colombiana”. Universidad del Valle, Cali-Colombia.

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

Baum, C.F., Schaffer, M. E. and Stillman, S. (2007) “Enhanced Routines for Instrumental Variables/Generalized Method of Moments Estimation and Testing”. *Stata Journal*, 7:465-506.

Becker, C., Williamson, J. and Mills, E. (1992) *Indian Urbanization and Economic Growth since 1960*. Baltimore, MD: Johns Hopkins University Press.

Calderón, V., Gáfaro, M. and Ibáñez, A (2011) “Forced Migration, Famale Labor Force Participation, and Intra-Household Bargaining: Does Conflict Empower Women”. *Documentos CEDE*, 2011-28, Bogotá.

Cameron, C., Gelbach, J. and Miller, D. (2008) “Bootstrap-Based Improvements for Inference with Clustered Errors”. *The Review of Economics and Statistics*, 90(3): 414-427.

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): 123–142.

Chakravorty, S., Koo, J. and Lall, S.V. (2003) “Metropolitan industrial clusters: patterns and processes”. *Policy Research Working Paper Series*, No 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.

Ciccone, A. and Hall, R. E. (1996) “Productivity and the Density of Economic Activity”. *American Economic Review*, 86: 54-70.

Ciccone, A. (2002) “Agglomeration Effects in Europe”. *European Economic Review*, 46:213–227.

CODHES (2011) “¿Consolidación de qué?. Informe sobre Desplazamiento, Conflicto Armado y Derechos Humanos en Colombia en 2010”. *Boletín informativo CODHES*, No. 77, Bogotá.

Combes, PP., Duranton, G., Gobillon, L. and Roux, S. (2010) “Estimating Agglomeration Economies with History, Geology, and Worker Effects”. In Agglomeration Economies, ed. E.L. Glaeser, 15-66. University of Chicago Press: Chicago.

CSPPDF (Commission for the Monitoring of the Public Policy on Forced Displacement) (2010) *Tercer Informe de Verificación sobre el Cumplimiento de Derechos de la Población en Situación de Desplazamiento*, Bogotá.

CSPPDF (Commission for the Monitoring of the Public Policy on Forced Displacement) (2011) *Cuantificación y Valoración de las Tierras y los Bienes Abandonados o Despojados a la Población Desplazada en Colombia*, Bogotá.

Da Mata, D., Deichmann, U., Henderson, J., Lall, S. and Wang, H. (2007) “Determinants of City Growth in Brazil.” *Journal of Urban Economics*, 62(2):252–72.

Daniels, PW. (2004) “Urban Challenges: the Formal and Informal Economies in Mega-Cities. *Cities*, 21(6): 501-511.

Davidson, R. and MacKinnon, J. G. (2000) “Bootstrap Tests: How Many Bootstraps?”. *Econometric Reviews*, 19: 55–68.

Davidson, R. and MacKinnon, J. G. (2010) “Wild Bootstrap Tests for IV Regression”. *Journal of Business and Economic Statistics*, 28(1):128-144.

Deichmann, U., Lall, S.V., Redding, S.J. and Venables, A.J. (2008) “Industrial Location in Developing Countries”. *The World Bank Research Observer*, 23(2): 219-46.

Dominicis, L., Arbia, G. and de Groot, H. (2007) “The Spatial Distribution of Economic Activities in Italy”. *Tinbergen Institute Discussion Paper*. 2007-094/3.

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. (2009) “Are Cities Engines of Growth and Prosperity for Developing Countries” In Urbanization and Growth, ed. M. Spence, P. Annez and R. Bukley, 67-114. The World Bank: Washington.

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

Duranton, G. and Puga, D (2004) “Micro-foundations of urban agglomeration economies” in Henderson, J.V. and Thisse, J. (Eds.) *Handbook of Regional and Urban Economics*, Vol. 4, Amsterdam: Elsevier Science. Pp. 2173-2242.

Ellison, G., Glaeser, E., and Kerr, W. (2010) “What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns”. *America Economic Review*, 100(3): 1195-1213.

Feitosa F., Câmara G., Monteiro A., Koschitzki T., and Silva, M. (2007), “Global and Local Spatial Indices of Urban Segregation”, *International Journal of Geographical Information Science*, 21: 299–323.

Henderson, J. V. (1986) “Efficiency of Resource Usage and City Size”. *Journal of Urban Economics*, 19:47–70.

Henderson, J. V. (2005) “Urbanization and Growth.” In Handbook of Economic Growth, vol. 1B, ed. Philippe Aghion and Steven N. Durlauf, 1543–91. Amsterdam: North-Holland.

Holl, A. (2008) “Production Subcontracting and Location”. *Regional Science and Urban Economics*, 38(3): 299-309.

IDMC (Internal Displacement Monitoring Centre) (2011) *Internal Displacement: Global Overview of Trends and Developments in 2010*, available at: http://www.internal-displacement.org/publications/global-overview-2010.pdf [accessed 26 February 2012].

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

Kleibergen, F. (2002) “Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression”. *Econometrica*, 70:1781–1803.

Kleibergen, F. and Paap, R. (2006) “Generalized Reduced Rank Tests Using the Singular Value Decomposition”. *Journal of Econometrics*, 133:97-126.

Lafourcade, M. and Mion, G. (2007) “Concentration, Agglomeration and the Size of Plants”. *Regional Science and Urban Economics*, 37(1): 46-68.

Lee, Y.J. and Zang, H. (1998) “Urbanization and Regional Productivity in Korean Manufacturing”. *Urban Studies*, 35:2085–99.

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.

Mills, E. and Becker, C. (1986) *Studies in Indian Urban Development*. New York: Oxford University Press.

Moreno-Monroy, A.I. (2012) “Critical commentary: Informality in Space: Understanding Agglomeration Economies During Economic Development”, *Urban Studies*, 49 (10): 2019-2030.

Murray, M.P. (2006) “Avoiding Invalid Instruments and Coping with Weak Instruments”. *Journal of Economic Perspectives*, 20:111–132.

Openshaw, S. and Taylor, P. (1979) “A Million or so Correlation Coefficients: Three Experiments on the Modifiable Area Unit Problem”, in Statistical Applications in the Spatial Sciences, ed. N. Wrigley, 127-144. London: Pion.

Overman, H.G. and Venables, A.J. (2005) “Cities in the Developing World”. *CEP Discussion Paper*, No 695, London.

Perry, G., Maloney, W., Arias, O., Fajnzylber, P., Mason, A. and Saavedra-Chanduvi, J. (2007) *Informality: Exit and Exclusion*, The World Bank.

Quigley, J. (2009) “Urbanization, Agglomeration, and Economic Development” In Urbanization and Growth, ed. M. Spence, P. Annez and R. Bukley, 115-132. The World Bank: Washington.

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

Rosenthal, S. and Strange, W. (2008) “The Attenuation of Human Capital Spillovers”. *Journal of Urban Economics*, 64:373-389.

Rogers, W.H. (1993) “sg17: Regression Standars Errores in Clustered Samples”. *Stata Technical Bulletin*, 13:19-23.

Sánchez, L. M. (2008) “Éxodos Rurales y Urbanización en Colombia. Perspectivas Históricas y Aproximaciones Teóricas”. *Bitacora*, 13(2): 57-72.

Shukla, V (1996) *Urbanization and Economic Growth*. Delhi: New York: Oxford University Press.

Sethuraman, S.V. (1997) “Urban Poverty and the Informal Sector: A Critical. Assessment of the Current Strategies. ILO Geneva.

Scott, A. (2002) “Regional Push: Towards a Geography of Development and Growth in Low-and Middle-Income Countries”. *Third World Quarterly*, 23(1):137-161.

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

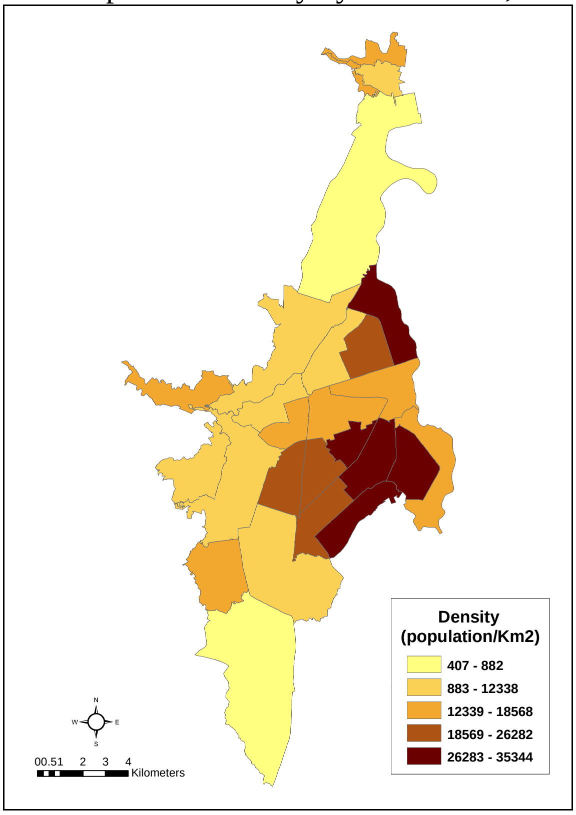
Stock, J.H. and Yogo, M. (2005) “Testing for Weak Instruments in IV Regression”, In Identification and Inference for Economic Models: A Festrschrift in Honor of Thomas Rothenberg, ed. D.W.K. Andrews and J.H. Stock, 80–108. Cambridge: Cambridge University Press.

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

Wu, F. (1999) "Intrametropolitan FDI enterprise location in Guangzhou, China: A Poisson and Negative Binomial Analysis". *Annals of Regional Science*, 33: 535-555.

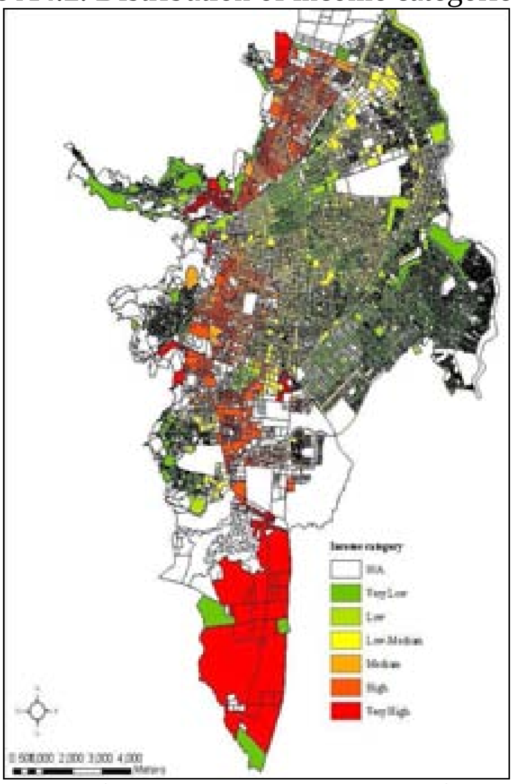
Appendix

Figure A4.1. Population density by Communes, Cali-Yumbo



Source: Population Census 2005, DANE

Figure A4.2. Distribution of income categories, Cali



Note: Data for Yumbo is not available. Levels of income based on economic stratification in six categories (1=Very Low, 6=Very high)

Table A4.1. 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: 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 A4.2. ISIC Revision 3, 3-digit level codes

Code	Industry	Code	Industry
151	Processed meat and fish	251	Rubber products
152	Processed fruit, vegetables and fats	252	Plastic products
153	Manufacture of milk products	261	Glass and glass products
154	Grain mill products; starches; animal feeds	269	Non-metallic mineral products, n.e.c
155	Bakery products, pasta and its products	271	Basic iron and steel
156	Manufacture of coffee	272	Basic precious and non-ferrous metals
157	Ingenious and sugar refineries	273	Casting of metals
158	Other food products	281	Structural metal products, tanks, steam generators
159	Beverages	289	Other metal products; metal working services
160	Tobacco products	291	General purpose machinery
171	Spinning of textiles	292	Special purpose machinery
172	Weaving of textiles	293	Domestic appliances, n.e.c.
173	Finishing of textiles	300	Office, accounting and computing machinery
174	Other textiles products	311	Electric motors, generators and transformers
175	Knitted and crocheted fabrics and articles	312	Electricity distribution and control apparatus
181	Wearing apparel, except fur apparel	314	Accumulators, primary cells and batteries
182	Dressing and dyeing of fur; processing of fur	315	Lighting equipment and electric lamps
191	Tanning, dressing and processing of leather	319	Other electrical equipment, n.e.c.
192	Footwear	321	Electronic valves, tubes, etc.
193	Manufacturing of leather products	322	TV/radio transmitters; line comm. apparatus
201	Sawmilling and planning of wood	323	TV and radio receivers and associated goods
202	Products of wood, cork, straw, etc.	333	Watches and clocks
203	Manufacture of joinery and carpentry for buildings and constructions	341	Motor vehicles
204	Manufacture of wood containers	342	Automobile bodies, trailers & semi-trailers
209	Other wood products	343	Parts/accessories for automobiles
210	Paper and paper products	351	Building and repairing of ships and boats
22	Printing/Editing and similar goods	353	Aircraft and spacecraft
224	Reproduction of recorded media	359	Transport equipment, n.e.c.
232	Refined petroleum products	361	Furniture
241	Manufacture of basic chemicals	369	Manufacturing, n.e.c.
242	Other chenical products	371	Recycling of metal waste and scrap
243	Man-made fibres	372	Recycling of non-metal waste and scrap

Table A4.3. Number of enterprises and employees by sectors and type of enterprise

	All enterprises		Formal small		Formal medium-large		Informal	
ISIC	Enterprises	Employees	Enterprises	Employees	Enterprises	Employees	Enterprises	Employees
151	49	622	31	136	3	453	15	33
152	124	1278	53	246	3	903	68	129
153	23	172	10	56	1	87	12	29
154	307	2387	144	854	9	1197	154	336
155	578	2585	363	1659	4	386	211	540
156	2	18	1	16			1	2
157	15	1137	10	91	2	1042	3	4
158	59	3867	37	259	6	3527	16	81
159	64	1299	19	197	3	1044	42	58
160	1	40	1	40				
171	50	498	24	93	1	366	25	39
172	13	60	8	51			5	9
173	77	314	37	244			40	70
174	81	562	42	249	2	240	37	73
175	22	106	10	89			12	17
181	662	8675	270	1672	23	5365	369	1638
182	3	7	3	7				
191	5	14	3	11			2	3
192	354	1692	139	895	3	304	212	493
193	120	789	56	191	4	477	60	121
201	26	97	20	88			6	9
202	11	44	6	25			5	19
203	205	625	70	282			135	343
204	32	132	15	48	1	50	16	34
209	158	610	77	316	2	105	79	189
210	122	2725	82	518	7	2130	33	77
22	259	1620	206	1157	4	350	49	113
224	2	98	1	3	1	95		
232	2	6	2	6				
241	36	217	29	200			7	17
242	100	8650	72	769	11	7839	17	42
243	2	18	2	18				
251	33	885	25	224	1	647	7	14
252	157	2595	112	963	11	1559	34	73
261	14	24	7	17			7	7
269	46	350	27	150	2	174	17	26
271	50	202	34	172			16	30
272	9	17	6	11			3	6
273	55	863	36	179	3	613	16	71
281	51	336	29	173	2	120	20	43
289	188	903	109	523	3	225	76	155
291	14	437	9	75	1	350	4	12
292	38	543	31	241	3	290	4	12
293	26	817	13	145	11	660	2	12
300	17	232	11	113	1	101	5	18
311	4	61	3	60			1	1
312	17	946	15	199	2	747		
314	1	4					1	4
315	4	225	2	7	1	216	1	2
319	14	257	12	93	1	163	1	1
321	2	15	2	15				
322	8	63	7	62			1	1
323	14	61	10	55			4	6
333	44	281	33	190	1	70	10	21
341	4	12	3	10			1	2
342	7	146	5	30	1	115	1	1
343	3	10	1	2			2	8
351	3	55	2	54			1	1
353	1	3	1	3				
359	13	1230	7	30	2	1194	4	6
361	259	1495	148	732	5	495	106	268
369	152	1205	86	554	4	524	62	127
371	14	42	5	19			9	23
372	36	169	16	114			20	55

Figure A4.3. Kernel density mapping of industries displaying significant clustering and co-clustering of formal small and informal enterprises

a) Footware **b) Mill products**

Formal small Informal Formal small Informal

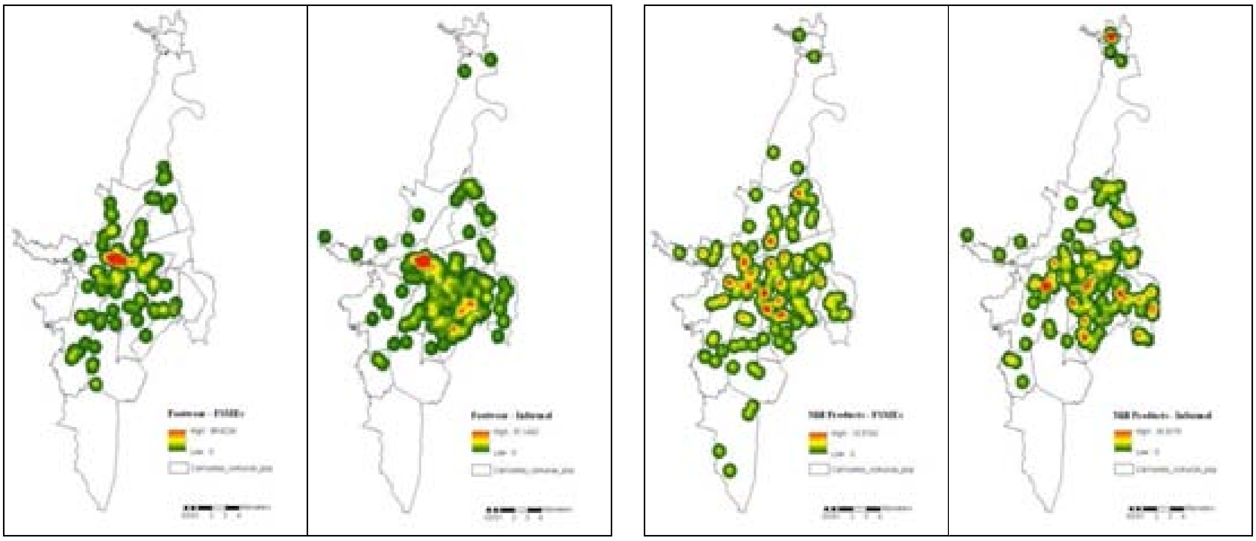
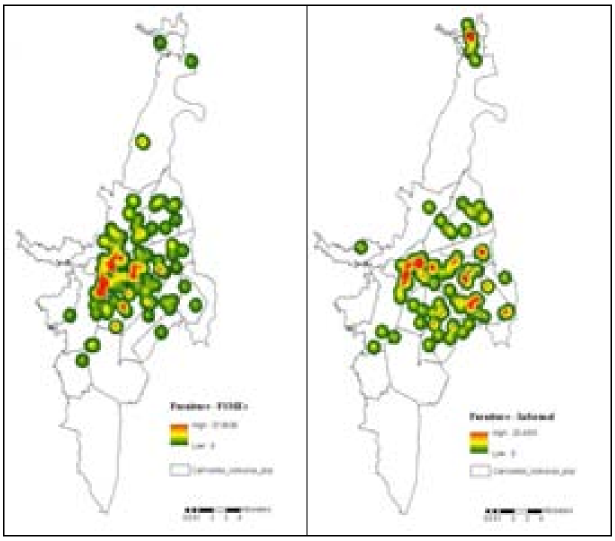


Figure A4.4. Kernel density mapping of industries displaying significant clustering formal small and informal enterprises but no co-clustering

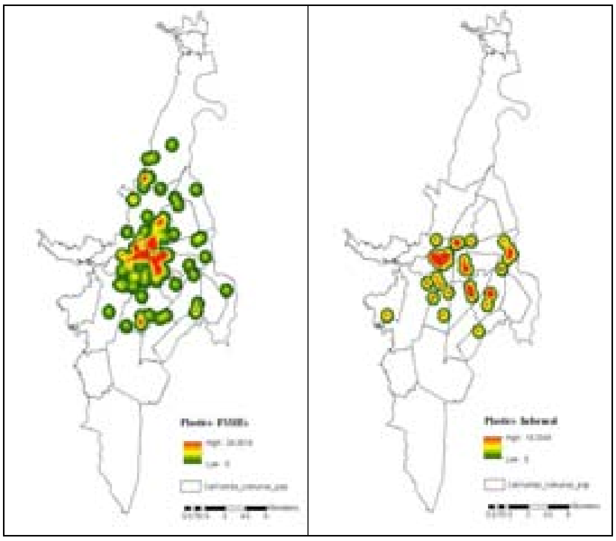
a) Furniture

Formal small | Informal



b) Plastic products

Formal small | Informal



b) Bags products

Formal small | Informal

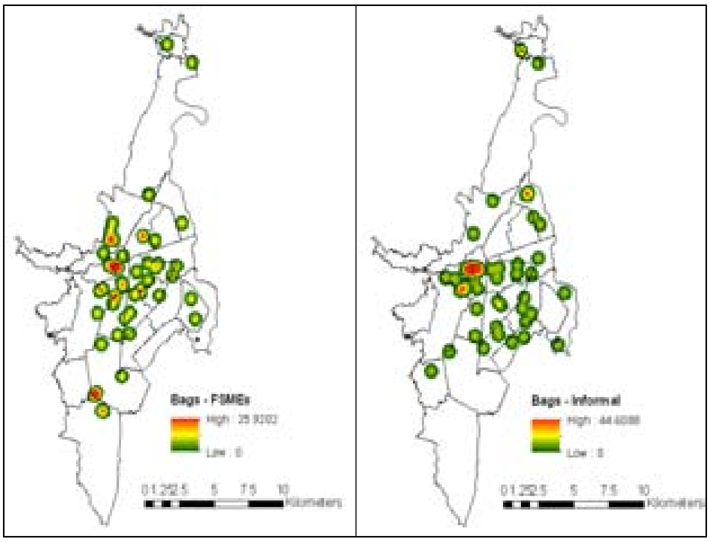


Table A4.4. Descriptive statistics

	Distance	# Pairwise industries	Mean	Std. Dev.	Min	Max
$M_{S_1S_2}$, $M_{S_2S_1}$	<1Km	253	1.455	1.235	0.000	6.738
	<5Km	253	1.171	0.212	0.603	2.212
$M_{FLEs,FLEs}$	<1Km	253	0.526	2.422	0.000	23.224
	<5Km	253	1.301	0.712	0.000	5.505
$M_{FSMEs,FLEs}$	<1Km	253	3.486	12.289	0.000	140.58
	<5Km	253	1.153	0.363	0.387	3.678
$M_{IEs,FLEs}$	<1Km	253	1.125	8.378	0.000	119.09
	<5Km	253	0.983	0.323	0.340	2.836
$M_{FSMEs,FSMEs}$	<1Km	253	2.436	3.272	0.000	24.834
	<5Km	253	1.275	0.183	0.807	1.862
$M_{IEs,FSMEs}$	<1Km	253	4.665	8.320	0.000	67.204
	<5Km	253	1.605	0.293	1.084	3.119
$M_{IEs,IEs}$	<1Km	253	8.265	21.189	0.000	209.99
	<5Km	253	2.213	0.704	1.017	5.618
LinkageIO		253	3.510	7.799	0.000	61.157
Labor correlation		253	0.703	0.276	0.107	0.999

Note: *FLEs*: Formal medium-large enterprises; *FSMEs*: Formal small enterprises; *IEs*: Informal enterprises

Table A4.5. Urbanization elasticity
dependent variable: log of real hourly wage

	Full sample		Formal		Informal	
	OLS	IV	OLS	IV	OLS	IV
Log of employment density	0.0489***	0.0564**	0.0877***	0.1150***	0.0412***	0.0430***
	(3.85)	(2.45)	(4.81)	(3.30)	(3.30)	(1.81)
Education (base: Less than primary)						
Primary	0.2926***	0.2916***	0.3452***	0.3453***	0.2675	0.2672***
	(15.65)	(14.79)	(15.14)	(15.91)	(14.66)	(13.86)
Secondary	0.6720***	0.6710***	0.7137***	0.7132***	0.6127	0.6124***
	(22.31)	(21.51)	(19.66)	(19.87)	(21.48)	(20.63)
Tertiary	1.4335***	1.4304***	1.2181***	1.2109***	1.3974	1.3966***
	(36.91)	(35.25)	(26.94)	(26.56)	(33.58)	(31.80)
Experience	0.0258***	0.0257***	0.0175***	0.0173***	0.0253	0.0252***
	(20.37)	(19.78)	(4.67)	(4.69)	(17.79)	(17.37)
Experience^2	-0.0003***	-0.0003***	-0.0002***	-0.0002***	-0.0003	-0.0003***
	(-19.74)	(-18.96)	(-3.12)	(-3.10)	(-19.32)	(-18.50)
Gender (1 = male)	0.1792***	0.1783***	0.2005***	0.2005***	0.1623	0.1621***
	(12.09)	(11.12)	(7.97)	(8.06)	(12.52)	(11.25)
Firm size (base: one employee)						
2 – 5 employees	0.2739***	0.2745***	0.3455***	0.3462***	0.1548	0.1550***
	(25.85)	(25.64)	(12.05)	(12.00)	(10.20)	(10.56)
6 – 10 employees	0.7232***	0.7230***	0.7156***	0.7148***	0.4496	0.4494***
	(15.21)	(15.47)	(10.18)	(10.45)	(6.00)	(6.08)
More than 11 employees	0.6180***	0.6195***	0.9051***	0.9045***	0.2987	0.2992***
	(9.10)	(9.38)	(6.06)	(6.23)	(5.67)	(6.01)
Observations	49209	49209	8903	8903	40306	40306
R2	0.3209	0.3208	0.3485	0.3474	0.2632	0.2632
Root MSE	0.8315	0.8312	0.8508	0.8491	0.8141	0.8136
Time, industry and region dummies	yes	yes	yes	yes	yes	yes

Standard errors are cluster-robust. Clustering is on municipalities.
Note: ***, **, *, denotes significance at 1%, 5% and 10%, respectively. ( ) t statistics.

Chapter 5

Conclusions

5.1 Summary and the main results

The thesis has undertaken an empirical examination of the informal sector and identified the factors that affect the decision to enter the informal sector and its inter-relation with formal activities, taking into consideration the heterogeneous nature of this sector in Colombia. This thesis has also sought to determine whether the structural constraints on the availability of formal jobs at the national level, as well as at the level of the local labor markets, are important in determining the composition of informal activities. The general empirical literature on this subject and especially that conducted in the context of Colombia is inconclusive on several vital questions related to the determinants of informal activities.

The findings are chapter specific and in this section we summarize the main empirical results. Chapter 2 sought to determine the factors that affect the decision to find employment in the formal or informal sectors taking into account that there is a rationing of formal sector jobs and, hence, a job queue to enter that sector. Although the theoretical literature on the informal sector has stressed the importance of rationing in the formal sector as a factor behind informality, there have been few applications of the job queue approach to examine this issue. The first chapter in this thesis aims to fill this gap.

The findings of this chapter can be summarized as follows. First, we confirm that formal job rationing does exist in Colombia and that it affects around 54% of the workers that find themselves in the informal labor market. These informal workers represent that segment of workers who see informality as a last resort. This restriction in the availability formal jobs is evidence of the labor segmentation afflicting the Colombian labor market, a situation that has made itself particularly manifest since the 1990s (Uribe *et al.*, 2007).

Second, the length of the queue in the formal sector is around 1.25, meaning that for each worker in the formal sector there are a further 1.25 workers queuing for formal jobs. This value is high in comparison to the values reported in neighboring countries such as Brazil and Chile where the lengths of the formal job queues are 1.19 and 1.06, respectively. We also found that females, non-heads of household, those that are not married, workers that were unemployed prior to finding their current job and individuals with a low level of education are those who once “in the queue” for a formal job have the lowest probability of being chosen from it.

Third, among the factors determining whether a worker finds employment in the formal or informal sectors, actively seeking formal employment is mainly influenced by personal traits and by the characteristics of human capital and of the household, as well as by the wage premiums associated with working in the formal sector. At the same time, among the factors that have the greatest influence on formal job offers are human capital endowment, the degree of commitment manifest by the workers for their job, the degree of responsibility presented by the worker and the individual’s employment history.

In Chapter 3 we examined the heterogeneous nature of the informal sector and observed differences at the local labor market level; in particular, we drew a distinction between more developed, central cities and their less developed, peripheral counterparts. The chapter makes three main contributions to the discussion. First, our results confirm that there is marked heterogeneity in Colombia’s informal sector characterized by at least two distinct segments of workers: those who are voluntarily informal as they can obtain higher benefits in this sector than in the formal sector, and those that are rationed out of formal employment opportunities and who see informality as providing their only form of employment and escape from unemployment.

Second, the results indicate marked differences in the kind of informal employment between groups of cities. In less developed cities the main share of informal employment is in the most traditional activities (75%) where much of the employment is low skilled and located in the personal services and commerce sectors. By contrast, in more developed cities the informal sector is associated with more modern activities and, therefore, there is a higher share of jobs in the modern informal segment (52%). This segment is characterized by the presence of more skilled and productive workers and the existence of higher intermediate linkages with the formal sector than are present in the traditional segment. In the case of Caribbean coast cities there is a marked contrast between a relatively high proportion of informal employment engaged in highly modern activities (12%) and that engaged in less modern and more traditional activities (31%). On the one hand, the export vocation of these cities and the high concentration of industries with important productive linkages promote the expansion of the modern informal segment; while, on the other hand, activities associated with tourism have a major influence on the creation of low-skill service jobs.

Third, two main findings can be drawn from the quantile decomposition of the formal/informal wage gap. First, in the group of cities with high levels of informality (Groups 2 and 3), the wage gap is smaller over the whole distribution than it is in the group of cities with low levels informality (Group 1). This suggests that in cities where informal activities are the main source of income, the informal sector is no longer considered the poor, marginal sector. The low level of regulations, the low degree of intensity with which labor laws are enforced and the greater tolerance for informal activities seem to imply higher wages and non-wage benefits for those in the informal sector.

A further result is that workers show different motivations for working in the informal sector. In the lower half of the distribution the informal sector represents the most disadvantaged segment, not only in terms of their more adverse characteristics, but also in terms of the lower rates of return to these characteristics with respect to the formal sector. Hence, these lower-paid, informal workers are rationed out of any formal employment opportunities and end up in informality as a last resort. By contrast, in the upper half of the distribution the informal sector represents the competitive segment, where the wage gap between the two sectors is much narrower than it is at the bottom and individuals receive similar returns to their characteristics to those received in the formal sector. In this higher-paid, informal segment workers voluntarily choose informality and accept lower wages in order to avoid the administrative costs of social security or to benefit from the greater labor flexibility that they could not enjoy in the formal sector.

Chapter 4 examined the spatial distribution of informal enterprises vis-à-vis that of formal enterprises within a metropolitan area of a developing country (Cali, Colombia) and the agglomeration effects in the presence of an informal sector. The results of this analysis allowed us to conclude that, in general, the clustering patterns of informal enterprises and formal small enterprises are similar and that there are different degrees of clustering depending on the industry being analyzed. The co-clustering analysis showed that formal and informal small enterprises seek each other out in terms of their spatial distribution, whereas formal large enterprises display different location patterns with regard to smaller enterprises. Analyzing each individual industry, we found that although there is a significant clustering of both formal and informal enterprises of similar sizes in the same industry, co-clustering does not seem to exist between these types of enterprise. Interestingly, this indicates that similarly sized formal and informal enterprises belonging to the same industry may locate in different parts of the city, which is an indication of spatial segmentation.

Finally, Chapter 4 showed that the formal sector achieves greater benefits from agglomeration economies than are obtained by the informal sector. This differential in returns may be due to the constraints faced by the informal sector in terms of accessing physical and human capital and their small-scale production, peripheral location, and the risks and vulnerability associated with operating outside legal regulations.

5.2 Policy implications

To combat informal employment our findings suggest it is necessary to understand the different realities characterizing individual cities and different groups identified within the informal sector. Here, it seems essential to distinguish between those that voluntarily choose informality (and who, therefore, are not necessarily worse-off than those working in the formal sector) and those that have no other choice than to remain in the informal sector and who are systematically excluded from the formal sector. It is this second group of individuals that makes a significant contribution to overall wage inequality and poverty in Colombia and policies need to be addressed to remedying this bottleneck.

Despite the Colombian Government's efforts to develop strategies to reduce informality, they are not having the anticipated impact. Policies in Colombia have mostly targeted facilitating formalization by reducing the costs and increasing the benefits of such process. These actions, though well-intentioned, have not provided a real stimulus to formalization and if they have had any impact it has been only in the short-run (OMTSS, 2011; Farné and Rodríguez, 2013).

This dissertation has shown that informality encompasses a wide range of activities and, depending on such factors as location, industry and type of worker, various benefits and costs can be associated with being engaged in the informal sector. Hence, policy packages require a careful balance of “carrots” and “sticks”, adapted the nature of informality being addressed. Policies based only on a series of economic incentives and on tax reduction to encourage formalization without any synchronization with policies aimed at increasing the aggregate productivity in the economy, so as to deter informal behavior and to change the pervasive culture of noncompliance that is observed in Colombia, may lead to worsening levels of informality (de Ferranti *et al*., 2003; Perry *et al*., 2007).

The findings of Chapter 2 show that the most vulnerable individuals (least educated, females, non-heads of household or those having suffered previous periods of unemployment) are those with the lowest probability of being chosen to fill a formal vacancy. Therefore, policy makers need to design strategies to encourage the hiring of workers with these characteristics. First and foremost, actions are needed to boost human capital levels, especially for individuals in these groups, which should increase their chances of finding formal work as well as raising their earnings. Second, policy makers should create incentives for the hiring of informal workers, for example, by offering tax reductions for firms that hire such workers.

As discussed, a common policy to encourage the formal sector to hire more workers is to reduce the contribution levied on the payroll and, hence, to reduce labor costs. However, if this policy is not accompanied by the enforcement of regulations and administrative reforms to strengthen information flows between the enforcing agencies, it can have unintentional negative effects. The immediate impact of a reduction in labor costs is a reduction in the wage gap between the formal and informal sector across the whole wage distribution. As we have seen in Chapter 3, among the upper half of the distribution such a reduction in the wage gap can increase incentives to be informal and increase the number of workers who find themselves on the threshold of indifference between formal and voluntary informal activities. This is particularly important in the case of less developed cities where there is an informal employment wage premium and a greater tolerance of informal activities.

This dissertation also provides policy makers with a number of important guidelines when addressing issues related to informality at the firm level and the consequences of their location. On the one hand, as stressed in previous studies (de Ferranti *et al*., 2003; Perry *et al*., 2007), actions are required to improve the aggregate productivity of the economy. Without such actions a large number of small firms will continue to exist and they will continue to be characterized by low productivity, geographical dispersed, weak growth and poor survival prospects; moreover, they would achieve few benefits from establishing relations with formal activities. Likewise, without an enabling investment climate, formal firms will not to be able to expand, and so the demand for more skilled workers and wages will not grow and, in general, the number of jobs in the formal sector will remain depressed.

On the other hand, policy makers should consider two consequences of the location of informal enterprises. First, this research has shown that in central areas informal enterprises tend to locate near similar formal enterprises. As such, they would seem to benefit from agglomeration economies, while at the same time generating additional locational competition and congestion costs for formal producers. Second, the results have shown that informal enterprises operate mostly in peripheral areas and hence might be denied access to the best final and intermediate markets available in the city. For informal enterprises this might result in fewer possibilities of establishing linkages with formal enterprises and fewer chances of growth and survival.

5.3 Future research lines

There are certain limitations that are worth recognizing in this research and which, at the same time, can serve to identify interesting lines for future research.

A limitation evident in Chapters 2 and 3 is that we did not consider the degree of mobility of workers between formal and informal jobs. This limitation is due to data constraint. In Colombia longitudinal information is not available to allow determining the dynamics of job status mobility. What we could do, based on available information (i.e., repeated cross-section surveys), is to compare different years and to analyze whether there have been any changes in the wage gap across the distribution. For instance, if there has been little or no change in the wage gap at the bottom of the distribution over time it might suggest that very low, informal wages at the bottom of the distribution have not provided any incentives movement to move towards the formal sector. This would be consistent with low mobility and, hence, significant labor market segmentation. This alternative approach has been taken by Arabsheibani and Henley (2009) for the Brazilian case.

As observed in Chapter 3, different characteristics of informality have been identified among the local labor markets, due mainly to the make-up of the production of the cities. It would have been useful to take this local heterogeneity into account in Chapter 2. Although in this later chapter we included city dummy controls in our estimates, an analysis for each city or group of cities would have allowed us to identify whether there are different factor effects in different cities influencing the decision being to enter the informal sector. Similarly, an estimate for each city would allow us to account for the different degrees of involuntary/voluntary informality which would be important for the design of local labor policies. Maloney (2004) and Kucera and

Roncolato (2008) argue that policy makers need to know how widespread voluntary informal employment is and how this can vary for regions at different levels of development and for different workers. The underlying idea is that the higher the level of voluntary informal employment, the greater the number of workers there are likely to be on the threshold of indifference between formal and voluntary informal employment. Consequently, this is of great importance in aligning policies with each reality so as to enhance the design and implementation of labor regulations and to avoid unintended negative consequences.

Measuring the amount of involuntary/voluntary informal employment could also have been particularly relevant in Chapter 3. Although our findings in this chapter show the existence of different groups of informal workers, we failed to determine how much of the informality can be attributed to rationing or a segmented labor market and how much can be attributed to the worker's voluntary decision. In this regard, we might have followed the methodological strategy proposed by Gunther and Launov (2012) which assumes that workers are earning maximizers and that they know the wage function and so can identify their expected earnings given their own characteristics for each segment of the labor market. In this way it is possible to compare the actual wage function with a hypothetical wage function for each segment and to determine whether an individual is in the sector where his earnings, given his personal characteristics, are the highest. The underlying assumption is that if we cannot reject the possibility that the actual wage function and the hypothetical wage function across sectors are equal, then this is an indication of perfect sectoral mobility, that is, such a market can be considered competitive. The rejection of the equality of these two functions would indicate the existence of some sort of entry barrier preventing certain individuals from being in the sector that pays the highest expected wage to them, i.e. that the market is segmented.

In Chapter 4 we analyzed the spatial distribution of enterprises by means of M-functions and we discussed the advantage of using this function over other distance-based methods, in particular Duranton and Overman's *Kd*-function (2005). However, as mentioned, calculating both functions (and other statistics) would be interesting, both academically and practically, in order to check the robustness of the results and to provide a complete picture of clustering and co-clustering patterns.

Finally, a further interesting issue to place on the future research agenda, drawing on the rich geocoded information available from the Cali census, would be a study of firms' informality phenomenon as a result of inter-firm spatial location, in other words, to analyze the impact of the “neighbor effect” on a firm’s decision to be informal or formal. As has been shown in this dissertation, the relationship between informality and location would seem to be evident, so obtaining more evidence for this relationship would help policy makers consider spatial issues in the design and implementation of policies that might be more effective in reducing informality.

5.4 References

Arabsheibani, G and Henley, A. (2009). “How are Informal Sector Workers Disadvantaged? Quantile Regression Evidence from Brazil”. Unpublished manuscript.

de Ferranti, D., Perry, G., Gill, I., Guasch, J. L., Maloney, W. F. , Sanchez-Paramo, C., and Schady, N. (2003). *Closing the Gap in Education and Technology*. Washington, DC: World Bank.

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

Farné, S. and Rodrígues, D (2013). “¿Bajar los Impuestos al Trabajo Genera Empleo? Ley 1607 de 2012 de Reforma Tributaria en Colombia”, *Cuadernos de Trabajo*, No 14, Universidad Externado de Colombia.

Gunther, I. and Launov, A (2012). “Informal Employment in Developing Countries. Opportunity or Last Resort?”, *Journal of Development Economics*, 97: 88-98.

Kucera, D. and Roncolato, L. (2008). “Informal Employment: Two Contested Policy Issue”, *International Labour Review*, 147(4): 321-348.

Maloney, W. (2004). “Informality Revisited”, *World Developed*, 32(7): 1159-1178.

OMTSS, Observatorio del Mercado de Trabajo y la Seguridad Social. (2011). “¿La Ley 1419 de 2010 ha Formalizado el Empleo en Colombia?, *Boletín del Observatorio del Mercado de Trabajo y la Seguridad Social*, No. 13, Universidad Externado de Colombia.

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.

Uribe, J., Ortiz, C. and García, G. (2007). “La Segmentación del Mercado Laboral Colombiano en la Década de los Noventa”, *Revista de Economía Institucional*, 9(16): 189-221.

UAB
Universitat Autònoma de Barcelona

Gustavo Adolfo García Cruz

Thesis in Economics