Review of Development Economics, 21(4), 985–1017, 2017
DOI:10.1111/rode.12317

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

*Gustavo A. García**

Abstract

The labor market in developing countries is remarkably heterogeneous, with a small productive formal sector characterized by high wages and attractive employment conditions, and a large informal sector characterized by low productivity and volatile wages. The informal sector is particularly diverse. In this paper, we examine the heterogeneity of the informal sector at the regional level in Colombia. In general, our findings suggest that both voluntary and involuntary informal employment co-exist by choice and as a consequence of labor market segmentation. We also find that there are striking differences in labor market characteristics across cities, particularly with respect to informal employment.

1. Introduction

One of the most notable features of developing countries is the remarkable heterogeneity of their urban labor markets. It is common to observe a small productive formal sector that offers attractive labor conditions and relatively high wages co-existing with a large informal sector that uses unskilled labor, is characterized by low earnings and productivity, and does not fully comply with established legal regulations (Dickens and Lang, 1985; Maloney, 1999, 2004; Jütting and De Laiglesia, 2009). Nevertheless, there is a considerable variety of workers within the large informal sector.

Why is there such diversity in the informal sector? Are there different kinds of informal workers—some who are informal by choice and others who have no alternative form of employment? Is labor informality a choice or the result of labor market segmentation?

The segmented labor market theory considers informality a survival alternative to involuntary unemployment for workers who are disadvantaged or rationed out of formal employment opportunities (Dickens and Lang, 1985). The result is an earnings dualism for individuals with similar characteristics that 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, whereas the informal sector has no institutional or efficiency–wage basis to regulate wages. In addition, low entry barriers and an abundant supply of unskilled workers lead to low wages. Thus, wages depend on the sector in which a worker is employed and not on his or her skills *per se* (Uribe et al., 2007).

In contrast, the orthodox neoclassical view of the human capital theory postulates that, as in any another market, price flexibility and free labor mobility lead to a full

*García: Research in Spatial Economics (RiSE-group) and Department of Economics, Universidad EAFIT, Medellín, Colombia. E-mail: ggarci24@eafit.edu.co. The author wishes to thank participants at the 16th IZA European Summer School in Labor Economics and thank José Luis Roig, Josep Lluís Raymond, the editor and two anonymous referees for their suggestions and comments.

employment equilibrium with equal remuneration for the same kind of work (De Soto, 1987; Saavedra and Chong, 1999; Maloney, 1999). Given this competitive market framework, participation in the informal sector may be a desirable choice for workers and firms, based on private cost–benefit calculations of inclusion in the sector. Informality can have desirable non-wage features and thus individuals maximize their utility rather than their earnings. Alternatively, certain workers have a comparative advantage in the informal sector that does not exist 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; that is, if the informal sector has its own internal duality. The literature has recognized the existence of "upper" and "lower" tiers or "voluntary" and "involuntary" entry of informal employees or firms (Fields, 1990, 2005; Cunningham and Maloney, 2001; Maloney, 2004). In this scenario, upper-tier employees are those who are voluntarily informal because, based on their specific characteristics, they expect to earn more in the informal sector than they would in the formal sector. In contrast, lower-tier employees are disadvantaged workers who see informality as a last resort.

From an empirical stance, this notion of dualism within the informal sector has not been satisfactorily addressed. For example, Magnac (1991), who tested the labor market of Colombia in the 1980s for competitiveness or segmentation, found evidence of a competitive labor market structure. Similarly, Gindling (1991) and Pratap and Quintin (2006) found evidence of segmentation in Costa Rica but a competitive structure in Argentina. However, in all of these studies, the authors assume homogeneity of the informal sector, thereby limiting their analysis.

Among the few studies that have tried to model the heterogeneous structure of the informal sector are Cunningham and Maloney (2001) and Günther and Launov (2012). Cunningham and Maloney (2001) model the informal sector as a mixture of "upper-tier" and "lower-tier" enterprises, using the econometric technique of factor and cluster analysis to allow for segmentation of the market. They find evidence of segmentation; however, they only consider informal firms and thus their model excludes the possibility of being a formal firm. Further, the authors do not take into account the selection bias induced by workers' decisions about type of employment.

Günther and Launov (2012) analyze the possible heterogeneous structure of the informal sector, estimating a finite mixture model that allows them to determine the number and size of segments that might compose the informal sector. This model uses minimal *a priori* assumptions to determine the segments and provides a new method for identifying the extent 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 comprises two segments: a highly paid segment and a low-paid segment. They also found that 45% of informal employment is involuntary and mainly located in the low-paid informal segment, whereas the remaining 55% of informal employment is voluntary and situated in the highly paid informal segment.

This paper provides new evidence on the heterogeneity of the informal sector in Colombia. The novelty of this study is twofold. First, we analyze the informal sector in Colombia, which provides rich evidence from a large and heterogeneous informal sector that presents a particularly interesting case for analyzing heterogeneity in the informal sector in developing countries. The extant literature reports mixed and inconclusive results on the heterogeneity of the informal sector; this paper uses the Colombian case to offer more evidence on the issue. Specifically, we analyze the

factors that affect wage differentials between formal and informal workers using quantile decomposition methodology, controlling for selectivity bias caused by correlated unobserved heterogeneity affecting wages. We employ the approach introduced by Albrecht et al. (2009) based on Buchinsky (1998), which is a non-parametric method to account for selection in quantile regression analysis. This methodology has two advantages. First, it allows us to analyze the heterogeneity of the informal sector by considering a decomposition of the wage differential between the formal and informal sectors throughout the entire distribution of wages. Second, it is one of the few approaches empirically available in the literature for selectivity correction in quantile regression on cross-sectional data.

Our second contribution is to conduct the analysis of formal–informal wage gap decomposition at a regional level. The extant literature typically considers a national context and does not take into account the possibility of regional heterogeneities among informal workers. Given the geographic, demographic and social conditions and the economic dynamics of Colombia, there are marked differences in the structures and dynamics of the local labor markets, which could imply that informality is not homogenous throughout the territory. In Colombia, roughly six out of ten employees work in the informal sector.[1] Cities such as Cúcuta or Montería have informality rates of approximately 75%; others, including Medellín and Bogotá, have rates of approximately 50% (García, 2011; Galvis, 2012).

The decomposition of the wage differential between the formal and informal sectors throughout the entire distribution of wages allows us to determine the proportion of the wage gap that is due to differences in prices related to individual characteristics and the proportion resulting from characteristics that differ between the formal and informal sectors. If the wage gap is mainly attributable to the former factor, it indicates that individuals in the informal sector earn less because they earn lower returns on their skills and therefore are members of the disadvantaged sector of a segmented market. In contrast, if the wage gap is primarily explained by the latter factor, labor segmentation is not as significant as it is in the former case and wage differences between sectors are due to differences in endowments. In the latter situation, being an informal worker is a choice, because these individuals can obtain non-wage benefits or earn more than they would in the formal sector.

Following this introduction, section 2 summarizes the literature on wage differences between formal and informal workers. Section 3 describes the data. In section 4, we discuss the estimation procedure. Section 5 presents the empirical findings, and conclusions are drawn in section 6.

2. Literature Review

Although the empirical evidence on wage differentials between the formal and informal sectors in various countries is considerable, the results are mixed and inconclusive. Most early studies analyzing the traditional segmented labor market theory find that workers in the formal sector are better rewarded than their counterparts in the informal sector. These studies include Mazumdar (1981) in Malaysia, Heckman and Hotz (1986) in Panama, Roberts (1989) in Mexico, Marcouiller et al. (1997) in El Salvador and Peru, Tansel (1999, 2000) in Turkey and Gong and Van Soest (2002) in Mexico. More recently, Tansel et al. (2015) use a fixed effects model with a quantile regression technique and find a persistent informal wage penalty in Egypt.

In contrast, other studies suggest that the formal–informal wage gap may not be a stylized fact. For instance, using data for urban Bolivia, Pradhan and Van Soest

(1997) find that formal–informal wage gaps tend to be negative rather than positive. Marcouiller et al. (1997) show a significant wage premium in the informal sector in Mexico. In Argentina, Pratap and Quintin (2006) find no difference between formal and informal earnings after controlling semiparametrically for individual and employer characteristics. Using quantile regression techniques on cross-sectional data, Tannuri-Pianto and Pianto (2002) find that in Brazil, the wage advantage of formal workers over informal workers completely disappears at high quantiles of the wage distribution. With the same methodology, Staneva and Arabsheibani (2014) find a significant informal employment wage premium across the entire earnings distribution. Using quantile regression techniques in the context of panel data, Bargain and Kwenda (2014) (in South Africa, Brazil and Mexico) and Nguyen et al. (2013) (in Vietnam) find that the informal sector wage penalty tends to disappear at the top of the earnings distribution after controlling for time-invariant unobservable characteristics. In the Russian case, Braithwaite (1995) and Kolev (1998) find that there is a higher wage premium in the informal sector than in the formal sector, whereas Lehmann and Zaiceva (2013) show a wage penalty for informal workers in the lower part of the distribution and no statistically significant wage gap between informal and formal workers in the upper half of the distribution.

Although there are several studies on wage gaps in Colombia, very few of them analyze the particular case of the formal–informal wage gap, its decomposition and regional differences. Among wage gap studies, we highlight those analyzing the gender wage gap, namely, Fernandez (2006), Badel and Peña (2010), Hoyos et al. (2010) and Galvis (2010). Regarding the formal–informal wage gap in Colombia, there is the work of Daza and Gamboa (2013). To our knowledge, this is the only work that studies the wage difference between formal and informal activities in Colombia. The authors use a non-parametric method based on a matching procedure proposed by Ñopo (2008). The results show that there is a positive wage gap in favor of formal workers and that this gap remains unexplained after controlling for demographic and labor variables. However, that study does not conduct a decomposition analysis to explain the factors affecting the sector wage gap and assumes that the wage gap is homogeneous throughout the territory. Additionally, that paper does not analyze the heterogeneity of the informal sector by distinguishing between low-pay and high-pay informal workers and comparing their corresponding wage gaps with those of their counterparts in the formal sector.

3. Data and Descriptive Evidence

In this study, we use data from Colombia, a medium-income country located in northwestern South America. Despite showing positive economic growth, Colombia is characterized by high levels of poverty, inequality and poor labor market conditions. The annual gross domestic product (GDP) growth rate in Colombia was 4.4% in 2013, whereas the percentage of people living below the poverty line was 5.6% and the Gini coefficient was 53.5% (United Nations Development Programme, 2015). In terms of labor market indicators, Colombia presents a marked heterogeneity characterized by high levels of informality and unemployment. According to the International Labor Organization (ILO, 2011), Colombia has an informality rate of approximately 57% and an unemployment rate of 10.1%, which are the highest in Latin America. See Table 1 for an international comparison of these primary economic and social indicators.

Table 1. Economic and Social Indicators in Colombia and Other Countries

	Colombia	*Ecuador*	*Brazil*	*Chile*	*Argentina*	*Mexico*	*USA*	*UK*
GDP per capita (2011 PPP, US$)[a]	12,025	10,541	14,555	21,714	15,501	16,291	51,340	37,017
GDP growth (annual, %)[b]	4.4	3.7	0.1	1.9	0.5	2.3	2.4	2.9
Gini coefficient[c]	53.5	46.6	52.7	50.8	43.6	48.1	41.1	38.0
Population living below US$1.25 PPP per day (%)[d]	5.6	4.0	3.8	1.4	1.4	1.0	–	–
Informality rate (%)[e]	56.8	52.2	38.4	–	46.9	54.2	–	–
Unemployment rate (%)[f]	10.1	4.6	6.8	6.4	8.2	4.9	6.2	6.3

Notes: [a]Data refer to 2013 (UNDP, 2015); [b]Refer to 2014 World Bank data; [c]Data refer to the most recent year available during the period 2005–2013 (UNDP, 2015); [d]Data refer to the most recent year available during the period 2002–2012 (UNDP, 2015); [e]Data refer to 2011 (ILO, 2011); [f]Refer to 2014 World Bank data.

The data used in this paper come from the Great Integrated Household Survey (GIHS) conducted by the National Administrative Statistics Department (DANE) and includes all of the monthly data from the year 2009. This cross-sectional survey has information at a micro-data level on the labor force, unemployment and informality of thirteen major Colombian cities and their greater metropolitan areas.[2]

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

Regarding definition of informality, in the literature there are at least two definitions regularly used: the productivity definition and the legalistic definition. The productive approach is based on the type of job and the firm size (usually firms with fewer than five persons), while the legalistic or social security approach informality corresponds to the lack of coverage of workers by mandated labor protections or firms avoiding taxation or other legal regulations. We opt by following the legalistic definition of informality for two reasons. First, given that the productivity definition of informality includes to all own-account workers (excluding administrative workers, professionals and technicians) and owner of very small firms, this definition overlaps with informal self-employment, which is not the target population of this study. Secondly, the legalistic definition is a much broader definition because it includes the possible presence of informal employment within large firms, which is not possible in the productivity definition. This can imply that using the productivity definition can yield lower informality rates than that using the legalistic definition.[4]

We define informal workers as those who are not covered by the social security system. More precisely, informal workers are those who are not covered by the health insurance and pension systems. Applying this condition, we have 36,293 (58.3%) formal workers and 25,985 (41.7%) informal workers. In Table 2, we provide descriptive statistics for the key variables for formal and informal workers.

As seen from Table 2, the average wage among formal workers is higher than the corresponding average among informal workers; on average, an informal worker earns 30% less than a formal worker. Regarding the variables that we can use to explain the variation in wages, there are several important differences between formal and informal employees. For example, although formal workers are on average similar in age to informal workers, years of tenure 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 workers. The informal sector has a higher percentage of individuals with primary and pre-primary education (52%), whereas the formal sector has a much higher percentage of individuals with secondary and tertiary education (74%). With respect to other personal characteristics, we can see that informal workers are less likely than formal workers to be male, head of the household and married. Finally, informal workers are more likely to work in firms with between 1 and 10 employees (77%), whereas formal workers are employed by firms with more than 51 employees (60%).

Figure 1 depicts the estimated kernel densities of the wages of formal and informal workers. Wage disparities between sectors are clearly visible, as wage

Table 2. Descriptive Statistics

	Formal workers		*Informal workers*		*Differences formal–informal*	*Total*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*Mean*	*SD*
Real hourly wage $	1.60	0.82	1.13	0.91	0.47	1.43	0.88
Age (years)	34.29	10.10	32.92	11.51	1.37	33.80	10.64
Education (years)	11.00	3.54	8.62	3.71	2.38	10.15	3.78
Tenure at job (years)	4.73	5.67	2.78	4.02	1.95	4.04	5.23
Education levels							
Less than primary	0.12	0.32	0.28	0.45	−0.16	0.18	0.38
Primary	0.14	0.34	0.26	0.44	−0.12	0.19	0.39
Secondary	0.62	0.48	0.42	0.49	0.20	0.55	0.49
Tertiary	0.12	0.32	0.03	0.17	0.08	0.08	0.28
Male	0.56	0.49	0.49	0.50	0.07	0.53	0.50
Head of household	0.43	0.49	0.35	0.47	0.08	0.40	0.49
Married	0.53	0.50	0.46	0.50	0.07	0.51	0.50
Size of firm							
1–10 employees	0.18	0.38	0.77	0.42	−0.59	0.39	0.49
11–50 employees	0.22	0.42	0.14	0.35	0.08	0.19	0.39
More than 51 employees	0.60	0.49	0.09	0.29	0.51	0.42	0.49
Sample size	36,293		25,985			62,278	

Notes: Author's calculations based on microdata from the 2009 GIHS. All differences in means between the formal and the informal sector are significant at the 1% level.

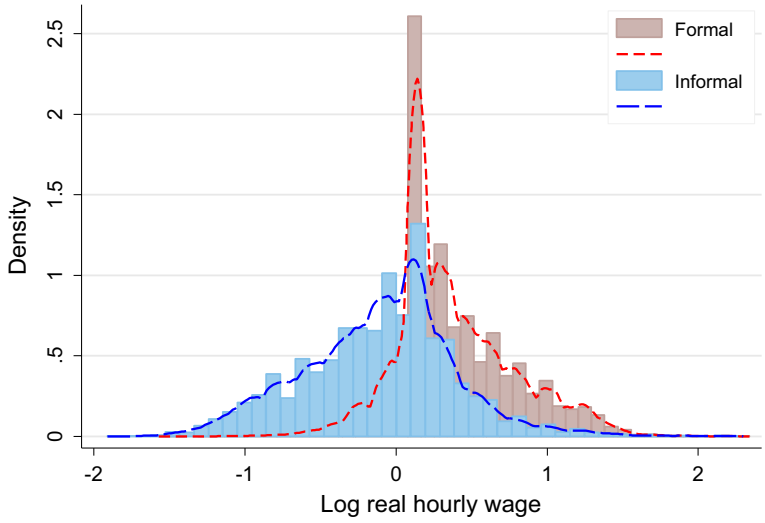

Figure 1. Kernel Density of Log Real Hourly Wage by Formal and Informal Sector [Colour figure can be viewed at wileyonlinelibrary.com]

distribution for formal workers is shifted to the right. The distribution of formal- and informal-sector wages and the wage gap between sectors by quantile—that is, the difference in log wages between formal workers and informal workers at each quantile of their respective distributions—is plotted in Figure 2. We can see that

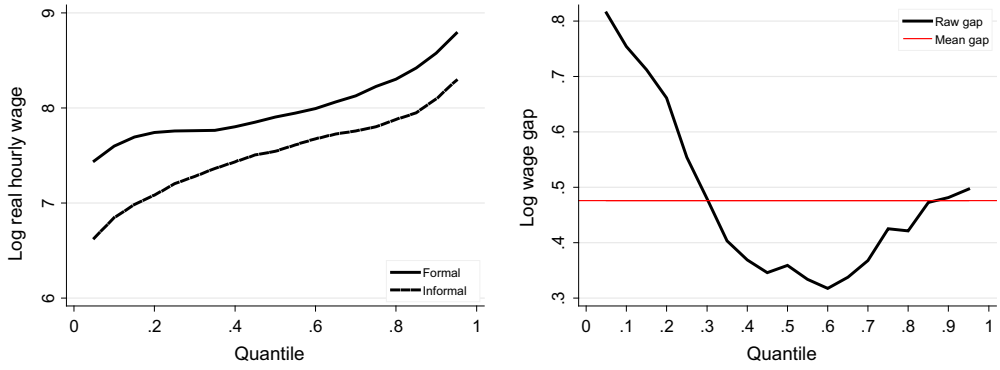

Figure 2. Wage Differentials between Formal and Informal Sector over Different Quantiles of the Wage Distribution [Colour figure can be viewed at wileyonlinelibrary.com]

the wage differential between sectors is positive throughout the entire wage distribution, with a large wage gap at the lower quantiles. The size of the wage differential decreases from 54% at the bottom end of the distribution to 30% at the median, then increases to approximately 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 substantial differences in human capital endowments and job opportunities at these points of the earnings distribution.

Similarly, at a city level, we can see that there are positive wage differences between sectors throughout the entire distribution and that the patterns differ across cities (see Figure 3). Pasto, Montería and Cartagena exhibit the largest wage gaps, with particularly large wage gaps at the lower quantiles. The common characteristic of these cities is that they present the highest levels of informality in Colombia (see Table 3) and therefore there is significant heterogeneity of jobs and workers in the informal sector. In these cities, the relative abundance of informal jobs is an important determinant for joining the informal sector. Regarding the largest and most developed cities, such as Bogotá, Medellín, and Cali, we can see that the wage differentials between sectors are smaller than those in the previously mentioned cities.

To account for these regional heterogeneities, we define three groups of cities. In the following section, we describe these groups and present descriptive statistics of their labor markets.

Group of Cities and their Labor Markets

We have divided the total sample into three sub-groups of cities: one group of central and more developed cities and two groups of peripheral cities with significant informal sectors.

The first group of cities (Group 1) includes Bogotá, Medellín, Cali, Bucaramanga, Manizales, Pereira and Ibagué. This group comprises the largest industrial and most dynamic cities in Colombia and forms the core of the country's economic activity. They represent 0.7% of the national territory and according to the 2005 Population Census, approximately 45% of the urban population is concentrated in these seven cities. 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.[5] It

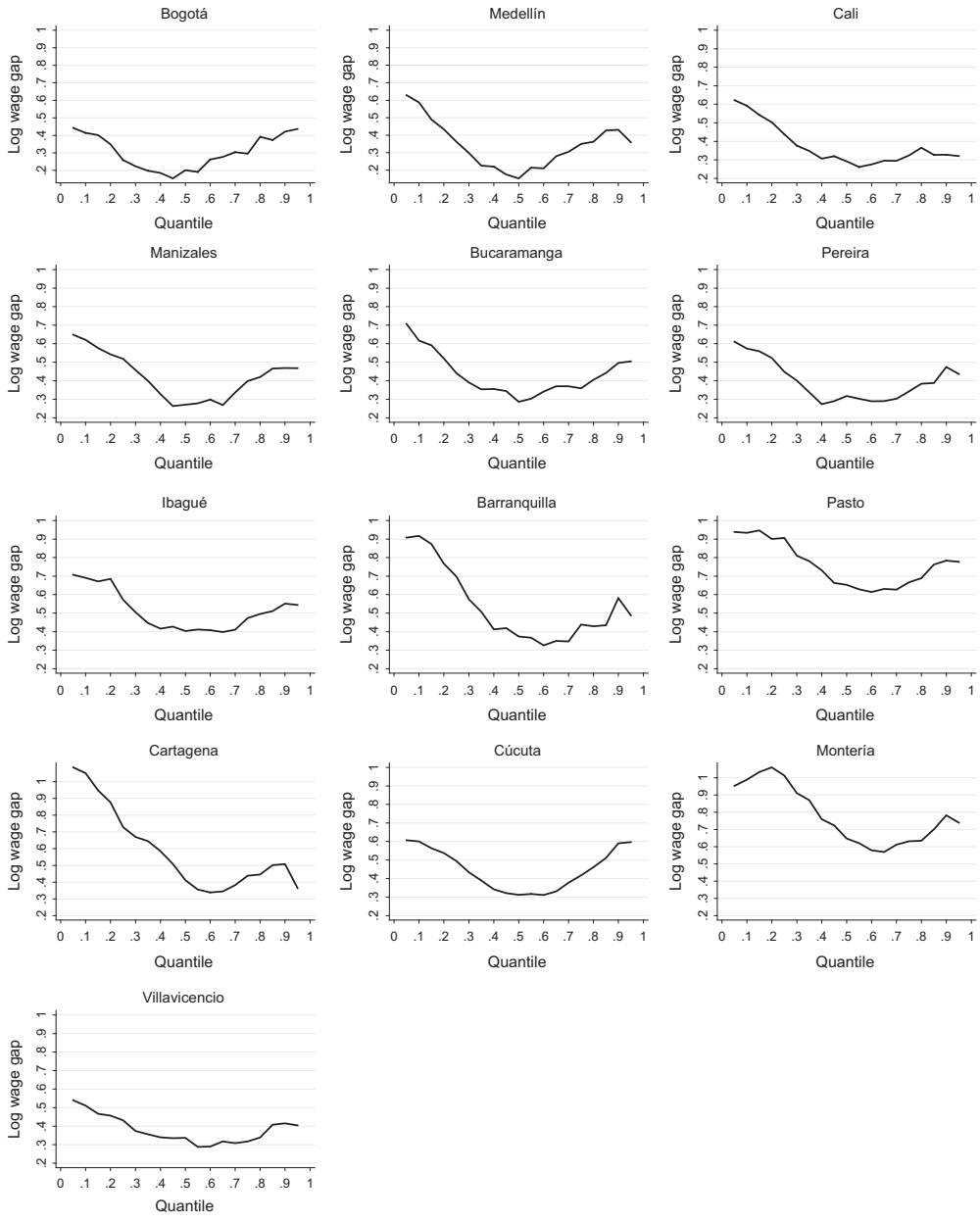

Figure 3. Wage Differentials between Formal and Informal Sector over Different Quantiles of the Wage Distribution by Metropolitan Area

is also worth highlighting that the rankings of these cities in terms of the extent of informality have been relatively stable over time. In this respect, García (2008, 2011) and Galvis (2012) find that from a regional perspective, these cities show consistently lower informality levels than cities outside of this region (see Figure 4).

The second group of cities (Group 2) is composed of Barranquilla and Cartagena. Although these cities are among the most urbanized cities in Colombia and represent an important economic dynamic, their tourist and export vocations

Table 3. Informality Rate by Metropolitan Area

	Social security approach	*ILO approach*
Colombia	57.59	51.43
Medellín	45.87	46.66
Bogotá	51.12	45.67
Manizales	51.66	48.18
Pereira	57.00	51.69
Cali	64.98	56.49
Bucaramanga	65.31	59.43
Ibagué	68.84	60.23
Barranquilla	70.03	56.74
Villavicencio	70.56	65.03
Cartagena	72.32	56.27
Pasto	72.26	62.00
Montería	74.77	68.10
Cúcuta	76.95	65.93
Correlation	0.655**	

Notes: Author's calculations based on microdata from the 2009 GIHS. We included government employees, employers and self-employees to calculate the informality rate. ILO (International Labor Organization) **Denotes significance at 5%.

distinguish them from other cities. The country's main ports are located in Barranquilla and Cartagena, and they have an important industrial cluster associated with the petrochemical–plastic sector.[6] Nevertheless, their socioeconomic and labor market indicators are unfavorable. Indeed, these two cities show some of the highest poverty, inequality and informality levels among the main cities of Colombia (Bonilla, 2008; Galvis, 2009). As shown in Figures 4 and 5, Cartagena and Barranquilla, along with Montería and Cúcuta, exhibit the highest levels of unsatisfied basic needs (UBN) and informality. The emphasis on tourism in the Caribbean region and the relatively low capacity for job creation in highly productive sectors, which is because most firms in these sectors are large companies with high capital intensity and significant export activities, have led to a process of economic tertiarization. The service sector has little impact on the competitiveness of the other sectors and generates many jobs, but these jobs are of low quality in terms of pay and working conditions (Bonet, 2005, 2007; Bonilla, 2010; Cepeda, 2011; Acosta, 2012).

The third group of cities includes Cúcuta, Montería, Pasto and Villavicencio (Group 3). These are the least developed cities; they are located in peripheral areas, and their activities are greatly influenced by agriculture, mining and commerce. Pasto and Cúcuta are border cities, the former sharing a border with Ecuador and the latter sharing a border with Venezuela, and this is a common characteristic that can influence the types of activity and employment generated, particularly with respect to jobs in the commerce (legal and illegal) and currency exchange areas (Bonet, 2007; García, 2008, 2011). Villavicencio is the capital of the department of Meta, which has the largest oil fields in the country [the department of Meta produces 47% of Colombia's oil (Departamento Administrativo Nacional de Estadística, 2011)]. Villavicencio and Montería are the capitals of the two main cattle farming regions of the country and therefore, their respective economies are

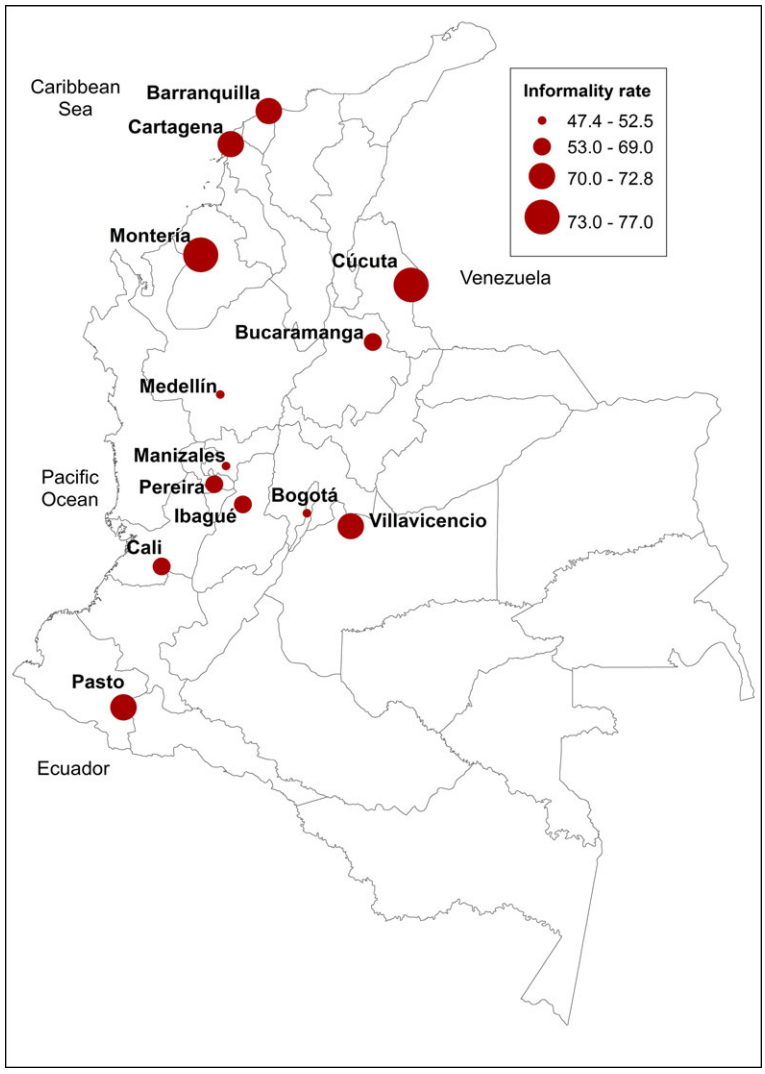

Figure 4. Informality Rate by City, 2009 [Colour figure can be viewed at wileyonlinelibrary.com]
Source: Table 3.

based mainly on these activities. Furthermore, these two regions are considered conflict zones owing to the presence of paramilitary groups, guerrillas and drug trafficking activities, which influences not only the economic activity but also the social, political and cultural make-up of the regions (Viloria de la Hoz, 2009; Sánchez et al., 2012). In contrast to the first group of cities, this group shows the highest levels of informality, with Cúcuta having the highest rate (77%) (see Figure 4). According to García (2008, 2011) and Galvis (2012), informality is more prevalent in less prosperous cities, which are usually located at the periphery of the country and have fewer resources and less industrial development than cities in the center of the country.

Table 4 shows descriptive statistics of the labor markets formed by the three groups of cities. As expected, there is a higher percentage of informal workers in

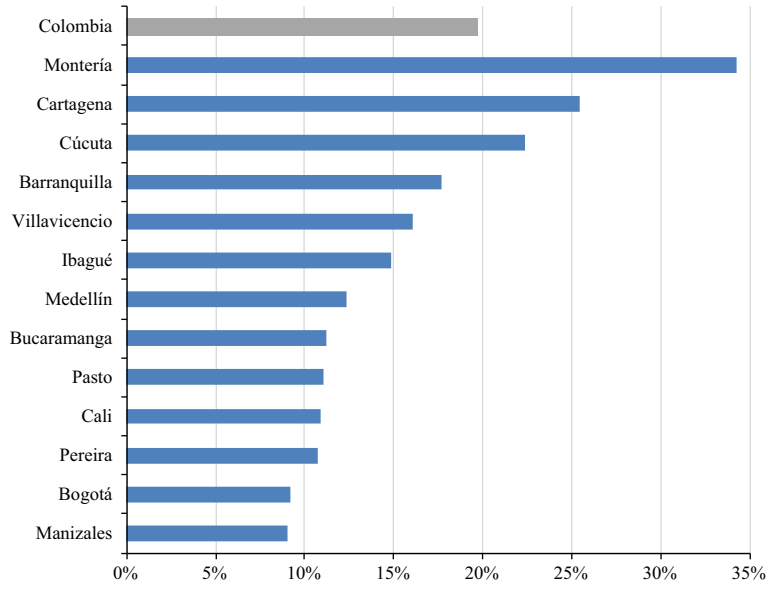

Figure 5. Percentage of Unsatisfied Basic Needs (UBN) by City [Colour figure can be viewed at wileyonlinelibrary.com]
Source: 2005 Population Census [Departamento Administrativo Nacional de Estadística (DANE), Bogotá].

City Groups 2 and 3 (47% and 56%, respectively) than in Group 1 (35%). We can also see that formal workers earn more than informal workers and that the differences are greater in City Group 2. The wage gap between sectors in City Group 2 is 61%, whereas in Groups 1 and 3, the wage differentials are 36% and 31%, respectively (see Table 5 for differences within groups of cities).

With regard to education, we 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, the group of less developed cities (Group 3) has a higher percentage of informal workers with primary and less than primary education (56%) than the two groups of more developed cities (53% in Group 1 and 43% in Group 2), whereas the two latter groups have a higher percentage of informal workers with tertiary education (16% and 21%, respectively) than Group 3 (10%). There are also striking differences in education levels in the formal sector between groups of cities. Interestingly, roughly one-half of the formal workers in the group of Caribbean coast cities (Group 2) have tertiary education, whereas this percentage barely reaches 37% in the group of more developed cities (Group 1). The reason for these results may relate to 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 regard, Arango (2011), who studied differences in the main labor market variables among major Colombian cities from 2001 to 2011, found that Barranquilla and Cartagena (along with Bogotá) are characterized by the highest worker education rates in Colombia.

Table 4. Mean of Variables by Sector and Groups of Cities

	Formal workers	*Informal workers*	*Total*	*Differences formal–informal*
Group 1				
Real hourly wage $	1.60	1.18	1.46	0.42
Age (years)	34.21	32.96	33.81	1.24
Education (years)	10.92	8.61	10.16	2.31
Tenure at job (years)	4.66	2.67	4.01	1.98
Education levels				
Less than primary	0.12	0.27	0.18	−0.15
Primary	0.15	0.27	0.19	−0.12
Secondary	0.62	0.42	0.55	0.20
Tertiary	0.11	0.03	0.08	0.08
Male	0.55	0.48	0.53	0.07
Head of household	0.43	0.36	0.41	0.07
Married	0.52	0.46	0.50	0.07
Size of firm				
1 – 10 employees	0.18	0.77	0.38	−0.59
11 – 50 employees	0.22	0.14	0.19	0.08
More than 51 employees	0.59	0.09	0.43	0.50
Sample size	25,368 (65%)	13,723 (35%)	39,091	
Group 2				
Real hourly wage $	1.59	0.99	1.31	0.60
Age (years)	35.38	33.90	34.71	1.48
Education (years)	11.89	9.24	10.67	2.65
Tenure at job (years)	5.73	3.91	4.90	1.82
Education levels				
Less than primary	0.06	0.25	0.15	−0.19
Primary	0.10	0.22	0.15	−0.12
Secondary	0.67	0.48	0.58	0.19
Tertiary	0.17	0.05	0.12	0.12
Male	0.62	0.46	0.55	0.16
Head of household	0.44	0.31	0.38	0.13
Married	0.65	0.50	0.58	0.15
Size of firm				
1–10 employees	0.09	0.67	0.36	−0.58
11–50 employees	0.23	0.17	0.21	0.06
More than 51 employees	0.67	0.16	0.44	0.51
Sample size	4,394 (53%)	3,832 (47%)	8,226	
Group 3				
Real hourly wage $	1.58	1.21	1.22	0.37
Age (years)	34.10	31.67	32.70	2.43
Education (years)	11.21	8.06	9.39	3.15
Tenure at job (years)	4.54	2.41	3.31	2.13
Education levels				
Less than primary	0.11	0.33	0.24	−0.22
Primary	0.11	0.25	0.19	−0.14
Secondary	0.62	0.38	0.49	0.24
Tertiary	0.16	0.28	0.08	0.13
Male	0.53	0.54	0.54	−0.01
Head of household	0.42	0.37	0.39	0.05

Table 4. Continued

	Formal workers	*Informal workers*	*Total*	*Differences formal–informal*
Married	0.56	0.46	0.50	0.10
Size of firm				
1–10 employees	0.21	0.81	0.56	−0.60
11–50 employees	0.22	0.14	0.17	0.08
More than 51 employees	0.57	0.05	0.27	0.52
Sample size	6,531 (44%)	8,430 (56%)	14,961	

Notes: Author's calculations based on microdata from the 2009 GIHS. All differences in means between the formal and the informal sector are significant at the 1% level.

Table 5. Mean of Variables by Groups of Cities

	Group 1	*Group 2*	*Group 3*	*Difference G1 – G2*	*Difference G1 – G3*	*Difference G2 – G3*
Real hourly wage ($)	1.46	1.31	1.22	0.15	0.24	0.09
Age (years)	33.81	34.71	32.70	−0.90	1.11	2.01
Education (years)	10.16	10.67	9.39	−0.51	0.77	1.28
Tenure at job (years)	4.01	4.90	3.31	−0.89	0.70	1.59
Education levels						
Less than primary	0.18	0.15	0.24	0.03	−0.06	−0.09
Primary	0.19	0.15	0.19	0.04	0.00	−0.04
Secondary	0.55	0.58	0.49	−0.03	0.06	0.09
Tertiary	0.08	0.12	0.08	−0.04	0.00	0.04
Male	0.53	0.55	0.54	−0.02	−0.01	0.01
Head of household	0.41	0.38	0.39	0.03	0.02	−0.01
Married	0.50	0.58	0.50	−0.08	0.00	0.08
Size of firm						
1–10 employees	0.38	0.36	0.56	0.02	−0.18	0.20
11–50 employees	0.19	0.21	0.17	−0.02	0.02	0.04
More than 51 employees	0.43	0.44	0.27	−0.01	0.26	0.17
Sample size	39,091	8226	14,961			

Notes: Author's calculations based on microdata from the 2009 GIHS. All differences in means between city groups are significant at the 1% level.

Another difference between the groups of cities is found in the variable for firm size. As shown in Table 4, there is a substantially higher proportion of informal workers conducting activities in medium and large firms in City Group 2 (approximately 33%) than in City Groups 1 and 3 (23% and 19%, respectively). This difference is due to the presence of large firms associated with the industrial cluster of petrochemical products and export activities in the Caribbean coast cities. These activities provide important productive linkages that benefit not only the formal sector but also the informal sector (Departamento Nacional de Paneación, 2007; Acosta, 2012).

Figure 6 shows the distribution of informal-sector employment across two-digit industries by city group. Most employment in the informal sector is in the service

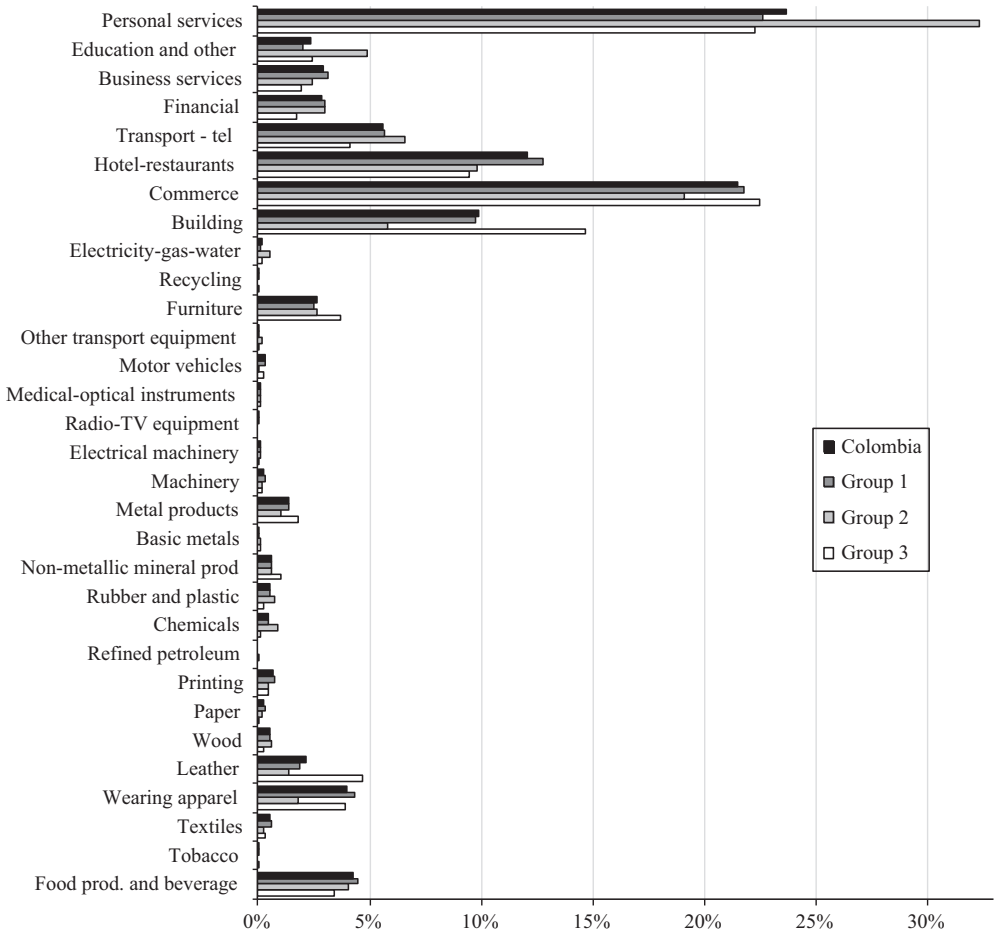

Figure 6. Distribution of Informal Employment across Sectors

industry (approximately 80%), with the greatest concentration of informal employment in the personal services and commerce areas. The group of Caribbean coast cities is notable in this respect because more than one-third of informal workers in these cities are employed in the personal services sector. This result reflects the significant influence of tourism-related activities on the economy of this region. Within the industrial sector overall, informal employment is concentrated in the food and beverages and apparel sectors, followed by the furniture, leather and metal products sectors. City Group 3 has a relatively high proportion of informal employment in the leather and apparel sectors, which may reflect the impact of border and cattle farming activities on the productive structures of these cities. As noted earlier, the sectoral composition of production in the cities is an important factor in explaining informality at a regional level.

Turning to the wage gap between sectors, Figure 7 presents the results at each quantile of the respective distributions by city group. This figure shows that the wage differential between the formal and informal sectors are considerably lower in City Group 1, especially at the bottom end of the distribution. Interestingly, the wage differential at the bottom of the distribution is higher in City Group 2 than in City Group 3. Again, this result may be due to the strong influence of personal

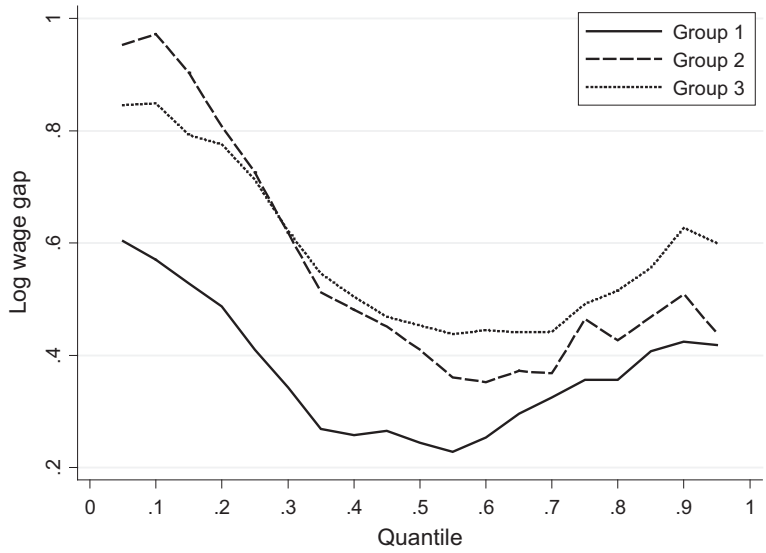

Figure 7. Wage Differentials between the Formal and Informal Sector over Different Quantiles of the Wage Distribution by Group of Cities

service activities associated with tourism in the Caribbean coast cities, the majority of which are informal activities with very low qualifications and wages.

4. Estimation Procedure

To determine which factors influence the wage gap between the formal and informal sectors, taking into account the heterogeneity of workers throughout the distribution and the differences that can exist between groups of cities, we use quantile decomposition methodology. Quantile regression methods are particularly useful for analyzing the decomposition of wage gaps at different points in the distribution in situations characterized by large disparities, as is the case in a country like Colombia (Bonilla, 2008, 2009). Furthermore, this methodology makes it possible to account for wage heterogeneity among groups of individuals and for differences in the impact of the determinants of wages and wage gaps by employment type at different points in the distribution (Machado and Mata, 2005). Thus, the results are more complete than those obtained by ordinary least squares (OLS).[7]

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

In our analysis, the potential selection bias in the estimation of wage equations may result from the self-selection of individuals into formal or informal employment. There are several observable and unobservable factors that may affect whether a worker participates in the formal or informal sector. 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 sectors. However, the methodology proposed by Buchinsky (1998) does not impose the restriction of normality and instead uses a semiparametric method developed by Ichimura (1993), which makes no assumptions about the distribution of the residuals.

Following Buchinsky (1998), we let I_i represent the sector in which worker i is employed, taking a value of 1 for the informal sector and 0 for the formal sector. 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 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, respectively, 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:

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

and

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

where $Q_\theta(Y|X = x)$ is the conditional quantile at θth quantile. The Machado–Mata procedure entails the generation of a random sample of size n from a uniform distribution $U[0,1]$:$u_1, u_2, \ldots, u_n$ and the calculation of 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 use the estimated result and a random sample of n 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 $\hat{y}_{inf} = \tilde{x}_{inf}'\hat{\beta}^{for}(u)$, which is the wage distribution of the informal sector that results from assigning the returns of the formal sector but keeping the observed characteristics of the informal sector unchanged. 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 θth quantile can be decomposed as:

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

The first term on the right-hand side of expression (5) refers to the characteristic 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 θth

quantile. The second term calculates the counterfactual value of the wage gap if the informal workers retain their observed characteristics but are paid for them like the formal workers. This term represents the coefficient effects. We use a bootstrap procedure to estimate standard errors for the reported components of the decomposition.

Because we only observe the wages of workers who actually work in the informal or formal sector, these workers are not drawn randomly from the distribution of individuals and thus a selection bias may exist when we estimate the wage equations. Consequently, to correct for selection and obtain unbiased estimates of β in the quantile wage equations, Buchinsky (1998) proposes introducing an extra term into the quantile regressions, namely,

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

and

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

The vector Z also includes the set of observable characteristics that influence wages (that is, the Xs), but for identification purposes 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 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 that are estimates of the constant and slope coefficients from the probit regression of I_i on the index $z'\hat{\gamma}$.

To estimate the coefficients γ in equation (1), Buchinsky (1998) proposes using the semiparametric least squares (SLS) method proposed by Ichimura (1993). Because we estimate a semiparametric 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; 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 proposed by Albrecht et al. (2009) to adjust for selection is the following:

(1) Estimate γ using a 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.[8]

Compare the simulated distributions to decompose the estimated wage gap between sectors.

5. 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 log wage distributions of formal and informal workers and to identify 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 the returns on those characteristics.[9]

SLS Estimation and the Quantile Regression Models

As mentioned in section 3, the first step involved the estimation of the SLS model for the probability of being informal and the second step entailed the estimation of the quantile regression models for the wage equation, including the power series expansion to address selection. In both the probability and quantile regression models, we included variables for education levels and gender, and dummies for firm size, industry and occupation. 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 in the formal sector; the average number of years of education of household members (as a measure of the educational environment of the household); whether the individual is the head of the household; and marital status. Table 6 shows the results for the probit and SLS probability models for the total sample and by city group.

To test whether the probability of being informal effectively relies on the assumption of normality for the residuals, we performed a Hausman test. As noted by Buchinsky (1998), the SLS estimate is consistent with and independent of the distribution of the residuals, whereas the probit estimate is efficient under normally distributed residuals; therefore, a Hausman-type test can be performed. Test statistics for the Hausman test, which are reported at the bottom of Table 6, clearly indicate that for the total sample and city groups, 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 6 indicate that overall, younger or older (there is a U-shaped pattern between age and informality), less educated, female, non-head of household and unmarried individuals are more likely to work in the informal sector. The higher probabilities of individuals in less important positions in the family may indicate that secondary household incomes are earned in the informal sector.

Regarding 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. The presence at home of relatives who work in the formal sector has a negative effect on the probability of working in the informal sector; this effect is greater in the group of

Table 6. 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	3.895***	3.928***	3.928	4.089***	4.089	3.559***	3.559	4.153***	4.153
	(50.46)	(42.23)	(·)	(39.52)	(·)	(15.53)	(·)	(25.15)	(·)
Age	−0.096***	−0.097***	−0.097	−0.107***	−0.107	−0.068***	−0.068	−0.097***	−0.097
	(−23.30)	(−19.52)	(·)	(−19.64)	(·)	(−5.74)	(·)	(−10.84)	(·)
Age^2	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	(18.97)	(16.01)	(57.82)	(16.54)	(64.13)	(4.42)	(28.52)	(8.36)	(32.92)
Education levels									
Primary	−0.133***	−0.124***	−0.129***	−0.097***	−0.100***	−0.155**	−0.027	−0.238***	−0.240***
	(−6.21)	(−4.83)	(−4.23)	(−3.47)	(−3.06)	(−2.27)	(−0.53)	(−5.00)	(−4.20)
Secondary	−0.469***	−0.444***	−0.449***	−0.430***	−0.460***	−0.583***	−0.361***	−0.657***	−0.602***
	(−23.58)	(−18.61)	(−11.76)	(−16.32)	(−11.36)	(−9.50)	(−5.71)	(−15.43)	(−8.26)
Tertiary	−0.697***	−0.687***	−0.859***	−0.647***	−0.800***	−0.896***	−0.667***	−0.977***	−0.946***
	(−20.13)	(−16.53)	(−11.08)	(−13.07)	(−9.51)	(−10.17)	(−5.71)	(−13.35)	(−7.25)
Male	−0.144***	−0.149***	−0.179***	−0.178***	−0.204***	−0.154***	−0.141***	−0.139***	−0.104***
	(−9.20)	(−7.95)	(−7.43)	(−8.43)	(−7.57)	(−3.66)	(−3.76)	(−4.05)	(−2.64)
Head of household	−0.147***	−0.169***	−0.198***	−0.147***	−0.183***	0.111**	−0.121***	−0.142***	−0.145***
	(−9.04)	(−8.68)	(−7.24)	(−6.73)	(−6.31)	(−2.48)	(−2.99)	(−4.07)	(−3.40)
Married	−0.041***	−0.046***	−0.062***	−0.049**	−0.051**	−0.060	−0.042	−0.135***	−0.143***
	(−2.79)	(−2.59)	(−2.76)	(−2.41)	(−2.06)	(−1.51)	(−1.27)	(−4.27)	(−3.55)
Presence of children at home	0.056***	0.067***	0.079***	−0.025	0.045*	0.052	0.018	0.076**	0.078**
	(3.81)	(3.79)	(3.68)	(1.24)	(1.86)	(1.31)	(0.58)	(2.45)	(2.29)
Other relatives working as formal	−0.372***	−0.373***	−0.386***	−0.228***	−0.242***	−0.479***	−0.302***	−0.332***	−0.366***
	(24.43)	(20.44)	(12.43)	(−11.25)	(−8.57)	(−11.55)	(−6.03)	(−9.59)	(−7.08)
Education of household	0.011***	0.011***	0.003	−0.012**	−0.022***	0.015*	0.001	0.005	−0.001
	(4.05)	(3.22)	(0.75)	(−3.27)	(−4.52)	(1.85)	(0.16)	(0.85)	(0.27)

Table 6. Continued

	Total sample			*Group 1*		*Group 2*		*Group 3*	
	Probit	*Probit*[a]	*SLS*	*Probit*	*SLS*	*Probit*	*SLS*	*Probit*	*SLS*
Size of firm									
11–50 employees	−0.979***	−0.986***	−1.018***	−0.961***	−0.988***	−1.166***	−0.681***	−1.043***	−1.184***
	(−57.40)	(−48.31)	(−15.71)	(−41.59)	(−15.58)	(−24.10)	(−7.39)	(−29.21)	(−9.75)
More than	−1.615***	−1.628***	−1.818***	−1.553***	−1.684***	−1.772***	−1.267***	−1.870***	−2.167***
51 employees	(−97.47)	(−82.34)	(−16.24)	(−68.67)	(−16.13)	(−38.67)	(−7.77)	(−51.10)	(−10.42)
Observations	62,278	43,595		39,091		8226		14,961	
Hausman test		398.0		439.8		934.3		56.2	
p-value		[0.000]		[0.000]		[0.000]		[0.000]	

Notes: Author's calculations based on microdata from the 2009 GIHS. ***,**,* Denote significance at 1%, 5% and 10% levels, 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. aGiven 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.

Caribbean coast cities. In addition, a household with a higher education level has a negative effect on the likelihood of working in the informal sector, especially in City Group 1. Clearly, family environment has a significant effect on the decision to work in the informal sector.

Finally, the variables for firm size are significant and show that as the size of the firm increases, the probability of participating in the informal sector decreases. This effect is greater in City Groups 2 and 3 than in City Group 1.

As described above, in the second step, we used the estimates from the SLS to calculate the power series expansion and introduced this term into the quantile regression models to correct for selectivity. To calculate the correction term, we included two terms of orthogonal polynomials in the series expansion.[10] At the same time, to identify the constant term in the wage equations, we used a subsample of workers with a high probability of being informal, namely, workers who are younger or older, less educated (less than primary education), with the presence of children at home and with relatives who work in the informal sector.

The results for corrected quantile regressions show that in City Groups 2 and 3, as well as in the total sample, most selection terms are statistically significant, whereas not all of these terms are significant in more developed cities (Group 1).[11] These results are confirmed by the *F*-test of joint significance of the selection terms, the results of which are reported at the bottom of the Table 7. These results indicate the presence of sample selection bias for individuals across the entire 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 7 summarizes the results for quantile regressions at the three representative quantiles.

Table 7 shows that in City Group 2, informal workers receive higher returns on education than formal workers, particularly at high quantiles. Similar results are found at the median and lower quantiles of the distribution in City Group 3. Regarding 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 that this effect is especially high at low quantiles in the informal sector and similar in magnitude across city groups. An additional year of tenure in a job has a positive impact on wages; this effect is relatively constant across the distribution in the formal sector, independent of the city group. An additional year of tenure also has a positive effect in the informal sector, but this effect decreases across the distribution.

Regarding the gender variable, the results reveal substantial 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 10th percentile are approximately 15% lower than a man's expected earnings. Similar results are found for City Groups 1 and 2, but at high quantiles, where the difference in wages between female and male informal workers is approximately 11%.

Decomposition Results

In this section, we present the results of the decomposition. Figure 8 plots the estimated wage gap, correcting for selection, for the total sample and by city group.

As shown in Figure 8, a significant positive wage gap remains across the entire distribution of the total sample, with a large gap at the bottom of the distribution.

This wage gap decreases as one moves from the bottom to the top of the distribution, which is consistent with greater freedom of choice between the formal and informal sectors as workers move up the distribution. At the top end, workers in the informal sector may to some extent be willing to accept lower earnings to avoid contributing to social protection systems that are perceived as ineffective.[12]

Regarding the contribution of each set of factors (coefficients and characteristics), much of the wage gap at the bottom of the distribution is due to informal workers being paid less than formal workers for the remunerated characteristics. The coefficient effects decrease over the distribution, whereas the characteristic effects increase, particularly toward the upper end of the distribution, where characteristic effects greatly exceed coefficient effects. These results indicate that low-paid informal workers earn less not only because they are less skilled but also because they earn lower returns on such skills, whereas highly paid informal workers earn less because formal workers have superior skills.

It is possible to distinguish at least two groups of informal workers that occupy very different positions relative to the formal sector. First, the group of informal workers at the bottom of the distribution represents the disadvantaged segment. In this segment, even workers with characteristics identical to those of formal workers earn lower rates of return on those characteristics. In Fields' (1990) formulation, this segment is deemed the "easy-entry informal sector" and is characterized by free-entry, low-wage employment that is undesirable relative to employment in the formal sector. This segment includes a large amount of residual and underemployed labor.

The second group, informal workers at the top of the distribution, includes workers with a higher wage and a lower wage differential than the informal workers at the bottom of the distribution. For informal workers in the second group, the rates of return on characteristics are similar to those in the formal sector. This group of informal workers constitutes 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 in the formal sector, they nonetheless find informal activities more profitable than formal activities. In this regard, Maloney (1999) argues that highly paid informal workers have specific characteristics or abilities that may imply a non-wage advantage over potential earnings in the formal sector. Additionally, he claims that the high administrative costs of social security combined with the low quality of its services may discourage some workers from employment in the formal sector. In this case, informality can be viewed as a deliberate choice to avoid administrative costs for services that are perceived to have a low relative value.

Different patterns in the wage gap and its determinants are evident at the city group level. The pattern in City Group 1 is similar to that of the total sample: the wage gap is positive over the entire distribution, the coefficient effect is greater at the bottom and median segments of the distribution, and the characteristic effect explains most of the wage gap at the top end of the distribution. In City Groups 2 and 3, the wage gap between sectors is smaller over the entire distribution; indeed, the wage gap trends toward zero at the top end of the distribution of City Group 2 and is negative between the 30th and 70th quantiles of the distribution of City Group 3. The lower wage gap may suggest that in cities where informal activities constitute the main source of income, the informal sector is no longer considered the poor, marginal sector. This result is in line with Marcouiller et al. (1997) and Staneva and Arabsheibani (2014), who find wage premiums associated with work in

Table 7. Quantile regressions (*y = Log real hourly wage*)

	Total sample					
	Formal			Informal		
	10%	50%	90%	10%	50%	90%
Constant	7.447***	7.559***	7.638***	6.514***	7.178***	7.638***
	(289.37)	(464.22)	(176.19)	(231.94)	(425.42)	(306.30)
λ	0.173***	0.027***	0.001	–0.064***	–0.022*	–0.049**
	(12.07)	(3.01)	(0.02)	(–2.91)	(–1.67)	(–2.48)
Education levels						
Primary	0.075***	0.042***	0.092***	0.075***	0.083***	0.092***
	(6.20)	(5.43)	(4.48)	(4.11)	(7.37)	(5.67)
Secondary	0.152***	0.121***	0.293***	0.206***	0.196***	0.214***
	(12.39)	(15.59)	(14.32)	(9.97)	(16.04)	(12.12)
Tertiary	0.330***	0.503***	0.658***	0.499***	0.520***	0.598***
	(19.16)	(48.48)	(23.81)	(11.05)	(18.76)	(14.92)
Experience	0.002	0.003***	0.006***	0.014***	0.010***	0.006***
	(0.30)	(6.13)	(4.14)	(8.39)	(9.30)	(4.27)
Experience^2	–0.001	–0.001***	–0.001*	–0.001***	–0.001***	–0.001**
	(–1.48)	(–5.22)	(–1.88)	(–6.64)	(–7.10)	(–2.30)
Tenure	0.008***	0.010***	0.015***	0.030***	0.017***	0.020***
	(6.07)	(11.57)	(6.42)	(8.35)	(7.83)	(6.60)
Tenure^2	–0.001*	0.001	–0.001	–0.001***	–0.001***	–0.001***
	(–1.87)	(0.11)	(–0.83)	(–6.56)	(–4.76)	(–5.08)
Male	–0.011*	0.036***	0.095***	0.108***	0.105***	0.119***
	(–1.86)	(8.98)	(9.21)	(6.74)	(10.74)	(8.48)
Size of firm						
11–50	0.025**	0.047***	0.061***	0.265***	0.151***	0.157***
employees	(2.09)	(6.19)	(3.11)	(12.38)	(11.38)	(8.20)
More than 51	0.024*	0.083***	0.152***	0.395***	0.230***	0.326***
employees	(1.79)	(9.49)	(6.66)	(11.42)	(11.29)	(11.31)
Observations		25,344			18,251	
Joint test on selection terms						
F-statistic		27.50			3.83	
p-value		0.0000			0.0090	

the informal sectors of Mexico and Tajikistan, respectively. These authors claim that the scarcity of regulations, weak enforcement of labor laws and higher tolerance for informal activities can imply higher wage benefits associated with employment in the informal sector.

Regarding the contribution of the coefficient and characteristic effects on the wage differential, we can see that at the low quantiles of the group of Caribbean coast cities, the coefficient effect is positive and makes an important contribution to the wage gap, whereas at the top half of the distribution, the characteristic effect is greater than the coefficient effect, which is negative.

These results suggest that at the lower quantiles, the levels of human capital and other remunerable characteristics are lower in the informal sector than in the formal sector, but more importantly, the rates of return on those characteristics are

Group 1					
Formal			Informal		
10%	50%	90%	10%	50%	90%
7.245***	7.525***	7.662***	6.702***	7.287***	7.719***
(384.30)	(709.99)	(262.73)	(201.81)	(478.11)	(250.72)
0.067***	0.038***	0.091***	0.071***	0.089***	0.091***
(4.83)	(4.84)	(4.20)	(3.35)	(8.79)	(4.50)
0.164***	0.119***	0.296***	0.179***	0.196***	0.216***
(12.20)	(15.72)	(14.20)	(7.74)	(18.43)	(10.27)
0.406***	0.536***	0.666***	0.388***	0.547***	0.765***
(21.89)	(51.48)	(22.74)	(7.37)	(21.57)	(15.11)
0.002***	0.004***	0.006***	0.012***	0.010***	0.005**
(2.22)	(7.42)	(3.58)	(6.57)	(10.47)	(2.55)
–0.001**	–0.001***	–0.001**	–0.001***	–0.001***	–0.001
(–2.32)	(–6.66)	(–1.96)	(–6.40)	(–8.29)	(–1.29)
0.011***	0.011***	0.019***	0.045***	0.019***	0.026***
(6.54)	(11.92)	(7.64)	(10.14)	(8.59)	(5.86)
–0.001***	0.001	–0.001**	–0.002***	–0.001***	–0.001***
(–3.06)	(0.29)	(–1.98)	(–10.50)	(–5.18)	(–4.02)
0.016***	0.052***	0.112***	0.083***	0.094***	0.105***
(2.06)	(12.16)	(9.96)	(4.68)	(10.84)	(6.21)
0.100***	0.059***	0.072***	0.205***	0.130***	0.134***
(9.09)	(9.52)	(4.31)	(9.60)	(12.12)	(6.31)
0.151***	0.106***	0.151***	0.189***	0.125***	0.221***
(15.78)	(19.75)	(10.46)	(7.09)	(9.51)	(8.69)
	25,368			13,723	
	1.58			1.49	
	0.1919			0.2154	

lower in the informal sector than in the formal sector. The presence of more disadvantaged individuals in this segment of informal workers might support the hypothesis of segmentation for these workers. It seems reasonable to conclude that for informal workers at these points in the wage distribution, there is no room in the formal sector, and informality is a last-resort strategy to escape unemployment.

In contrast, at the top end of the distribution, informal workers obtain higher returns on their characteristics than formal workers do and the wage gap is almost zero—indeed, it is only 2% at the 75th quantile, albeit it increases to 10% at the 95th quantile. The improved conditions at this part of distribution stem from the presence of informal workers with particular skills who, despite having job opportunities in the formal sector because of such skills, prefer the combination of monetary rewards and greater flexibility (in terms of working hours, work

Table 7. (Continued)

	Group 2					
	Formal			Informal		
	10%	50%	90%	10%	50%	90%
Constant	7.553***	7.643***	7.907***	6.464***	7.040***	7.445***
	(84.36)	(127.50)	(62.81)	(129.17)	(212.71)	(144.47)
λ	0.148***	0.080***	0.188***	0.212***	0.083*	0.113*
	(3.07)	(2.60)	(2.72)	(3.26)	(1.80)	(1.71)
Education levels						
Primary	0.108***	0.047**	0.067	0.068**	0.056**	0.110***
	(3.36)	(2.11)	(1.43)	(2.08)	(2.49)	(3.39)
Secondary	0.143***	0.096***	0.186***	0.150***	0.204***	0.230***
	(4.52)	(4.27)	(3.83)	(3.95)	(7.90)	(6.10)
Tertiary	0.234***	0.424***	0.505***	0.405***	0.517***	0.653***
	(5.38)	(14.70)	(8.07)	(5.79)	(10.02)	(8.89)
Experience	–0.005**	0.001	0.003	0.010***	0.007***	0.003
	(–2.21)	(0.53)	(1.19)	(3.32)	(3.26)	(0.87)
Experience^2	0.001*	–0.001	–0.001	–0.001**	–0.001**	–0.001
	(1.88)	(–0.44)	(–0.39)	(–2.49)	(–2.15)	(–0.20)
Tenure	0.007***	0.015***	0.015***	0.028***	0.025***	0.023***
	(2.64)	(7.43)	(3.24)	(6.11)	(7.19)	(4.94)
Tenure^2	–0.001	–0.001***	–0.001	–0.001***	–0.001***	–0.001***
	(–1.53)	(–3.20)	(–0.70)	(–4.80)	(–4.35)	(–2.78)
Male	–0.019	0.022**	0.057***	0.042	0.104***	0.112***
	(–1.29)	(2.25)	(2.83)	(1.54)	(5.56)	(4.18)
Size of firm						
11–50 employees	0.093***	0.058***	0.036	0.272***	0.182***	0.140***
	(2.70)	(2.39)	(0.74)	(6.62)	(6.20)	(3.15)
More than 51 employees	0.044	0.033	–0.001	0.289***	0.200***	0.209***
	(0.93)	(1.00)	(–0.13)	(3.93)	(3.71)	(2.64)
Observations		4,394			3,832	
Joint test on selection terms						
F-statistic		5.73			3.16	
p-value		0.0007			0.0238	

Notes: Author's calculations based on microdata from the 2009 GIHS. ***,**,*denote significance at 1%, 5% and 10% levels, 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.

relationships, responsibilities, etc.) in the informal sector (Fields, 1990). As discussed in section 2, informal workers in the Caribbean coast cities show higher education and productivity levels than informal workers in the other two city groups. Therefore, the narrow wage gap can be easily compensated for by the cost savings and non-pecuniary benefits of being an unregistered worker and hence there will be incentives to voluntarily select employment in the informal sector.

Finally, in City Group 3, a positive wage gap exists only at the extremes of the distribution, and this gap is largely explained by the characteristic effects. In contrast, at the median of the distribution, there is an informal employment wage

Group 3					
Formal			Informal		
10%	50%	90%	10%	50%	90%
7.410***	7.569***	7.573***	6.417***	6.994***	7.438***
(125.23)	(256.39)	(100.10)	(169.60)	(335.88)	(185.23)
0.191***	0.049***	–0.029	–0.221***	–0.177***	–0.219***
(7.09)	(3.50)	(–0.80)	(–6.13)	(–8.93)	(–5.93)
0.078**	0.069***	0.127***	0.127***	0.106***	0.098***
(2.52)	(4.44)	(3.21)	(5.24)	(7.67)	(3.85)
0.174***	0.142***	0.299***	0.242***	0.250***	0.284***
(5.64)	(9.44)	(7.87)	(8.63)	(16.61)	(10.33)
0.366***	0.505***	0.692***	0.759***	0.696***	0.693***
(8.94)	(25.94)	(13.85)	(11.95)	(19.26)	(10.21)
0.001	0.004***	0.008***	0.017***	0.015***	0.017***
(0.58)	(3.80)	(3.16)	(7.05)	(11.73)	(6.97)
–0.001	–0.001***	–0.001**	–0.001***	–0.001***	–0.001***
(–0.38)	(–2.66)	(–1.99)	(–5.69)	(–9.27)	(–5.43)
0.005	0.003**	0.010**	0.023***	0.017***	0.022***
(1.29)	(2.11)	(2.46)	(4.62)	(5.88)	(4.37)
–0.001	–0.001***	0.001	–0.001***	–0.001***	–0.001**
(–0.25)	(–3.14)	(0.39)	(–2.85)	(–3.12)	(–2.49)
–0.026*	0.019**	0.064***	0.147***	0.137***	0.142***
(–1.76)	(2.56)	(3.31)	(6.61)	(10.88)	(5.96)
0.014	0.020	0.074**	0.291***	0.187***	0.230***
(0.51)	(1.49)	(2.07)	(9.93)	(11.35)	(7.82)
0.005	0.062***	0.209***	0.501***	0.480***	0.586***
(0.17)	(3.84)	(4.98)	(8.24)	(16.06)	(11.10)
	6,531			8,430	
	7.38			34.64	
	0.0001			0.0000	

premium, which is explained mainly by the negative coefficient effect. The significant rationing of formal jobs and relative abundance of informal workers—particularly those with very low qualification levels (see Table 4)—are important determinants of the labor market conditions in this group of cities. In this context, only formal workers who have significant advantages in terms of characteristics and/or rates of return earn higher wages than informal workers, whereas at the median of the distribution, where there is a notable concentration of informal workers, the benefits of formality are undermined. From the perspective of informal workers, the relatively high disadvantage at the bottom of the distribution implies that workers in this group have no alternative to employment in the informal sector, whereas at high quantiles, the relatively low wages can be compensated by the benefit of greater flexibility, which may allow these informal workers to obtain greater enjoyment from their work.

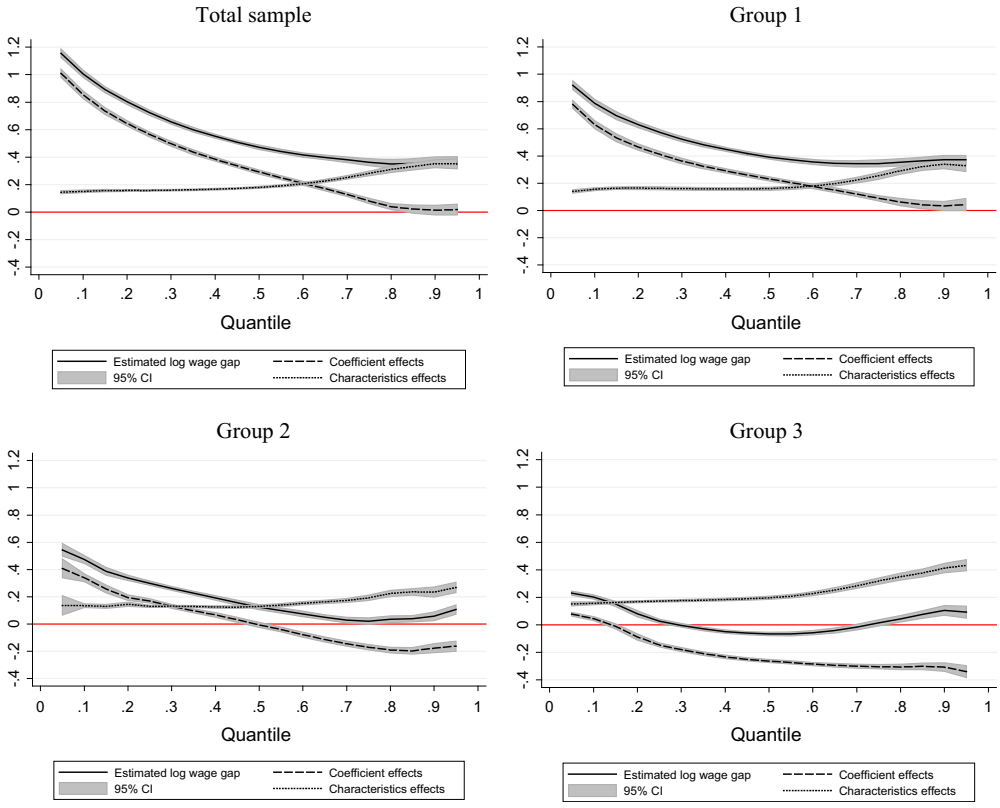

Figure 8. Quantile Decomposition of the Wage Gap between the Formal Sector and Informal Sector [Colour figure can be viewed at wileyonlinelibrary.com]

6. Conclusions

In this paper, 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 sectors. We used the quantile regression decomposition method and corrected for selectivity with semiparametric methods. This econometric model allowed us to analyze individuals across the entire distribution of wages and to determine whether the informal sector has its own internal duality.

Our results show that there is a marked heterogeneity in the informal sector of Colombia. We find that in general, there are two distinct segments of workers in the informal sector and that each segment has a unique motivation for informal employment. On the one hand, there is a lower-paid informal segment in which informal workers are at a significant disadvantage relative to formal workers, not only in terms of characteristics but also in terms of the rate of return on those characteristics. Individuals in this segment are rationed out of the formal labor market, and informality is perceived as the only option for employment. On the other hand, there is a higher-paid informal segment that indicates a competitive area in which the wage gap between the formal and informal sectors is much narrower than at the bottom of the distribution, and individuals receive returns on

their characteristics similar to those received in the formal sector. In this segment, although informal workers earn less than formal workers, informality is preferred because the benefits of formality are less attractive. These results suggest that just as formal and informal activities co-exist, voluntary and involuntary informal employment also co-exist. Informality may be a choice rather than the result of labor market segmentation. Certainly, these are two concurrent scenarios within the same phenomenon.

We also find that there are striking differences in labor market characteristics between groups of cities, particularly with respect to the nature of informal employment. Once the difference in the returns on observed characteristics between the formal and informal sectors has been taken into account, the results show that the wage gap over the entire distribution is much narrower in City Groups 2 and 3 than in City Group 1. This result supports the idea that a relatively high abundance of informal activities can undermine the ability of the state to provide employee protection; therefore, the benefits of working in the informal sector are greater.

Regarding the decomposition, we find that the wage gap at the bottom of the distribution is largely explained by the differential in returns on individual characteristics, particularly in City Groups 1 and 2. In this segment, levels of human capital and other remunerable characteristics are low, and because the differential between rates of return on characteristics has a greater impact on the wage gap between the formal and informal sectors, there is a marked segmentation effect. This result indicates that at these points in the distribution, the informal sector represents a disadvantaged sector in which workers turn to informal employment as a last-resort option for a paid job.

At the upper half of the distribution, the characteristic effect dominates the coefficient effect, and the wage gap is positive. These findings suggest that the choice to work in the informal sector at these points of the distribution may be due in part to the acceptance of lower wages to avoid the administrative cost of social security, which is perceived as costly and ineffective, or to gain greater flexibility in terms of responsibilities or work schedule. For example, the results show that the size of the estimated wage differential in City Groups 2 and 3 are 10% and 9%, respectively, at the 95th quantile, which can be easily compensated by the non-wage benefits associated with informal employment.

References

Acosta, J., "Cartagena, entre el Progreso Industrial y el Rezago Social [Cartagena, between Industrial Progress and Social Lag]," *Documentos de Trabajo sobre Economía Regional*, no. 178 (December 2012). Banco de la República, Cartagena.

Albrecht, J., A. van Vuuren, and S. Vroman, "Counterfactual Distributions with Sample Selection Adjustments: Econometric Theory and an Application to the Netherlands," *Labour Economics* 16 (2009):383–96.

Arango, L., "Mercados de Trabajo de Colombia: Suma de Partes Heterogéneas [Colombia's Labor Market: Sum of Heterogeneous Parts]," *Borradores de Economía*, no. 671 (September 2011), Banco de la República, Bogotá.

Badel, A. and X. Peña, "Decomposing the Gender Wage Gap with Sample Selection Adjustment: Evidence from Colombia," *Revista de Análisis Económico* 25 (2010):169–91.

Bargain, O. and P. Kwenda, "The Informal Sector Wage Gap: New Evidence Using Quantile Estimations on Panel Data," *Economic Development and Cultural Change* 63 (2014):117–53.

Bernal, R., "The Informal Labor Market in Colombia: Identification and Characterization," *Desarrollo y Sociedad* 63 (2009):145–208.

Bonet, J., "Desindustrialización y Terciarización Espuria en el Departamento del Atlántico, 1990–2005 [Deindustrialization and Spurious Tertiarization in the Department of Atlántico: 1990-2005]," *Documentos de Trabajo sobre Economía Regional*, no. 60 (July 2005), Banco de la República, Cartagena.

——, "La Terciarización de las Estructuras Económicas Regionales en Colombia [The Tertiarization of Regional Economic Structures in Colombia]," *Revista de Economía del Rosario* 10 (2007):1–19.

Bonet, J. and A. Meisel, "La Convergencia Regional en Colombia: Una Visión de Largo Plazo, 1926–1995 [The Regional Convergence in Colombia: A Long-Term Vision, 1926–1995]," *Coyuntura económica* 29 (1999):15–51.

Bonilla, L., "Diferencias Regionales en la Distribución del Ingreso en Colombia [Regional Differences in the Distribution of Income in Colombia]," *Documentos de Trabajo sobre Economía Regional*, no. 108 (December 2008), Banco de la República, Cartagena.

——, "Determinantes de las Diferencias Regionales en la Distribución del Ingreso en Colombia, un Ejercicio de Microdescomposición [Determinants of the Regional Differences in the Distribution of Income in Colombia, a Micro-decomposition Exercise]," *Ensayos sobre Política Económica* 27(59) (2009):46–82.

——, "El Sector Industrial de Barranquilla en el siglo XXI: ¿Cambian Finalmente las Tendencias? [The Industrial Sector of Barranquilla in the 21st Century: Do Trends Finally Change?]" *Documentos de Trabajo sobre Economía Regional*, no. 136 (December 2010), Banco de la República, Cartagena.

Braithwaite, J., "From Second Economy to Informal Sector: The Russian Labor Market in Transition," World Bank working paper 21228, Washington, DC (1995).

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

Cepeda, L., "Los Sures de Barranquilla: La Distribución Espacial de la Pobreza [Southern Barranquilla: The Spatial Distribution of Poverty]," *Documentos de Trabajo sobre Economía Regional*, no. 142 (April 2011), Banco de la República, Cartagena.

Cunningham, W. and W. Maloney, "Heterogeneity among Mexico's Microenterprises: An Application of Factor and Cluster Analysis," *Economic Development and Cultural Change* 50 (2001):131–56.

Daza, N. and L. Gamboa, "An Approximation to the Informal–Formal Wage Gap in Colombia 2008–2012," *Serie Documentos de Trabajo*, no. 138, Universidad del Rosario (2013).

De Soto, H., *El Otro Sendero. La Revolución Informal* [The Other Path. The Informal Revolution], Instituto Libertad y Democracia, Lima, Peru (1987).

Departamento Administrativo Nacional de Estadística (DANE), *Informe de Coyuntura Económica Regional*, *Meta 2011* [Report on the Regional Economic Situation, Meta 2011]. Banco de la República, Bogotá, Colombia (2011).

Departamento Nacional de Paneación (DNP), "Cadena Petroquímica-plásticos, Cauchos, Pinturas, Tintas y Fibras [Petrochemical-Plastics Chain, Rubbers, Paints, Inks and Fibers]," Documento sectorial, DNP, Bogotá, Colombia (2007).

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

Fernández, M., "Determinantes del Diferencial Salarial por Género en Colombia, 1997–2003 [Determinants of the Salary Differentials by Gender in Colombia: 1997–2003]," *Desarrollo y Sociedad* 58 (2006):168–208.

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

——, "A Guide to Multisector Labor Market Models," Social Protection discussion paper series 0505, World Bank, Washington, DC (April 2005).

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

Galvis, L., "La Topografía Económica de Colombia [The Economic Topography of Colombia], " in J. Bonet (ed.), *Geografía Económica y Análisis Espacial en Colombia*, Bogotá: Banco de la República (2007).

——, "Geografía Económica del Caribe Continental [Economic Geography of the Continental Caribbean], " *Documentos de Trabajo sobre Economía Regional*, no. 119 (December 2009), Banco de la República, Cartagena.

——, "Diferencias Salariales por Género y Región en Colombia: Una Aproximación con Regresión por Cuantiles [Wage Differences by Gender and Region in Colombia: An Approach with Quantile Regression], " *Revista de Economía del Rosario* 13 (2010):235–77.

——, "Informalidad Laboral en las Áreas Urbanas de Colombia [Labor Informality in the Urban Areas of Colombia], " *Coyuntura Económica* 42 (2012):15–51.

García, G. A., "Informalidad Regional en Colombia: Evidencia y Determinantes [Regional Informality in Colombia: Evidence and Determinants]," *Desarrollo y Sociedad* 61 (2008):43–86.

——, "Determinantes Macro y Efectos Locales de la Informalidad Laboral en Colombia [Macro Determinants and Local Effects of Labor Informality in Colombia], " *Sociedad y Economía* 21 (2011):69–98.

Gindling, T., "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 (1991):585–603.

Gong, X. and A. van Soest, "Wage Differentials and Mobility in the Urban Labour Market: A Panel Data Analysis for Mexico," *Labour Economics* 9 (2002):513–29.

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

Heckman, J., "Sample Selection Bias as a Specification Error," *Econometrica* 47 (1979):153–61.

Heckman, J. and V. Hotz, "An Investigation of Labor Market Earnings of Panamanian Males," *Journal of Human Resources* 21 (1986):507–42.

Hoyos, A., H. Ñopo, and X. Peña, "The Persistent Gender Earnings Gap in Colombia, 1994–2006," IDB working papers series IDB-WP-174, Washington, DC (2010).

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

International Labor Organization, *2011 Labour Overview*, Latin America and the Caribbean, Lima: ILO (2011).

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

Kolev, A., *"Labour Supply in the Informal Economy in Russia during Transition", CEPR discussion paper 2024*, London: Centre for Economic Policy Research (1998).

Lehmann, H. and A. Zaiceva, "Informal Employment in Russia: Incidence, Determinants and Labor Market Segmentation," working paper DSE 930, University of Bologna (2013).

Machado, J. and J. Mata, "Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression," *Journal of Applied Econometrics* 20 (2005):445–65.

Magnac, T., "Segmented or Competitive Labor Markets," *Econometrica* 59 (1991):165–87.

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

——, "Informality Revisited," *World Development* 32 (2004):1159–78.

Marcouiller, D., V. Ruiz de Castilla, and C. Woodruff, "Formal Measures of the Informal-sector Wage Gap in Mexico, El Salvador, and Peru," *Economic Development and Cultural Change* 45 (1997):367–92.

Mazumdar, D., *The Urban Labor Market Income Distribution: A Study of Malaysia*, Oxford: Oxford University Press (1981).

Nguyen, H., C. Nordaman, and F. Roubaud, "Who Suffers the Penalty? A Panel Data Analysis of Earnings Gaps in Vietnam," *Journal of Development Studies* 49 (2013):1694–710.

Ñopo, H., "Matching as a Tool to Decompose Wage Gaps," *The Review Economics and Statistics* 90 (2008):290–99.
Pradhan, M. and A. van Soest, "Household Labor Supply in Urban Areas of Bolivia," *The Review of Economics and Statistics* 79 (1997):300–310.
Pratap, S. and E. Quintin, "Are Labor Markets Segmented in Developing Countries? A Semiparametric Approach," *European Economic Review* 50 (2006):1817–41.
Roberts, B. R., "Employment Structure Life Cycle and Life Chances: Formal and Informal Sectors in Guadalajara," in A. Portes, M. Castells, and L. A. Benton (eds), *The Informal Economy: Studies in Advanced and Less Developed Countries*, Baltimore, MD: Johns Hopkins University Press (1989).
Saavedra, J. and A. Chong, "Structural Reform, Institutions and Earnings: Evidence from the Formal and Informal Sectors in Urban Peru," *Journal of Development Studies* 35 (1999):95–116.
Sánchez, A., A. Días, A. Peláez, O. Castelblanco, J. Tautiva, C. González, and L. Ángel, "Evolución Geográfica del Homicidio en Colombia [Geographical Evolution of Homicide in Colombia]," Documentos de Trabajo Sobre Economía Regional, no. 169 (June 2012), Banco de la República, Cartagena.
Staneva, A. and G. Arabsheibani, "Is there an Informal Employment Wage Primium? Evidence from Tajikistan," *IZA Journal of Labor and Development* 3 (2014):1–24.
Tannuri-Pianto, M. and D. Pianto, "Informal Employment in Brazil—A Choice at the Top and Segmentation at the Bottom: A Quantile Regression Approach," Serie de Textos para Discussao 236, Universidade de de Brasilia (2002).
Tansel, A., "Formal versus Informal Sector Choice of Wage Earners and Their Wages in Turkey," Economic Research Forum working paper 9927, Cairo (1999).
——, *"Wage Earners, Self Employed and Gender in the Informal Sector in Turkey", Policy Research Report on Gender and Development 24*, Washington, DC: The World Bank (2000).
Tansel, A., H. Keskin, and Z. Ozdemir, "Is There an Informal Employment Wage Penalty in Egypt?" ERC Working Papers in Economics 15/08, Economic Research Center, Ankara (2015).
United Nations Development Programme (UNDP). *Human Development Report 2015, Work for Human Development*. New York: UNDP (2015).
Uribe, J., C. Ortiz, and G. García, "La Segmentación del Mercado Laboral Colombiano en la Década de los Noventa [The Segmentation of the Colombian Labor Market in the Decade of the Ninety]," *Revista de Economía Institucional* 9(16) (2007):189–221.
Viloria de la Hoz, J., "Geografía Económica de la Orinoquia," *Documentos de Trabajo Sobre Economía Regional* [Economic Geography of Orinoquia], no. 113 (June 2009), Banco de la República, Cartagena.

Notes

1. According to estimates of the International Labor Organization (2011), Colombia has the fourth highest rate of informality in South America, following Paraguay (70.4%), Peru (70.3%) and Bolivia (69.5%).
2. Namely, Barranquilla, Bogotá, Bucaramanga, Cali, Cartagena, Cúcuta, Ibagué, Manizales, Medellín, Montería, Pasto, Pereira and Villavicencio. According to the 2005 Population Census, these metropolitan areas represent 45% of the total population and approximately 60% of the urban population.
3. Note that we exclude government employees, employers and the self-employed. Because of these exclusions, the informality rate may differ from that reported by the ILO.
4. In fact, we calculated the informality rate using the productivity definition proposed by the ILO, which expands informal-sector employment to include all own-account workers (excluding administrative workers, professionals and technicians), unpaid family workers, and employers and employees working in establishments with fewer than five people, and found

that this informality rate is approximately six percentage points lower than the rate calculated based on the legal definition (see Table 3). However, the differences in informality rates across cities remain, and there is a high positive and significant correlation between the two measures of informality, which is consistent with the results of previous studies that compare different measures of informality in Colombia (see Bernal, 2009; Galvis, 2012).

5. Colombia comprises 32 departments plus the capital district of Bogotá. The departments are country subdivisions similar to US states; they are granted a certain degree of autonomy, and each has its own capital city. Galvis (2007) conducted a study that identified the economic regions in Colombia at a city level and used per capita bank deposits and local tax revenues as measures of city economic activity (according to Bonet and Meisel (1999), there is a correlation between GDP and bank deposits of approximately 0.8). The author reports that the region formed by these cities accounts for 80% of the total economic activity of the country.

6. 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 beverage, chemical, 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.

7. A more detailed review of the literature on this methodology can be found in Fortin et al. (2011).

8. Our estimations are based on $m = 1000$.

9. We estimated the quantile decomposition of the formal/informal wage gap using the definition proposed by the ILO, and the main results are generally very similar to those obtained using the social security definition. The results using the ILO definition are not presented here owing to limitations of space but are available upon request.

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

11. Results are available from the author upon request.

12. 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 owing to cost and 20% of salaried workers were not covered owing to excessive red tape. That survey included questions about the quality and use of the health system, which are omitted from the GIHS.