# Quantifying the heterogeneity of the Colombian Informal Labor Market

Jesús Daniel Cañizares[1] Brian Feld[2]

Understanding labor informality is a central issue in policymakers' agendas across the developing world. For instance, recognizing its remarkable heterogeneity is a must in that duty. This paper studies the heterogeneity of Colombian informal labor market at the city level. Our results suggest the existence of a large heterogeneity both inside and between the main urban areas of Colombia. While in some cities informal jobs are predominantly the last resort for individuals to escape from unemployment, in others informality seems to be a choice. In addition, we observe in almost all urban areas that informal workers tend to be self-selected into informality at the top of the income distribution.

[1] Colombian Institute for the Assessment of Education. jcanizares@icfes.gov.com
[2] Department of Economics, Universidad EAFIT. bhfeld@eafit.edu.co
We would like to thank Bruno Van der Linden and Juan Camilo Chaparro Cardona for their helpful comments. All errors and omissions are our own.

# Introduction

Labor informality is a salient characteristic of developing countries (ILO, 2018). Early debates around this phenomenon revolved around whether informal workers were excluded from the formal sector due to their characteristics (Lewis, 1954; Harris & Todaro, 1970; De Soto, 1987; Fields 1990), or instead informality was the result of an optimization process where individuals compare the costs and benefits of working in each sector (Magnac, 1991; Maloney, 1999; Pratap and Quentin, 2006). This distinction is relevant to understand the outside option of informal workers (e.g., a job in the formal sector or unemployment) and implement policies better suited to reduce the size of the informal sector under each scenario.

Recent studies consider a more nuanced view, in which both informality by exclusion and by choice coexist in the labor market (Cunningham & Maloney, 2001; Maloney, 2004; Fields, 2005; Paulson & Townsend, 2005; Perry et.al, 2007). Empirical studies have given support to this view by quantifying the extent of each form of informality at the national level (Günther & Launov, 2012; Alcaraz et al., 2015), but less is known about the importance of each type of informality at the local labor market level.

In this paper, we aim to fill this gap by estimating the composition of informal labor across the main urban areas of Colombia. To do this, we first take a stated preference approach and compute the observed preference for informality at the city level using individuals' survey responses about their satisfaction with their current job benefits. By doing this, we can approximate to the share of workers who might be voluntary informal in each of the main urban areas of Colombia.

Secondly, we use the quantile decomposition methodology proposed by Albrecht et al. (2009) to analyze the wage differential among formal and informal sectors across the entire wage distribution controlling for selectivity bias for each Colombian urban area. This decomposition provides an estimation of how much of the wage differential is explained by differences in returns on observed individuals' characteristics and how much is due to the differences in observable characteristics between formal and informal workers. If at a certain point of the entire wage distribution, the wage differential is mainly explained by differences in prices related to individual characteristics, it can be a sign of segmented markets at that point of the wage distribution, since informal workers earn lower returns on their skills. On the other hand, if the wage gap is mainly explained by characteristics that differ among workers in each sector, the market segmentation hypothesis would not bet relevant to explain the wage differential and, therefore, the gap might be explained by differences in endowments and there can be an important presence of voluntary informality.

Results obtained from survey responses suggests the existence of a large heterogeneity in the preference for informality of informal workers across cities, which can be partially related to the size of the formal-informal wage gap. The preference for informality of informal workers tends to be higher in cities where earnings in the informal sector are similar to those of the formal one. Similarly, the quantile decomposition results also indicate the presence of significant heterogeneity with respect to the causes driving individuals into informality across and within cities. In some of them, the formal-informal wage gap is better explained by differences in returns on individual's characteristics, indicating informality by exclusion and segmented markets since individuals receive different payments on similar characteristics; and in some other cases where the wage differential is mainly determined by

differences in individual characteristics rather than in the returns to those characteristics, indicating that the payments they receive from their characteristics are comparable with those received by formal workers. These workers might find informal work more valuable than a formal job. These results are also compatible with the ones observed across the income distribution of the wage gap for each city, where informal workers tend to be excluded from formality at the bottom of income distribution and self-selected into informality at the top.

This paper contributes to the study of the determinants of labor informality. In addition to the papers cited above, several studies have focused on how taxes and regulations (Kugler & Kugler, 2009; Mondragón-Vélez, et.al, 2010; Almeida & Carneiro, 2012; Fernández & Villar, 2017), social programs (Camacho et al., 2014; Bérgolo & Cruces, 2014; Saavedra-Caballero & Londoño, 2018) and macroeconomic fluctuations (Fields et.al, 2010) affect the size of the informal sector. By differentiating between informality by exclusion and by choice and quantifying the size of each reason for informality, we expect to guide policymakers into determining which measures could be better suited to address informality depending on the relative type of informality prevalent in each case.

We also make a contribution to the study of the heterogeneity of the informal sector at the subnational level. While most studies regarding the composition of the informal sector have taken a national (Cunnigham & Maloney, 2001; Tannuri-Pianto & Pianto; 2002; Günther & Launov, 2012) or multi-country (Bargain & Kwenda, 2014) approach, to our knowledge only García (2017) looks at the relative prevalence of informality by choice and by exclusion across different areas within a country (Colombia).

Even though we follow Garcia (2017)'s approach to decompose the formal-informal wage gap, we make numerous contributions. First, we include a stated preference method to

estimate the extent of heterogeneity of the informal sector. Since stated-preference methods are relatively cheaper and easier to implement, it is convenient to understand how these methods perform with respect to revealed preference approaches. Second, we focus on each urban area separately rather than grouping them into regions to obtain a more granular perspective of the heterogeneity within the informal sector. Finally, we use a more recent version of the survey, which allows us to understand whether and how the composition of the informal sector evolved after several years of robust economic growth and the implementation of policies aimed towards reducing informal labor.

The rest of the paper is structured as follows: Section 2 presents a description of the data and the empirical strategies and estimation methods used in this work. Section 3 presents the results for the entire country and a group of selected cities (Bogotá, Medellin, Barranquilla, Cúcuta, and Quibdó). Finally, section 4 concludes.

# Data and Empirical Methods

## Data

We use data from the Colombian Great Integrated Household Survey, -*Gran Encuesta Integrada de Hogares*- (GEIH, 2019), a monthly survey conducted by the National Administrative Statistics Department of Colombia -*Departamento Administrativo Nacional de Estadisticas*- (DANE). It is carried out in Colombia's main 23 metropolitan areas, where 70% of the population lives, and is representative at the metropolitan area level. It is the main source for the country's socioeconomic indicators, including labor force participation, unemployment, earnings, and poverty status.

The survey has specific questions that allow us to determine whether a person is employed in the formal or informal sector. That is, wage workers are asked whether their employer makes contributions to the social security system on their behalf.[3] We use the legalistic definition of labor informality and consider informal workers those who answer to this question in the negative (Tornarolli et al., 2014).

Our sample is composed of all wage workers surveyed during 2019. Table 1 presents summary statistics of the sample of formal and informal workers. Only 43% of the workers in our sample (122,365 out of 283,755) are considered formal under our definition. Formal and informal workers are different in terms of their socio-economic and job characteristics: formal workers are younger and better educated, and they are less likely to be married than informal workers. Compared to informal workers, those employed in the formal sector tend to be employed in larger firms, earn more than twice as much (1.6 vs. 0.7 million Colombian Pesos per month), enjoy more fringe benefits, and work more hours.

Our main variable of interest is hourly income, calculated as each worker's monthly income divided by the number of hours they worked in that month. Figure 1 shows the kernel density of log hourly wage for formal and informal workers. Log hourly wages for informal workers follow a normal distribution, centered closed to but to the left of the minimum wage, while that of formal workers shows bunching right at the minimum wage level, suggesting that the minimum wage in Colombia is binding and could be a reason for the existence of a large informal sector in the country.

[3] Employers must pay 20.5% of a worker's gross monthly salary in the form of contributions for the pension and health insurance systems.

In turn, Figure 2 presents the hourly wage gap between formal and informal workers across the entire distribution (Panel A), and the hourly wage distribution for formal and informal workers at the national level (Panel B). The dotted line represents the minimum wage. The larger gap is at the lower quantiles, which could be due to the presence of the minimum wage in the formal sector. The wage differential is high at lower quantiles of the wage distribution (probably due to the compliance with the minimum wage in the formal sector), and decreases towards the median, when the wage gap starts rising again.

Even though this pattern of wage differentials at the national level is similar across all cities covered in the survey, there is large heterogeneity in the size of the wage gap. While areas like Quibdó, Riohacha and Montería exhibit large wage gaps across the entire wage distribution, others like Medellin, Cali, Pereira, Manizales and Barranquilla present a relatively lower wage gap than the rest of the cities in almost all points of the wage distribution.

These regional differences motivate us to quantify the informal sector at the metropolitan area instead of just the national level. Table 2 presents a general description of these urban area's hourly wages by formal and informal sectors. From this table, we can see the high correlation ($\rho$=0.42) between the informality rate and formal-informal wage gap across metropolitan areas.

## Empirical Strategy

This work uses two different approaches to quantify the proportion of workers in the informal sector because of a rational choice and those who are excluded from the formal sector because of the lower return on their characteristics. The first approach follows Fernandez & Villar (2016), which identifies the "preference for informality" for individuals using surveys, while

the second approach combines the Machado-Mata quantile decomposition accounting for self-selection method used by García (2017). In the following subsections we provide a brief description of each method.

## The preference for informality among informal workers

Following the method used by Fernández & Villar (2016) and using the data from the Colombian Household Survey from 2019, we take a "stated preference" approach and identify workers who are informal by choice and by exclusion based on their answer to the following survey question:

> **P7170S5:** *Are you satisfied with the benefits and social protection that you are receiving in your current job?*

We assume that individuals answer this question by comparing the social protection and benefits they get in their current informal job and those they could get by law mandate in a formal job.

Informal workers do not make contributions to the pensions and health systems. One reason for this is that they may not value the benefits they obtain for their contributions to those systems (e.g., because they have a high discount rate, or low risk aversion), and hence informality would be a choice. On the other hand, some individuals might prefer to save for their retirement using the pension system or might be willing to contribute to the health system to receive a better service than what is provided free of charge by the Government. In this case, informality would not be voluntary since workers prefer the benefits of being

formal. Similar arguments can be made about other benefits associated with formal employment, such as unemployment insurance, paid vacations, etc.

Using the legal definition of informality and the answers to these questions, we estimate the proportion of informal workers who express a preference to remain informal and those who prefer the benefits and social protection they would get from a formal job. We should note that we assume that workers who are not satisfied with this dimension of their job characteristics are also not satisfied with being informal. Additionally, we use the data available in the survey to study which characteristics determine the preference of workers for the informal sector.

## Quantile Regression Decomposition: The predominance of Coefficients or Characteristics.

While stated preference approaches are convenient because they use direct information from individuals, this approach presents a series of shortcomings. Besides the strong assumption mentioned above regarding how satisfaction with one's job benefits translate into satisfaction with sector of employment, this question is not incentive compatible. Individuals may not report truthfully if they think their answers may be used to design and implement policies that could affect them, or if they suspect their answers could be shared with third parties like their employers.

To overcome these problems, we estimate the degrees of informality by choice and exclusion in each metropolitan area using the quantile regression decomposition method proposed by Machado and Mata (2005); we account for sample selection bias as proposed by Albretch

et.al (2009). This technique allows us to decompose the difference between formal and informal wage distributions into two components: the differences in the distribution of observable characteristics (e.g., firm size, job tenure, gender, education) between formal and informal workers, and the difference in how these characteristics are remunerated in each sector.[4]

Using this approach, we can estimate which effect dominates at each point of the distribution at the city level. If the coefficient effect explains a larger proportion of the formal-informal wage gap in a determined city, it means that in that city informal workers earn less than their formal counterparts because they receive lower returns on their characteristics. This is consistent with the idea that informal workers are "excluded" from the formal labor market. In contrast, if the effect of characteristics dominates, then labor informality is due to workers having different endowments than those in formal employment, which can be a sign of informality by choice.

One drawback of this approach is that, if unobserved worker characteristics that determine the sector of employment are correlated with wages (e.g., ability, motivation, social capital), we would observe these differences as a "coefficient effect". Thus, we are likely to overestimate the coefficient effect and the share of workers who are informal by exclusion. That is why it is important to compare the results from both stated and revealed preference approaches.

[4] A detailed explanation of the estimation process can be found in the methodological appendix.

# Results

## Satisfaction with current job's benefits

Table 3 presents the results from the survey question regarding informal workers' satisfaction with their current job for each metropolitan area under analysis. Preference for informality is taken as a proxy of voluntary informality or informality by choice. No preference for informality is taken as a proxy of informality by exclusion. The preference for informality across all urban areas is 51%, and in half of the 23 metropolitan areas this preference is 50% or higher.

While there is no correlation between the size of the informal sector in each city and its volume of voluntary informal workers, there is a relatively high and negative correlation of -0.36 between the preference for informality and the formal-informal wage gap. This might indicate that workers tend to self-select into the informal sector when the gap between formal and informal earnings is low, indicating a response to an income effect. On the other hand, the metropolitan areas with the lower preference for informality are Cúcuta and Quibdó, cities with two of the larger formal-informal wage gaps in the country and with the larger informality rates.

These results suggest there is a large heterogeneity in the preference for informality across cities, which can be partially related to the size of the formal-informal wage gap. Table 4 shows the average wage for formal workers and informal ones, the latter divided between those with and without preference for informality. The wage gaps across these groups show that formal workers earn more than their informal counterparts regardless of their preference

for the informal sector across all cities, with the only exception of Bogotá where formal workers and informal ones who are satisfied with their job earn similar wages.

Table 4 also shows that there are significant differences among the wages of both groups of informal workers. Those satisfied with the benefits of their current employment earn significantly more than those who are not satisfied. This difference is significantly bigger in two of the principal cities of the country: Bogotá and Cali. In these cities the wage gap between formal workers and informal workers satisfied with their employment is the smallest among all metropolitan areas. This suggest that workers tend to prefer an informal job when the wages they earn in this sector is closer to the one they would earn in the formal sector.

## Results from the Quantile Decomposition

To estimate the share of the formal-informal wage gap that is explained by differences in characteristics and differences in how these characteristics are rewarded, we first need to account for selection into informality. We do this using a semiparametric least squares (SLS) model where the dependent variable is an indicator that takes the value of 1 if the person is employed in the informal sector. In this model, we control for gender, age, marital status, literacy, educational levels, years of schooling, firm size, and whether the individual is the head of the household.

Table 5 shows the results of this exercise for the total sample and five cities that capture various features of the economy: Bogotá (the capital and most important city in economic sense), Medellín (the second largest city in the country and economically, characterized by a large industrial sector), Barranquilla (where the largest port of the country is located), Cúcuta (across the border with Venezuela and the city with the largest share of informal employment

in the country) and Quibdó (one of the poorest cities in the country located in the Pacific region).[5]

For all cities, men, household heads (except for the case of Cúcuta), older individuals, as well as those with more education and those employed in big firms are less likely to be informal. For most cities, except for Bogotá and Medellin, there is a negative relationship between being married and holding a job in the informal sector. These results are in line with those from García (2017).

Despite the similarities in the determinants of labor informality, there is considerable heterogeneity in the size of certain coefficients across cities. For example, the gender effect is bigger in Cúcuta and smaller in Bogotá, which may reflect differences in the labor force participation of women across these cities (DANE, 2020). Similarly, the coefficients on age and years of education are smaller in Quibdó than in other cities, possibly due to the higher informality rate in the city compared to the rest of the country.

From the SLS estimated models, we calculated the power series of expansion as a term to include in the quantile regressions to correct for selection. These selection terms are statistically significant at the nation-wide level in most of the cases, but their significance is mixed at the metropolitan area level. Table 6 present the results of the quantile regressions correcting for selection at country level, while results for each urban area are shown in Appendix Tables 8 through 34.

Formal and informal workers have similar, positive, and concave returns on job tenure. These results (which are similar across all urban areas) are in line with classical economic theory

[5] Results for the rest of the cities can be found in the appendix.

(Heckman et.al, 1979; 2003). We also find that returns to tenure tend to be smaller at the lower quantiles of the distribution of earnings for formal workers, a result that could be explained by the rigidity of salaries at this part of the wages distribution.

Concerning gender differences in earnings, it is interesting to note that while the wage gap is increasing along the earnings distribution in the formal sector, the opposite is true in the informal sector. These patterns are true for most cities and could reflect differences in occupations held by men and women in each sector.

Regarding the returns to education, these are increasing along the earnings distribution. The returns of secondary education tend to be higher for informal than formal workers along the entire distribution, while the penalty for having no education is also higher in the informal sector. On the other hand, the returns to tertiary education and above tend to be larger for the formal sector. These patterns, however, hide significant differences across cities: in some cases (like Popayán and Riohacha) the returns to secondary education are similar for those in the $10^{th}$, $50^{th}$ and $90^{th}$ percentiles of the distribution of earnings, while in others (like Bucaramanga and Pasto) the returns to tertiary education tend to be similar across sectors.

Finally, there is a positive relation across all urban areas between earnings and firm size. Workers employed in smaller firms earn less the fewer employees the firm has, and this penalty is higher at higher percentiles of the income distribution and among informal workers. This potentially reflects lower productivity levels of smaller firms, especially in the informal sector.

Based on these results, figure 3 present the decomposition results for the total sample of urban areas and the selected cities. The vertical axis shows the difference in log wages

between formal and informal workers, while the horizontal axis expresses the different points of the wage distribution.

Wage gaps are significantly large across the entire wage distribution for the country as a whole, as well as for the selected cities. While there is a common U-shaped pattern across all metropolitan areas, there are significant differences in the levels of the wage gaps and its causes. In some of them like Medellín and Quibdó, the wage gap does not increase much at the top of the distribution, whereas in others like Barranquilla and Cúcuta, it shows much larger increases, even registering an income gap like that registered at the lower percentiles of their respective distributions. These differences in the shape of the wage gap distribution curves provide a first approach to the heterogeneity of informality in the different territories of Colombia.

Regarding the contribution of coefficients and characteristics to the wage gap in the different points of the distribution, the wage gap at the bottom of the distribution tends to be explained by differences in coefficients in the cases of the total sample, Bogotá, and Medellin. This means that the formal-informal wage gap for workers at the bottom of the distribution in these cities can be explained by differences in the payment to characteristics of workers. In the case of Barranquilla, the coefficient effect dominates the effect of the characteristics in explaining the wage gap in almost all points of the income distribution. These results indicate that informal workers at these segments of the distribution are paid less than formal workers because they earn lower returns on their skills and not only because they have differences in terms of those skills. In contrast, in Cúcuta and Quibdó, the wage differential between the formal and the informal sector is mainly driven by differences in their characteristics

themselves. These informal workers do not earn less than their formal counterparts because they are paid less for their skills but because they have different skills than formal workers.

As we move along the wage distribution, the predominance of the coefficients' effect decline, and the effects of characteristics becomes more important for explaining the formal-informal wage gap. The fact that at the top of the distribution the characteristics' effect takes predominance means that the returns that informal workers in this segment get on their characteristics are similar to the ones obtained by formal workers. This segment of workers still earns less money than their formal counterparts but might receive certain non-wage amenities from remaining in informality because of their characteristics. This group of workers might prefer to be informal to avoid social security payments that they consider unnecessary or inefficient (Maloney, 2009; Fields, 1990).

From the decomposition results, we can say that there are two kinds of informal workers in different positions compared to their formal counterparts. The first group of informal workers are those at the bottom of the distribution in the total sample, Bogotá and Medellín, and in most deciles of the distribution in the case of Barranquilla. These workers can be classified as informal by exclusion since their lower pay is mainly explained by differences in the return to their skills (education, experience, job tenure, age, etc.) compared those received by formal workers. Hence, this group of workers is primarily constituted by residual labor, rationed out from the formal sector due to entry barriers (Fields, 1990).

The second group of workers are those at the right tail of the wage distribution in the total sample, Bogotá, Medellín, Barranquilla, and along the entire wage distribution in the case of Cúcuta and Quibdó. For these workers, the payments they receive from their characteristics are comparable with those received by formal workers. These workers might find informal

work more valuable than a formal job. For them, informality could be a rational choice looking to avoid administrative costs of formality, taxes, and considered costly payments to social security services, or due to specific abilities that could give them non-wage advantages over similar positions in the formal sector (Magnac, 1991).

It is useful to compare the results from the different approaches used in this study to have a sense of how reliable survey responses can be as a proxy for the preference for informality. Table 7 presents a comparison between the predominant effect quantified by the analysis of surveys and the one estimated by quantile decomposition. We consider an effect as predominant from the analysis of surveys when the reported proportion of workers with a preference for informality was higher than the proportion of workers without it. Similarly, we consider the estimated effect of characteristics from the decomposition as predominant when it explained the larger part of the wage gap.

The results show that, for 12 urban areas as well as for the total sample, both methods estimated a similar predominance of voluntary informality for their informal workers. For eight of the 24 territories analyzed strong disparities among results from both methods were found. Finally, for two metropolitan areas, small disparities in the magnitude were found. Disparities among both methods can be due to two things. First, although using the same individuals, both methods are carried out using different sets of variables. This could lead to disparities in the reported information for both sets of variables, which could signal different outputs. Second, the first method employed was due to give an approximation to the number of voluntary informal workers in each metropolitan area, whereas the second method gives an estimation of the predominance of a condition that could be caused by voluntary informality or not throughout different levels of income for each urban area.

# Conclusions

Informality is pervasive in developing countries' labor markets. Informal work is often associated with low wages and productivity, job instability and lack of welfare benefits. In addition, it reduces the tax revenue that Governments can collect (Ulyssea, 2018) and makes it difficult to target welfare benefits (Gerard & Gonzaga, forthcoming). Thus, policymakers continuously make efforts to reduce the size of the informal sector.

While some scholars suggest informality is the product of a segmented labor market, where informal workers are "excluded" from the formal sector, another strand of the literature argues that informal workers could have a formal job but choose not to do so as a result of a cost-benefit analysis. Understanding which of these representations dominates in the labor market is crucial for the design of policies to encourage the registration of workers.

In this paper, we estimate the heterogeneity of the informal sector in the Colombian labor market. Using data from 2019 representative of each of the 23 largest metropolitan areas, we first take a stated preference approach to analyze workers' preference for informality using their answers regarding job satisfaction. Second, we study the extent and magnitude of the formal-informal wage gap along the entire wage distribution, and we use the Machado-Mata decomposition (with a correction for selection into informality) to distinguish between a "coefficient effect" that proxies for informality by exclusion and a "characteristics effect" which can be related to informality by choice. By combining these two methods we can characterize the predominant type of informality for the major metropolitan areas in Colombia.

Our results show an enormous heterogeneity inside the Colombian informal labor market, not only across metropolitan areas, but also within local labor markets. This results in a dynamic co-existence among formal, voluntary informal workers and individuals who end up in informality as a last resort to escape from unemployment after being excluded from the formal sector. These two types of informal workers are driven into informality by different incentives, and hence cannot be taken as a homogenous policy target when promoting formality. We find that cities like Villavicencio, Armenia, Riohacha, Pasto, Florencia, Bucaramanga, and Cali show preference and estimated results from the decomposition to have a predominant voluntary informality, while urban areas like Valledupar, Santa Marta, Montería, Medellín, and Barranquilla might tend to have more informal workers who end up in this condition because they are excluded from formality. This heterogeneity across urban areas deserves a disaggregated and focused set of labor policies that push some individuals into formality while attracts others. Future work should focus on developing similar exercises across a set of years and bring evidence about how voluntary informality changes over time.

# References


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

Alcaraz, C., Chiquiar, D., & Salcedo, A. (2015). *Informality and segmentation in the Mexican labor market* (No. 2015-25). Working Papers.

Almeida, R., & Carneiro, P. (2012). Enforcement of labor regulation and informality. *American Economic Journal: Applied Economics*, *4*(3), 64-89.

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

Bérgolo, M., & Cruces, G. (2014). Labor informality and the incentive effects of social protection systems: Evidence from a health reform in Uruguay. *Social Insurance, Informality, and Labour Markets: How to Protect Workers While Creating Good Jobs*, (7879), 58.

Camacho, A., Conover, E., & Hoyos, A. (2014). Effects of Colombia's social protection system on workers' choice between formal and informal employment. *The World Bank Economic Review*, *28*(3), 446-466.

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

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

Fernández, C., & Villar, L. (2016). A taxonomy of Colombia’s informal labor market.

Fernández, C., & Villar, L. (2017). The impact of lowering the payroll tax on informality in Colombia. Economía, 18(1), 125-155.

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

Fields, Gary S., 2005. "A guide to multisector labor market models," Policy Research Working Paper Series 32547, The World Bank.

Fields, N. M., Fugazza, M., & Maloney, W. F. (2010). Informal self-employment and macroeconomic fluctuations. Journal of Development Economics, 91(2), 211-226.

Gerard, F., & Gonzaga, G. (forthcoming). Informal labor and the efficiency cost of social programs: Evidence from unemployment insurance in Brazil. American Economic Journal: Economic Policy.

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

García Cruz, G. A. (2017). Labor Informality: Choice or Sign of Segmentation? A Qunatile Regression Approach at the Regional Level for Colombia. Review of Development Economics, 21 (4), 985 – 1017, 2017.

Harris, John R & Todaro, Michael P, 1970. "Migration, Unemployment & Development: A Two-Sector Analysis," American Economic Review, American Economic Association, vol. 60(1), pages 126-142.

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

Heckman, James, 2013. "Sample selection bias as a specification error," Applied Econometrics, Publishing House "SINERGIA PRESS", vol. 31(3), pages 129-137.

Kugler, A., & Kugler, M. (2009). Labor market effects of payroll taxes in developing countries: evidence from Colombia. Economic development and cultural change, 57(2), 335-358.

LEWIS, W.A. (1954), Economic Development with Unlimited Supplies of Labour. The Manchester School, 22: 139-191.

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

Magnac, T. (1991). Segmented or competitive labor markets. Econometrica: journal of the Econometric Society, 165-187.

Maloney, William F, 1999. "Does Informality Imply Segmentation in Urban Labor Markets? Evidence from Sectoral Transitions in Mexico," World Bank Economic Review, World Bank Group, vol. 13(2), pages 275-302.

Maloney, William F., 2004. "Informality Revisited," World Development, Elsevier, vol. 32(7), pages 1159-1178.

Maloney, William F., 2009. "How will Labor Markets Adjust to the Crisis? A Dynamic View," World Bank Other Operational Studies 10995, The World Bank.

Mondragón-Vélez, C., Peña, X., Wills, D., & Kugler, A. (2010). Labor market rigidities and informality in Colombia [with comment]. Economía, 11(1), 65-101.

Perry, Guillermo, William Maloney, Omar Arias et al., 2007, “Informality: Exit and Exclusion,” International Bank for Reconstruction and Development.

Pratap, Sangeeta & Quintin, Erwan, 2006. "Are labor markets segmented in developing countries? A semiparametric approach," European Economic Review, Elsevier, vol. 50(7), pages 1817-1841.

Saavedra-Caballero, F., & Londoño, M. O. (2018). Social Assistance and Informality: Examining the Link in Colombia. Revista de Economía del Rosario, 21(1), 81-120.

Tannuri-Pianto, M., & Pianto, D. (2002). Informal employment in Brazil-a choice at the top and segmentation at the bottom: a quantile regression approach. Anais do XXIV Encontro Brasileiro de Econometria, 2(15), 77.

Tornarolli, L., Battistón, D., Gasparini, L., & Gluzmann, P. (2014). Exploring trends in labor informality in Latin America, 1990-2010 (tech. rep.). Documento de Trabajo.

Ulyssea, G. (2018). Firms, informality, and development: Theory and evidence from Brazil. The American Economic Review, 108 (8), 2015-47.

# Tables and figures

**Table 1: Descriptive Statistics**

| Variables | Formal | | Informal | | Difference |
|---|---|---|---|---|---|
| | Mean | Std. Dev. | Mean | Std. Dev. | |
| Wage (COP) | 1,649,145.00 | 1,788,575 | 712,822.70 | 974,286.70 | 936,322.30 |
| Worked hours (monthly) | 191.1 | 44.62 | 173.80 | 73.20 | 17.34 |
| Hourly wage (COP) | 9,338.66 | 12,020.49 | 4,380.86 | 7,741.49 | 4,957.8 |
| Male (share) | 0.54 | 0.50 | 0.54 | 0.50 | 0.01 |
| Age | 39.20 | 12.27 | 41.24 | 14.82 | -2.04 |
| Married (share) | 0.27 | 0.44 | 0.18 | 0.38 | 0.09 |
| Literacy (share) | 1.00 | 0.05 | 0.97 | 0.18 | 0.03 |
| Years of Schooling | 12.86 | 3.73 | 9.05 | 4.14 | 3.81 |
| Has labor contract (share) | 0.91 | 0.29 | 0.34 | 0.47 | 0.57 |
| Paid vacations (share) | 0.75 | 0.43 | 0.03 | 0.18 | 0.72 |
| Paid Christmas (share) | 0.26 | 0.44 | 0.01 | 0.10 | 0.25 |
| Job Tenure | 78.45 | 98.21 | 81.83 | 113.49 | -3.38 |
| **Education Level** | | | | | |
| None | 0,00 | 0,05 | 0,03 | 0,18 | -0,03 |
| Primary | 0,06 | 0,24 | 0,24 | 0,42 | -0,17 |
| Highschool | 0,35 | 0,48 | 0,38 | 0,49 | -0,03 |
| Technical/ Undergraduate | 0,41 | 0,49 | 0,14 | 0,35 | 0,27 |
| Postgraduate degree | 0,11 | 0,31 | 0,01 | 0,10 | 0,10 |
| **Firm Size** | | | | | |
| Work alone | 0.07 | 0.26 | 0.61 | 0.49 | -0.53 |
| 2 to 10 Employees | 0.14 | 0.34 | 0.32 | 0.47 | -0.18 |
| 11 to 50 Employees | 0.16 | 0.37 | 0.04 | 0.21 | 0.12 |
| 51 to 100 Employees | 0.05 | 0.22 | 0.00 | 0.07 | 0.05 |
| More than 100 Employees | 0.58 | 0.49 | 0.02 | 0.15 | 0.55 |
| ***observations*** | **122.365** | | **161.390** | | |

Notes: Wages are presented as the monthly average income before taxes in Colombian Pesos. Worked hours are expressed as the average worked hours by individuals in a one-month period. Hourly wage is the monthly wage divided on the number of monthly worked hours. Male, Married, Literacy, Contract, Paid Vacations and Christmas, and Unemployment savings, are dummy variables which takes de value of 0 in the case of non-presence of each condition and 1 in the opposite case. Contract denotes whether a worker has a written contract with his employer, Paid Vacations, paid Christmas, and unemployment savings are a set of benefits outside from the monthly salary. **All mean differences are significant at the 1% level.**

Source: Own calculations with monthly data from GEIH, 2019.

**Table 2: Hourly wage differentials between formal and informal workers by each Metropolitan Area**

| City | Hourly Wage | | | | Wage Differential | | Informality Rate |
|---|---|---|---|---|---|---|---|
| | Formal | SD | Informal | SD | COP | (%) | |
| Armenia | 8,527.21 | 9,632.95 | 4,192.18 | 6,220.08 | 4,335.03 | 50.84% | 51.0% |
| Barranquilla | 8,960.02 | 13,492.17 | 4,416.20 | 6,094.07 | 4,543.82 | 50.71% | 63.1% |
| Bogotá | 10,480.92 | 14,309.65 | 7,768.03 | 23,735.24 | 2,712.89 | 25.88% | 43.0% |
| Bucaramanga | 8,520.64 | 11,479.67 | 4,586.84 | 5,024.21 | 3,933.80 | 46.17% | 53.5% |
| Cali | 9,182.78 | 15,871.28 | 6,213.12 | 10,321.35 | 2,969.66 | 32.34% | 51.0% |
| Cartagena | 8,015.37 | 7,904.42 | 4,273.28 | 3,414.73 | 3,742.09 | 46.69% | 59.5% |
| Cúcuta | 7,966.43 | 7,731.22 | 3,222.27 | 3,112.26 | 4,744.16 | 59.55% | 70.6% |
| Florencia | 10,353.37 | 10,931.75 | 3,913.40 | 4,767.66 | 6,439.97 | 62.20% | 57.4% |
| Ibagué | 9,009.73 | 15,043.29 | 4,401.98 | 5,390.83 | 4,607.75 | 51.14% | 51.9% |
| Manizales | 8,140.44 | 8,408.01 | 4,715.92 | 4,420.06 | 3,424.52 | 42.07% | 34.3% |
| Medellín | 9,102.93 | 10,037.59 | 4,139.30 | 4,956.19 | 4,963.62 | 54.53% | 39.9% |
| Montería | 9,032.83 | 88,371.82 | 3,851.00 | 4331.82 | 5,181.84 | 57.37% | 63.2% |
| Neiva | 10,717.82 | 11,431.56 | 4,287.51 | 6,174.82 | 6,430.31 | 60.00% | 55.1% |
| Pasto | 7,522.72 | 8,060.85 | 4,235.44 | 3,695.02 | 3,287.28 | 43.70% | 59.9% |
| Pereira | 10,506.48 | 19,232.20 | 4,158.35 | 4,593.61 | 6,348.13 | 60.42% | 47.9% |
| Popayán | 12,662.81 | 11,038.05 | 4,037.51 | 9,165.13 | 8,625.30 | 68.12% | 58.6% |
| Quibdó | 11,599.72 | 13,654.25 | 3,536.20 | 4,815.61 | 8,063.52 | 69.51% | 56.0% |
| Riohacha | 8,890.31 | 9,730.82 | 3,733.56 | 4,257.87 | 5,156.75 | 58.00% | 66.4% |
| Santamarta | 10,066.05 | 11,297.61 | 3,622.83 | 3,498.14 | 6,443.22 | 64.01% | 63.5% |
| Sincelejo | 8,890.31 | 97,308.15 | 3,733.56 | 4,257.87 | 5,156.75 | 58.00% | 70.4% |
| Tunja | 10,632.69 | 14,570.09 | 4,694.85 | 7,109.29 | 5,937.84 | 55.85% | 44.2% |
| Valledupar | 9,056.75 | 11,542.15 | 4,092.70 | 5,120.74 | 4,964.05 | 54.81% | 67.6% |
| Villavicencio | 9,956.21 | 12,869.20 | 4,993.89 | 8,233.22 | 4,962.32 | 49.84% | 56.4% |
| Correlation wage differential-informality rate | | | | | | | 0.42 |

Note: Columns 1 and 3 show the mean hourly wage of formal and informal workers, respectively, for each urban area included in the survey. Columns 2 and 4 present the corresponding standard deviations. Column 5 shows the average wage differential between formal and informal workers for each urban area, and column 6 reports the wage differential as a percentage of the average hourly wage of formal workers. Column 7 presents the informality rate at the urban area. The last row presents the correlation coefficient between the wage differentials (%) and the informality rate. Hourly wages presented in Colombian pesos.

## Table N°3: Quantification of the preference for informality across Colombian Metropolitan Areas

| Metropolitan Area | Preference for Informality | | No preference for Informality | |
|---|---|---|---|---|
| | Proportion | SD | Proportion | SD |
| **Armenia** | **0.56** | **0.50** | **0.44** | **0.50** |
| Barranquilla | 0.50 | 0.50 | 0.50 | 0.50 |
| Bogotá | 0.50 | 0.50 | 0.50 | 0.50 |
| **Bucaramanga** | **0.62** | **0.49** | **0.38** | **0.49** |
| **Cali** | **0.58** | **0.49** | **0.42** | **0.49** |
| **Cartagena** | **0.52** | **0.50** | **0.48** | **0.50** |
| Cúcuta | 0.35 | 0.48 | 0.65 | 0.48 |
| **Florencia** | **0.61** | **0.49** | **0.39** | **0.49** |
| Ibagué | 0.44 | 0.50 | 0.56 | 0.50 |
| Manizales | 0.49 | 0.50 | 0.51 | 0.50 |
| Medellín | 0.47 | 0.50 | 0.53 | 0.50 |
| Montería | 0.45 | 0.50 | 0.55 | 0.50 |
| Neiva | 0.44 | 0.50 | 0.56 | 0.50 |
| **Pasto** | **0.78** | **0.42** | **0.22** | **0.42** |
| **Pereira** | **0.60** | **0.49** | **0.40** | **0.49** |
| Popayán | 0.48 | 0.50 | 0.52 | 0.50 |
| Quibdó | 0.34 | 0.47 | 0.66 | 0.47 |
| **Riohacha** | **0.61** | **0.49** | **0.39** | **0.49** |
| Santamarta | 0.41 | 0.49 | 0.59 | 0.49 |
| **Sincelejo** | **0.61** | **0.49** | **0.39** | **0.49** |
| Tunja | 0.44 | 0.50 | 0.56 | 0.50 |
| Valledupar | 0.45 | 0.50 | 0.55 | 0.50 |
| **Villavicencio** | **0.64** | **0.48** | **0.36** | **0.48** |
| **All Urban Areas** | **0.51** | **0.49** | **0.49** | **0.49** |

Note: The table shows, for each urban area, the share and standard deviation of informal workers who manifested to be satisfied (columns 1 and 2) and dissatisfied (columns 3 and 4) with the benefits provided by their current job. Cities where a preference for informality is observed as predominant are **bolded**.

## Table N° 4: Wage differentials among different types of informal workers and formals

| Metropolitan Area | Formal | | Preference for Informality | | No preference for Informality | | Differences | |
|---|---|---|---|---|---|---|---|---|
| | Hourly wage | SD | Hourly wage | SD | Hourly wage | SD | F-IC | IC-IE |
| **Armenia** | 8,527.21 | 9,632.95 | 4,747.87 | 7,011.70 | 3,480.23 | 4,937.86 | 3,779.34 | 1,267.64 |
| **Barranquilla** | 8,960.02 | 13,492.17 | 4,983.38 | 5,836.10 | 3,871.48 | 6,284.44 | 3,976.64 | 1,111.90 |
| **Bogotá** | 10,480.92 | 14,309.65 | 10,660.14 | 32,629.27 | 4,901.83 | 7,085.83 | - 179.22 | 5,758.31 |
| **Bucaramanga** | 8,520.64 | 11,479.67 | 5,091.31 | 5,646.79 | 3,770.73 | 3,664.07 | 3,429.33 | 1,320.58 |
| **Cali** | 9,182.78 | 15,871.28 | 7,435.47 | 12,215.36 | 4,551.58 | 6,614.16 | 1,747.31 | 2,883.89 |
| **Cartagena** | 8,015.37 | 7,904.42 | 4,974.37 | 4,106.19 | 3,533.08 | 2,259.69 | 3,041.00 | 1,441.29 |
| **Cúcuta** | 7,966.43 | 7,731.22 | 4,005.80 | 3,888.03 | 2,798.85 | 2,498.63 | 3,960.63 | 1,206.95 |
| **Florencia** | 10,353.37 | 10,931.75 | 4,409.19 | 5,688.12 | 3,165.04 | 2,698.56 | 5,944.18 | 1,244.15 |
| **Ibagué** | 9,009.73 | 15,043.29 | 5,234.41 | 7,085.74 | 3,735.92 | 3,339.79 | 3,775.32 | 1,498.49 |
| **Manizales** | 8,140.44 | 8,408.01 | 5,255.81 | 5,090.23 | 4,200.66 | 3,594.51 | 2,884.63 | 1,055.15 |
| **Medellín** | 9,102.93 | 10,037.59 | 4,846.61 | 6,570.90 | 3,520.04 | 2,730.45 | 4,256.32 | 1,326.57 |
| **Montería** | 9,032.83 | 88,371.82 | 4,589.96 | 5,465.97 | 3,228.93 | 2,927.30 | 4,442.87 | 1,361.04 |
| **Neiva** | 10,717.82 | 11,431.56 | 5,164.02 | 8,050.74 | 3,603.41 | 4,030.51 | 5,553.81 | 1,560.61 |
| **Pasto** | 7,522.72 | 8,060.85 | 4,481.00 | 3,868.26 | 3,382.83 | 2,859.30 | 3,041.72 | 1,098.18 |
| **Pereira** | 10,506.48 | 19,232.20 | 4,452.81 | 4,893.31 | 3,727.79 | 4,079.21 | 6,053.67 | 725.03 |
| **Popayán** | 12,662.81 | 11,038.05 | 4,415.32 | 12,327.29 | 3,682.90 | 4,471.91 | 8,247.49 | 732.42 |
| **Quibdó** | 11,599.72 | 13,654.25 | 4,282.10 | 6,400.88 | 3,162.07 | 3,722.83 | 7,317.62 | 1,120.03 |
| **Riohacha** | 8,890.31 | 9,730.82 | 4,247.79 | 4,906.60 | 2,956.01 | 2,852.99 | 4,642.52 | 1,291.79 |
| **Santamarta** | 10,066.05 | 11,297.61 | 4,170.63 | 3,857.64 | 3,249.85 | 3,177.32 | 5,895.42 | 920.78 |
| **Sincelejo** | 8,890.31 | 97,308.15 | 4,247.79 | 4,906.60 | 2,956.01 | 2,852.99 | 4,642.52 | 1,291.79 |
| **Tunja** | 10,632.69 | 14,570.09 | 5,350.66 | 9,620.42 | 4,190.21 | 4,202.24 | 5,282.03 | 1,160.45 |
| **Valledupar** | 9,056.75 | 11,542.15 | 4,546.29 | 4,452.80 | 3,717.24 | 5,586.63 | 4,510.46 | 829.05 |
| **Villavicencio** | 9,956.21 | 12,869.20 | 5,478.48 | 9,182.06 | 4,124.95 | 6,083.15 | 4,477.73 | 1,353.53 |

Note: All differences are significant at the 0,05 level, except for the case of Bogotá. Hourly wage represents the average hourly wage for group for each metropolitan area. Preference for informality is taken as the share of the workforce from each city which answered to the in a negative way to the question about the satisfaction with the benefits they were obtaining from their current jobs. F – IC represents the hourly wage gap between formal workers and informal workers with revealed preference for informality. IC – IE represents the hourly wage gap between informal workers with revealed preference for informality and informal workers with revealed preference for formality.

## Table N° 5: SLS model estimates for the probability of being informal

| y = 1 informal; 0 formal | **Total Sample** | **Bogota** | **Medellin** | **Barranquilla** | **Cucuta** | **Quibdó** |
|---|---|---|---|---|---|---|
| Constant | 4.02 | 3.98 | 4.12 | 3.99 | 5.11 | 4.45 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Male | -0.122*** | -0.0529*** | -0.155*** | -0.150*** | -0.234*** | -0.172*** |
| | (-4.69) | (-1.79) | (-2.20) | (-2.26) | (-3.91) | (-3.70) |
| Age | -0.0337*** | -0.0400*** | -0.0534*** | -0.0329*** | -0.0332*** | -0.0171** |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Age Sq. | 0.0002*** | 0.0003*** | 0.0005*** | 0.0002*** | 0.00024*** | 4.03e-05 |
| | (1.03e-05) | (4.44e-05) | (6.33e-05) | (5.11e-05) | (8.15e-05) | (8.44e-05) |
| Head of Household | -0.0684*** | -0.0722*** | -0.102*** | -0.0679*** | -0.0210 | -0.105*** |
| | (-4.71) | (-1.84) | (-1.95) | (-2.18) | (-3.38) | (-3.78) |
| Married | -0.0833*** | -0.0106 | -0.0308 | -0.136*** | -0.0758** | -0.205*** |
| | (-5.58) | (-2.22) | (-2.13) | (-2.48) | (-3.71) | (-5.61) |
| Literacy | -0.0143 | 0.196 | -0.290 | -0.0662 | -0.682 | -0.370 |
| | (-3.54) | (-1.25) | (-3.18) | (-1.89) | (-4.26) | (-2.58) |
| Years of Schooling | -0.0299*** | -0.00619 | -0.00162 | -0.0447*** | -0.00959 | -0.0467*** |
| | (-0.14) | (-0.05) | (-0.05) | (-0.76) | (-0.10) | (-0.01) |
| **Education levels** | | | | | | |
| No education | 0.238*** | 0.588*** | 0.0637 | 0.0605 | -0.122 | -0.106 |
| | (3.87) | (1.51) | (2.84) | (2.59) | (2.94) | (2.93) |
| Secondary | -0.124*** | -0.151*** | -0.198*** | -0.102** | -0.257*** | -0.154** |
| | (-0.91) | (-3.43) | (-3.64) | (-4.38) | (-6.87) | (-7.33) |
| Terciary | -0.359*** | -0.386*** | -0.459*** | -0.408*** | -0.691*** | -0.443*** |
| | (-1.39) | (-5.29) | (-5.70) | (-7.05) | (-1.15) | (-1.05) |
| Post-tertiary | -0.644*** | -0.370*** | -0.479*** | -0.660*** | -1.010*** | -1.016*** |
| | (-2.06) | (-6.93) | (-7.91) | (-9.61) | (-1.48) | (-1.48) |
| **Firm size** | | | | | | |
| One worker | 2.049*** | 1.643*** | 1.682*** | 2.268*** | 3.035*** | 2.218*** |
| | (2.81) | (1.15) | (1.31) | (1.39) | (4.60) | (1.98) |
| 2 - 10 employees | 1.597*** | 1.330*** | 1.281*** | 1.723*** | 2.326*** | 1.696*** |
| | (2.49) | (1.04) | (1.03) | (1.17) | (4.18) | (1.66) |
| 11 - 50 employees | 0.704*** | 0.594*** | 0.465*** | 0.788*** | 0.980*** | 0.710*** |
| | (1.65) | (7.26) | (5.17) | (7.11) | (2.80) | (10.90) |
| **Observations** | 305,199 | 16,074 | 18,907 | 19,087 | 10,322 | 6,863 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size.

## Table N° 6 Quantile Regressions (Total Sample)

| *y = log hourly wage* | *Formal* | | | *Informal* | | |
|---|---|---|---|---|---|---|
| | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.140*** | 8.337*** | 8.622*** | 7.364*** | 8.225*** | 8.877*** |
| | (0.00657) | (0.00297) | (0.00943) | (0.0177) | (0.00731) | (0.0273) |
| λ | -15.46 | -11.64** | -9.298*** | 2.759*** | 0.879 | -0.847*** |
| | (12.19) | -5.093 | -1.912 | (0.116) | (0.672) | (0.112) |
| Male | 0.00152 | 0.0506*** | 0.155*** | 0.394*** | 0.141*** | 0.0501*** |
| | (0.00236) | (0.00238) | (0.00611) | (0.00894) | (0.00337) | (0.00581) |
| Tenure (months) | 0.000393*** | 0.00135*** | 0.00219*** | 0.00255*** | 0.00160*** | 0.00209*** |
| | (4.22e-05) | (4.12e-05) | (8.84e-05) | (9.81e-05) | (4.28e-05) | (7.34e-05) |
| Tenure squared | 5.18e-07*** | -5.58e-07*** | -1.79e-06*** | -4.99e-06*** | -2.75e-06*** | -3.08e-06*** |
| | (1.35e-07) | (1.16e-07) | (2.36e-07) | (2.58e-07) | (1.07e-07) | (1.68e-07) |
| **Educational levels** | | | | | | |
| No education | -0.0274 | -0.0672*** | -0.105** | -0.464*** | -0.304*** | -0.209*** |
| | (0.0302) | (0.0183) | (0.0423) | (0.0259) | (0.0114) | (0.0142) |
| Secondary | 0.117*** | 0.0612*** | 0.196*** | 0.258*** | 0.138*** | 0.208*** |
| | (0.00617) | (0.00249) | (0.00867) | (0.00927) | (0.00355) | (0.00621) |
| Tertiary | 0.224*** | 0.460*** | 0.957*** | 0.460*** | 0.370*** | 0.757*** |
| | (0.00627) | (0.00432) | (0.00958) | (0.0122) | (0.00520) | (0.0131) |
| Post Tertiary | 0.907*** | 1.324*** | 1.471*** | 0.866*** | 1.350*** | 1.782*** |
| | (0.0101) | (0.00596) | (0.0120) | (0.0433) | (0.0285) | (0.0468) |
| **Firm size** | | | | | | |
| One worker | -0.642*** | -0.126*** | 0.104*** | -0.812*** | -0.438*** | -0.405*** |
| | (0.0187) | (0.00784) | (0.0145) | (0.0159) | (0.00737) | (0.0269) |
| 2 - 10 employees | -0.150*** | -0.0886*** | -0.0397*** | -0.440*** | -0.251*** | -0.317*** |
| | (0.00561) | (0.00312) | (0.0112) | (0.0158) | (0.00728) | (0.0272) |
| 11 - 50 employees | -0.0325*** | -0.0660*** | -0.104*** | -0.135*** | -0.117*** | -0.237*** |
| | (0.00256) | (0.00266) | (0.00845) | (0.0200) | (0.00875) | (0.0303) |
| Observations | 122,363 | 122,363 | 122,363 | 161,377 | 161,377 | 161,377 |

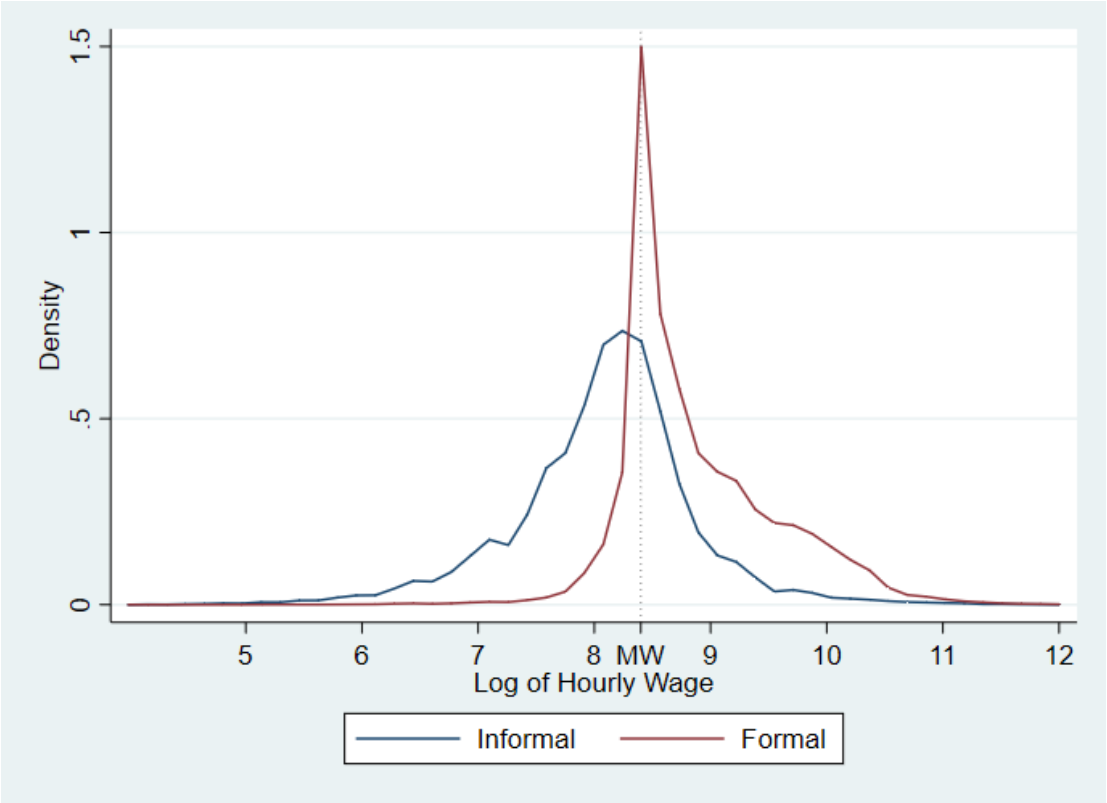
Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Tenure refers to months of work in the current job.

## Table N° 7: Comparison of results

| Metropolitan Area | Prefference for Informality Proportion | Predominance of the effect of characteristics |
|---|---|---|
| **Armenia** | 0.56 | 0.67 |
| **Barranquilla** | 0.49 | 0.11 |
| **Bogotá** | 0.50 | 0.55 |
| **Bucaramanga** | 0.62 | 0.53 |
| **Cali** | 0.58 | 0.6 |
| **Cartagena** | 0.52 | 0.31 |
| **Cúcuta** | 0.35 | 0.6 |
| **Florencia** | 0.61 | 0.75 |
| **Manizales** | 0.49 | 0.53 |
| **Medellín** | 0.47 | 0.49 |
| **Montería** | 0.45 | 0.48 |
| **Neiva** | 0.44 | 0.54 |
| **Pasto** | 0.78 | 1 |
| **Pereira** | 0.60 | 0.47 |
| **Popayán** | 0.48 | 0.69 |
| **Quibdó** | 0.34 | 0.95 |
| **Riohacha** | 0.61 | 0.85 |
| **Santamarta** | 0.41 | 0.25 |
| **Sincelejo** | 0.61 | 0.39 |
| **Tunja** | 0.44 | 1 |
| **Valledupar** | 0.45 | 0.37 |
| **Villavicencio** | 0.64 | 0.61 |
| **Total Sample** | 0.51 | 0.51 |

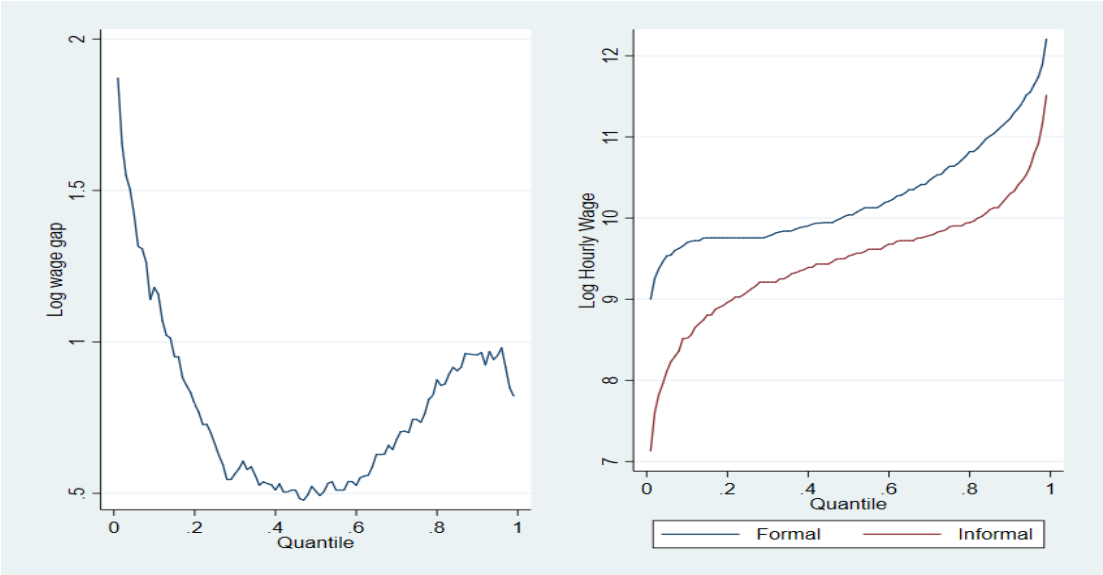
**Note:** Predominance of the effect of characteristics is taken as the number of deciles of the decomposed wage gap where the characteristics effect dominated the coefficients effect (*See step 9 in the methodological appendix).* Matching results are underlined in green. Contradictory results are underlined in orange. In the case of Manizales and Bogotá, both results were contradictory but extremely close, so it is not underlined with green or orange.

**Figure 1: Kernel Density of Log Hourly Wage by Formal and Informal Workers**



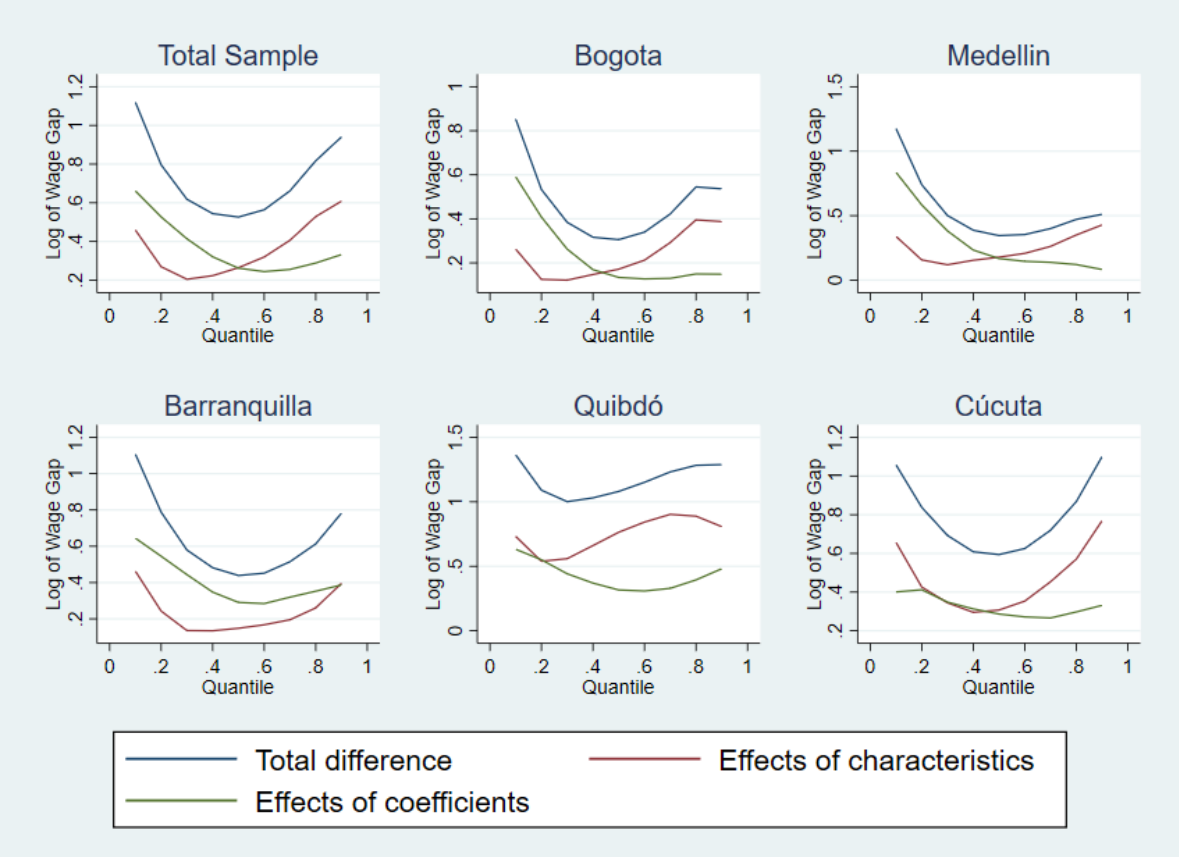

Note: The Figure the kernel density estimates of log hourly wages for workers in the formal (blue) and informal (red) sectors. Dotted line on "MW" denotes the Log of the Hourly Minimum Wage. Data was obtained the 2019 GEIH produced by DANE.

**Figure 2: Wage gap for formal and informal sector across the wage distribution**




Note: The figure shows, for each quantile of the distribution of log hourly wages, the distribution of the wage gap between formal and informal sectors (1) and the distribution of log hourly wages (2). Data was obtained the 2019 GEIH produced by DANE.

**FIGURE N° 3: QUANTILE DECOMPOSITION OF THE WAGE GAP AMONG FORMAL AND INFORMAL SECTOR**




Note: The figure shows the quantile decomposition of the formal-informal wage gap (blue) on the effects of coefficients (green) and the effects of characteristics (red) for Colombia as well as for the cities of Bogotá, Medellín, Barranquilla, Quibdó and Cúcuta. Data was obtained the 2019 GEIH produced by DANE.

# Appendix A

## Additional tables and figures

**Table N° A1: SLS model estimates for the probability of being informal (Cali, Cartagena, Tunja, Manizales and Florencia)**

| y=1 informal; 0 formal | Cali | Cartagena | Tunja | Manizales | Florencia |
|---|---|---|---|---|---|
| Male | -0.124*** | -0.0423 | -0.133*** | -0.183*** | -0.105*** |
| | (-2.10) | (-2.75) | (-4.82) | (-2.93) | (-3.17) |
| Age | -0.0385*** | -0.0407*** | -0.0636*** | -0.0523*** | -0.0385*** |
| | (0.00) | (0.00) | (0.01) | (0.00) | (0.00) |
| Age Sq. | 0.000311*** | 0.0003*** | 0.0005*** | 0.0004*** | 0.0003*** |
| | (4.32e-05) | (7.22e-05) | (0.000140) | (6.99e-05) | (8.00e-05) |
| Head of Household | -0.0892*** | -0.00904 | -0.0613 | -0.106*** | -0.123*** |
| | (-2.08) | (-2.72) | (-4.60) | (-2.60) | (-3.38) |
| Married | -0.00993 | -0.0674* | -0.0881* | -0.139*** | -0.113*** |
| | (-2.49) | (-3.70) | (-5.09) | (-2.92) | (-4.06) |
| Literacy | 0.0817 | 0.883*** | -0.555 | -0.316 | -0.263 |
| | (-1.25) | (-2.45) | (-7.04) | (-2.07) | (-2.96) |
| Years of Schooling | -0.0153** | -0.0469*** | -0.0258** | -0.0296*** | -0.0523*** |
| | (-0.65) | (-0.10) | (-0.12) | (-0.77) | (-1.00) |
| **Education levels** | | | | | |
| No education | 0.261 | 0.932*** | 0.187 | 0.0230 | -0.169 |
| | (1.69) | (3.10) | (9.31) | (1.99) | (2.77) |
| Secondary | -0.229*** | -0.0233 | -0.145 | -0.0780* | -0.197*** |
| | (-3.80) | (-5.70) | (-9.04) | (-4.57) | (-6.43) |
| Terciary | -0.430*** | -0.272*** | -0.658*** | -0.296*** | -0.440*** |
| | (-6.35) | (-8.73) | (-1.64) | (-6.86) | (-9.39) |
| Post-tertiary | -0.375*** | -1.012*** | -1.180*** | -0.649*** | -0.981*** |
| | (-9.07) | (-1.67) | (-2.17) | (-1.12) | (-1.71) |
| **Firm size** | | | | | |
| One worker | 1.761*** | 2.380*** | 3.729*** | 1.913*** | 2.444*** |
| | (1.01) | (1.88) | (7.42) | (1.80) | (2.23) |
| 2 - 10 employees | 1.407*** | 1.753*** | 3.018*** | 1.542*** | 1.808*** |
| | (9.21) | (1.43) | (6.30) | (1.52) | (1.86) |
| 11 - 50 employees | 0.719*** | 0.721*** | 1.617*** | 0.772*** | 0.705*** |
| | (7.08) | (8.86) | (4.46) | (1.01) | (11.40) |
| **Observations** | 14,177 | 12,143 | 8,355 | 11,35 | 8,172 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size.

**Table N° A2: SLS model estimates for the probability of being informal (Popayan, Valledupar, Monteria, Neiva and Riohacha)**

| y=1 informal; 0 formal. | **Popayan** | **Valledupar** | **Monteria** | **Neiva** | **Riohacha** |
|---|---|---|---|---|---|
| Male | -0.0885*** | -0.116*** | -0.0740* | -0.272*** | -0.0357 |
| | (-2.37) | (-3.61) | (-3.86) | (-4.27) | (-3.25) |
| Age | -0.0178*** | -0.0436*** | -0.0494*** | -0.0439*** | -0.0307*** |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Age Sq. | 2.05e-05 | 0.0003*** | 0.0003*** | 0.0002*** | 0.0001** |
| | (5.22e-05) | (9.01e-05) | (8.65e-05) | (9.20e-05) | (7.83e-05) |
| Head of Household | 0.0197 | -0.0989*** | -0.148*** | -0.0627 | -0.0282 |
| | (-2.37) | (-3.44) | (-4.58) | (-4.06) | (3.23) |
| Married | -0.0844*** | -0.156*** | 0.0122 | -0.117** | -0.0789* |
| | (-2.93) | (-4.01) | (-4.71) | (-4.81) | (-4.19) |
| Literacy | 0.333** | -0.0767 | -1.181*** | -0.789** | 0.580* |
| | (-1.67) | (-2.39) | (-2.94) | (-3.17) | (-3.15) |
| Years of Schooling | -0.0456*** | -0.0693*** | -0.0714*** | -0.0853*** | -0.0453*** |
| | (-0.751) | (-0.01) | (-0.01) | (-0.01) | (-0.10) |
| **Education levels** | | | | | |
| No education | 0.906** | -0.452* | -0.767*** | -0.324 | 0.463 |
| | (4.41) | (2.60) | (2.82) | (3.40) | (3.35) |
| Secondary | -0.194*** | -0.0847 | -0.172** | -0.0206 | -0.207*** |
| | (-4.93) | (-7.05) | (-8.40) | (-7.99) | (-7.20) |
| Terciary | -0.391*** | -0.316*** | -0.453*** | -0.267** | -0.498*** |
| | (-7.01) | (-1.03) | (-1.37) | (-1.17) | (-1.03) |
| Post-tertiary | -0.646*** | -0.820*** | -1.008*** | -0.721*** | -0.844*** |
| | (-1.01) | (-1.40) | (-1.81) | (-1.75) | (-1.25) |
| **Firm size** | | | | | |
| One worker | 1.816*** | 2.936*** | 3.726*** | 4.442*** | 2.415*** |
| | (1.07) | (3.14) | (6.80) | (3.23) | (2.28) |
| 2 - 10 employees | 1.424*** | 2.417*** | 3.111*** | 3.437*** | 1.846*** |
| | (8.97) | (2.79) | (6.17) | (3.55) | (1.89) |
| 11 - 50 employees | 0.553*** | 1.464*** | 1.584*** | 0.815*** | 0.890*** |
| | (6.00) | (2.18) | (4.02) | (11.71) | (11.93) |
| **Observations** | 10,508 | 11,142 | 10,983 | 10,15 | 10,561 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size.

**Table N° A2: SLS model estimates for the probability of being informal (Santamarta, Villavicencio, Pasto, Armenia and Pereira)**

| y=1 informal; 0 formal. | **Santamarta** | **Villavicencio** | **Pasto** | **Armenia** | **Pereira** |
|---|---|---|---|---|---|
| Male | -0.153*** | -0.233*** | -0.144*** | -0.463*** | -0.200*** |
| | (-2.53) | (-3.29) | (-2.92) | (-5.40) | (-2.77) |
| Age | -0.0294*** | -0.0286*** | -0.0348*** | -0.0826*** | -0.0447*** |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Age Sq. | 0.0002*** | 0.0002*** | 0.0002*** | 0.000676*** | 0.000415*** |
| | (5.39e-05) | (6.19e-05) | (5.79e-05) | (9.29e-05) | (6.41e-05) |
| Head of Household | -0.0248 | -0.0463* | -0.0598** | -0.0252 | -0.0670*** |
| | (-2.42) | (-2.76) | (-2.99) | (-4.73) | (-2.45) |
| Married | -0.0737** | -0.0706** | -0.162*** | -0.109** | -0.0788*** |
| | (-3.07) | (-3.36) | (-3.25) | (-5.52) | (-2.90) |
| Literacy | -0.178 | 0.184 | -0.0645 | -0.800** | 0.0131 |
| | (-2.85) | (-3.04) | (-2.23) | (-3.81) | (-2.60) |
| Years of Schooling | -0.0271*** | -0.0436*** | -0.0298*** | -0.0576*** | -0.0265*** |
| | (-0.08) | (-0.01) | (-0.00) | (-0.13) | (-0.07) |
| **Education levels** | | | | | |
| No education | 0.0972 | 0.279 | 0.420 | -0.638* | 0.270 |
| | (2.94) | (3.17) | (3.56) | (3.69) | (2.77) |
| Secondary | -0.0692 | -0.156*** | -0.177*** | -0.317*** | -0.175*** |
| | (-4.99) | (-5.36) | (-0.84) | (-8.96) | (-4.55) |
| Terciary | -0.319*** | -0.369*** | -0.567*** | -0.595*** | -0.398*** |
| | (-7.85) | (-8.92) | (-9.21) | (-1.35) | (-6.99) |
| Post-tertiary | -1.061*** | -0.567*** | -0.845*** | -1.149*** | -0.558*** |
| | (-1.27) | (-1.16) | (-1.02) | (-2.55) | (-1.18) |
| **Firm size** | | | | | |
| One worker | 2.338*** | 2.509*** | 2.488*** | 4.264*** | 2.061*** |
| | (1.90) | (2.64) | (1.82) | (4.02) | (1.84) |
| 2 - 10 employees | 1.737*** | 2.069*** | 1.966*** | 2.916*** | 1.649*** |
| | (1.63) | (2.39) | (1.65) | (3.33) | (1.62) |
| 11 - 50 employees | 0.745*** | 0.965*** | 0.913*** | 1.127*** | 0.817*** |
| | (1.01) | (1.60) | (12.41) | (1.50) | (10.87) |
| **Observations** | 13,687 | 10,765 | 10,767 | 10,215 | 11,094 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size.

**Table N° A3: SLS model estimates for the probability of being informal (Bucaramanga, Sincelejo and Ibagué)**

| | Bucaramanga | Sincelejo | Ibague |
|---|---|---|---|
| Male | -0.145*** | -0.0988*** | -0.422*** |
| | (-2.42) | (-3.03) | (-5.42) |
| Age | -0.0404*** | -0.0321*** | -0.0712*** |
| | (0.00) | (0.00) | (0.00) |
| Age Sq. | 0.000315*** | 0.000211*** | 0.000602*** |
| | (5.13e-05) | (6.33e-05) | (9.47e-05) |
| Head of Household | -0.0503** | -0.106*** | -0.143*** |
| | (-2.25) | (-3.19) | (-4.07) |
| Married | -0.0396 | -0.187*** | -0.0496 |
| | (-2.51) | (-3.48) | (-4.67) |
| Literacy | -0.405* | -0.00993 | -1.167*** |
| | (-2.31) | (-2.48) | (-4.25) |
| Years of Schooling | -0.0227*** | -0.0694*** | -0.0592*** |
| | (-0.72) | (-0.10) | (-0.12) |
| **Education levels** | | | |
| No education | -0.115 | 0.137 | -0.268 |
| | (1.85) | (2.67) | (4.50) |
| Secondary | -0.171*** | 0.00473 | -0.140* |
| | (-4.61) | (-6.09) | (-7.67) |
| Terciary | -0.481*** | -0.292*** | -0.635*** |
| | (-7.24) | (-8.41) | (-1.25) |
| Post-tertiary | -0.770*** | -0.786*** | -0.939*** |
| | (-1.08) | (-1.25) | (-1.74) |
| **Firm size** | | | |
| One worker | 2.029*** | 2.571*** | 3.492*** |
| | (1.47) | (1.99) | (3.83) |
| 2 - 10 employees | 1.573*** | 1.966*** | 3.088*** |
| | (1.25) | (1.68) | (3.52) |
| 11 - 50 employees | 0.674*** | 1.075*** | 1.172*** |
| | (8.08) | (1.21) | (1.68) |
| **Observations** | 12,381 | 13,437 | 10,114 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size.

## Table N° A4: Quantile Regressions (Bogotá)

| *y = log hourly wage* | *Formal* | | | *Informal* | | |
|---|---|---|---|---|---|---|
| | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.213*** | 8.353*** | 8.611*** | 7.719*** | 8.421*** | 9.248*** |
| | (0.0147) | (0.0127) | (0.0335) | (0.0648) | (0.0360) | (0.0944) |
| λ | -365.2** | -73.68 | 426.5 | -360.8 | -184.1 | 89.68 |
| | (160.8) | (103.0) | (532.1) | (345.2) | (167.8) | (307.8) |
| Male | 0.00831 | 0.0764*** | 0.189*** | 0.222*** | 0.0694*** | 0.0915*** |
| | (0.00829) | (0.0103) | (0.0243) | (0.0394) | (0.0186) | (0.0290) |
| Job Tenure | 0.000529*** | 0.00139*** | 0.00237*** | 0.00336*** | 0.00190*** | 0.00185*** |
| | (0.000205) | (0.000218) | (0.000634) | (0.000559) | (0.000284) | (0.000371) |
| Job Tenure Sq. | -2.18e-07 | -9.12e-07 | -1.46e-06 | -6.31e-06*** | -3.05e-06*** | -1.08e-06 |
| | (7.89e-07) | (7.33e-07) | (2.51e-06) | (1.58e-06) | (7.48e-07) | (7.46e-07) |
| **Educational levels** | | | | | | |
| No education | -0.139 | -0.124 | -0.249** | -0.306* | -0.250*** | -0.160** |
| | (0.117) | (0.109) | (0.122) | (0.159) | (0.0858) | (0.0732) |
| Secondary | 0.0939*** | 0.0899*** | 0.294*** | 0.221*** | 0.116*** | 0.181*** |
| | (0.0134) | (0.0107) | (0.0296) | (0.0438) | (0.0202) | (0.0290) |
| Tertiary | 0.205*** | 0.571*** | 1.153*** | 0.380*** | 0.416*** | 1.034*** |
| | (0.0154) | (0.0186) | (0.0363) | (0.0661) | (0.0307) | (0.0754) |
| Post Tertiary | 0.913*** | 1.468*** | 1.714*** | 1.148*** | 1.761*** | 1.928*** |
| | (0.0433) | (0.0276) | (0.0551) | (0.123) | (0.0713) | (0.109) |
| **Firm size** | | | | | | |
| One worker | -0.493*** | -0.0133 | 0.192*** | -0.896*** | -0.377*** | -0.517*** |
| | (0.0525) | (0.0300) | (0.0522) | (0.0602) | (0.0368) | (0.0918) |
| 2 - 10 employees | -0.144*** | -0.0781*** | -0.0141 | -0.544*** | -0.279*** | -0.579*** |
| | (0.0164) | (0.0146) | (0.0362) | (0.0549) | (0.0342) | (0.0937) |
| 11 - 50 employees | -0.0644*** | -0.0756*** | -0.0695** | -0.274*** | -0.173*** | -0.419*** |
| | (0.00986) | (0.0122) | (0.0341) | (0.0655) | (0.0410) | (0.103) |
| Observations | 8,421 | 8,421 | 8,421 | 6,356 | 6,356 | 6,356 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A5: Quantile Regressions (Medellín)

| *y = log hourly wage* | *Formal* | | | *Informal* | | |
|---|---|---|---|---|---|---|
| | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.248*** | 8.334*** | 8.586*** | 7.419*** | 8.205*** | 8.985*** |
| | (0.0101) | (0.0106) | (0.0277) | (0.0675) | (0.0341) | (0.110) |
| λ | -15.66 | -8.897* | 4.974 | 3.290*** | 0.345 | -1.442*** |
| | (39.49) | -4.967 | -6.596 | (0.404) | (0.542) | (0.359) |
| Male | 0.00926 | 0.0637*** | 0.174*** | 0.324*** | 0.143*** | 0.0942*** |
| | (0.00630) | (0.00804) | (0.0200) | (0.0388) | (0.0187) | (0.0318) |
| Job Tenure | 0.000340*** | 0.00157*** | 0.00270*** | 0.00472*** | 0.00254*** | 0.00283*** |
| | (0.000104) | (0.000144) | (0.000278) | (0.000653) | (0.000262) | (0.000426) |
| Job Tenure Sq. | 1.69e-07 | -1.06e-06** | -2.42e-06*** | -9.20e-06*** | -4.85e-06*** | -4.51e-06*** |
| | (3.32e-07) | (4.13e-07) | (7.27e-07) | (2.08e-06) | (6.58e-07) | (9.33e-07) |
| **Educational levels** | | | | | | |
| No education | -0.218** | -0.0210 | -0.0787 | -0.509*** | -0.271*** | -0.292*** |
| | (0.101) | (0.0491) | (0.0760) | (0.168) | (0.0742) | (0.0814) |
| Secondary | 0.0690*** | 0.0637*** | 0.175*** | 0.412*** | 0.188*** | 0.243*** |
| | (0.00865) | (0.00835) | (0.0240) | (0.0494) | (0.0210) | (0.0338) |
| Tertiary | 0.117*** | 0.468*** | 1.026*** | 0.657*** | 0.465*** | 0.921*** |
| | (0.0107) | (0.0139) | (0.0284) | (0.0529) | (0.0308) | (0.0535) |
| Post Tertiary | 0.930*** | 1.432*** | 1.663*** | 1.445*** | 1.699*** | 1.695*** |
| | (0.0419) | (0.0242) | (0.0474) | (0.0940) | (0.0649) | (0.126) |
| **Firm size** | | | | | | |
| One worker | -0.545*** | -0.0213 | 0.164*** | -1.091*** | -0.359*** | -0.310*** |
| | (0.0477) | (0.0263) | (0.0344) | (0.0552) | (0.0342) | (0.106) |
| 2 - 10 employees | -0.110*** | -0.0520*** | 0.0442 | -0.617*** | -0.166*** | -0.259** |
| | (0.0110) | (0.0105) | (0.0309) | (0.0547) | (0.0308) | (0.110) |
| 11 - 50 employees | -0.0176*** | -0.0373*** | -0.0163 | -0.257*** | -0.0360 | -0.132 |
| | (0.00626) | (0.00944) | (0.0261) | (0.0694) | (0.0384) | (0.132) |
| Observations | 11,021 | 11,021 | 11,021 | 7,318 | 7,318 | 7,318 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A6: Quantile Regressions (Barranquilla)

| *y = log hourly wage* | *Formal* | | | *Informal* | | |
|---|---|---|---|---|---|---|
| | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 7.971*** | 8.304*** | 8.607*** | 7.104*** | 8.081*** | 8.526*** |
| | (0.0396) | (0.0192) | (0.0426) | (0.0743) | (0.0293) | (0.0433) |
| λ | -10.81 | -286.5*** | 237.3 | 30.84 | -91.59 | -82.36 |
| | (77.41) | (108.9) | (551.2) | (101.8) | (78.82) | (100.0) |
| Male | -0.00844 | 0.0275** | 0.127*** | 0.455*** | 0.133*** | -0.000940 |
| | (0.0120) | (0.0114) | (0.0262) | (0.0364) | (0.0146) | (0.0237) |
| Job Tenure | 6.02e-05 | 0.00141*** | 0.00246*** | 0.00326*** | 0.00192*** | 0.00255*** |
| | (0.000219) | (0.000210) | (0.000340) | (0.000409) | (0.000162) | (0.000274) |
| Job Tenure Sq. | 1.66e-06** | -7.60e-07 | -2.37e-06*** | -5.46e-06*** | -3.07e-06*** | -3.95e-06*** |
| | (6.90e-07) | (6.79e-07) | (8.02e-07) | (1.08e-06) | (3.62e-07) | (5.78e-07) |
| Educational level | | | | | | |
| No education | 0.0677 | -0.0699 | -0.283*** | -0.403*** | -0.291*** | -0.122** |
| | (0.160) | (0.143) | (0.0840) | (0.0870) | (0.0522) | (0.0585) |
| Secondary | 0.205*** | 0.0595*** | 0.104*** | 0.227*** | 0.152*** | 0.200*** |
| | (0.0377) | (0.0164) | (0.0384) | (0.0356) | (0.0153) | (0.0241) |
| Terciary | 0.394*** | 0.432*** | 0.949*** | 0.522*** | 0.423*** | 0.823*** |
| | (0.0384) | (0.0210) | (0.0433) | (0.0468) | (0.0242) | (0.0426) |
| Post Tertiary | 0.908*** | 1.408*** | 1.709*** | 1.091*** | 1.060*** | 1.295*** |
| | (0.0530) | (0.0332) | (0.0619) | (0.234) | (0.114) | (0.0951) |
| **Firm size** | | | | | | |
| One worker | -0.533*** | 0.0126 | 0.309*** | -0.671*** | -0.276*** | 0.0471 |
| | (0.0846) | (0.0446) | (0.0670) | (0.0677) | (0.0274) | (0.0419) |
| 2 - 10 employees | -0.116*** | -0.0275 | 0.0637 | -0.230*** | -0.149*** | 0.119*** |
| | (0.0265) | (0.0200) | (0.0655) | (0.0665) | (0.0276) | (0.0455) |
| 11 - 50 employees | -0.0242* | -0.0266* | -0.0828*** | -0.0504 | -0.0118 | 0.140** |
| | (0.0142) | (0.0145) | (0.0313) | (0.0823) | (0.0331) | (0.0667) |
| Observations | 6,300 | 6,300 | 6,300 | 10,783 | 10,783 | 10,783 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A7: Quantile Regressions (Cúcuta)

| *y = log hourly wage* | *Formal* | | | *Informal* | | |
|---|---|---|---|---|---|---|
| | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.126*** | 8.298*** | 8.635*** | 7.687*** | 8.216*** | 8.528*** |
| | (0.0288) | (0.0259) | (0.0551) | (0.0861) | (0.0487) | (0.0728) |
| λ | -1,124 | 570.6 | 403.6 | 1,810* | 637.5 | 2,795 |
| | -1,632 | (844.3) | -1,043 | (962.8) | (612.9) | -1,786 |
| Male | -0.00652 | 0.0555*** | 0.122*** | 0.332*** | 0.131*** | 0.0620*** |
| | (0.0150) | (0.0176) | (0.0344) | (0.0375) | (0.0155) | (0.0233) |
| Job Tenure | 0.000526** | 0.000944*** | 0.00238*** | 0.00236*** | 0.00182*** | 0.00250*** |
| | (0.000258) | (0.000300) | (0.000486) | (0.000483) | (0.000256) | (0.000281) |
| Job Tenure Sq. | 1.00e-07 | 6.30e-07 | -2.81e-06** | -4.96e-06*** | -3.43e-06*** | -3.78e-06*** |
| | (8.54e-07) | (9.29e-07) | (1.31e-06) | (1.44e-06) | (8.13e-07) | (5.60e-07) |
| **Educational levels** | | | | | | |
| No education | -0.157 | 0.0327 | 0.239 | -0.384** | -0.188*** | -0.0918 |
| | (0.193) | (0.161) | (0.295) | (0.154) | (0.0536) | (0.0680) |
| Secondary | 0.0531** | 0.0674*** | 0.179*** | 0.229*** | 0.116*** | 0.181*** |
| | (0.0237) | (0.0204) | (0.0436) | (0.0381) | (0.0157) | (0.0239) |
| Tertiary | 0.224*** | 0.456*** | 0.902*** | 0.293*** | 0.279*** | 0.681*** |
| | (0.0264) | (0.0325) | (0.0540) | (0.0667) | (0.0314) | (0.0618) |
| Post Tertiary | 0.861*** | 1.352*** | 1.326*** | 0.366*** | 1.260*** | 1.600*** |
| | (0.0746) | (0.0492) | (0.0573) | (0.142) | (0.223) | (0.215) |
| **Firm size** | | | | | | |
| One worker | -0.747*** | -0.239*** | -0.126** | -0.968*** | -0.528*** | -0.306*** |
| | (0.0789) | (0.0532) | (0.0628) | (0.0729) | (0.0479) | (0.0748) |
| 2 - 10 employees | -0.136*** | -0.132*** | -0.143* | -0.657*** | -0.272*** | -0.0923 |
| | (0.0237) | (0.0261) | (0.0745) | (0.0825) | (0.0500) | (0.0804) |
| 11 - 50 employees | -0.0168 | -0.0777*** | -0.263*** | -0.289*** | -0.107* | -0.133 |
| | (0.0165) | (0.0200) | (0.0468) | (0.103) | (0.0614) | (0.0900) |
| Observations | 2,659 | 2,659 | 2,659 | 6,389 | 6,389 | 6,389 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A8: Quantile Regressions (Quibdó)

| *y = log hourly wage* | *Formal* | | | *Informal* | | |
|---|---|---|---|---|---|---|
| | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.127*** | 8.374*** | 8.721*** | 7.180*** | 8.107*** | 8.790*** |
| | (0.0439) | (0.0402) | (0.0588) | (0.157) | (0.0682) | (0.174) |
| λ | 284.0 | -671.0 | 1,602 | 4,005*** | 858.3 | 88.74 |
| | (824.5) | -1,807 | -1,415 | (761.8) | (623.6) | (680.5) |
| Male | 0.0243 | 0.0443* | 0.0772** | 0.358*** | 0.254*** | 0.298*** |
| | (0.0241) | (0.0238) | (0.0353) | (0.0560) | (0.0249) | (0.0357) |
| Job Tenure | 0.000868*** | 0.00158*** | 0.00115*** | 0.000144 | 0.000686** | 0.000781 |
| | (0.000331) | (0.000300) | (0.000399) | (0.000639) | (0.000338) | (0.000503) |
| Job Tenure Sq. | -3.37e-07 | -4.83e-07 | -4.20e-08 | -1.46e-06 | -2.23e-06** | -9.77e-07 |
| | (8.98e-07) | (6.90e-07) | (9.30e-07) | (1.64e-06) | (9.38e-07) | (1.29e-06) |
| Educational level | | | | | | |
| No education | -0.0580 | -0.0780 | -0.380*** | -0.458*** | -0.295*** | -0.162* |
| | (0.0929) | (0.120) | (0.0892) | (0.148) | (0.0748) | (0.0856) |
| Secondary | 0.0903** | 0.153*** | 0.328*** | 0.296*** | 0.144*** | 0.193*** |
| | (0.0396) | (0.0374) | (0.0657) | (0.0611) | (0.0271) | (0.0377) |
| Terciary | 0.330*** | 0.742*** | 1.099*** | 0.460*** | 0.362*** | 0.599*** |
| | (0.0406) | (0.0396) | (0.0598) | (0.0793) | (0.0362) | (0.0536) |
| Post Tertiary | 1.140*** | 1.185*** | 1.256*** | 1.042*** | 1.317*** | 1.912*** |
| | (0.0364) | (0.0364) | (0.0568) | (0.205) | (0.212) | (0.555) |
| **Firm size** | | | | | | |
| One worker | -1.131*** | -0.371*** | -0.180* | -0.713*** | -0.475*** | -0.484*** |
| | (0.268) | (0.0777) | (0.0966) | (0.158) | (0.0674) | (0.172) |
| 2 - 10 employees | -0.247*** | -0.214*** | 0.0623 | -0.322** | -0.217*** | -0.417** |
| | (0.0403) | (0.0384) | (0.0856) | (0.159) | (0.0675) | (0.172) |
| 11 - 50 employees | -0.160*** | -0.189*** | -0.0805 | 0.188 | -0.00449 | -0.272 |
| | (0.0282) | (0.0342) | (0.0520) | (0.172) | (0.0744) | (0.179) |
| Observations | 2,915 | 2,915 | 2,915 | 3,709 | 3,709 | 3,709 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A9: Quantile Regression (Cali)

| | Cali | | | | | |
|---|---|---|---|---|---|---|
| | *Formal* | | | *Informal* | | |
| *y = log hourly wage* | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.076*** | 8.383*** | 8.737*** | 7.443*** | 8.341*** | 9.018*** |
| | (0.0287) | (0.0169) | (0.0470) | (0.0751) | (0.0344) | (0.0822) |
| λ | -762.0 | -1,873*** | -1,63 | 345.5 | -1,709** | -120.7 |
| | (781.0) | (629.9) | -1,505 | -1,422 | (727.8) | -1,765 |
| Male | 0.0158 | 0.0571*** | 0.185*** | 0.294*** | 0.0966*** | 0.0408 |
| | (0.0119) | (0.0116) | (0.0273) | (0.0402) | (0.0188) | (0.0339) |
| Job Tenure | 0.00104*** | 0.00159*** | 0.00247*** | 0.00381*** | 0.00204*** | 0.00235*** |
| | (0.000210) | (0.000197) | (0.000350) | (0.000423) | (0.000229) | (0.000464) |
| Job Tenure Sq. | -1.43e-06** | -2.23e-06*** | -3.15e-06*** | -8.54e-06*** | -3.62e-06*** | -3.36e-06*** |
| | (7.02e-07) | (6.20e-07) | (9.34e-07) | (1.08e-06) | (5.34e-07) | (1.04e-06) |
| Educational level | | | | | | |
| No education | 0.105 | 0.0510 | -0.125* | -0.424 | -0.233*** | 0.0152 |
| | (0.142) | (0.135) | (0.0717) | (0.285) | (0.0816) | (0.139) |
| Secondary | 0.173*** | 0.0974*** | 0.197*** | 0.317*** | 0.174*** | 0.246*** |
| | (0.0258) | (0.0141) | (0.0406) | (0.0438) | (0.0205) | (0.0333) |
| Terciary | 0.341*** | 0.598*** | 0.945*** | 0.689*** | 0.533*** | 1.097*** |
| | (0.0289) | (0.0215) | (0.0467) | (0.0595) | (0.0357) | (0.0895) |
| Post Tertiary | 1.002*** | 1.384*** | 1.582*** | 1.198*** | 1.499*** | 1.777*** |
| | (0.0483) | (0.0327) | (0.0552) | (0.144) | (0.0800) | (0.116) |
| **Firm size** | | | | | | |
| One worker | -0.662*** | -0.107*** | 0.178** | -0.806*** | -0.370*** | -0.230*** |
| | (0.126) | (0.0341) | (0.0711) | (0.0672) | (0.0347) | (0.0782) |
| 2 - 10 employees | -0.141*** | -0.103*** | -0.0720 | -0.543*** | -0.263*** | -0.282*** |
| | (0.0218) | (0.0160) | (0.0457) | (0.0664) | (0.0321) | (0.0796) |
| 11 - 50 employees | -0.0631*** | -0.105*** | -0.145*** | -0.132* | -0.132*** | -0.226** |
| | (0.0140) | (0.0141) | (0.0334) | (0.0684) | (0.0370) | (0.0924) |
| Observations | 6,245 | 6,245 | 6,245 | 6,507 | 6,507 | 6,507 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A10: Quantile Regression (Cartagena)

| | Cartagena | | | | | |
|---|---|---|---|---|---|---|
| | *Formal* | | | *Informal* | | |
| *y = log hourly wage* | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.369*** | 8.304*** | 8.508*** | 7.292*** | 8.455*** | 9.060*** |
| | (0.0108) | (0.0170) | (0.0297) | (0.0810) | (0.0582) | (0.0702) |
| λ | -242.4*** | -261.3** | 42.89 | -27.66 | -558.7* | 189.9 |
| | (83.46) | (121.4) | (258.3) | (257.2) | (316.1) | (656.1) |
| Male | 0.000164 | 0.0619*** | 0.154*** | 0.688*** | 0.211*** | 0.0766*** |
| | (0.00703) | (0.0114) | (0.0230) | (0.0397) | (0.0142) | (0.0178) |
| Job Tenure | 3.69e-07 | 0.00176*** | 0.00283*** | 0.00322*** | 0.00141*** | 0.00140*** |
| | (0.000136) | (0.000208) | (0.000278) | (0.000279) | (0.000139) | (0.000197) |
| Job Tenure Sq. | 6.83e-09 | -1.96e-06*** | -4.33e-06*** | -5.27e-06*** | -2.18e-06*** | -2.05e-06*** |
| | (4.74e-07) | (6.48e-07) | (6.01e-07) | (6.13e-07) | (2.92e-07) | (4.27e-07) |
| Educational level | | | | | | |
| No education | 0.00178 | -0.00185 | -0.304*** | -0.463** | -0.236*** | -0.0969 |
| | (0.0522) | (0.0761) | (0.0580) | (0.182) | (0.0484) | (0.0701) |
| Secondary | 0.000182 | 0.0535*** | 0.138*** | 0.195*** | 0.0636*** | 0.0873*** |
| | (0.00798) | (0.0141) | (0.0252) | (0.0310) | (0.0137) | (0.0178) |
| Terciary | 0.000416 | 0.385*** | 0.787*** | 0.387*** | 0.281*** | 0.588*** |
| | (0.0107) | (0.0170) | (0.0315) | (0.0395) | (0.0206) | (0.0445) |
| Post Tertiary | 0.977*** | 1.326*** | 1.607*** | 1.296*** | 1.254*** | 1.559*** |
| | (0.0430) | (0.0367) | (0.0991) | (0.130) | (0.141) | (0.245) |
| **Firm size** | | | | | | |
| One worker | -0.200*** | 0.00450 | 0.106 | -0.585*** | -0.545*** | -0.633*** |
| | (0.0579) | (0.0385) | (0.0906) | (0.0695) | (0.0573) | (0.0652) |
| 2 - 10 employees | -0.0341*** | -0.0892*** | -0.0834** | -0.366*** | -0.427*** | -0.470*** |
| | (0.0107) | (0.0201) | (0.0419) | (0.0704) | (0.0580) | (0.0693) |
| 11 - 50 employees | 2.79e-05 | -0.0509*** | -0.156*** | -0.0781 | -0.290*** | -0.437*** |
| | (0.00696) | (0.0120) | (0.0250) | (0.0769) | (0.0605) | (0.0748) |
| Observations | 4,499 | 4,499 | 4,499 | 6,602 | 6,602 | 6,602 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A11: Quantile Regression (Tunja)

| | Tunja | | | | | |
|---|---|---|---|---|---|---|
| | *Formal* | | | *Informal* | | |
| *y = log hourly wage* | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.147*** | 8.363*** | 8.752*** | 7.500*** | 8.216*** | 8.822*** |
| | (0.0332) | (0.0257) | (0.0477) | (0.151) | (0.0755) | (0.184) |
| λ | 137.1 | -216.8 | -815.9* | -347.5 | -531.9 | 11.73 |
| | (276.4) | (303.3) | (432.9) | (604.9) | (328.3) | (491.5) |
| Male | -0.00714 | 0.0866*** | 0.227*** | 0.403*** | 0.169*** | 0.0903** |
| | (0.0163) | (0.0169) | (0.0299) | (0.0474) | (0.0227) | (0.0363) |
| Job Tenure | 0.000232 | 0.00166*** | 0.00150*** | 0.00260*** | 0.00129*** | 0.00190*** |
| | (0.000254) | (0.000229) | (0.000423) | (0.000561) | (0.000276) | (0.000487) |
| Job Tenure Sq. | 1.37e-06* | -9.49e-07* | -9.68e-07 | -5.07e-06*** | -1.69e-06*** | -2.75e-06*** |
| | (7.97e-07) | (5.72e-07) | (1.16e-06) | (1.39e-06) | (6.13e-07) | (1.00e-06) |
| Educational level | | | | | | |
| No education | -0.117 | -0.255 | -0.596*** | -0.325* | -0.236** | -0.232*** |
| | (0.106) | (0.161) | (0.0900) | (0.196) | (0.100) | (0.0723) |
| Secondary | 0.101*** | 0.0383* | 0.224*** | 0.135*** | 0.117*** | 0.248*** |
| | (0.0300) | (0.0232) | (0.0450) | (0.0501) | (0.0243) | (0.0398) |
| Terciary | 0.220*** | 0.387*** | 0.835*** | 0.255*** | 0.311*** | 0.567*** |
| | (0.0311) | (0.0271) | (0.0467) | (0.0633) | (0.0331) | (0.0639) |
| Post Tertiary | 0.766*** | 1.197*** | 1.228*** | 0.109 | 0.455*** | 0.682*** |
| | (0.0409) | (0.0264) | (0.0453) | (0.250) | (0.121) | (0.106) |
| **Firm size** | | | | | | |
| One worker | -0.744*** | -0.272*** | -0.139** | -0.688*** | -0.367*** | -0.267 |
| | (0.109) | (0.0489) | (0.0706) | (0.144) | (0.0748) | (0.184) |
| 2 - 10 employees | -0.205*** | -0.148*** | -0.193*** | -0.261* | -0.133* | -0.236 |
| | (0.0291) | (0.0230) | (0.0508) | (0.142) | (0.0729) | (0.185) |
| 11 - 50 employees | -0.0518*** | -0.123*** | -0.245*** | -0.0990 | -0.0163 | 0.00724 |
| | (0.0173) | (0.0224) | (0.0474) | (0.165) | (0.0856) | (0.198) |
| Observations | 4,398 | 4,398 | 4,398 | 3,479 | 3,479 | 3,479 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A12: Quantile Regression (Manizales)

| | Manizales | | | | | |
|---|---|---|---|---|---|---|
| | *Formal* | | | *Informal* | | |
| *y = log hourly wage* | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.217*** | 8.322*** | 8.522*** | 7.614*** | 8.342*** | 8.867*** |
| | (0.0123) | (0.0119) | (0.0322) | (0.139) | (0.0624) | (0.154) |
| λ | -1,714*** | -777.3 | -874.4** | 94.92 | -731.4* | -502.5 |
| | (642.0) | (520.3) | (355.2) | (396.2) | (373.2) | (360.7) |
| Male | 0.000231 | 0.0337*** | 0.143*** | 0.102** | 0.0516** | 0.0743** |
| | (0.00598) | (0.00867) | (0.0243) | (0.0463) | (0.0211) | (0.0332) |
| Job Tenure | -1.61e-05 | 0.00102*** | 0.00238*** | 0.00235*** | 0.000528* | 0.00170*** |
| | (0.000111) | (0.000167) | (0.000327) | (0.000472) | (0.000296) | (0.000383) |
| Job Tenure Sq. | 6.01e-07 | 2.03e-07 | -1.92e-06** | -4.44e-06*** | -4.94e-07 | -2.16e-06** |
| | (3.92e-07) | (5.30e-07) | (8.34e-07) | (1.01e-06) | (7.11e-07) | (8.51e-07) |
| Educational level | | | | | | |
| No education | -0.0447 | -0.0686 | -0.166* | -0.644*** | -0.458*** | -0.294*** |
| | (0.0614) | (0.0559) | (0.0912) | (0.125) | (0.0924) | (0.0905) |
| Secondary | 0.120*** | 0.0332*** | 0.178*** | 0.288*** | 0.156*** | 0.288*** |
| | (0.0111) | (0.00960) | (0.0294) | (0.0517) | (0.0228) | (0.0396) |
| Terciary | 0.152*** | 0.337*** | 0.922*** | 0.319*** | 0.306*** | 0.595*** |
| | (0.0124) | (0.0172) | (0.0361) | (0.0728) | (0.0334) | (0.0566) |
| Post Tertiary | 0.808*** | 1.382*** | 1.490*** | 0.271** | 0.637*** | 1.072*** |
| | (0.0396) | (0.0273) | (0.0502) | (0.117) | (0.231) | (0.255) |
| **Firm size** | | | | | | |
| One worker | -0.405*** | -0.0560** | 0.118*** | -0.634*** | -0.236*** | -0.184 |
| | (0.0426) | (0.0272) | (0.0356) | (0.133) | (0.0636) | (0.152) |
| 2 - 10 employees | -0.132*** | -0.0325*** | 0.0405 | -0.271** | -0.186*** | -0.292* |
| | (0.0133) | (0.0124) | (0.0427) | (0.127) | (0.0616) | (0.152) |
| 11 - 50 employees | -0.0114 | -0.0223* | -0.00769 | -0.105 | -0.169** | -0.255 |
| | (0.00698) | (0.0115) | (0.0376) | (0.137) | (0.0684) | (0.169) |
| Observations | 7,127 | 7,127 | 7,127 | 3,722 | 3,722 | 3,722 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A13: Quantile Regression (Florencia)

| | Florencia | | | | | |
|---|---|---|---|---|---|---|
| | *Formal* | | | *Informal* | | |
| *y = log hourly wage* | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.019*** | 8.410*** | 8.783*** | 7.310*** | 8.093*** | 8.740*** |
| | (0.0469) | (0.0326) | (0.0516) | (0.137) | (0.0711) | (0.265) |
| λ | 186.2 | -202.7 | -1,236*** | 452.5 | 18.92 | -860.0*** |
| | (397.8) | (345.5) | (367.7) | (491.6) | (255.5) | (235.5) |
| Male | 0.0849*** | 0.0918*** | 0.126*** | 0.338*** | 0.129*** | 0.0285 |
| | (0.0190) | (0.0185) | (0.0324) | (0.0485) | (0.0175) | (0.0280) |
| Job Tenure | 0.000233 | 0.00141*** | 0.00186*** | 0.00274*** | 0.00101*** | 0.000772 |
| | (0.000237) | (0.000248) | (0.000442) | (0.000682) | (0.000226) | (0.000578) |
| Job Tenure Sq. | 2.35e-06*** | -7.35e-07 | -1.30e-06 | -6.51e-06*** | -1.47e-06** | 9.72e-08 |
| | (5.45e-07) | (6.04e-07) | (1.08e-06) | (2.09e-06) | (5.76e-07) | (2.03e-06) |
| Educational level | | | | | | |
| No education | -0.168 | 0.00950 | 0.0969 | -0.373*** | -0.233*** | -0.158** |
| | (0.174) | (0.291) | (0.206) | (0.130) | (0.0697) | (0.0688) |
| Secondary | 0.193*** | 0.0857*** | 0.147*** | 0.244*** | 0.106*** | 0.191*** |
| | (0.0418) | (0.0258) | (0.0407) | (0.0478) | (0.0189) | (0.0311) |
| Terciary | 0.350*** | 0.448*** | 0.825*** | 0.391*** | 0.343*** | 0.616*** |
| | (0.0433) | (0.0302) | (0.0508) | (0.0731) | (0.0269) | (0.0769) |
| Post Tertiary | 0.922*** | 1.128*** | 1.296*** | 0.940*** | 1.574*** | 2.333*** |
| | (0.0357) | (0.0331) | (0.0520) | (0.186) | (0.199) | (0.344) |
| **Firm size** | | | | | | |
| One worker | -0.506*** | -0.274*** | 0.0287 | -0.554*** | -0.281*** | -0.250 |
| | (0.0818) | (0.0677) | (0.0981) | (0.136) | (0.0704) | (0.265) |
| 2 - 10 employees | -0.154*** | -0.198*** | -0.167*** | -0.233* | -0.0755 | -0.227 |
| | (0.0237) | (0.0257) | (0.0560) | (0.135) | (0.0701) | (0.266) |
| 11 - 50 employees | -0.0555*** | -0.180*** | -0.293*** | 0.0126 | -0.0589 | -0.252 |
| | (0.0186) | (0.0226) | (0.0357) | (0.142) | (0.0776) | (0.274) |
| Observations | 3,317 | 3,317 | 3,317 | 4,473 | 4,473 | 4,473 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A14: Quantile Regression (Popayan)

| | Popayán | | | | | |
|---|---|---|---|---|---|---|
| | *Formal* | | | *Informal* | | |
| *y = log hourly wage* | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.054*** | 8.327*** | 8.664*** | 7.392*** | 8.140*** | 8.694*** |
| | (0.0529) | (0.0286) | (0.0534) | (0.0988) | (0.0501) | (0.0814) |
| λ | -887.3** | -1,359*** | -1,469** | -65.35 | 61.65 | -722.7** |
| | (351.0) | (382.4) | (581.3) | -2,873 | (279.4) | (345.8) |
| Male | -0.00498 | 0.0429** | 0.0864*** | 0.154*** | 0.111*** | 0.0231 |
| | (0.0182) | (0.0173) | (0.0331) | (0.0372) | (0.0179) | (0.0273) |
| Job Tenure | 0.000393* | 0.00126*** | 0.00137*** | 0.00150*** | 0.00139*** | 0.00189*** |
| | (0.000204) | (0.000243) | (0.000413) | (0.000551) | (0.000219) | (0.000308) |
| Job Tenure Sq. | 3.47e-07 | -9.66e-07 | -3.25e-07 | -4.32e-06*** | -2.83e-06*** | -3.21e-06*** |
| | (4.90e-07) | (6.32e-07) | (1.07e-06) | (1.46e-06) | (5.12e-07) | (6.22e-07) |
| Educational level | | | | | | |
| No education | 0.362*** | -0.113 | -0.345** | -0.387*** | -0.264*** | -0.133 |
| | (0.129) | (0.158) | (0.140) | (0.109) | (0.0752) | (0.115) |
| Secondary | 0.162*** | 0.116*** | 0.262*** | 0.233*** | 0.174*** | 0.240*** |
| | (0.0505) | (0.0231) | (0.0483) | (0.0431) | (0.0198) | (0.0297) |
| Terciary | 0.359*** | 0.512*** | 0.994*** | 0.490*** | 0.406*** | 0.716*** |
| | (0.0503) | (0.0264) | (0.0481) | (0.0518) | (0.0256) | (0.0536) |
| Post Tertiary | 0.955*** | 1.314*** | 1.411*** | 0.760*** | 1.163*** | 1.322*** |
| | (0.0473) | (0.0293) | (0.0583) | (0.202) | (0.112) | (0.114) |
| **Firm size** | | | | | | |
| One worker | -0.639*** | -0.211*** | -0.00484 | -0.545*** | -0.319*** | -0.172** |
| | (0.0694) | (0.0493) | (0.0757) | (0.0952) | (0.0498) | (0.0837) |
| 2 - 10 employees | -0.182*** | -0.149*** | -0.197*** | -0.265*** | -0.228*** | -0.235*** |
| | (0.0291) | (0.0228) | (0.0519) | (0.0903) | (0.0487) | (0.0827) |
| 11 - 50 employees | -0.0848*** | -0.114*** | -0.111** | -0.0989 | -0.0765 | -0.0566 |
| | (0.0190) | (0.0223) | (0.0455) | (0.107) | (0.0567) | (0.0988) |
| Observations | 4,027 | 4,027 | 4,027 | 5,69 | 5,69 | 5,69 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A15: Quantile Regression (Valledupar)

| | Valledupar | | | | | |
|---|---|---|---|---|---|---|
| | *Formal* | | | *Informal* | | |
| *y = log hourly wage* | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.172*** | 8.278*** | 8.598*** | 7.293*** | 8.061*** | 8.523*** |
| | (0.0322) | (0.0315) | (0.0606) | (0.0969) | (0.0399) | (0.0747) |
| λ | -692.0 | 95.24 | -1,334 | -61.96 | 178.2 | 671.9 |
| | (810.5) | (921.9) | (867.9) | (664.4) | (410.5) | (716.2) |
| Male | 3.32e-05 | 0.0831*** | 0.238*** | 0.430*** | 0.193*** | 0.0939*** |
| | (0.0130) | (0.0181) | (0.0336) | (0.0377) | (0.0148) | (0.0246) |
| Job Tenure | 0.000375 | 0.00176*** | 0.00295*** | 0.00314*** | 0.00217*** | 0.00358*** |
| | (0.000315) | (0.000285) | (0.000455) | (0.000354) | (0.000174) | (0.000321) |
| Job Tenure Sq. | 2.19e-08 | -1.07e-06 | -2.97e-06*** | -5.30e-06*** | -3.84e-06*** | -5.62e-06*** |
| | (1.14e-06) | (7.61e-07) | (1.10e-06) | (8.49e-07) | (4.25e-07) | (7.21e-07) |
| Educational level | | | | | | |
| No education | 0.235** | 0.204* | -0.128 | -0.241*** | -0.205*** | -0.182*** |
| | (0.0930) | (0.107) | (0.100) | (0.0821) | (0.0452) | (0.0562) |
| Secondary | 0.125*** | 0.0876*** | 0.107* | 0.261*** | 0.113*** | 0.146*** |
| | (0.0291) | (0.0283) | (0.0592) | (0.0379) | (0.0146) | (0.0256) |
| Terciary | 0.189*** | 0.453*** | 0.827*** | 0.373*** | 0.330*** | 0.687*** |
| | (0.0308) | (0.0334) | (0.0622) | (0.0494) | (0.0249) | (0.0491) |
| Post Tertiary | 0.769*** | 1.095*** | 1.252*** | 0.407 | 1.138*** | 1.593*** |
| | (0.0498) | (0.0445) | (0.0575) | (0.481) | (0.192) | (0.146) |
| **Firm size** | | | | | | |
| One worker | -0.679*** | -0.135* | 0.0342 | -0.605*** | -0.333*** | -0.273*** |
| | (0.0980) | (0.0780) | (0.0774) | (0.0905) | (0.0384) | (0.0728) |
| 2 - 10 employees | -0.186*** | -0.109*** | 0.0776 | -0.466*** | -0.192*** | -0.155** |
| | (0.0343) | (0.0309) | (0.0901) | (0.0904) | (0.0387) | (0.0741) |
| 11 - 50 employees | -0.0421*** | -0.0856*** | -0.121*** | -0.116 | -0.0722* | -0.119 |
| | (0.0156) | (0.0238) | (0.0439) | (0.0968) | (0.0425) | (0.0847) |
| Observations | 3,184 | 3,184 | 3,184 | 6,646 | 6,646 | 6,646 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A16: Quantile Regression (Montería)

| | Montería | | | | | |
|---|---|---|---|---|---|---|
| | *Formal* | | | *Informal* | | |
| *y = log hourly wage* | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.159*** | 8.314*** | 8.462*** | 7.131*** | 8.044*** | 8.406*** |
| | (0.0286) | (0.0237) | (0.0402) | (0.104) | (0.0511) | (0.0591) |
| λ | -1,248 | -257.7 | -1,358*** | -24.18 | 92.93 | -192.8 |
| | -1,181 | (345.8) | (513.9) | (562.1) | (327.0) | (528.6) |
| Male | 0.00217 | 0.0223 | 0.133*** | 0.544*** | 0.127*** | 0.0154 |
| | (0.00966) | (0.0136) | (0.0269) | (0.0425) | (0.0170) | (0.0273) |
| Job Tenure | -0.000143 | 0.000804*** | 0.00299*** | 0.00264*** | 0.00125*** | 0.00235*** |
| | (0.000222) | (0.000250) | (0.000340) | (0.000388) | (0.000203) | (0.000356) |
| Job Tenure Sq. | 1.70e-06* | 1.60e-06** | -2.91e-06*** | -4.59e-06*** | -1.89e-06*** | -3.26e-06*** |
| | (9.05e-07) | (7.57e-07) | (7.32e-07) | (1.03e-06) | (5.03e-07) | (8.72e-07) |
| Educational level | | | | | | |
| No education | 0.0288 | -0.0863 | -0.239** | -0.362*** | -0.271*** | -0.159** |
| | (0.0956) | (0.0932) | (0.0937) | (0.106) | (0.0474) | (0.0632) |
| Secondary | 0.160*** | 0.0501*** | 0.208*** | 0.275*** | 0.120*** | 0.189*** |
| | (0.0271) | (0.0192) | (0.0372) | (0.0359) | (0.0172) | (0.0265) |
| Terciary | 0.211*** | 0.435*** | 1.050*** | 0.562*** | 0.408*** | 0.864*** |
| | (0.0289) | (0.0293) | (0.0419) | (0.0491) | (0.0293) | (0.0577) |
| Post Tertiary | 0.919*** | 1.386*** | 1.479*** | 0.958*** | 1.315*** | 1.593*** |
| | (0.0437) | (0.0322) | (0.0585) | (0.301) | (0.183) | (0.0979) |
| **Firm size** | | | | | | |
| One worker | -0.778*** | -0.130** | 0.0378 | -0.740*** | -0.309*** | -0.0887 |
| | (0.108) | (0.0506) | (0.0826) | (0.0924) | (0.0488) | (0.0587) |
| 2 - 10 employees | -0.152*** | -0.0573*** | -0.111** | -0.319*** | -0.0960** | 0.109* |
| | (0.0229) | (0.0211) | (0.0473) | (0.0899) | (0.0490) | (0.0607) |
| 11 - 50 employees | -0.0100 | -0.0375** | -0.0863** | -0.0196 | 0.0291 | 0.225** |
| | (0.0103) | (0.0159) | (0.0431) | (0.0977) | (0.0552) | (0.0990) |
| Observations | 3,864 | 3,864 | 3,864 | 6,622 | 6,622 | 6,622 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A17: Quantile Regression (Neiva)

| | Neiva | | | | | |
|---|---|---|---|---|---|---|
| | *Formal* | | | *Informal* | | |
| *y = log hourly wage* | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.023*** | 8.356*** | 8.613*** | 7.558*** | 8.187*** | 8.533*** |
| | (0.0448) | (0.0229) | (0.0367) | (0.0787) | (0.0421) | (0.0734) |
| λ | 1,338** | -424.9 | 1,823** | 37.33 | 303.2 | -774.7 |
| | (658.4) | (757.6) | (712.2) | (934.5) | (723.6) | (802.4) |
| Male | -0.00420 | 0.0782*** | 0.186*** | 0.409*** | 0.157*** | 0.0424 |
| | (0.0153) | (0.0156) | (0.0279) | (0.0424) | (0.0179) | (0.0276) |
| Job Tenure | 0.000135 | 0.00139*** | 0.00252*** | 0.00173*** | 0.00133*** | 0.00176*** |
| | (0.000229) | (0.000256) | (0.000380) | (0.000492) | (0.000206) | (0.000309) |
| Job Tenure Sq. | 2.16e-06*** | -5.94e-07 | -2.80e-06*** | -3.66e-06*** | -1.95e-06*** | -2.56e-06*** |
| | (6.44e-07) | (7.57e-07) | (8.66e-07) | (1.21e-06) | (4.65e-07) | (6.54e-07) |
| Educational level | | | | | | |
| No education | -0.187 | 0.0697 | -0.0752 | -0.493*** | -0.378*** | -0.269*** |
| | (0.260) | (0.0991) | (0.0809) | (0.166) | (0.0796) | (0.0935) |
| Secondary | 0.221*** | 0.0915*** | 0.259*** | 0.186*** | 0.150*** | 0.209*** |
| | (0.0423) | (0.0197) | (0.0358) | (0.0463) | (0.0193) | (0.0297) |
| Terciary | 0.346*** | 0.391*** | 0.837*** | 0.396*** | 0.296*** | 0.589*** |
| | (0.0437) | (0.0232) | (0.0398) | (0.0563) | (0.0289) | (0.0478) |
| Post Tertiary | 0.812*** | 1.204*** | 1.370*** | 0.429** | 0.933*** | 1.219*** |
| | (0.0484) | (0.0371) | (0.0551) | (0.194) | (0.149) | (0.185) |
| **Firm size** | | | | | | |
| One worker | -0.686*** | -0.147*** | -0.0586 | -0.843*** | -0.349*** | -0.0648 |
| | (0.0934) | (0.0431) | (0.0678) | (0.0634) | (0.0410) | (0.0721) |
| 2 - 10 employees | -0.170*** | -0.129*** | -0.174*** | -0.334*** | -0.139*** | 0.156* |
| | (0.0267) | (0.0217) | (0.0442) | (0.0631) | (0.0419) | (0.0803) |
| 11 - 50 employees | -0.0361** | -0.118*** | -0.185*** | -0.178* | -0.0394 | 0.00774 |
| | (0.0160) | (0.0189) | (0.0424) | (0.0915) | (0.0549) | (0.0881) |
| Observations | 4,358 | 4,358 | 4,358 | 5350 | 5350 | 5350 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A18: Quantile Regression (Riohacha)

| | Riohacha | | | | | |
|---|---|---|---|---|---|---|
| | *Formal* | | | *Informal* | | |
| *y = log hourly wage* | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.066*** | 8.276*** | 8.773*** | 6.644*** | 8.030*** | 8.689*** |
| | (0.0491) | (0.0432) | (0.0617) | (0.147) | (0.0517) | (0.0881) |
| λ | -716.2 | -278.9 | -1,073 | 2,747*** | 1,697** | 2,995** |
| | -2,254 | (909.1) | (953.8) | -1,062 | (832.2) | -1,337 |
| Male | -0.0439* | 0.0928*** | 0.0992*** | 0.615*** | 0.135*** | -0.0251 |
| | (0.0250) | (0.0247) | (0.0325) | (0.0559) | (0.0224) | (0.0319) |
| Job Tenure | 0.00120*** | 0.00145*** | 0.00111* | 0.00314*** | 0.00151*** | 0.00223*** |
| | (0.000364) | (0.000335) | (0.000572) | (0.000542) | (0.000251) | (0.000312) |
| Job Tenure Sq. | -3.01e-07 | 4.70e-07 | 1.96e-06 | -5.60e-06*** | -2.33e-06*** | -3.33e-06*** |
| | (1.17e-06) | (9.30e-07) | (1.70e-06) | (1.45e-06) | (6.04e-07) | (5.93e-07) |
| Educational level | | | | | | |
| No education | -0.174 | -0.157 | -0.397*** | -0.532*** | -0.314*** | -0.202*** |
| | (0.124) | (0.159) | (0.112) | (0.116) | (0.0598) | (0.0620) |
| Secondary | 0.156*** | 0.206*** | 0.287*** | 0.277*** | 0.159*** | 0.181*** |
| | (0.0465) | (0.0401) | (0.0675) | (0.0554) | (0.0229) | (0.0316) |
| Terciary | 0.299*** | 0.691*** | 0.870*** | 0.637*** | 0.434*** | 0.687*** |
| | (0.0465) | (0.0403) | (0.0603) | (0.0595) | (0.0279) | (0.0479) |
| Post Tertiary | 0.946*** | 1.157*** | 1.276*** | 0.910*** | 1.027*** | 1.484*** |
| | (0.0474) | (0.0367) | (0.0652) | (0.181) | (0.162) | (0.246) |
| **Firm size** | | | | | | |
| One worker | -1.146*** | -0.552*** | 0.172 | -0.698*** | -0.502*** | -0.381*** |
| | (0.146) | (0.0812) | (0.146) | (0.134) | (0.0487) | (0.0879) |
| 2 - 10 employees | -0.241*** | -0.274*** | 0.0133 | -0.476*** | -0.333*** | -0.277*** |
| | (0.0541) | (0.0429) | (0.101) | (0.135) | (0.0495) | (0.0883) |
| 11 - 50 employees | -0.0489 | -0.182*** | -0.216*** | 0.0436 | -0.0897 | -0.121 |
| | (0.0319) | (0.0334) | (0.0535) | (0.172) | (0.0631) | (0.102) |
| Observations | 3,247 | 3,247 | 3,247 | 6,407 | 6,407 | 6,407 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A19: Quantile Regression (Santa Marta)

| | Santa Marta | | | | | |
|---|---|---|---|---|---|---|
| | *Formal* | | | *Informal* | | |
| *y = log hourly wage* | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.122*** | 8.315*** | 8.556*** | 7.268*** | 8.050*** | 8.570*** |
| | (0.0329) | (0.0231) | (0.0400) | (0.0778) | (0.0463) | (0.0789) |
| λ | -732.2 | -202.1 | -2,371*** | 2,504*** | 640.4 | -74.80 |
| | (777.5) | (723.4) | (542.7) | (573.2) | (411.6) | -1,006 |
| Male | -0.0109 | 0.0222 | 0.0871*** | 0.457*** | 0.174*** | 0.0407 |
| | (0.0129) | (0.0142) | (0.0301) | (0.0394) | (0.0170) | (0.0248) |
| Job Tenure | 0.000378* | 0.00124*** | 0.00297*** | 0.00274*** | 0.00194*** | 0.00207*** |
| | (0.000223) | (0.000204) | (0.000441) | (0.000338) | (0.000192) | (0.000285) |
| Job Tenure Sq. | 6.42e-07 | -2.71e-07 | -2.80e-06*** | -4.89e-06*** | -3.20e-06*** | -3.15e-06*** |
| | (6.93e-07) | (5.18e-07) | (1.01e-06) | (8.05e-07) | (4.56e-07) | (6.55e-07) |
| Educational level | | | | | | |
| No education | -0.560* | -0.130 | 0.687** | -0.519*** | -0.254*** | -0.132** |
| | (0.329) | (0.0983) | (0.291) | (0.158) | (0.0622) | (0.0622) |
| Secondary | 0.0848*** | 0.0523*** | 0.240*** | 0.220*** | 0.163*** | 0.213*** |
| | (0.0296) | (0.0183) | (0.0364) | (0.0405) | (0.0177) | (0.0253) |
| Terciary | 0.233*** | 0.414*** | 0.973*** | 0.566*** | 0.425*** | 0.671*** |
| | (0.0309) | (0.0241) | (0.0480) | (0.0474) | (0.0244) | (0.0567) |
| Post Tertiary | 0.926*** | 1.449*** | 1.458*** | 0.650*** | 0.684** | 1.143*** |
| | (0.0812) | (0.0402) | (0.0534) | (0.251) | (0.303) | (0.179) |
| **Firm size** | | | | | | |
| One worker | -0.496*** | -0.0936** | 0.150** | -0.856*** | -0.401*** | -0.199** |
| | (0.0704) | (0.0437) | (0.0720) | (0.0677) | (0.0438) | (0.0776) |
| 2 - 10 employees | -0.165*** | -0.0658*** | -0.0534 | -0.464*** | -0.183*** | -0.0532 |
| | (0.0277) | (0.0232) | (0.0597) | (0.0727) | (0.0456) | (0.0838) |
| 11 - 50 employees | -0.0158 | -0.0299* | -0.122*** | -0.106 | -0.0199 | 0.00642 |
| | (0.0156) | (0.0172) | (0.0415) | (0.0873) | (0.0521) | (0.0951) |
| Observations | 4,517 | 4,517 | 4,517 | 7,843 | 7,843 | 7,843 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A20: Quantile Regression (Villavicencio)

| | Villavicencio | | | | | |
|---|---|---|---|---|---|---|
| | *Formal* | | | *Informal* | | |
| *y = log hourly wage* | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.132*** | 8.374*** | 8.638*** | 7.343*** | 8.154*** | 8.649*** |
| | (0.0363) | (0.0228) | (0.0443) | (0.104) | (0.0531) | (0.115) |
| λ | -2,635*** | -2,539** | -3,018*** | -937.7 | -415.6 | 1,887* |
| | (798.5) | -1,169 | (755.1) | -1,558 | -1,146 | -1,126 |
| Male | -0.0210 | 0.0759*** | 0.217*** | 0.423*** | 0.117*** | 0.0811*** |
| | (0.0165) | (0.0148) | (0.0319) | (0.0451) | (0.0183) | (0.0310) |
| Job Tenure | 0.00110*** | 0.00186*** | 0.00270*** | 0.00273*** | 0.00159*** | 0.00300*** |
| | (0.000317) | (0.000276) | (0.000342) | (0.000604) | (0.000247) | (0.000358) |
| Job Tenure Sq. | -1.85e-07 | -1.56e-06* | -3.21e-06*** | -4.33e-06** | -2.67e-06*** | -5.03e-06*** |
| | (1.13e-06) | (8.67e-07) | (6.50e-07) | (1.77e-06) | (6.15e-07) | (7.05e-07) |
| Educational level | | | | | | |
| No education | -0.520*** | -0.223 | -0.293** | -0.475*** | -0.382*** | -0.210*** |
| | (0.191) | (0.234) | (0.118) | (0.131) | (0.0818) | (0.0791) |
| Secondary | 0.0963*** | 0.0898*** | 0.249*** | 0.314*** | 0.157*** | 0.300*** |
| | (0.0326) | (0.0183) | (0.0418) | (0.0476) | (0.0187) | (0.0324) |
| Terciary | 0.318*** | 0.571*** | 1.083*** | 0.582*** | 0.423*** | 0.813*** |
| | (0.0358) | (0.0275) | (0.0479) | (0.0687) | (0.0353) | (0.0688) |
| Post Tertiary | 0.929*** | 1.255*** | 1.471*** | 0.601*** | 1.143*** | 1.240*** |
| | (0.0498) | (0.0347) | (0.0624) | (0.186) | (0.140) | (0.121) |
| **Firm size** | | | | | | |
| One worker | -0.660*** | -0.116** | 0.130* | -0.763*** | -0.213*** | -0.105 |
| | (0.0845) | (0.0458) | (0.0688) | (0.0988) | (0.0528) | (0.115) |
| 2 - 10 employees | -0.158*** | -0.109*** | -0.0855* | -0.344*** | -0.0676 | -0.0964 |
| | (0.0258) | (0.0190) | (0.0490) | (0.0958) | (0.0515) | (0.114) |
| 11 - 50 employees | -0.0563*** | -0.0766*** | -0.0834* | -0.0591 | 0.0500 | 0.194 |
| | (0.0187) | (0.0181) | (0.0436) | (0.113) | (0.0603) | (0.155) |
| Observations | 4,374 | 4,374 | 4,374 | 5,666 | 5,666 | 5,666 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A21: Quantile Regression (Pasto)

| | Pasto | | | | | |
|---|---|---|---|---|---|---|
| | *Formal* | | | *Informal* | | |
| *y = log hourly wage* | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.035*** | 8.326*** | 8.530*** | 7.589*** | 8.224*** | 8.721*** |
| | (0.0415) | (0.0239) | (0.0352) | (0.0776) | (0.0456) | (0.0969) |
| λ | -203.7 | -78.45 | -785.0 | 98.08 | 567.5 | -1,299 |
| | (876.8) | (837.0) | (816.8) | -1,357 | (468.2) | (795.3) |
| Male | -0.00440 | 0.0760*** | 0.156*** | 0.252*** | 0.151*** | 0.0625** |
| | (0.0162) | (0.0161) | (0.0301) | (0.0336) | (0.0170) | (0.0251) |
| Job Tenure | 0.000592** | 0.00119*** | 0.00242*** | 0.00225*** | 0.00122*** | 0.00172*** |
| | (0.000247) | (0.000219) | (0.000397) | (0.000356) | (0.000216) | (0.000260) |
| Job Tenure Sq. | 1.12e-06 | -4.14e-07 | -3.26e-06*** | -5.29e-06*** | -2.68e-06*** | -2.92e-06*** |
| | (7.00e-07) | (5.45e-07) | (8.91e-07) | (8.05e-07) | (5.45e-07) | (4.95e-07) |
| Educational level | | | | | | |
| No education | -0.00642 | -0.00933 | 0.969*** | -0.479*** | -0.333*** | -0.167** |
| | (0.208) | (0.126) | (0.258) | (0.105) | (0.0843) | (0.0726) |
| Secondary | 0.192*** | 0.0791*** | 0.268*** | 0.250*** | 0.156*** | 0.267*** |
| | (0.0366) | (0.0200) | (0.0427) | (0.0375) | (0.0181) | (0.0267) |
| Terciary | 0.368*** | 0.547*** | 1.097*** | 0.491*** | 0.516*** | 1.002*** |
| | (0.0384) | (0.0272) | (0.0405) | (0.0556) | (0.0322) | (0.0525) |
| Post Tertiary | 0.910*** | 1.315*** | 1.458*** | 0.969*** | 1.144*** | 1.195*** |
| | (0.0361) | (0.0287) | (0.0486) | (0.109) | (0.103) | (0.106) |
| **Firm size** | | | | | | |
| One worker | -0.885*** | -0.249*** | 0.132* | -0.906*** | -0.464*** | -0.339*** |
| | (0.108) | (0.0516) | (0.0744) | (0.0751) | (0.0455) | (0.0965) |
| 2 - 10 employees | -0.180*** | -0.114*** | -0.0124 | -0.473*** | -0.296*** | -0.255*** |
| | (0.0239) | (0.0237) | (0.0659) | (0.0723) | (0.0448) | (0.0985) |
| 11 - 50 employees | -0.0809*** | -0.0960*** | -0.0630 | -0.110 | -0.161*** | -0.277*** |
| | (0.0181) | (0.0209) | (0.0481) | (0.0815) | (0.0521) | (0.105) |
| Observations | 4,118 | 4,118 | 4,118 | 6,153 | 6,153 | 6,153 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A22: Quantile Regression (Armenia)

| | Armenia | | | | | |
|---|---|---|---|---|---|---|
| | *Formal* | | | *Informal* | | |
| *y = log hourly wage* | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.002*** | 8.280*** | 8.626*** | 7.625*** | 8.176*** | 8.511*** |
| | (0.0312) | (0.0210) | (0.0432) | (0.0789) | (0.0510) | (0.0677) |
| λ | 963.8 | -754.4 | -1,275 | -2,95 | -1,694* | 1,066 |
| | (992.3) | (834.0) | -3,59 | -2,16 | (899.1) | -1,062 |
| Male | -0.0120 | 0.0377*** | 0.117*** | 0.371*** | 0.193*** | 0.100*** |
| | (0.0148) | (0.0138) | (0.0295) | (0.0452) | (0.0203) | (0.0302) |
| Job Tenure | 0.000748*** | 0.00139*** | 0.00270*** | 0.00227*** | 0.000899*** | 0.00127*** |
| | (0.000175) | (0.000253) | (0.000450) | (0.000503) | (0.000265) | (0.000404) |
| Job Tenure Sq. | -6.86e-07 | -4.41e-07 | -2.94e-06** | -4.48e-06*** | -1.68e-06** | -7.25e-07 |
| | (4.61e-07) | (8.15e-07) | (1.22e-06) | (1.23e-06) | (7.07e-07) | (8.88e-07) |
| Educational level | | | | | | |
| No education | 0.00386 | -0.273*** | -0.383*** | -0.919*** | -0.546*** | -0.286** |
| | (0.0823) | (0.0691) | (0.0912) | (0.149) | (0.118) | (0.117) |
| Secondary | 0.171*** | 0.0771*** | 0.147*** | 0.320*** | 0.191*** | 0.276*** |
| | (0.0289) | (0.0163) | (0.0394) | (0.0505) | (0.0213) | (0.0336) |
| Terciary | 0.325*** | 0.448*** | 0.985*** | 0.519*** | 0.369*** | 0.717*** |
| | (0.0302) | (0.0242) | (0.0459) | (0.0611) | (0.0306) | (0.0527) |
| Post Tertiary | 0.971*** | 1.354*** | 1.446*** | 0.573 | 1.152*** | 1.212*** |
| | (0.0505) | (0.0300) | (0.0652) | (0.378) | (0.305) | (0.166) |
| **Firm size** | | | | | | |
| One worker | -0.615*** | -0.138*** | -0.0258 | -0.960*** | -0.325*** | -0.0129 |
| | (0.103) | (0.0328) | (0.0641) | (0.0665) | (0.0507) | (0.0672) |
| 2 - 10 employees | -0.106*** | -0.0769*** | -0.136*** | -0.535*** | -0.138*** | 0.183** |
| | (0.0187) | (0.0173) | (0.0444) | (0.0723) | (0.0524) | (0.0779) |
| 11 - 50 employees | -0.0693*** | -0.0583*** | -0.179*** | -0.234*** | -0.0592 | 0.0460 |
| | (0.0174) | (0.0169) | (0.0384) | (0.0727) | (0.0621) | (0.0895) |
| Observations | 4,666 | 4,666 | 4,666 | 4,848 | 4,848 | 4,848 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A23: Quantile Regression (Pereira)

| | Pereira | | | | | |
|---|---|---|---|---|---|---|
| | *Formal* | | | *Informal* | | |
| *y = log hourly wage* | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.151*** | 8.354*** | 8.433*** | 7.774*** | 8.205*** | 8.402*** |
| | (0.0221) | (0.0101) | (0.0208) | (0.0693) | (0.0336) | (0.0493) |
| λ | -2,471** | -1,382*** | -1,153 | -1,105 | -223.2 | 1,778 |
| | -1,242 | (363.0) | -1,439 | -2,013 | (570.8) | -1,22 |
| Male | -0.00128 | 0.0142* | 0.155*** | 0.178*** | 0.117*** | 0.0858*** |
| | (0.00857) | (0.00757) | (0.0223) | (0.0438) | (0.0162) | (0.0215) |
| Job Tenure | 0.000129 | 0.000403*** | 0.00227*** | 0.00199*** | 0.00164*** | 0.00206*** |
| | (0.000150) | (0.000144) | (0.000406) | (0.000679) | (0.000252) | (0.000291) |
| Job Tenure Sq. | 1.09e-06** | 7.31e-07 | -1.39e-06 | -4.82e-06*** | -3.69e-06*** | -3.75e-06*** |
| | (4.81e-07) | (5.28e-07) | (1.07e-06) | (1.74e-06) | (7.09e-07) | (6.54e-07) |
| Educational level | | | | | | |
| No education | 0.105** | 0.000831 | -0.118** | -0.508** | -0.252*** | -0.148*** |
| | (0.0509) | (0.0380) | (0.0560) | (0.254) | (0.0608) | (0.0495) |
| Secondary | 0.106*** | 0.0140* | 0.152*** | 0.183*** | 0.113*** | 0.183*** |
| | (0.0199) | (0.00765) | (0.0237) | (0.0512) | (0.0167) | (0.0253) |
| Terciary | 0.219*** | 0.373*** | 0.966*** | 0.426*** | 0.247*** | 0.689*** |
| | (0.0211) | (0.0178) | (0.0328) | (0.0527) | (0.0255) | (0.0608) |
| Post Tertiary | 0.989*** | 1.415*** | 1.523*** | 1.077*** | 0.947*** | 1.488*** |
| | (0.0499) | (0.0260) | (0.0553) | (0.284) | (0.100) | (0.157) |
| **Firm size** | | | | | | |
| One worker | -0.512*** | -0.0138 | 0.232*** | -1.181*** | -0.227*** | 0.100** |
| | (0.105) | (0.0298) | (0.0582) | (0.0710) | (0.0345) | (0.0502) |
| 2 - 10 employees | -0.0845*** | -0.0142 | 0.00634 | -0.329*** | -0.0913*** | 0.0633 |
| | (0.0162) | (0.0106) | (0.0356) | (0.0505) | (0.0311) | (0.0498) |
| 11 - 50 employees | -0.000928 | -0.00753 | -0.00327 | -0.314*** | -0.0831** | 0.0521 |
| | (0.00947) | (0.00905) | (0.0283) | (0.0730) | (0.0380) | (0.0603) |
| Observations | 5,596 | 5,596 | 5,596 | 5,154 | 5,154 | 5,154 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A24: Quantile Regression (Bucaramanga)

| | Bucaramanga | | | | | |
|---|---|---|---|---|---|---|
| | *Formal* | | | *Informal* | | |
| *y = log hourly wage* | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.159*** | 8.351*** | 8.612*** | 7.528*** | 8.259*** | 9.049*** |
| | (0.0252) | (0.0179) | (0.0412) | (0.107) | (0.0424) | (0.129) |
| λ | -683.0 | -240.6 | 1,382 | 1,099 | 62.16 | -196.3 |
| | (887.3) | -1,025 | -2,412 | (896.6) | (566.6) | -1,006 |
| Male | 0.00599 | 0.0465*** | 0.173*** | 0.315*** | 0.0953*** | 0.0642** |
| | (0.0111) | (0.0118) | (0.0281) | (0.0393) | (0.0171) | (0.0294) |
| Job Tenure | 0.00100*** | 0.00156*** | 0.00232*** | 0.00346*** | 0.00202*** | 0.00272*** |
| | (0.000283) | (0.000232) | (0.000373) | (0.000658) | (0.000215) | (0.000420) |
| Job Tenure Sq. | -1.82e-06 | -6.46e-07 | -8.86e-07 | -7.13e-06*** | -3.48e-06*** | -4.43e-06*** |
| | (1.19e-06) | (7.55e-07) | (8.95e-07) | (2.05e-06) | (4.78e-07) | (1.09e-06) |
| Educational level | | | | | | |
| No education | 0.0467 | -0.101 | -0.209 | -1.017*** | -0.497*** | -0.324*** |
| | (0.0753) | (0.133) | (0.142) | (0.263) | (0.0845) | (0.0664) |
| Secondary | 0.0977*** | 0.0700*** | 0.144*** | 0.187*** | 0.100*** | 0.160*** |
| | (0.0231) | (0.0152) | (0.0367) | (0.0418) | (0.0180) | (0.0291) |
| Terciary | 0.200*** | 0.388*** | 0.816*** | 0.279*** | 0.313*** | 0.751*** |
| | (0.0246) | (0.0198) | (0.0455) | (0.0726) | (0.0327) | (0.0701) |
| Post Tertiary | 0.820*** | 1.240*** | 1.506*** | 0.0852 | 0.889*** | 1.542*** |
| | (0.0492) | (0.0341) | (0.0755) | (0.168) | (0.247) | (0.279) |
| **Firm size** | | | | | | |
| One worker | -0.580*** | -0.123*** | 0.138** | -0.696*** | -0.322*** | -0.483*** |
| | (0.0640) | (0.0322) | (0.0702) | (0.100) | (0.0420) | (0.126) |
| 2 - 10 employees | -0.122*** | -0.0712*** | 0.0548 | -0.341*** | -0.130*** | -0.354*** |
| | (0.0206) | (0.0166) | (0.0451) | (0.101) | (0.0411) | (0.127) |
| 11 - 50 employees | -0.0396*** | -0.0540*** | -0.0284 | -0.221* | -0.0713 | -0.339** |
| | (0.0136) | (0.0149) | (0.0335) | (0.126) | (0.0508) | (0.149) |
| Observations | 5,379 | 5,379 | 5,379 | 6,179 | 6,179 | 6,179 |

Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A25: Quantile Regression (Sincelejo)

| | Sincelejo | | | | | |
|---|---|---|---|---|---|---|
| | *Formal* | | | *Informal* | | |
| *y = log hourly wage* | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.060*** | 8.229*** | 8.435*** | 6.698*** | 7.967*** | 8.582*** |
| | (0.0429) | (0.0324) | (0.0439) | (0.120) | (0.0498) | (0.0654) |
| λ | -469.0 | -2,532** | -3,206** | 1,624 | 126.5 | -1,033 |
| | (899.1) | -1,099 | -1,528 | -1,016 | (413.3) | (784.9) |
| Male | 0.00375 | 0.0817*** | 0.223*** | 0.608*** | 0.209*** | 0.0639*** |
| | (0.0126) | (0.0189) | (0.0306) | (0.0363) | (0.0150) | (0.0215) |
| Job Tenure | -0.000118 | 0.00172*** | 0.00265*** | 0.00346*** | 0.00159*** | 0.00196*** |
| | (0.000193) | (0.000260) | (0.000483) | (0.000372) | (0.000167) | (0.000263) |
| Job Tenure Sq. | 2.29e-06*** | -6.23e-07 | -2.72e-06* | -6.56e-06*** | -2.29e-06*** | -2.33e-06*** |
| | (6.11e-07) | (6.62e-07) | (1.41e-06) | (1.04e-06) | (4.07e-07) | (6.27e-07) |
| Educational level | | | | | | |
| No education | -0.148 | 0.0313 | 0.792** | -0.462*** | -0.126*** | -0.110*** |
| | (0.257) | (0.139) | (0.392) | (0.0912) | (0.0307) | (0.0380) |
| Secondary | 0.193*** | 0.0740*** | 0.156*** | 0.294*** | 0.145*** | 0.200*** |
| | (0.0395) | (0.0277) | (0.0435) | (0.0345) | (0.0146) | (0.0222) |
| Terciary | 0.306*** | 0.406*** | 1.032*** | 0.466*** | 0.323*** | 0.608*** |
| | (0.0407) | (0.0309) | (0.0469) | (0.0379) | (0.0208) | (0.0394) |
| Post Tertiary | 0.997*** | 1.323*** | 1.553*** | 1.047*** | 1.042*** | 1.127*** |
| | (0.0431) | (0.0338) | (0.0525) | (0.123) | (0.0986) | (0.0909) |
| **Firm size** | | | | | | |
| One worker | -0.699*** | 0.0181 | 0.166* | -0.275** | -0.310*** | -0.299*** |
| | (0.141) | (0.0707) | (0.0899) | (0.118) | (0.0488) | (0.0638) |
| 2 - 10 employees | -0.152*** | -0.0728** | 0.0168 | -0.111 | -0.192*** | -0.203*** |
| | (0.0263) | (0.0285) | (0.0585) | (0.118) | (0.0492) | (0.0657) |
| 11 - 50 employees | -0.0463*** | -0.0955*** | -0.0936* | 0.159 | -0.0923* | -0.0819 |
| | (0.0148) | (0.0231) | (0.0491) | (0.125) | (0.0540) | (0.0963) |
| Observations | 3,775 | 3,775 | 3,775 | 8,986 | 8,986 | 8,986 |

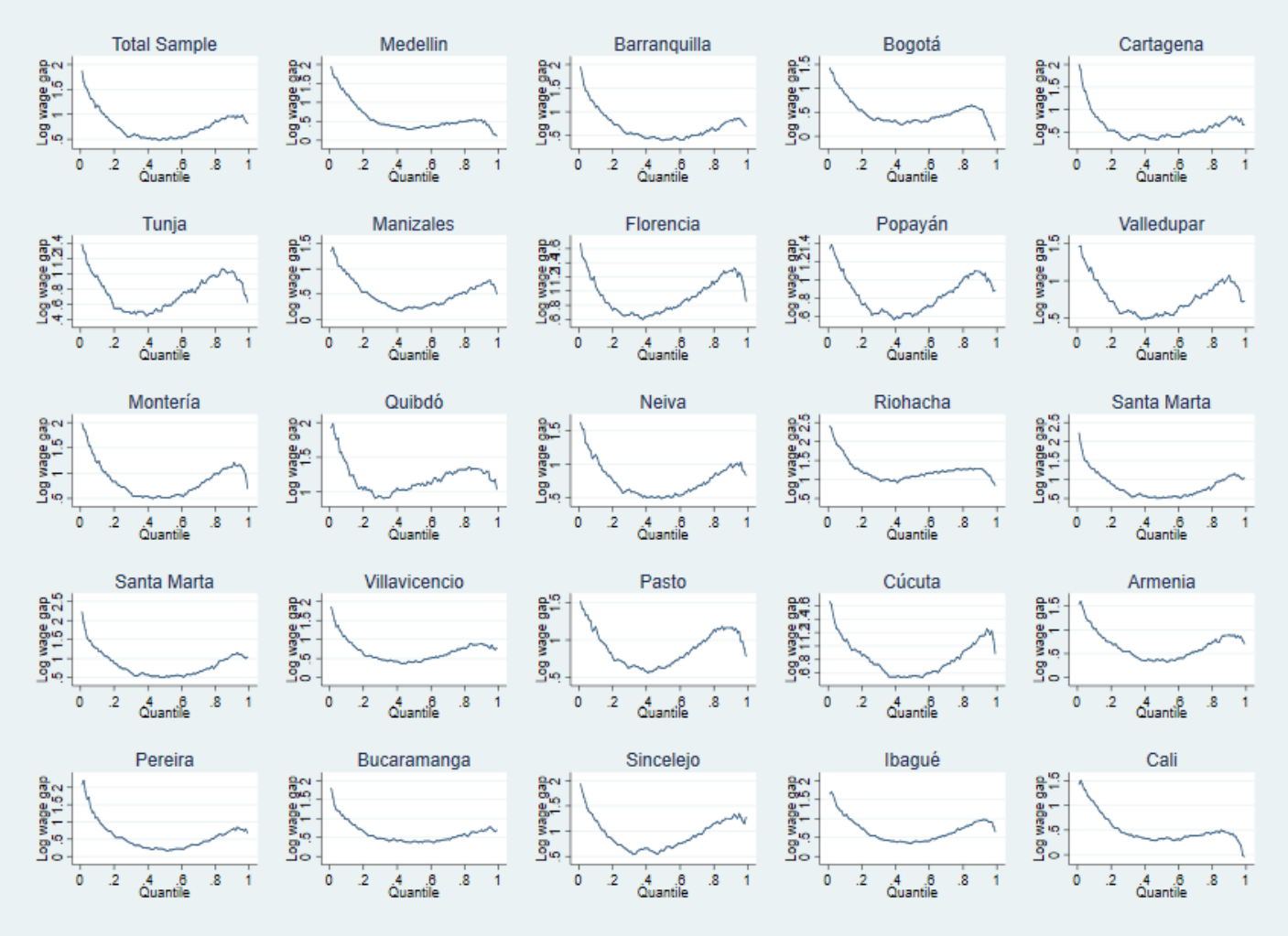
Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

## Table N° A26: Quantile Regression (Ibagué)

| | Ibagué | | | | | |
|---|---|---|---|---|---|---|
| | *Formal* | | | *Informal* | | |
| *y = log hourly wage* | 10% | 50% | 90% | 10% | 50% | 90% |
| Constant | 8.111*** | 8.361*** | 8.642*** | 7.287*** | 8.137*** | 8.557*** |
| | (0.0296) | (0.0207) | (0.0395) | (0.100) | (0.0456) | (0.0752) |
| λ | 598.1 | 574.5 | 9,328*** | 1,51 | -1,514 | 406.9 |
| | -2,04 | -1,222 | -2,432 | -1,698 | (941.9) | -1,514 |
| Male | 0.0114 | 0.0508*** | 0.148*** | 0.361*** | 0.0985*** | 0.0605** |
| | (0.0132) | (0.0146) | (0.0288) | (0.0519) | (0.0191) | (0.0265) |
| Job Tenure | 0.000663*** | 0.00112*** | 0.00195*** | 0.00345*** | 0.00194*** | 0.00246*** |
| | (0.000207) | (0.000231) | (0.000547) | (0.000509) | (0.000256) | (0.000450) |
| Job Tenure Sq. | -4.59e-07 | -7.82e-08 | -1.65e-06 | -6.79e-06*** | -3.75e-06*** | -3.82e-06*** |
| | (6.41e-07) | (6.44e-07) | (1.77e-06) | (1.19e-06) | (5.93e-07) | (1.16e-06) |
| Educational level | | | | | | |
| No education | -0.246 | -0.276* | 0.152 | -0.470 | -0.371*** | -0.0611 |
| | (0.151) | (0.154) | (0.259) | (0.297) | (0.0976) | (0.102) |
| Secondary | 0.118*** | 0.0574*** | 0.196*** | 0.321*** | 0.145*** | 0.200*** |
| | (0.0271) | (0.0163) | (0.0366) | (0.0538) | (0.0205) | (0.0287) |
| Terciary | 0.235*** | 0.378*** | 0.928*** | 0.501*** | 0.310*** | 0.745*** |
| | (0.0278) | (0.0220) | (0.0425) | (0.0684) | (0.0318) | (0.0603) |
| Post Tertiary | 0.917*** | 1.299*** | 1.443*** | 0.558** | 0.768*** | 1.852*** |
| | (0.0462) | (0.0347) | (0.0547) | (0.230) | (0.199) | (0.172) |
| **Firm size** | | | | | | |
| One worker | -0.874*** | -0.332*** | -0.192*** | -0.881*** | -0.250*** | -0.0623 |
| | (0.0923) | (0.0383) | (0.0506) | (0.0930) | (0.0452) | (0.0763) |
| 2 - 10 employees | -0.175*** | -0.124*** | -0.150*** | -0.349*** | -0.0714* | -0.0450 |
| | (0.0279) | (0.0206) | (0.0445) | (0.0901) | (0.0418) | (0.0764) |
| 11 - 50 employees | -0.0341** | -0.0691*** | -0.144*** | -0.0203 | 0.0157 | 0.116 |
| | (0.0138) | (0.0161) | (0.0419) | (0.101) | (0.0501) | (0.100) |
| Observations | 4,597 | 4,597 | 4,597 | 4,958 | 4,958 | 4,958 |

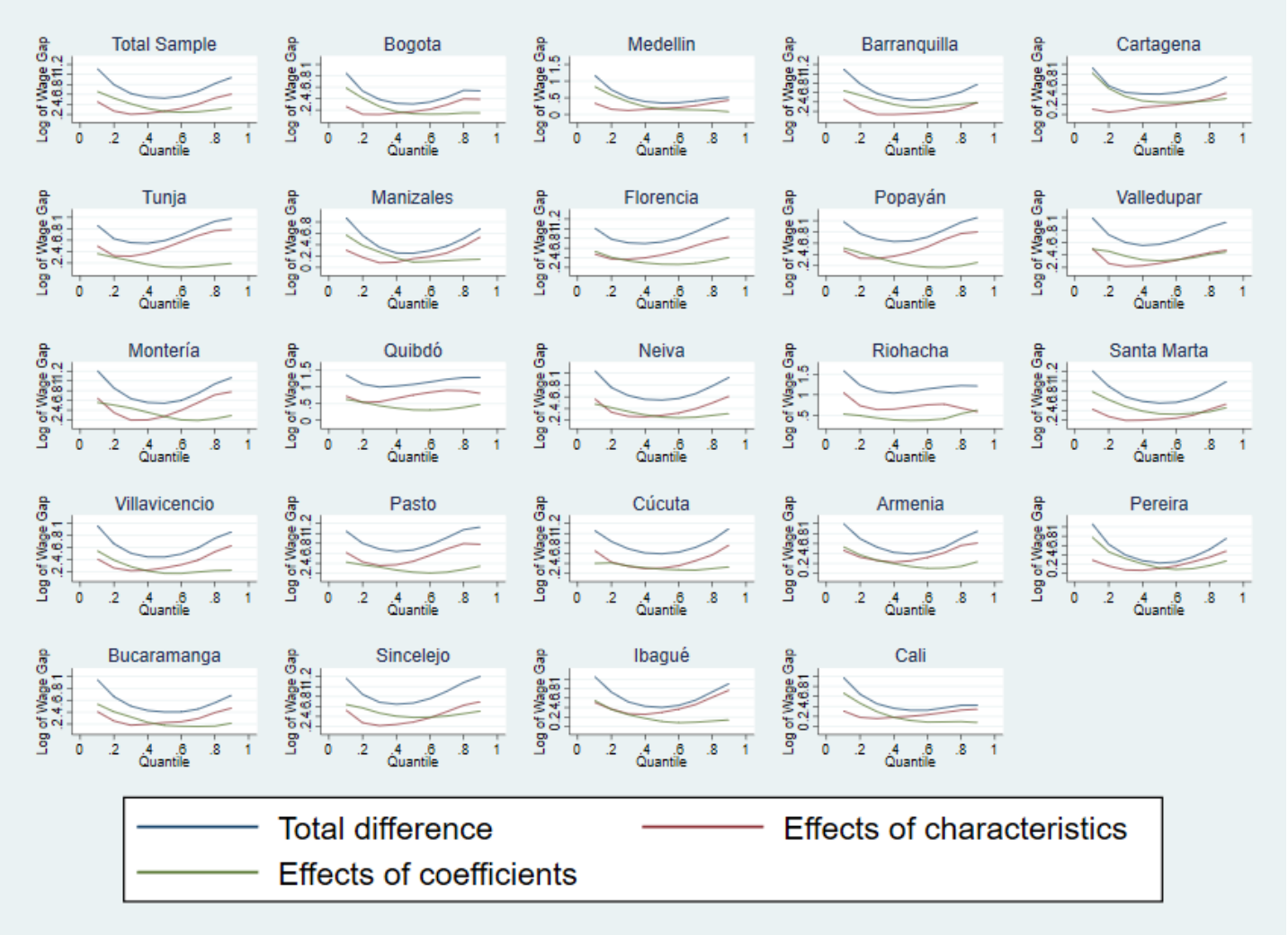
Source: Own calculations with monthly data from GEIH, 2019. Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0. Dummies for primary education and more than 51 employees were excluded categories in education levels and firm size. Job tenure is measured in months.

**FIGURE N° A1: Quantile distribution of the Formal - Informal Wage gap for all cities**




Note: The figure shows the quantile distribution of the formal-informal wage gap for each urban area of the country. Data was obtained the 2019 GEIH produced by DANE.

**FIGURE N° A2: Quantile decomposition of the Formal – Informal wage gap for all cities**




Note: The figure shows, for each urban area of the country, the quantile decomposition of the formal-informal wage gap (blue) on effects of coefficients (green) and effects of characteristics (red). Data was obtained the 2019 GEIH produced by DANE.

## Appendix B: Quantile decomposition method

Following García (2017) and Tannuri-Pianto & Pianto (2002) both based on Buchinsky (1998) and a semi-parametric method developed by Ichimura (1993), the step-by-step procedure of estimation of the Macho-Mata quantile regression decomposition accounting for selection bias as proposed by Albretch et.al (2009) , is explained as follows:

1. Let $I_i$ be a binary variable which takes the value of 1 if worker $i$ belongs to the informal sector, and 0 if they belong to the formal sector.

2. Assume the existence of latent or index variable, $I^*{}_i$, with the following equation:

$$I^*{}_i = z_i'\gamma + \mu_i$$

3. Where $z_i'$ is a set of individual characteristics which determine the likelihood of individual $i$ working in the informal sector, and $\gamma$ is a vector of coefficients to estimate.

4. Now considering:
$$I^*{}_i \begin{cases} 1 \; if \; I^*{}_i > 0 \\ 0 \; if \; I^*{}_i \leq 0 \end{cases}$$

5. Let $X_{inf}$ and $X_{for}$ be stochastic vectors of individual's characteristics for both informal and formal individuals, with distribution functions given as $D_{Xi}$ and $D_{Xf}$ respectively. The log wage for informal workers is $W_{inf}$ and $W_{for}$ for the group of formals, with unconditional distribution functions. The realization of the stochastic vectors $X_{inf}$ and $X_{for}$ can be expressed $x_{inf}$ and $x_{for}$

6. Then, the conditional quantile regression at the $\theta$ th quantile can be written for both formal and informal sectors as follows:

For the formal sector: $$Q_\theta(W_{for}|X_{for} = x_{for}) = x_{for}'\beta^{for}(\theta)$$
For the informal sector: $$Q_\theta(W_{inf}|X_{inf} = x_{inf}) = x_{inf}'\beta^{inf}(\theta)$$

7. Now, in the same line of García (2017), we follow the strategy proposed by Machado & Mata (2005) and adapted by Albrecht et.al (2009). This procedure entails the generation of a random sample of size $n$ from a uniform distribution $U[0,1]$: $u_1, u_2, u_3, \dots, u_n$ and the estimation of the conditional quantile regression for both groups of workers, which give us $n$ estimations of the Beta coefficients from the quantile regressions presented above.

8. Those results combined with a random sample of $n$ vectors of covariates $x$ from a uniform distribution can be used to predict the values of $w_{for} = x_{for}{}'\beta^{for}(u)$ and its counterfactual $w_{inf} = x_{inf}{}'\beta^{for}(u)$. These estimations are repeated $m$ times.

9. At the end, the gap between the wages of formal workers and the wages from their informal counterfactuals at the $\theta$th quantile can be presented as:

$$\begin{aligned} Q_\theta(W_{for}|X_{for} = x_{for}) - Q_\theta(W_{inf}|X_{inf} = x_{inf}) \\ = \left[Q_\theta\left(x_{for}{}'\beta^{for}(u)\right) - Q_\theta\left(x_{inf}{}'\beta^{for}(u)\right)\right] \\ + \left[Q_\theta\left(x_{inf}{}'\beta^{for}(u)\right) - Q_\theta\left(x_{inf}{}'\beta^{inf}(u)\right)\right] \end{aligned}$$

   The first term keeps the coefficient effects unchanged while changing the characteristics of individuals. The second term works in the opposite way, it keeps characteristics unchanged while changing coefficients. The first term shows how much of the wage gap is described by individual characteristics. The second term is a counterfactual considered as the formal payout of an informal worker, and it is expressed as the coefficient effect.

10. Since selection bias might exists because, different from the counterfactuals, the wages of formal and informal workers -the observable ones- are not drawn randomly from the distribution wage distribution. For solving this, Buchinsky (1998) introduces an extra term into the equations presented in numeral 6:

    For the formal sector: $Q_\theta(W_{for}|Z = z) = x_{for}{}'\beta^{for}(\theta) + h_\theta(z'\gamma)$
    For the informal sector: $Q_\theta(W_{inf}|Z = z) = x_{inf}{}'\beta^{inf}(\theta) + h_\theta(z'\gamma)$

11. Following Buchinsky (1998) power series suggestion to have an approximation to the last term in equations from numeral 10, we get:

    For the formal sector:
    $$Q_\theta(W_{for}|Z = z) = x_{for}{}'\beta^{for}(\theta) + \sum_{k=1}^{K}(\rho(\mu + \sigma\, z'\gamma))^{k-1}\delta_{k\theta}$$

    For the informal sector:
    $$Q_\theta(W_{inf}|Z = z) = x_{inf}{}'\beta^{for}(\theta) + \sum_{k=1}^{K}(\rho(\mu + \sigma\, z'\gamma))^{k-1}\delta_{k\theta}$$

    The term $\rho(\mu + \sigma z'\gamma)$ contains a constant ($\mu$) and a slope ($\sigma$) coefficients from the probit regression $I^*{}_i = z_i{}'\gamma + \mu_i$. To estimate the coefficient $\gamma$ in this regression, a semiparametric least square method is proposed -SLS-.