ARCHIVOS DE ECONOMÍA

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

Gustavo A. GARCIA

Documento 412
Dirección de Estudios Económicos
1 de Abril 2014

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

Gustavo A. García†

March 25, 2014

Abstract

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

JEL classification: *O17, J42, J31, C21*

Keywords: Informality, local labor markets, quantile regression, selection bias, formal/informal wage gap decomposition

*This paper is part of my Ph.D. dissertation thesis at Autonomous University of Barcelona. I wish to thank participants at the 16th IZA European Summer School in Labor Economics, XVI Encuentro de Economía Aplicada and X Jornadas de Economía Laboral. I thank José Luis Roig and Josep Lluís Raymond for their valuable suggestions and comments. Part of the research was funded by the Spanish Ministerio de Ciencia e Innovación, reference number ECO2010-20718, and by AGAUR (FI-DGR 2010).

†Economic Consultant (STC) at The World Bank; Email: gusgarciacruz@gmail.com

1 Introduction

One of the most notable features of developing countries is the great heterogeneity in 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.

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 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 paper 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 allows 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 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 2 proceeds with the description of the data. In Section 3 we discuss the estimation procedure. Section 4 describes the empirical findings, and finally conclusions are drawn in Section 5.

2 Data and descriptive evidence

The data used in this paper 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

[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%).

[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.

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]

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 1 we provide some descriptive statistics for the key variables for formal and informal workers.

Table 1. Descriptive statistics

	Formal workers		**Informal workers**		**Total**	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Real hourly wage	3269.2	1670.7	2311.9	1858.5	2927.4	1799.6
Age (years)	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 levels						
Less than primary	0.12	0.32	0.28	0.45	0.18	0.38
Primary	0.14	0.34	0.26	0.44	0.19	0.39
Secondary	0.62	0.48	0.42	0.49	0.55	0.49
Tertiary	0.12	0.32	0.04	0.17	0.08	0.28
Male	0.56	0.49	0.49	0.50	0.53	0.50
Head of household	0.43	0.49	0.35	0.47	0.40	0.49
Married	0.53	0.50	0.46	0.50	0.50	0.50
Size of firm						
1-10 employees	0.17	0.38	0.77	0.42	0.39	0.49
11-50 employees	0.22	0.42	0.14	0.35	0.19	0.39
More than 51 employees	0.60	0.49	0.09	0.29	0.41	0.49
Sample size	36,293		25,985		62,278	

Note: We used person sampling weight available in the database. The wages are in Colombian pesos

[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.

As can be seen from Table 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 (52%), while the formal sector has a much higher percentage of individuals with secondary and tertiary education (74%). 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%).

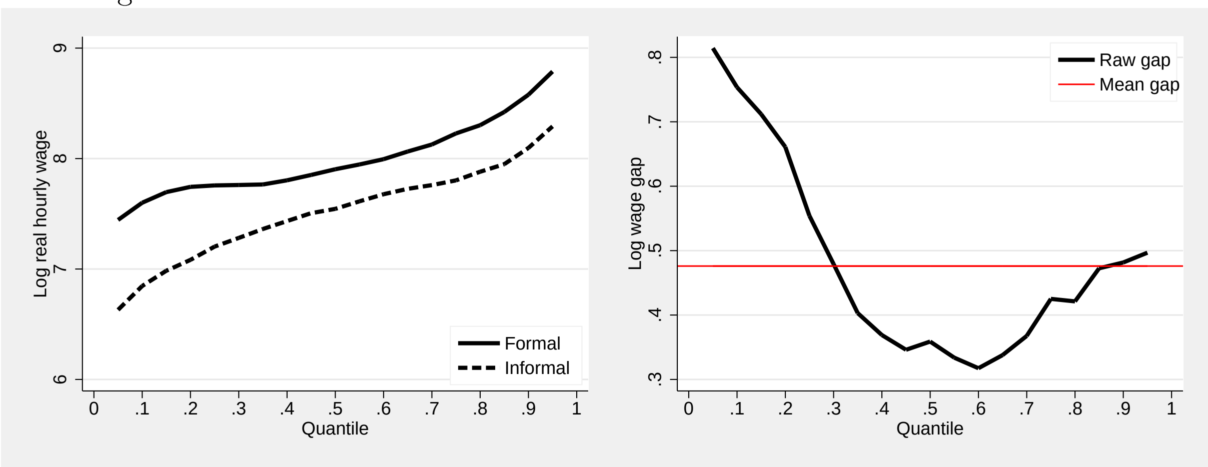
Figure 1 depicts the estimated kernel densities of the wages of formal and informal workers. Wage disparities between sectors are clearly visible, as wage distribution for formal workers is shifted to the right. The distribution of formal and informal sector wage and the wage gap between sectors by quantile, i.e., the difference in log wages between formal workers and informal workers at each quantile of their respective distributions, is plotted in Figure 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.

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 A1 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 A1 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.

Figure 1. Kernel density of log real hourly wage by formal and informal sector

Figure 2. Wage differentials between formal and informal sector over different quantiles of the wage distribution

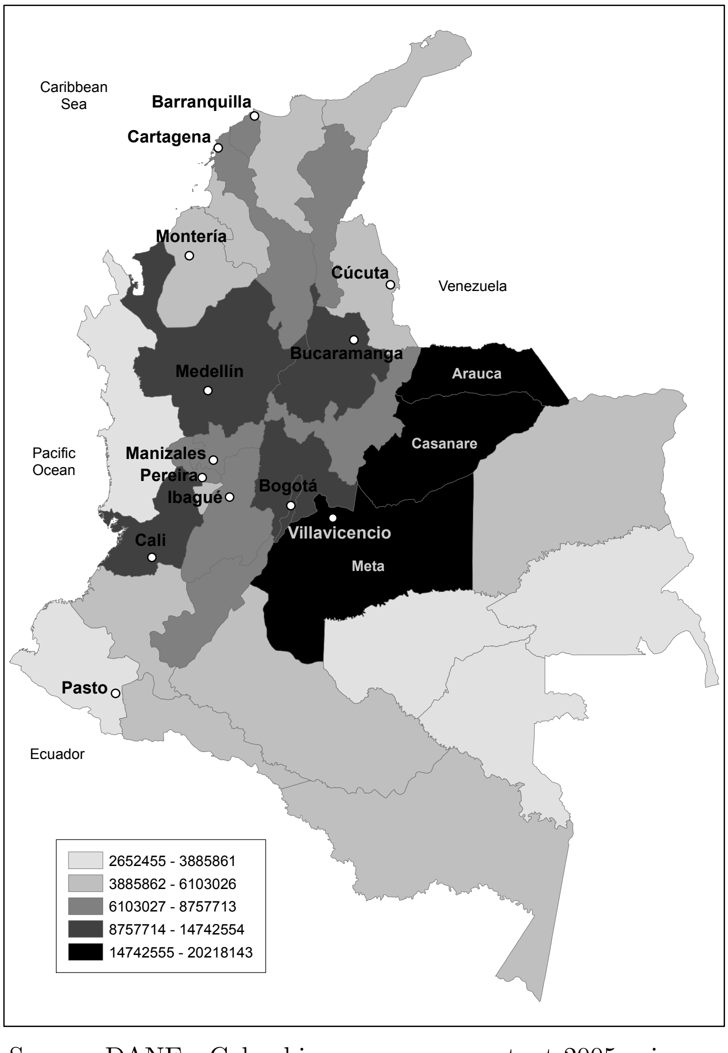


2.1 Group of cities and their labor markets

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.

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.[5][6] Figure 3 shows the spatial distribution of the real GDP per capita at a department level in 2009.[7]

Figure 3. Real Gross Domestic Product (GDP) per capita at departmental level, 2009



Source: DANE - Colombian currency, constant 2005 prices.

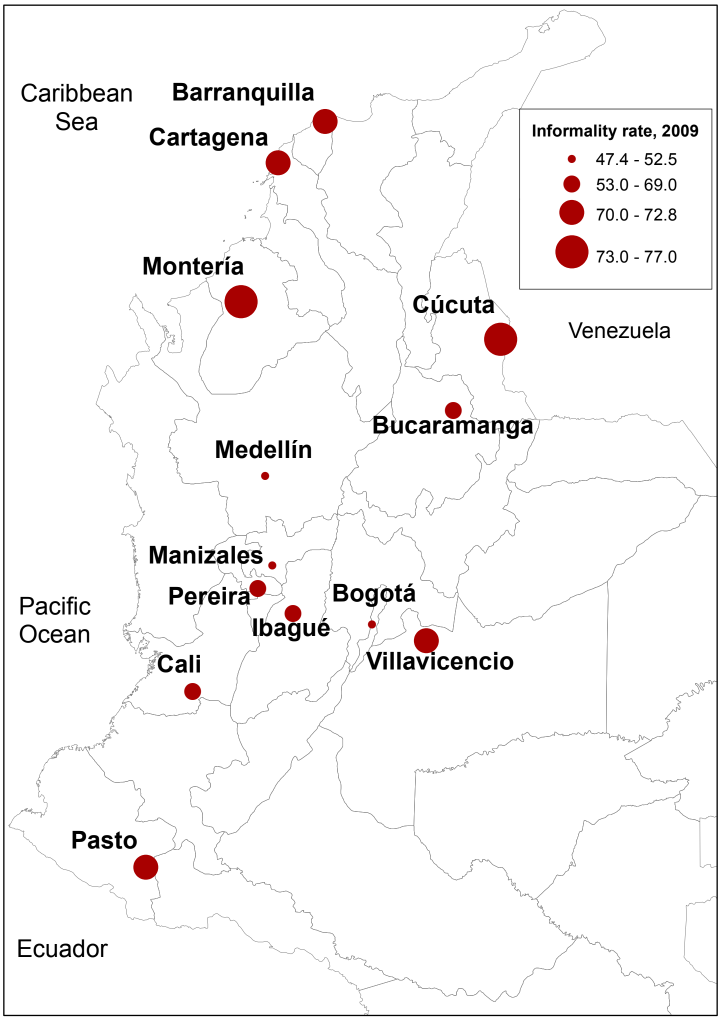
[5]Colombia is made up of thirty-two departments and Bogotá, the Capital District. The departments are country subdivisions similar to US states, are granted a certain degree of autonomy and each has its own capital city.

[6]Galvis (2007) carried out a study identifying the economic regions in Colombia at city level and used the bank deposits and the local tax collections per capita as measures of economic activity of the cities (according to Bonet and Meisel (1999) there is a correlation between GDP and bank deposits of around 0.8). The author reports that the region formed by these cities account for 80% of total economic activity of the country.

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

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 4).

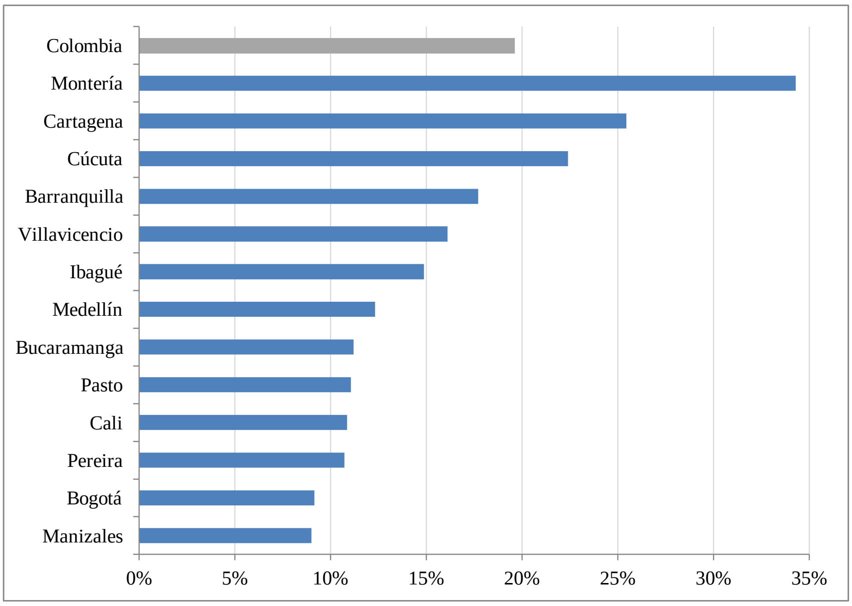
Figure 4. Informality rate by city, 2009



Source: Table A1 in the Appendix.

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), 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.[8] 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 4 and 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 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 5. Percentage of Unsatisfied Basic Needs (UBN) by city



Source: 2005 Population Census - DANE.

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). 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)

[8]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.

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 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.

Table 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.

Table 2. Descriptive statistics by groups of cities (mean)

	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	2941.5	1946.7	2492.6
Age (years)	34.2	33	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	46.2	54.8	53.4	54.2	53.8
Head of household	43	35.8	40.7	43.6	31	37.9	42.4	36.5	39
Married	52.3	45.7	50.1	64.8	49.9	58	56.4	45.6	50.1
Size of firm									
1 10 employees	18.5	77.3	37.7	9	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

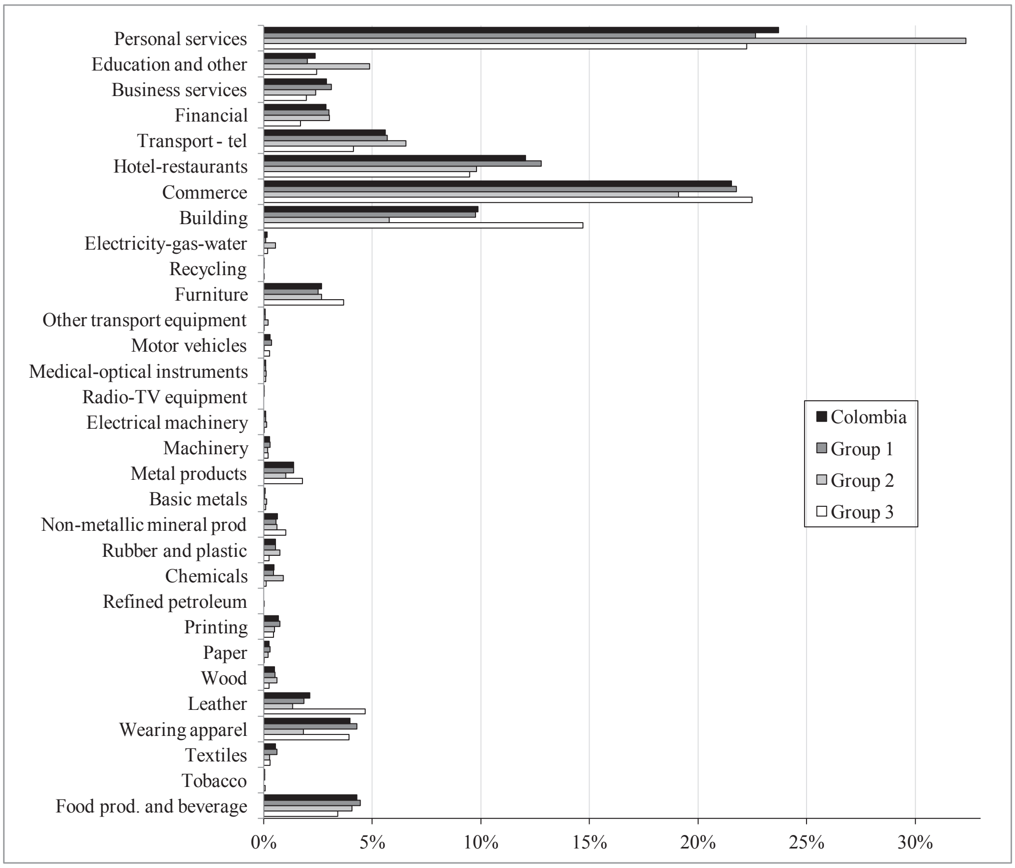
Note: We used person sampling weight available in the database. The wages are in Colombian pesos.

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.

Another difference between the groups of cities can be found in the firm size variable. As can been seen from Table 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 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 reveals 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 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.[9] 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 A2 in the Appendix). We calculated this index for each 2-digit industry and city.

In Figure 7 and Table 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

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

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 7. Informal sector employment by modernity quartile

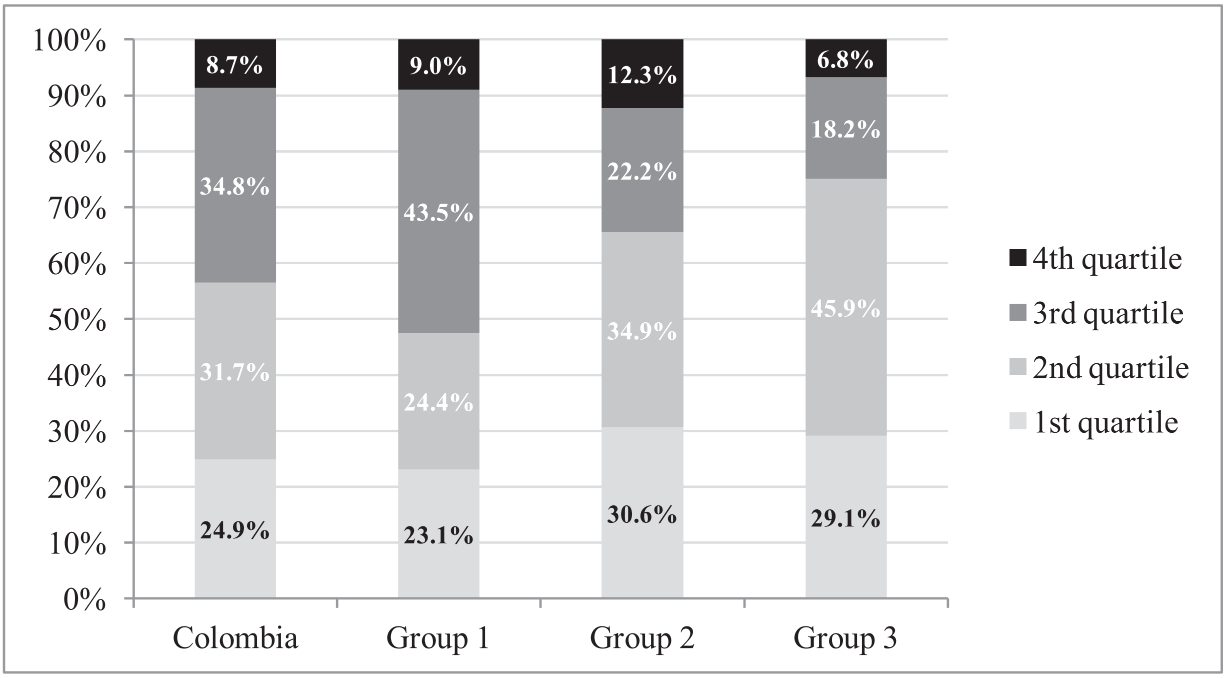


Table 3. Informal sector characteristics by modernity quartile

	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.

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.

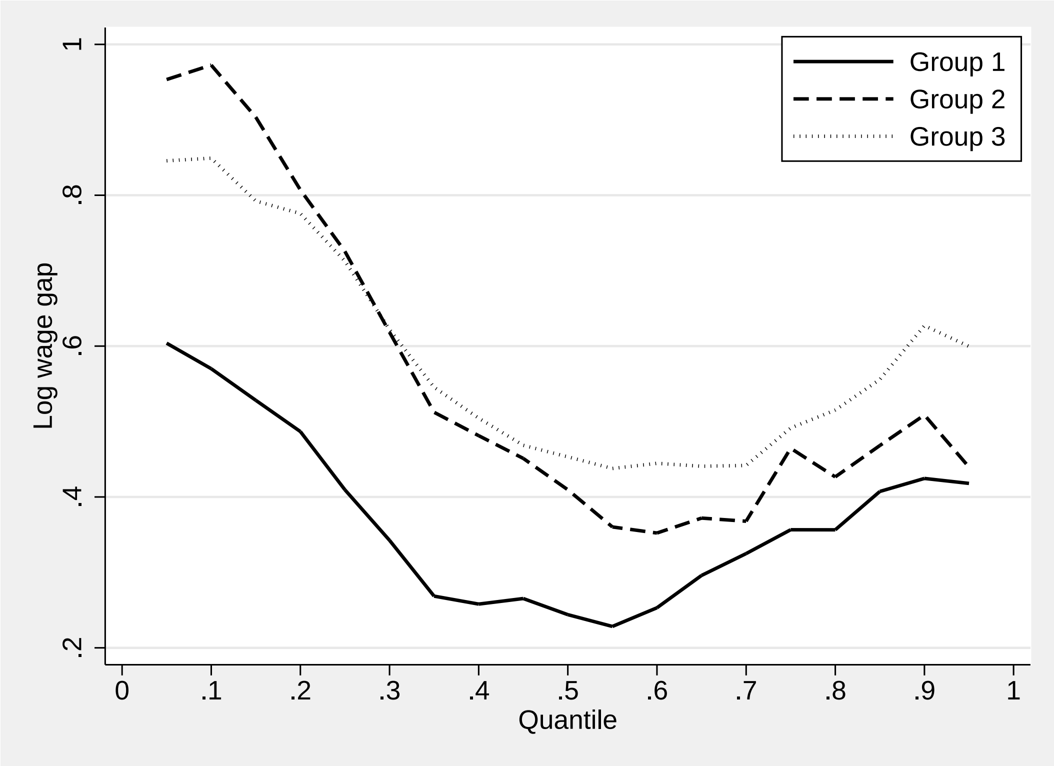
Interestingly, from Figure 7 we can see that in City Group 2 there are a relatively high proportion of informal workers in sectors in the top quartile of the modernity index distribution (12%). This degree of modernity of the informal sector contrasts with the high participation of the more traditional 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 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.

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.

Figure 8. Wage differentials between the formal and informal sector over different quantiles of the wage distribution by group of cities



The decomposition methods have been extensively used to analyze the gender and union wage gap, and temporal change in wages.[10] 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).

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 calcu-

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

late 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:

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

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 $G_{X_{inf}}$, 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 consists of generating a random sample of size n from a uniform distribution $U[0,1] : u_1, u_1, ...u_n$, 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 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 θth quantile can be decomposed as:

$$Q_\theta(Y_{for}|X_{for}=\widetilde{x}_{for})-Q_\theta(Y_{inf}|X_{infor}=\widetilde{x}_{inf})=\underbrace{Q_\theta(\widetilde{x}'_{for}\widehat{\beta}^{for}(u))-Q_\theta(\widetilde{x}'_{inf}\widehat{\beta}^{for}(u))}_{\text{Characteristics effects}}$$
$$+\underbrace{Q_\theta(\widetilde{x}'_{inf}\widehat{\beta}^{for}(u))-Q_\theta(\widetilde{x}'_{inf}\widehat{\beta}^{inf}(u))}_{\text{Coefficients effects}} \tag{5}$$

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 θth 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,

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

and

$$Q_\theta(Y_{inf}|X_{inf}=x_{inf})=x'_{inf}(\theta)+h_\theta(z'\gamma) \tag{7}$$

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.
7. Compare the simulated distributions to decompose the estimated wage gap between sectors.[11]

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.

4.1 SLS estimation and the quantile regression models

As mentioned in Section 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 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 Hausman's test, reported at the bottom of Table 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 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.

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

[11] Our estimations are based on $m = 1000$.

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.

Table 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.

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.[12] 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 to A6 in the Appendix we present results for corrected quantile regressions for the 5th, 10th, 25th, 50th, 75th, 90th and 95th quantiles.

[12] 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).

It can be seen from data in Tables A3 to A6 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 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 A7 in the Appendix.

From Table 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 10th 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 5. Marginal effects for quantile regressions by group of cities (y = Log real hourly wage)

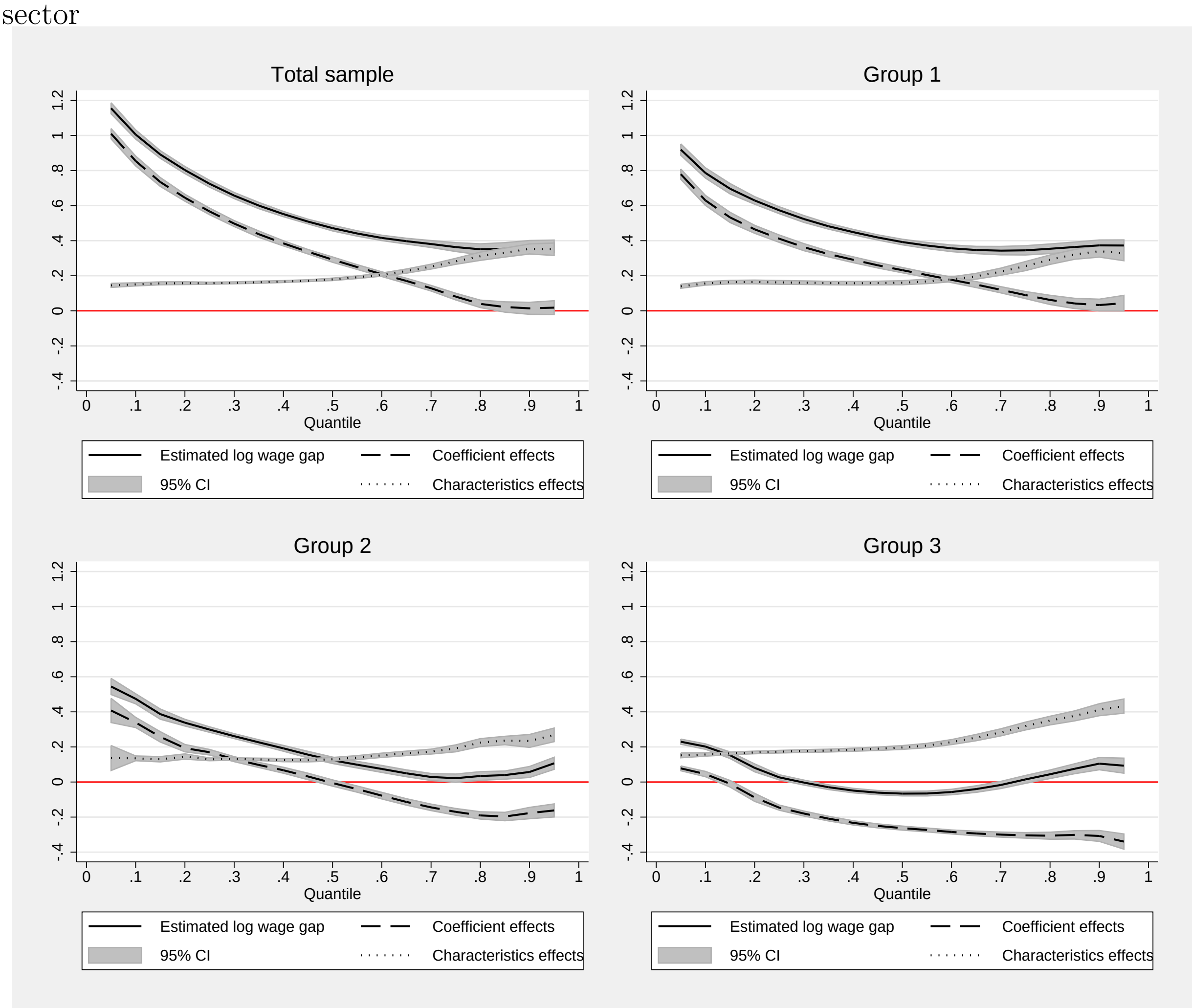
	Group 1						**Group 2**						**Group 3**					
	Formal			**Informal**			**Formal**			**Informal**			**Formal**			**Informal**		
Percentile:	**10%**	**50%**	**90%**	**10%**	**50%**	**90%**	**10%**	**50%**	**90%**	**10%**	**50%**	**90%**	**10%**	**50%**	**90%**	**10%**	**50%**	**90%**
Constant	7.245***	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)
Experience2	-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)
Tenure2	-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**	0.04	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***	-0.01	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.

4.2 Decomposition results

In this section we present the results of the decomposition. Figure 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 9. Quantile decomposition of the wage gap between the formal sector and informal sector



Source: Table A8 and A9 in the Appendix

As can be seen from Figure 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.[13]

[13]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.

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 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 level 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 7). It seems fair to think that at these points in the wages distribution 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 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 7 and Table 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.

5 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 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%. In the more developed cities the informal sector is associated with more modern activities through intermediate linkages which can 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.

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 Reloaded”, 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.

7 Appendix

Table A1. Informality rate by metropolitan area (%)

Colombia	58.28
Medellín	47.38
Bogotá	52.13
Manizales	52.5
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 A1. Wage differentials between formal and informal sector over different quantiles of the wage distribution by metropolitan area

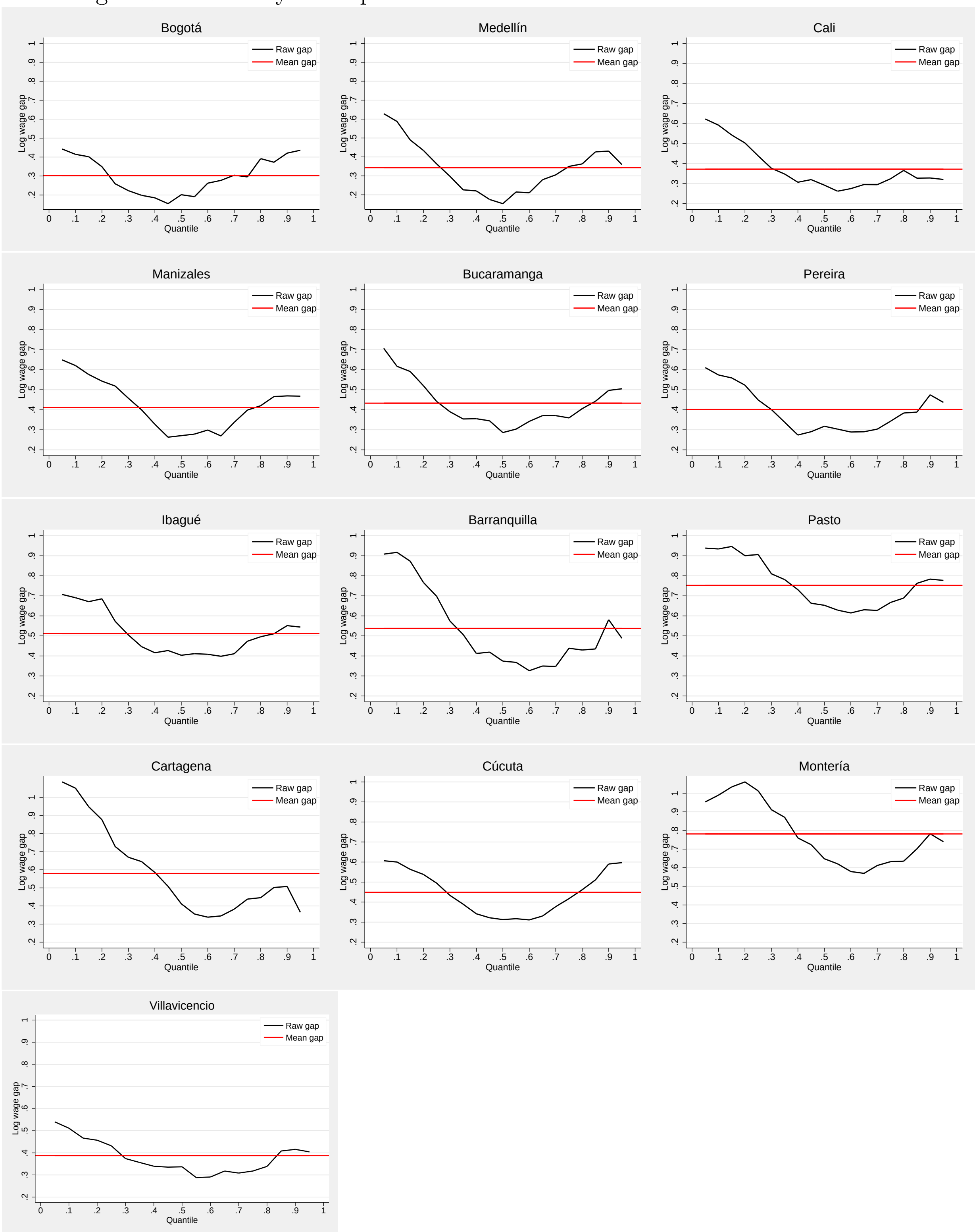


Table A2. Sector characteristics by modernity quartile

	Total sample	**Group 1**	**Group 2**	**Group 3**
Size of firm (> 50 empoyees)				
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	7.0
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. Marginal effects for quantile regressions for 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	0.04	-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																
Primary	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)
Secondary	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)
Tertiary	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)
Experience2	-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)
Tenure2	-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***	0.02	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 A4. Marginal effects for quantile regressions for the group of cities 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																
Primary	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)
Secondary	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)
Tertiary	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)
Experience2	-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)
Tenure2	-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***	0.01	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 A5. Marginal effects for quantile regressions for the group of cities 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)
Experience2	-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)
Tenure2	-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	0.04	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 A6. Marginal effects for quantile regressions for the group of cities 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***	-0.02	-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)
Experience2	-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)
Tenure2	-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*	-0.01	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*	-0.01	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 A7. Marginal effects for quantile regressions for the group of cities 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																
Primary	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)
Secondary	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)
Tertiary	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)
Experience2	-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)
Tenure2	-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 A8. Decomposition results for the total sample

Q	**Raw log wage gap**	**Estimated log wage gap**	**Characteristics**	**Coefficients**
0.05	0.812	1.155	0.145	1.01
		(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	0.16	0.498
		(0.009)	(0.003)	(0.009)
0.35	0.404	0.6	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	0.18	0.292
		(0.008)	(0.004)	(0.008)
0.55	0.335	0.441	0.192	0.25
		(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	0.04
		(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 A9. Decomposition results by group of cities

	Group 1				Group 2				Group 3			
Q	**Raw log wage gap**	**Estimated log wage gap**	**Characteristics**	**Coefficients**	**Raw log wage gap**	**Estimated log wage gap**	**Characteristics**	**Coefficients**	**Raw log wage gap**	**Estimated log wage gap**	**Characteristics**	**Coefficients**
0.05	0.604	0.919	0.14	0.779	0.953	0.545	0.137	0.408	0.846	0.23	0.152	0.078
		(0.017)	(0.006)	(0.015)		(0.023)	(0.036)	(0.035)		(0.007)	(0.007)	(0.007)
0.10	0.57	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	0.13	0.257	0.792	0.153	0.163	-0.01
		(0.015)	(0.005)	(0.015)		(0.015)	(0.008)	(0.015)		(0.010)	(0.004)	(0.010)
0.20	0.487	0.63	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	0.17	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	0.13	0.622	-0.003	0.177	-0.18
		(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	-0.03	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	0.19	-0.25
		(0.008)	(0.005)	(0.008)		(0.009)	(0.005)	(0.009)		(0.006)	(0.005)	(0.006)
0.50	0.244	0.392	0.16	0.232	0.41	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	0.36	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	0.18	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	0.05	0.163	-0.114	0.441	-0.04	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	-0.3
		(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	-0.17	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	0.35	-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	0.34	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	0.44	0.107	0.269	-0.162	0.6	0.093	0.433	-0.34
		(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.