

# **Regional wage gaps, education, and informality in an emerging country. The case of Colombia.**

**Paula Herrera-Idárraga\***

Pontificia Universidad Javeriana and AQR–University of Barcelona

**Enrique López-Bazo**

AQR–University of Barcelona

**Elisabet Motellón**

Universitat Oberta de Catalunya and AQR–University of Barcelona

## **Abstract**

This paper uses Colombian micro-data to analyze the role of education and informality on regional wage differentials. Our hypothesis is that apart from differences in the endowment of human capital across regions, regional heterogeneity in the incidence of informality is another important source of regional wage inequality in developing and emerging countries. This is confirmed by the evidence from Colombia, which in addition reveals remarkable heterogeneity across territories in the wage return to individuals' characteristics. Regional heterogeneity in returns to education is especially intense in the upper part of the wage distribution. In turn, heterogeneity in the informal pay penalty is more relevant in the lower part.

**Keywords:** Regions, Wage differentials, Quantile-based decompositions, Formal/Informal Jobs, Economic Development

**JEL codes:** C21, J31, J38

---

\* **Corresponding author:** Paula Herrera-Idárraga. Mailing address: Calle 40 N 6-23 Piso 7 Edificio Gabriel Giraldo, S.J., Bogotá-Colombia; telephone number: +0057-13208320; e-mail: pherrera@javeriana.edu.co

**Acknowledgements:** The authors gratefully acknowledge financial support from the Spanish Ministry of Science and Innovation, *Plan Nacional de I+D+i*, Project ECO2011-30260-C03-03, and the European Commission under the 7<sup>th</sup> Framework Programme project SEARCH, SSH-2010-2.2-266834. Paula Herrera-Idárraga also acknowledges financial support from the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) (grant FI-DGR 2012) and from Pontificia Universidad Javeriana.

## 1. Introduction

Over the past decade, several studies have registered the decline in income inequality for Latin America countries (López-Calva & Lustig, 2009; Gasparini, Cruces and Tornarolli, 2011). While this trend in income inequality has received special attention at the national level, studies on regional disparities in the components of individuals' income are still scarce for Latin America countries. Analyses focusing on the regional dimension are of great relevance, because even in the presence of declining income inequality at national level, important inter-regional disparities may persist. This is so, because socio economic indicators at the national level can often hide significant variances between territories of the same country. This study considers the case of Colombia, a country that despite a decrease in income inequality in the past decade presents one of the highest Gini coefficients of Latin America countries and faces large geographical inequalities. Colombia shows important disparities in economic and social development among its regions. This implies that an important part of inequality between Colombian individuals may be the consequence of disparities between regions of the country (Bonet and Meisel, 2008; Joumard and Londoño, 2013). In particular, differences in wages deserve attention from a regional perspective as, for example, in 2010 the average gross hourly wage of a small city, such as Cucuta, was only 66% of that paid in Bogota, the country's capital.

To explain large spatial wage disparities, a number of arguments have been proposed. One of them emphasizes that wage disparities across areas are caused by differences in amenities. For instance, certain areas may have a favorable climate and easier access to natural resources. Under this context, wage differentials may be seen as compensated differentials, meaning that some areas may have higher wages to attract workers so as to compensate for the lack of amenities (Greenwood et al. 1991). Another explanation is related to the point that differences in wages across regions could reflect spatial differences in the skill composition of the workforce (Combes, Duranton and Gobillon, 2008). Workers with better labor market characteristics tend to sort themselves in areas that concentrate industries with high skill requirements where wages tend to be higher. Associated to this last explanation, the third one is based on agglomeration economies. A larger pool of high skill workers in an area may provide a source of important knowledge spillovers that can lead to productivity gains (Glaeser et al., 1992). Also, labor pooling improves the matching between firms and workers, which could also

increase economic efficiency and lead to higher wages (Andersson, Burgess and Lane, 2007).

A number of studies have been devoted at measuring the degree of regional wage gaps and identifying their origin. For instance, Blackaby and Murphy (1995), Duranton and Monastiriotis (2002), and Dickey (2014) analyze the case of Great Britain; García and Molina (2002), Motellón, López-Bazo and El-Attar (2011), and López-Bazo and Motellón (2012) that of Spain; and Galego and Pereira (2014) and Pereira and Galego (2014) the one of Portugal. These studies center their analysis on the estimation of wage equations and on the decomposition of regional wage gaps. The decomposition analysis is based on the idea that regional wage differentials are the result of how characteristics that determine wages are distributed across regions (the endowment component) and by how different these characteristics are rewarded across space (the wage structure component). The extent to which these two components explain regional wage differentials has been of great interest in past studies and their importance in explaining regional wage gaps differ considerably across and within countries. Some studies conclude that the regional wage differentials are mostly due to differences in individual characteristics between regions (Blackaby and Murphy, 1995). Other studies found that a significant part of wage differentials are explained by differences in returns, (Motellón, López-Bazo and El-Attar, 2011; Galego and Pereira, 2014; and Pereira and Galego 2014), while some others point that both components play an important role (García and Molina, 2002).

Less evidence is available on the amount and origin of regional disparities using individual-level data in the Latin America countries. Azzoni and Servo (2002) using micro-data for the 10 largest metropolitan regions in Brazil found that wage differentials were lower after adding controls for worker and job characteristics and cost of living, though they remain sizeable. With regard to the factors that explain regional wage disparities, they found education as the most important variable for explaining such differences. Romero (2008) pursued a similar study for the Colombian case, and concluded that a significant part of regional labor income differences disappeared after adding controls for worker and firm characteristics. His results also indicate that the contribution attributed to regional differences in the cost of living is negligible for explaining labor income inequality across regions, while differences in education is the most important source of the observed regional labor income disparities. Quiñones and

Rodriguez (2011) reach the same conclusion when assessing the contribution of differences in education in explaining wage differentials across Colombian regions. So it can be concluded from past studies that differences in the endowment of human capital and in its returns has been the most important factor for explaining regional wage differentials.

In accordance with the previous evidence, this study pays special attention to spatial imbalances in the endowment of human capital, and to what extent these differences and the regional heterogeneity in the return to this type of capital may help to explain regional wage gaps. But unlike most previous studies for developing countries, as a new and major contribution, this paper will focus not only on regional differences in human capital, but will go further by exploring one important feature of almost all developing and emerging countries: the large proportion of workers employed in informal jobs. Interestingly, recent studies for Colombia have emphasized that informal jobs are not equally distributed across the main metropolitan areas of the country (Galvis, 2012). As an example, some Colombian cities have informality rates of around 60% while the incidence of informality in other territories is about 20%.

In addition, we build on the results in the study by Ortiz, Uribe and Badillo (2008), which indicates that the Colombian labor market is segmented in two dimensions. An intra-regional or scale segmentation, which is mainly due to the restrictions on the access to physical and human capital that limited the possibility of expansion of firms to a larger scale. This type of segmentation may imply that workers and employers in the informal sector, usually associated with small establishments, face significant barriers in the transition to the formal sector, with higher productivity and income. The second type is inter-regional segmentation, which is mainly due to the barriers of mobility of labor and other factors between regions. Accordingly, the hypothesis of our study is that regional wage inequality may be explained by regional differences in the availability of good jobs that generate higher wages. Meaning that, apart from the differences in the endowment of human capital across Colombian regions, regional heterogeneity in the incidence of informality is likely to be another important source of regional wage disparities. As far as we are aware this issue has not been considered in any of the previous studies on regional wage disparities.

Our empirical analysis consists of examining the returns to education and the pay penalty of informal jobs across Colombian regions by using mean and quantile regression models in order to analyze the effect of observed characteristics along the wage distribution. Then, regional wage gaps are decomposed into the contribution of differences in the regional distribution of characteristics, and into the contribution of differences in wage structures (heterogeneity in prices or returns to characteristics). In doing so, we apply the standard Blinder-Oaxaca decomposition at the mean and the decomposition for unconditional quantile regression (UQR) models proposed by Firpo, Fortin and Lemieux (2009; 2011) at selected quantiles. With both of these approaches it is possible to isolate the particular contribution of education and informality to the regional wage gap at different quantiles, in contrast with other procedures (Machado and Mata, 2005; Melly, 2005). Galego and Pereira (2014) applied this method in the case of regional wage differentials for Portugal. As far as we know, our study represents the first application of this method for the analysis of regional wage differentials of a developing country.

Results for Colombia show that regions not only differed in earning relevant characteristics, but also display sizeable regional variability in the returns to these characteristics. Particularly, heterogeneity in returns to education across regions play an important role in explaining regional wage gaps. Additionally, workers face different informal pay penalties throughout the territory and it affects mostly individuals at the lower part of the wage distribution. Therefore, its contribution in explaining regional wage gaps is limited to this part. Our results confirm previous evidence on the existence of significant regional wage differences between the Golden Triangle region, conformed by the cities of Cali, Medellin and Bogota, and other regions in the country. The difference is particularly wide for those regions with a large share of labor in the informal sector. Moreover, it seems that the distribution of education is generating an equalizing effect of wages across some regions, whereas the returns to education continue to be a source of wage inequality across Colombian territories.

The results of this study point to the conclusion that some public policies aiming at reducing human capital differences among regions will help to decrease regional wage gaps, especially at the higher parts of the wage distribution. However, equalizing years of education of workers across regions would not be enough to reduce regional wage

differences due to the sizeable differences in returns to years of education at higher quantiles. Similar results have been found in previous studies, albeit in the context of developed countries. Meanwhile policies that point towards the reduction of informality will help to minor regional wage gaps at the lower part of the wage distribution, particularly for those regions with sizable informality.

The remainder of the paper is organized as follows. The next section presents a description of the data used. Section 3 outlines the methodology applied in this study. Then, sections 4 and 5 report and discuss the results. Finally, conclusions are presented in section 6.

## **2. Data and descriptive analysis**

We use data from the second quarter of 2010 of the Colombian Household Survey (CHS), a repeated cross-section conducted by the National Statistics Department (DANE). The survey gathers information about employment conditions for population aged 12 or more including income, occupation, and industry sector at two digit level, in addition to the general population characteristics such as sex, age, marital status and educational attainment. The CHS is representative for the thirteen major metropolitan areas in Colombia, composed of a main city and its associated municipalities.

The analysis was restricted to salary workers that were not carrying formal studies aged between 15 and 60 years and who report working more than 16 hours per week. We do not include self-employed and employers workers in the analysis because their source of income is a combination of labor and physical capital and therefore may not be compared with earnings of other employees. We also exclude public employees from the sample. Public wages are fixed at the national level for all the public administration along the territory so that the regional wage differentials may be artificially lower if public employees are included in the analysis. After excluding observations with missing values or inconsistencies, 13796 individuals remained in our sample.

As for the measure of wages used in the empirical exercise, we have combined information from gross monthly wage earnings and worked hours to obtain gross hourly wages. A first look at the degree of regional wage differentials in Colombia is obtained from a simple inspection of Table 1, which in the second column of data displays the average gross hourly wage. Large differences in average wages across the thirteen

metropolitan areas are observed. For instance, the average wage in Cucuta, the metropolitan area with the lowest level, was 66.15% the average wage in Bogota, the metropolitan area with the highest level. As in previous studies, we attempt to control for price differentials by adjusting the nominal gross hourly wage using the deflator from the consumer price index of each city. Consumer price indices for the main city of each metropolitan area were obtained from DANE. Averages of this *adjusted* gross hourly wages are shown in the third column of Table 1. It is observed that the position in the regional ranking of wages is fairly the same and that the metropolitan areas in the top and the bottom of the ranking remain unchanged. The fact that the consumer price index is built with a base year fairly recent, 2008, may explain the small variation obtained after controlling for difference in prices across the metropolitan areas. However, as far as we know this is the only information on regional prices available for Colombia.<sup>1</sup>

The regional wage gap may be due to differences across metropolitan areas in workers' characteristics. In particular, they are known to differ in the workers' endowment of education, which is one of the essential determinants of wages. Table 1 shows the average years of education of workers in each metropolitan area. As it can be seen, there are notable differences in education. On average, workers in Cartagena have more than two years of education than those in Cucuta. On the other hand, as has already been mentioned, previous studies for Colombia have shown that the incidence of informality varies considerable between regions. Since informal workers earn considerably lower wages than their formal counterparts, a metropolitan area with a higher proportion of informal workers may have lower wages than a metropolitan area with a low fraction of informal workers. In this paper, we define workers as formal if they contribute both to health and old-age insurance, as proposed by the International Labor Organization (ILO). According to this legal definition, an informal job is an activity that is unregulated by the formal institutions and regulations of a country. Importantly, since data comes from a household survey and therefore the information relates only to workers and not to firms, the informal sector term is related to the nature of the job and not of the firm in which the worker is employed.

The percentage of informal workers in each of the metropolitan areas is included in the last column of Table 1. In accordance with what has been found in previous

---

<sup>1</sup> A similar procedure has been applied in, for instance, Dickey (2014) and Pereira and Galego (2014).

studies, the incidence of informality is very different across metropolitan areas. While Cucuta displays an informality of around 59%, the share of informal workers in Medellin is about 19%. Interestingly, some metropolitan areas with the lowest average wages have the highest levels of informality (Villavicencio, Pasto and Cucuta). So, as expected, these simple descriptive figures suggest a negative correlation between the incidence of informality and hourly wages in the Colombian metropolitan areas.

In order to make the analysis more manageable and for the sake of brevity, metropolitan areas were grouped into regions, following the classification suggested by DANE based on geographical proximity and natural characteristics. However, it should be mentioned that we grouped Bogota, Medellin, and Cali into one region that we will refer to as the Golden Triangle.<sup>2</sup> These metropolitan areas are the most productive and dynamic of the country. The most productive firms, most of R&D investments and the highest skill workers are concentrated in these three areas. Besides Golden Triangle, the other metropolitan areas in the dataset are grouped into five regions: Atlantic (Barranquilla, Cartagena and Monteria), Oriental (Cucuta, Bucaramanga, and Villavicencio), Central (Manizales, Pereira and Ibague), and Pacific, only composed by Pasto.

Table 2 provides a description of hourly wages for the five regions. Clearly, average hourly wages differ between regions, although the magnitude of the differences is lower than the one found for the thirteen metropolitan areas. The average hourly wage of the region with the lowest level, Pacific, is 74% that in the region with the highest level, Golden Triangle. So by grouping metropolitan areas into regions the amount of disparities is attenuated, but they still remain sizable. Apart from the differences in the mean, the wage distributions of these five regions present other interesting variations. For instance, Table 2 shows that regional wage distributions vary in terms of the degree of dispersion. The standard deviation of the logarithm of gross hourly wages and the Gini index for the region with the lowest level of wages, Pacific, are far above those for the region with the highest wage level, Golden Triangle, suggesting that regions also differ in terms of the amount of intra-regional inequality. Finally, from the value of wages at the

---

<sup>2</sup> Colombia's Golden Triangle refers to an urban region, limited by a triangle whose vertexes are defined by the three largest cities: Bogotá, Medellin and Cali. In our particular case, we are not referring to the region, but only to the three cities that demarcates the triangle.



quartiles of the distribution (25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> percentiles),<sup>3</sup> reported in the last block of columns of Table 2, it can be concluded that regional wage differentials are far from constant over the entire wage distribution, with symptoms of a non-monotonic behavior. Summing up, evidence from Table 2 confirms that there are noticeable differences across regions in the entire wage distribution, and not just on average wages. To account for these differences, in the rest of this paper we provide results for the average and the quartiles.

As it has been already mentioned, regional wage differentials might be caused by the spatial distribution of human capital and other earning relevant determinants, as informality. A simple description of the amount of regional variability in worker and firm characteristics is reported in Table 3. It is observed that regions with high levels of wages have workers employed in relatively larger firms and with a permanent contract. The proportion of workers employed in the sectors of industry and financial intermediation is larger in high wage regions. One aspect that also worth mentioning is the low proportion of women working in Atlantic region, 39%, compare to 45% in Golden Triangle. Interestingly, informality also differs considerably between regions. For instance, the incidence of informality is 49% in Pacific while in the Golden Triangle is 23%. These differences in the proportion of informal workers across regions might intensify regional wage differentials, since formal jobs usually entail higher wages than informal jobs.

Therefore, we should conclude that there are differences between regions in characteristics that may result in regional wage differentials. In particular, data confirm that Colombian regions differ markedly in the endowment of education and in the share of informal jobs. Nevertheless, the key point is if these differences account for the bulk of regional wage disparities, or if part of the wage gap is produced by differences across regions in how these characteristics are rewarded. If regional wage gaps were completely explained by differences in the distribution of observable characteristics across regions, then under such circumstances, similar workers employed in similar firms but located in different regions would earn the same wage. On the contrary, if part of the wage gap could be explained by differences in how characteristics are rewarded, this could be associated to failures in regional labor markets, as similar workers in comparable firms but

---

<sup>3</sup> In order to save space we do not reproduce here results for other percentiles, although they are available upon request. In any case, including results corresponding to more percentiles does not modify the general conclusions regarding regional disparities over the entire wage distribution.

in different regions would be earning different wages. In the sections that follow we aim to shed more light on this issue, paying particular attention to the role of differences in education and informality.

### 3. Empirical strategy

#### 3.1. Specification of the wage equation

The empirical strategy is based on a model in which the wage of individual  $i$  in region  $r$  is given by:

$$W_{ir} = X_{ir}\boldsymbol{\beta}_r + \varepsilon_{ir} \quad (1)$$

where  $W_{ir}$  denotes the log of the hourly wage of individual  $i$  in region  $r$ .  $X_{ir}$  denotes the set of characteristics that affect the wage of this individual, including years of education, experience (and its square), tenure (and its square), gender, sector of employment, marital status, head of household, hours worked, type of contract, size of the firm and firm sector.  $\boldsymbol{\beta}_r$  is the vector of prices or returns at region  $r$  associated to the characteristics in  $X_{ir}$ . Equation (1) is estimated for each region, so that an estimate of the effect of education and informality is obtained for each region rather than imposing the same effect for all regions. Therefore, the wage equation specified in (1) is consistent with inter-regional segmentation, that is to say, with workers with similar characteristics obtaining different returns across regions.

The analysis from equation (1) is based on the mean distribution of wages. However, the descriptive in the previous section showed that regional disparities are far from uniform over the entire wage distribution. Therefore, it is of interest to know the effects of variables such as education and informality at different points of the distribution of wages. This can be done by means of the conditional quantile regression (CQR) model introduced by Koenker and Bassett (1978). It can be written as:

$$W_{ir} = X_{ir}\boldsymbol{\beta}_{\tau r} + \varepsilon_{\tau ir} \quad \text{with} \quad Q_{\tau}(W_{ir}|X_{ir}) = X_{ir}\boldsymbol{\beta}_{\tau r} \quad (2)$$

where  $Q_{\tau}(W_{ir}|X_{ir})$  denotes the  $\tau$ -th conditional quantile of wages given the set of characteristics in  $X_{ir}$ .

The coefficients  $\beta_{\tau r}$  are estimated by already standard procedures (Koenker, 2005; Koenker and Bassett, 1978), and may be interpreted as marginal or partial effects (depending on whether the corresponding covariate is continuous or binary) on the conditional quantile of interest. If  $\beta_{\tau r}$  is a consistent estimator of the conditional and unconditional quantile of  $W_r$ , the underlying data generating process follows a linear-in-parameters additive model structure, i.e. is a pure parallel location-shift data generating process for every covariate. However, if the conditional effect of a specific variable in  $X_r$  varies over the levels of other covariates in  $X_r$ ,  $\beta_{\tau r}$  may be a consistent estimator of the conditional effect of a variable at the mean values of the other  $k-1$  remaining covariates, but is not a consistent estimator of the effect of  $X_r$  on the unconditional wage distribution (see e.g. Borah and Basu, 2013). Meaning that, for example, the 90<sup>th</sup> percentile of the unconditional distribution of wages may not be the same as the 90<sup>th</sup> percentile of conditional distribution, conditioned on the covariates.

It is possible to estimate the unconditional quantile effect of  $X_r$  using the approach proposed by Firpo, Fortin and Lemieux (2009) based on the influence function (IF) and recentered influence function (RIF). In the context of wages, the IF is:

$$IF(W_r; q_\tau) = (\tau - I\{Y \leq q_\tau\})/f_{W_r}(q_\tau) \quad (3)$$

where  $q_\tau$  refers to the  $\tau$ -th unconditional quantile of wages,  $f_{W_r}(q_\tau)$  is the probability density function of  $W_r$  evaluated at  $q_\tau$ , and  $I\{Y \leq q_\tau\}$  is an indicator variable to denote whether an outcome value is less than  $q_\tau$  or not. By definition the RIF is equal to:

$$RIF(W_r; q_\tau) = q_\tau + IF(W_r; q_\tau) \quad (4)$$

Firpo, Fortin and Lemieux (2009), demonstrate that the implementation of the UQR is straightforward and similar to the OLS regression. For a specific quantile  $\tau$ , the first step is to estimate the RIF of the  $\tau$ -th quantile of  $W_i$  following eq. (3) and eq. (4). The second step is to run OLS regression of the  $RIF(W_{ir}; q_\tau)$  on the observed covariates,  $X_{ij}$ :

$$E[RIF(W_{ir}; q_\tau | X_{ir})] = X_{ir} \boldsymbol{\beta}_{\tau r} \quad (5)$$

Coefficients  $\boldsymbol{\beta}_{\tau r}$  represents the approximate marginal effects of the explanatory variables on the unconditional quantile  $q_\tau$  of wages for workers in region  $r$ .

### 3.2. Decomposition of regional wage gaps

The Blinder-Oaxaca method is formulated for decomposing mean differences in log wages between two groups after the estimation of the wage equation in (1).<sup>4</sup> In our particular case, the wage gap between a high wage region ( $r = h$ ) and a low wage region ( $r = l$ ) can be specified as:

$$\bar{W}_h - \bar{W}_l = (\bar{X}_h - \bar{X}_l) \hat{\boldsymbol{\beta}}_h - \bar{X}_l (\hat{\boldsymbol{\beta}}_h - \hat{\boldsymbol{\beta}}_l) \quad (6)$$

The first term in the RHS of this expression corresponds to the difference in the average values of observed worker and firm characteristics between regions  $h$  and  $l$ , whereas the second term is the part of the wage gap attributable to differences in the estimated coefficients; i.e. differences in the wage structure.

It is possible to obtain a decomposition of the wage differential at quantile  $\tau$ , similar to the classical Blinder-Oaxaca decomposition, for any two regions using the RIF regression approach by Firpo, Fortin and Lemieux (2009). Any distributional parameter, for example a wage quantile, can be written as a function  $q_\tau(F_W)$  of the cumulative distribution of wages,  $F_W(W)$ . For example, the difference in a wage quantile  $\tau$ ,  $\Delta^{q_\tau}$ , between a high wage region and a low wage region, can be written as:

$$\begin{aligned} \Delta^{q_\tau} &= q_\tau(F_{W_h|r=h}) - q_\tau(F_{W_l|r=l}) \\ \Delta^{q_\tau} &= [q_\tau(F_{W_h|r=h}) - q_\tau(F_{W_l|r=h})] + [q_\tau(F_{W_l|r=h}) - q_\tau(F_{W_l|r=l})] \\ \Delta^{q_\tau} &= \Delta_S^{q_\tau} + \Delta_X^{q_\tau} \end{aligned} \quad (7)$$

---

<sup>4</sup> The Blinder-Oaxaca decomposition has been applied in several studies analyzing wage gaps by gender, race, public and private sector, etc. It has also been applied recently for understanding regional wage gaps (García and Molina, 2002; Motellón, López-Bazo and El-Attar, 2011; Pereira and Galego, 2014; Galego and Pereira 2014).

where  $q_\tau(F_{w_h|r=h})$  indicates the actual wage quantile of workers belonging and rewarded under the wage structure of region  $r = h$ .  $q_\tau(F_{w_l|r=h})$  represents the counterfactual wage quantile, that is the wage quantile that would prevailed if workers observed in the region with high wages,  $r = h$ , had been paid under the wage structure of workers in the low wage region,  $r = l$ . Using the actual and counterfactual wage quantile for each region it is possible to decompose the wage gap at any quantile,  $\Delta^{q_\tau}$ , in two terms, one which captures the wage structure effect,  $\Delta_S^{q_\tau}$ , and another that represents the endowments effect  $\Delta_X^{q_\tau}$ .

However, as in the case of the Blinder-Oaxaca decomposition for the mean, if the true conditional expectation is not linear, the decomposition based on a linear regression may be biased (Barsky et al., 2002). A reweighted procedure coupled with the RIF-regressions can solve this problem (Fortin, Lemieux and Firpo, 2011). First a reweighting factor has to be calculated as:

$$\Psi(X) = \frac{\Pr(r = h|X) / \Pr(r = h)}{\Pr(r = l|X) / \Pr(r = l)} \quad (8)$$

Then RIF-regressions are computed for workers in regions  $l$ ,  $h$  and for the counterfactual  $l^c$  region, using the weights in  $\Psi(X)$ , to later calculate the next decomposition:

$$\begin{aligned} \hat{\Delta}^{q_\tau} &= (\bar{X}_h \hat{\beta}_{th} - \bar{X}_l^c \hat{\beta}_{tl}^c) + (\bar{X}_l^c \hat{\beta}_{tl}^c - \bar{X}_l \hat{\beta}_{tl}) \\ \hat{\Delta}^{q_\tau} &= \hat{\Delta}_S^{q_\tau} + \hat{\Delta}_X^{q_\tau} \end{aligned} \quad (9)$$

where  $\bar{X}_r$  denotes the mean of wages in region  $r$  ( $=l$  and  $h$ ), and  $\bar{X}_l^c$  is the counterfactual mean for region  $l$  using the reweighting factor in (8) so to make the distribution of the characteristics,  $X$ , in the region with low wages similar to that of region with high wages.

The wage structure effect can be divided into a pure wage structure effect and a component measuring the reweighting error, as follows:

$$\hat{\Delta}_S^{q_\tau} = \bar{X}_h (\hat{\beta}_{th} - \hat{\beta}_{tl}^c) + (\bar{X}_h - \bar{X}_l^c) \hat{\beta}_{tl}^c \quad (10)$$

$$\hat{\Delta}_S^{q\tau} = \hat{\Delta}_{S,p}^{q\tau} + \hat{\Delta}_{S,e}^{q\tau}$$

The reweighting error goes to zero as  $\bar{X}_l^c \rightarrow \bar{X}_h$ . Similarly, the composition effect can be divided into a pure composition effect and a component for the specification error as:

$$\begin{aligned} \hat{\Delta}_X^{q\tau} &= (\bar{X}_l^c - \bar{X}_l) \hat{\beta}_{\tau l}^c + \bar{X}_l (\hat{\beta}_{\tau l}^c - \hat{\beta}_{\tau l}) \\ \hat{\Delta}_X^{q\tau} &= \hat{\Delta}_{X,p}^{q\tau} + \hat{\Delta}_{X,e}^{q\tau} \end{aligned} \quad (11)$$

## 4. Results

### 4.1. OLS and quantile regressions estimates

Table 4 reports the results of Mincer wage equations estimated at the mean (OLS) and at the quartiles (25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> percentiles) for the five regions and Colombia as a whole. Since the particular focus is on the effect of education and informal work, the results are shown only for the estimates of the coefficients associated to years of schooling and informality, though all the variables presented in Table 3 were included as control variables.<sup>5</sup> The first column in Table 4 contains the estimates at the mean, that is to say the results of the OLS. The estimated returns to schooling for each region and the entire country are displayed in the upper panel of the table. It can be observed how at the country level, investments in education are quite profitable, since the estimated return is 7.42%, and highly significant. This is so as well in the five regions under analysis although, as expected, there are significant differences across regions in the return to years of education. For example, a higher return to schooling is observed in those regions with the highest level of wages. The returns to schooling in Atlantic and Golden Triangle are 8.14% and 8.26% respectively. On the other hand, those regions with the lowest levels of hourly wages display the lowest returns to schooling; 5.57% in the Oriental region and 6.82% in Pacific. Thus, in addition to differences in the endowment of education, returns to schooling may be thought to be an important factor in explaining wage gaps across regions.

The OLS estimates of the informal pay penalty, reported in the lower panel of Table 4, show a more complex pattern. The Pacific region, which is the region with the

---

<sup>5</sup> The full set of estimates is available upon request.

lowest wage level, is the one with the highest pay penalty; an informal worker earns 26.8% less than an otherwise similar formal worker in that region. However, the next region in the pay penalty ranking is that with the highest wage level, Golden Triangle, with an estimated penalty of 13.56%. Even though the pay penalty is considerably larger in the region with the lowest level of wages compared to the region with the highest, there seems not to be a clear pattern between the informal pay penalty and the regional wage gap. In any case, the OLS results suggest that Colombian regions differ not only in the incidence of informality (the share of the informal sector) but also in the mean wages earned by otherwise similar formal and informal workers.

As for the results of the quantile regressions, the second block of columns in Table 4 show the conditional quantile estimates (CQR) of the returns to schooling and the informality pay penalty. More precisely, results are shown for the three quartiles of the conditional wage distribution.<sup>6</sup> Consistently with previous literature (e.g. Galego and Pereira, 2014), returns to schooling are heterogeneous and increasing along the quartiles for all regions and the country as a whole. CQR results for Colombia confirm that the return increases along the conditional wage distribution. It is almost 70% higher in the third than in the first quartile. The same applies to all regions. However, results reveal some interesting features. For instance, returns to schooling in the Golden Triangle range from 4.62% for the first quartile to 8.99% for the last quartile of the conditional distribution of wages. In turn, in the Pacific region, returns to schooling in the first and last quartiles are 5.16% and 7.29%, respectively. Interestingly, the returns to schooling are therefore higher for Pacific compared to those in Golden Triangle at the bottom of the conditional distribution, they are fairly the same at the middle, and lower at the third quartile. Another interesting aspect is that the increase in the return along the wage distribution for all regions indicates that education has an unequalizing effect in the conditional wage distribution, i.e. for groups of individuals of similar characteristics. This unequalizing effect is especially strong for the Golden Triangle and Atlantic regions, where the returns to education increase substantially between the first and third quartiles.

However, interpreting conditional quantile regression results must be done cautiously. A common difficulty associated with interpreting these results is that, as has

---

<sup>6</sup> Estimates were also obtained for other percentiles. They are not reported here to save space but are available from the authors upon request.

already been mentioned, a given percentile of the unconditional distribution of wages may not be the same as the same percentile of the conditional distribution. Thus, the positive and heterogeneous CQR effects do not imply that education has a stronger effect for the highest wage earners, contributing therefore to increase inequality. Instead, it means that it has a stronger effect for the conditionally rich, that is, for those with the highest wages conditioned to the other covariates. As mentioned in section 3.1, in contrast to the CQR, the UQR allow studying the effects directly on the distribution of wages.

The estimates of the UQR are summarized in the last panel of columns in Table 4. They also reveal a heterogeneous pattern of the returns to schooling along the unconditional wage distribution, which is even more pronounced than that observed for the conditional distribution. The estimated return in the country as a whole is as low as 1.39% in the first quartile, increases to 3.74% in the middle of the distribution, and rises sharply in the third quartile, up to 12.54%. This means that the wage increase caused by an additional year of education in Colombia is ten-fold higher at the upper part of the wage distribution than at the bottom. In other words, education seems to contribute clearly to increase wage inequality. A similar pattern is observed in all regions, although the increase in the return is more pronounced in some regions and less in others. More precisely, the difference in returns along the distribution is more intense in the regions with the highest wage levels. They range from 1.18% to 16.17% in Golden Triangle, and from 0.87% to 13.19% in Atlantic. In contrast, the return is 4.19% in the first quartile and 8.99% in the third in the Pacific region. Another interesting feature derived from the UQR results is that in regions with low wage levels, the return to schooling at the middle of the distribution is similar, and even lower in the case of Pacific, to that at the bottom. Therefore, increasing education does not raise intra-regional inequality at the middle-bottom part of the distribution in low-wage regions. This is not observed for regions with higher wages, in which the return increases monotonically along the three quartiles.

As for the effect of informality along the conditional distribution (CQR), the pay penalty in the country as a whole decreases sharply from the first quartile (-19.27%) to the middle part (-8.91%) of the conditional distribution of wages. In turn, the reduction in wages associated to informal jobs is similar in the middle and at the upper part (-8.56%). The same pattern is observed in all five regions under analysis, although interesting variability in the strength of the effect of informality is observed. It is more intense in the



Pacific region, regardless of the point of the distribution in which it is measured, and it is also particularly high at the first and second quartiles of the Golden Triangle conditional distribution. In the other three regions, the effect of informality seems to be limited to the bottom part, since the wage penalty is moderate, and even not significant, at the middle and upper parts.

However, as mentioned before in the case of education, the interpretation of the informality pay penalty in the entire distribution of wages is more straightforward when based on the UQR results. They suggest that, in the country as a whole, working in the informal sector reduces wages of workers with the lowest earnings by almost 19%, while the reduction is of about 9% for workers with median wages. In turn, the pay penalty is only marginally significant for workers at the upper part of the distribution of wages. A similar pattern is observed in all regions with the exception of Pacific. In the Colombian region with the highest incidence of informality, the pay penalty is roughly similar all along the distribution of wages (about -30%). This means that in Pacific, informality reduces wages of workers with low, medium and high wage levels. It is also worthy to note that the pay penalty in that region is higher than in any other region all over the distribution. In all, the UQR estimates suggest that reducing informality would contribute to decrease within-region inequality, by increasing wages at the bottom and middle part of the distribution more than for workers with higher wages. The strength of this effect varies somehow across regions, being more intense in those in which informal workers are more abundant. The exception to this general pattern is Pacific, the region ranking first in incidence of informality that, in any case, is the region with the strongest effect of informality on wages.

Summing up, the estimates in this section confirm, on the one hand, the positive effect of education on wages, which increases along the wage distribution, and the existence of substantial regional variability in the returns to schooling. On the other hand, results confirm that workers face different informal pay penalties throughout the territory, that affects mostly individuals at the lower part of the wage distribution. This supports the hypothesis that regional differences in the effect of education may explain regional disparities mostly at the upper part of the wage distribution, whereas differences in the informal pay penalty would be behind those observed at the bottom of the distribution.

#### 4.2. *Decomposition of regional wage gaps*

The evidence presented so far confirms that regions not only differ in the endowment of earning relevant characteristics, such as education and the incidence of informality, but also points to sizeable regional variability in their returns. The contribution of this variability in characteristics and returns to the wage gap across regions is assessed next.

Following the method sketched in section 3.2, the decomposition of regional wage differentials in Colombia is analyzed by considering the difference between Golden Triangle, the region with the highest level of wages, and the other regions. The regional wage differentials relative to Golden Triangle for the mean and the quartiles are reported in the first row of information for each region in Table 5. It also contains the global decomposition, in which wage gaps are decomposed in two terms, one that accounts for the contribution attributable to differences in observable characteristics (labeled *Total explained by characteristics*) and another that corresponds to differences in the wage structure (labeled *Total wage structure*). Both of these two components can in turn be decomposed in the specific contribution of each factor that determine wages, by using the detailed decomposition. Given our goal in this paper, the details of the specific contribution of education and informality are presented in Table 5, while the contribution of the rest of observable characteristics have been grouped in the term labeled *rest*.

Wage differentials between Golden Triangle and each of the other four regions, calculated at the mean, are all statistically significant. The highest wage gap is found in the Pacific region, 36%, while the lowest one is that of Atlantic, 9%. Interestingly, differences in the size of the gap along the distribution are observed between Pacific and Oriental, the regions with lowest wage levels, and Atlantic and Central, which are the regions with wages close to those in Golden Triangle. Wage differentials follow a sort of U-shape in the case of the first two regions, whereas they increase monotonically over the distribution in the case of the latter group.<sup>7</sup> The decomposition shows that these two groups also differ in the origin of the gap. Results from the global decomposition reveal that 20.7 percentage points (pp) out of the 36.2pp of the mean gap for Pacific are attributable to differences in observed characteristics between this region and Golden Triangle. The contribution of this component is even larger in the case of the Oriental

---

<sup>7</sup> Motellón, López-Bazo and El-Attar (2011) found an increasing wage differential across the distribution for Spain and Pereira and Gallego (2014) found the same pattern for Portugal. Such a non-monotonic increase of the gap along the distribution is only observed for the most developed areas in Colombia.

region, where 17.1pp of the gap of 19pp correspond to differences in characteristics. The decomposition of the gap at the different quartiles indicates that the role of differences in characteristics is especially strong at the bottom and at the top of the distribution, particularly in Pacific.

In sharp contrast, the bulk of the wage gap between Atlantic and Central regions, and the Golden Triangle can not be explained by differences in observed characteristics. Only 3.6pp out of 12pp of the wage gap in the Central region is explained by characteristics. In the case of Atlantic, results even suggest that the average wage would have been higher than in Golden Triangle (by 2.1pp), if the two regions had had the same wage structure. The analysis of the global decomposition at the quartiles for these regions indicates that differences in wage structures widen the gap at the middle and, particularly, at the top of the wage distribution (0.4pp in the first versus 14.4pp in the last quartile in Atlantic, and 0.7pp and 13.8pp respectively in Central).

Therefore, the global decomposition reveals that the origin of the much lower wages in Pacific and Oriental is essentially on their lower endowment of characteristics that favor high wages, whereas wage differentials in regions with wage levels closer to those in Golden Triangle can be explained almost completely by differences in returns to characteristics (wage structure), which are higher in the benchmark region. The specific contribution of the two factors under analysis in this paper, education and informal work, is obtained from the results of the detailed gap decomposition, also included in Table 5. It is observed how differences in years of schooling and in the incidence of informality greatly contributed to widen the gap observed in the Pacific region. To be sure, 6.5pp of the mean wage gap between this region and Golden Triangle correspond to the higher level of education of the working population in the latter region, whereas differences in the share of informal work account for 6.6pp. A similar portion (6.7pp) is attributable to education in the case of the Oriental region, though the contribution of informality is lower in this case (1.7pp). As for the regions in which the gap is narrower, the contribution of differences in education and informality is much less intense. Actually, the better endowment of education in Atlantic with respect to Golden Triangle reduces the magnitude of the wage gap in 2.1pp.

Finally, it should be stressed that the detailed decomposition at the different quartiles supports one of the major hypothesis in this paper, which is that differences

across regions in the level of education provoke regional disparities at the upper part of the distribution, whereas differences in informality explain a big deal of the gap at the bottom. This feature is particularly intense in the regions with the widest wage gaps. In the Pacific region, differences in the endowment of education with respect to Golden Triangle account for 10.5pp of the gap at the third quartile and only 2.3pp at the first. The same applies to Oriental (11.3pp in the third and 1.4pp in the first quartile). Remarkably, differences in returns to schooling, reported in section 4.1, also contribute greatly to the wage gap at the upper part of the distribution in these two regions. Actually, the joint effect attributable to differences in the endowment and in the return to education (54.8pp in Pacific and 56.3pp in Oriental) exceeds by large the observed wage gap in the upper quartile, meaning that in the absence of other mechanisms, it would have been even wider in these two regions.

Regarding the effect of differences in the share of informal jobs, it is observed how it concentrates at the bottom part of the distribution in all regions, with almost no effect for median and top wages. In this respect, the results for the Pacific region are of particular interest, since it shows the highest incidence of informality and the widest wage gap among Colombian regions. One third of the wage gap at the first quartile in Pacific is explained by differences in informality between this region and Golden Triangle. In turn, the contribution of this component is a bit less than 5pp at the median, and negligible at the third quartile. In addition, the higher pay penalty suffered by informal workers in Pacific, in comparison with their counterparts in Golden Triangle, increases the wage gap by 11.7pp at the first quartile, but only 3.4pp at the median and in a non-significant amount at the third quartile. In all, the total effect linked to informality at the bottom quartile in Pacific amounts to as much as 28.2pp, which represents more than 56% of the gap for workers earning the lower wages.

Summing up, results of the gap decomposition confirm that differences across regions in both education and informality play a prominent role in explaining regional wage gaps. However, and beyond this general statement, the evidence reported in this section probes that the effect of differences in education on regional wage gaps is concentrated in the upper part of the wage distribution, whereas the one of informality basically affects workers at the bottom of the distribution.

## 5. Conclusions

Results in this paper confirmed that regional wage disparities, which vary over the wage distribution, exist in an emerging country such as Colombia. They also indicate that beyond differing in terms of the endowment of workers' education and the incidence of informality, regions show a remarkable dispersion on the wage effect of these two characteristics. Actually, the decomposition approach proposed by Firpo, Fortin and Lemieux (2009) has allowed us to confirm the main hypothesis in the paper, which was that spatial differences in education and informal work explain a big deal of regional wage disparities in Colombia. Also, the analysis in the entire wage distribution permitted us to probe that differences in education across Colombian regions account for gaps at the upper part of the wage distribution. In contrast, our results suggest that the effect of differences in informal work is limited to workers with medium and low wages.

The evidence from Colombia lead to the conclusion that policies aiming at stimulating investments in human capital in the less developed regions will help to decrease regional wage gaps, especially in the upper part of the wage distribution. However, equalizing years of education across regions would not be enough to reduce regional wage disparities due to the sizeable differences in returns to schooling at higher quantiles. Meanwhile, policies that point towards the reduction of informality will help to reduce regional wage gaps at the bottom part of the wage distribution, particularly for those regions with sizable informality. In addition, evidence has been obtained suggesting that improvements of the level of education will lead to increasing within-region inequality, due to the fact that the return is higher for high wage levels than for workers with medium and low wages. Interestingly, the lesson from the Colombian case is that successful policies to reduce informality in the labour market will contribute to narrowing regional wage gaps, particularly at the bottom of the distribution while, simultaneously, helps decreasing within-region inequality. This is so since the wage effect of decreasing informality is stronger for low than for high wage levels.

Finally, we must admit that our empirical analysis faces some potential caveats. It might be argued that some sources of bias in the estimates of the wage equations are likely to exist. One is related with the sample selection on wages caused by the probability of employment. It arises because some unobserved characteristics could be correlated with the likelihood of employment and wages. Another source of sample selection comes

from the probability of being a migrant, as for example, there could be unobservable spatial factors that affect the probability to migrate and correlate with wages. Although both sources of selection may lead to biased results, there are two reasons why they are not addressed in this study. The first one is that previous analyses that have controlled for employment selection in Colombia have found that the results are not strongly affected (Quiñones and Rodríguez, 2011). On the other hand, internal migration in Colombia has been found to be relatively low, so that this source of selection does not seem to be especially relevant (Ortiz, Uribe and Badillo, 2008).<sup>8</sup>

A last source of bias is the well-known endogeneity of the measure of education caused by unobserved characteristics, such as ability and quality of education, and/or measurement errors. As in previous studies, this problem is hard to address due to the lack of appropriate instruments in the dataset, and the impossibility to control for individual unobserved effects in a cross-section setting. Accordingly, one should be cautious in interpreting the estimates as causal effects. In any case, it is worth taking into account that most studies using estimation methods that account for endogeneity have provided estimates of the returns to schooling that exceed somewhat those obtained when no controlling for endogeneity. Therefore, we could expect an increase in the estimate of the return to education in all regions that, in any case, would not change dramatically the difference in the estimate between regions. Actually, the estimates in this paper are consistent with an explanation of the regional heterogeneity in the returns to schooling based on the effect of unobserved ability and quality of education. It is sensible thinking that the most productive and prosperous territories offer higher opportunities, and thus attract, the ablest individuals and also those whose education is of superior quality. If the wage effect of these unobserved characteristics is incorporated in the estimated return to schooling, one would expect higher estimated returns in the most developed regions, which is what our results reveal. In addition, the increasing estimate of the return to schooling along the conditional wage distribution that we have reported for all Colombian regions is consistent with this estimate incorporating the effect of unobserved ability and quality of education. This is so because of the traditionally assumed correspondence between the conditional distribution of wages and the distribution of these unobserved characteristics.

---

<sup>8</sup> See Pereira and Galego (2014) for similar arguments in the case of Portugal.

At any rate, we believe the magnitude of territorial disparities in the estimated effect of education and informality is large enough to allow us to conclude that they exert a substantial contribution in explaining regional wage gaps in an emerging country such as Colombia.

## References

- Andersson F., Burgess S. & Lane J.I. (2007). Cities, matching and the productivity gains of agglomeration. *Journal of Urban Economics*, 61, 112–128.
- Azzoni C. R. & Servo L.M. (2002). Education, cost of living and regional wage inequality in Brazil. *Papers in Regional Science*, 81, 157-75.
- Barsky R., Bound J., Charles K. & Lupton J. (2002). Accounting for the black-white wealth gap: A nonparametric approach. *Journal of the American Statistical Association*, 97, 663-673.
- Blackaby D. & Murphy P. (1995). Earnings, unemployment and Britain's North–South divide: real or imaginary? *Oxford Bulletin of Economics and Statistics*, 57, 487–512.
- Bonet J. & Meisel A. (2008). Regional economic disparities in Colombia. *Investigaciones Regionales*, 14, 61–80.
- Borah B. and Basu A. (2013). Highlighting differences between conditional and unconditional quantile regression approaches through an application to assess medication adherence. *Health Economics*, 22, 1052-1070.
- Combes P., Duranton G. & Gobillon L. (2008). Spatial wage disparities: sorting matters! *Journal of Urban Economics*, 63, 723–742.
- Dickey H. (2014) The impact of migration on regional wage inequality: A semiparametric approach. *Journal of Regional Science*, 54, 893–915.
- Duranton G. & Monastiriotis V. (2002). Mind the gaps: the evolution of regional inequalities in the UK, 1982–1997. *Journal of Regional Science*, 42, 219–256.
- Firpo S., Fortin M. & Lemieux T. (2009). Unconditional quantile regressions. *Econometrica*, 77, 953- 973.
- Fortin N.M., Lemieux T. & Firpo S. (2011). Decomposition methods in economics, in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Vol. IV, Part A, 1–102. Amsterdam: Elsevier.
- Galego A. & Pereira J. (2014). Decomposition of regional wage differences along the wage distribution in Portugal: the importance of covariates. *Environment and Planning A*, 46, 2514–2532.
- Galvis L. (2012). Informalidad laboral en las áreas urbanas de Colombia. *Coyuntura Económica*, 42, 15–51.
- García I. & Molina J. (2002). Inter-regional wage differentials in Spain. *Applied Economic Letters*, 9, 209–215.
- Gasparini L., Cruces G. & Tornarolli L. (2011). Recent trends in income inequality in Latin America. *Economía*, 11, 147–201.
- Glaeser E., Kallal H., Scheinkman J. & Shleifer A. (1992). Growth in cities. *Journal of Political Economy*, 100, 1126–1152.



- Greenwood M. J., Hunt, G. L., Rickman D. S. & Treyz G. I. (1991). Migration, regional equilibrium, and the estimation of compensating differentials. *The American Economic Review*, 81, 1382–1390.
- Joumard I. & Londoño Vélez J. (2013). Income inequality and poverty in Colombia - Part 1. The role of the labour market. *OECD Economics Department Working Papers*, No. 1036. Paris: OECD Publishing.
- Koenker R. (2005). *Quantile regression*. Econometric Society Monographs. Cambridge: Cambridge University Press.
- Koenker R. & Basset G. (1978). Regression quantiles. *Econometrica*, 46, 33–50.
- López-Bazo E. & Motellón E. (2012). Human capital and regional wage gaps. *Regional Studies*, 46, 1347–1365.
- López-Calva L.F. & Lustig N. (eds). (2010). Declining inequality in Latin America: A decade of progress? Washington DC: Brookings Institution Press.
- Machado J. & Mata J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, 20, 445–465.
- Melly B. (2005). Decomposition of differences in distribution using quantile regression. *Labour Economics*, 12, 577–590.
- Motellón E., López-Bazo E. & El-Attar M. (2011). Regional heterogeneity in wage distributions: Evidence from Spain. *Journal of Regional Science*, 51, 558–584.
- Pereira J. & Galego A. (2014). Inter-regional wage differentials in Portugal: An analysis across the wage distribution. *Regional Studies*, 48, 1529–1546.
- Ortiz C., Uribe J. & Badillo E. (2008). Segmentación inter e intrarregional en el mercado laboral urbano de Colombia, 2001-2006. *Ensayos sobre Política Económica*, 27, 194–231.
- Quiñones-Domínguez M. & Rodríguez-Sinisterra J. (2011). Rendimiento de la educación en las regiones colombianas: Un análisis usando la descomposición Oaxaca-Blinder. *Sociedad y Economía*, 20, 37–68.
- Romero J. (2008). Diferencias sociales y regionales en el ingreso laboral de las principales ciudades colombianas, 2001-2004. *Revista de Economía del Rosario*, 1, 165–201.

**Table 1.** Hourly wage, informality and human capital in the largest Colombian metro-areas.

	Number of Observations	Nominal Gross Hourly wage (pesos)	Adjusted Hourly wage (pesos)	Schooling (years)	Informality (%)
<b>Barranquilla</b>	1037	3663.16 (2947.25)	3510.73 (2824.61)	11.31 (3.45)	35.29
<b>Cartagena</b>	809	3760.54 (2518.59)	3605.99 (2415.08)	11.74 (3.44)	22.00
<b>Monteria</b>	759	3650.30 (3218.13)	3493.12 (3079.56)	11.26 (3.59)	36.89
<b>Cucuta</b>	754	2825.23 (1837.99)	2634.22 (1713.73)	9.39 (4.07)	59.15
<b>Bucaramanga</b>	988	3662.94 (2562.04)	3442.25 (2407.68)	10.65 (3.87)	31.88
<b>Villavicencio</b>	862	3306.05 (2464.41)	3141.81 (2341.98)	10.11 (3.48)	43.85
<b>Manizales</b>	1109	3506.84 (2680.53)	3402.62 (2600.87)	11.19 (3.74)	20.83
<b>Pereira</b>	1014	3351.98 (2547.55)	3230.37 (2455.12)	10.24 (3.89)	28.60
<b>Ibague</b>	869	3678.27 (2913.20)	3501.31 (2773.05)	11.06 (3.73)	36.02
<b>Pasto</b>	733	2981.61 (2668.21)	2885.20 (2581.93)	10.53 (4.14)	49.39
<b>Medellin</b>	1913	3903.84 (2904.72)	3718.43 (2766.76)	10.96 (3.76)	18.98
<b>Bogota</b>	1754	4305.70 (3566.44)	4132.05 (3422.61)	11.33 (3.96)	23.95
<b>Cali</b>	1195	3872.52 (3147.60)	3745.43 (3044.30)	10.68 (3.83)	28.62
<hr style="border-top: 1px dashed black;"/>					
<b>Colombia</b>	13796	3662.54 (2894.79)	3504.48 (2773.67)	10.86 (3.82)	31.05

**Note:** Sample means. Standard deviation is shown in parenthesis for continuous variables.

**Table 2.** Descriptive of *adjusted* hourly wages (pesos) in the five regions of Colombia.

	Average	Std. Dev. of Logs	Gini	Quartiles		
				25th	50th	75th
<b>Atlantic</b>	3535.18	0.57	0.33	2395.67	2617.42	3727.07
<b>Oriental</b>	3108.82	0.54	0.31	2000.76	2489.83	3321.36
<b>Central</b>	3372.9	0.54	0.32	2144.57	2467.86	3489.06
<b>Pacific</b>	2885.19	0.69	0.39	1458.48	2325.62	3010.51
<b>Golden Triangle</b>	3874.31	0.57	0.34	2384.57	2778.14	4167.22

**Table 3.** Descriptive of worker and firm characteristics.

	<b>Atlantic</b>	<b>Oriental</b>	<b>Central</b>	<b>Pacific</b>	<b>Golden Triangle</b>
<i>Adjusted Hourly Wage</i> (pesos)	3535.18	3108.82	3372.9	2885.19	3874.28
<b>Informal job</b>	0.32	0.44	0.28	0.49	0.23
<i>Worker's characteristics</i>					
<b>Schooling</b> (years)	11.43	10.10	10.83	10.53	11.03
<b>Experience</b> (years)	18.02	17.09	18.55	17.99	18.05
<b>Tenure</b> (months)	53.91	36.92	48.57	44.74	50.21
<b>Women</b>	0.39	0.43	0.42	0.43	0.45
<b>Married</b>	0.60	0.51	0.49	0.52	0.51
<b>Head of household</b>	0.43	0.40	0.43	0.43	0.44
<i>Type of contract</i>					
<b>No-contract</b>	0.25	0.44	0.26	0.43	0.23
<b>Temporary</b>	0.21	0.21	0.24	0.28	0.24
<b>Permanent</b>	0.54	0.36	0.50	0.29	0.52
<i>Firm size</i>					
<b>Micro</b>	0.27	0.44	0.32	0.50	0.28
<b>Small</b>	0.21	0.21	0.18	0.16	0.20
<b>Medium</b>	0.06	0.05	0.05	0.02	0.07
<b>Large</b>	0.46	0.31	0.44	0.32	0.45
<i>Sector</i>					
<b>Mining, electricity, gas and water</b>	0.04	0.04	0.03	0.02	0.02
<b>Industry</b>	0.21	0.19	0.24	0.16	0.26
<b>Construction</b>	0.05	0.11	0.08	0.07	0.06
<b>Sales, Hotels and Restaurants</b>	0.29	0.34	0.26	0.38	0.27
<b>Transportation</b>	0.10	0.08	0.10	0.08	0.07
<b>Financial Intermediation</b>	0.11	0.09	0.11	0.07	0.15
<b>Social Services</b>	0.20	0.16	0.18	0.22	0.17
<i>Observations</i>	2605	2604	2992	733	4862

**Notes:** Sample means. All figures in %, excepting those indicated in parenthesis.

**Table 4.** Estimated returns to education and informality for the five regions of Colombia.

	OLS	CQR			UQR		
		25th	50th	75th	25th	50th	75th
<i>Years of education</i>							
<b>Atlantic</b>	0.0826** [0.0028]	0.0553** [0.0020]	0.0697** [0.0029]	0.0873** [0.0035]	0.0087** [0.0012]	0.0435** [0.0025]	0.1319** [0.0056]
<b>Oriental</b>	0.0557** [0.0027]	0.0353** [0.0024]	0.0440** [0.0025]	0.0557** [0.0035]	0.0215** [0.0036]	0.0253** [0.0022]	0.0740** [0.0046]
<b>Central</b>	0.0752** [0.0024]	0.0412** [0.0016]	0.0569** [0.0023]	0.0779** [0.0034]	0.0214** [0.0024]	0.0306** [0.0016]	0.1148** [0.0048]
<b>Pacific</b>	0.0682** [0.0050]	0.0516** [0.0051]	0.0659** [0.0053]	0.0729** [0.0083]	0.0419** [0.0099]	0.0288** [0.0051]	0.0899** [0.0079]
<b>Golden Tr.</b>	0.0814** [0.0020]	0.0462** [0.0014]	0.0674** [0.0017]	0.0899** [0.0032]	0.0118** [0.0011]	0.0519** [0.0019]	0.1617** [0.0047]
<b>Colombia</b>	0.0742** [0.0012]	0.0460** [0.0009]	0.0597** [0.0011]	0.0778** [0.0020]	0.0139** [0.0009]	0.0374** [0.0010]	0.1254** [0.0024]
<i>Informality</i>							
<b>Atlantic</b>	-0.1023** [0.0257]	-0.1691** [0.0192]	-0.0435+ [0.0264]	-0.0475+ [0.0265]	-0.1137** [0.0138]	-0.0874** [0.0258]	-0.0472 [0.0525]
<b>Oriental</b>	-0.0991** [0.0257]	-0.1341** [0.0224]	-0.0516* [0.0234]	-0.0599* [0.0302]	-0.2710** [0.0355]	-0.0810** [0.0231]	0.0123 [0.0445]
<b>Central</b>	-0.0951** [0.0274]	-0.1704** [0.0183]	-0.0515* [0.0263]	0.0159 [0.0341]	-0.2389** [0.0326]	-0.0572** [0.0215]	0.0414 [0.0493]
<b>Pacific</b>	-0.2680** [0.0558]	-0.2959** [0.0573]	-0.2422** [0.0595]	-0.1737+ [0.0893]	-0.3085* [0.1200]	-0.3499** [0.0642]	-0.2939** [0.0868]
<b>Golden Tr.</b>	-0.1356** [0.0227]	-0.1662** [0.0169]	-0.1091** [0.0195]	-0.0726* [0.0298]	-0.1473** [0.0147]	-0.0470+ [0.0249]	-0.0215 [0.0487]
<b>Colombia</b>	-0.1430** [0.0125]	-0.1927** [0.0096]	-0.0891** [0.0116]	-0.0856** [0.0186]	-0.1881** [0.0109]	-0.0917** [0.0118]	-0.0471+ [0.0242]

**Notes:** experience (and its square), tenure (and its square), gender, marital status, head of household, hours worked, type of contract, size of the firm and firm sector are included as controls. Standard errors in []. +  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ . OLS refers to the ordinary least square estimates of the wage equation, whereas CQR and UQR denote the results of the conditional and unconditional quantile regressions, respectively.

**Table 5.** Regional wage gap decomposition.

	ATLANTIC				ORIENTAL			
	Mean	Quartiles			Mean	Quartiles		
		25th	50th	75th		25th	50th	75th
<b>Overall wage gap</b>	<b>0.087 **</b>	<b>0.006</b>	<b>0.068 **</b>	<b>0.114 **</b>	<b>0.190 **</b>	<b>0.187 **</b>	<b>0.118 **</b>	<b>0.238 **</b>
<i>Composition Effect attributable to</i>								
Education	-0.020 **	-0.002 *	-0.009 *	-0.033 *	0.067 **	0.014 **	0.045 **	0.113 **
Informality	0.012 **	0.013 **	0.007 **	-0.006	0.017 **	0.035 **	0.004	-0.018
Rest	-0.023 **	-0.008 **	-0.013 **	-0.017 +	0.090 **	0.057 **	0.075 **	0.114 **
Error	0.010	-0.002	-0.007	0.027	-0.004	0.058 **	-0.008	0.013
<b>Total explained by characteristics</b>	<b>-0.021 **</b>	<b>0.002</b>	<b>-0.022 *</b>	<b>-0.030</b>	<b>0.171 **</b>	<b>0.165 **</b>	<b>0.115 **</b>	<b>0.221 **</b>
<i>Wage structure effects attributable to</i>								
Education	0.052	0.048 **	0.190 **	0.378 **	0.100 **	-0.037 **	0.046	0.450 **
Informality	0.000	0.001	0.009	-0.022	-0.012	0.005	-0.006	-0.026
Rest	0.137	0.004	0.034	-0.200	0.139	0.174 **	0.118	-0.211
Constant	-0.063	-0.044	-0.130	0.025	-0.203 +	-0.118	-0.151	-0.188
Error	-0.018 +	-0.005	-0.013 *	-0.037 *	-0.004	0.024	0.007	-0.009
<b>Total wage structure</b>	<b>0.109 **</b>	<b>0.004</b>	<b>0.091 **</b>	<b>0.144 **</b>	<b>0.019</b>	<b>0.022</b>	<b>0.003</b>	<b>0.017</b>
	CENTRAL				PACIFIC			
	Mean	Quartiles			Mean	Quartiles		
		25th	50th	75th		25th	50th	75th
<b>Overall wage gap</b>	<b>0.119 **</b>	<b>0.111 **</b>	<b>0.127 **</b>	<b>0.189 **</b>	<b>0.362 **</b>	<b>0.499 **</b>	<b>0.180 **</b>	<b>0.334 **</b>
<i>Composition Effect attributable to</i>								
Education	0.021 **	0.005 **	0.010 **	0.035 **	0.065 **	0.023 **	0.040 **	0.105 **
Informality	0.005 **	0.010 **	0.003 *	-0.004	0.066 **	0.165 **	0.049 **	0.022
Rest	0.007	0.011 *	0.006	0.007	0.078 **	0.098 **	0.049 *	0.075 *
Error	0.003	0.017 *	0.021 **	0.012	-0.002	0.018	-0.091 **	0.080 +
<b>Total explained by characteristics</b>	<b>0.036 **</b>	<b>0.043 **</b>	<b>0.040 **</b>	<b>0.051 **</b>	<b>0.207 **</b>	<b>0.303 **</b>	<b>0.047 +</b>	<b>0.283 **</b>
<i>Wage structure effects attributable to</i>								
Education	0.029	-0.061 **	0.152 **	0.339 **	0.068	-0.169 *	0.054	0.443 **
Informality	-0.005	0.014 *	0.004	-0.025	0.028 +	0.117 **	0.034 *	0.016
Rest	0.025	0.222 **	0.005	-0.069	-0.093	0.319	-0.099	-0.231
Constant	0.042	-0.104	-0.069	-0.093	0.179	-0.063	0.164	-0.123
Error	-0.007	-0.002	-0.005	-0.014	-0.028 +	-0.008	-0.021 +	-0.054 +
<b>Total wage structure</b>	<b>0.083 **</b>	<b>0.069 **</b>	<b>0.087 **</b>	<b>0.138 **</b>	<b>0.155 **</b>	<b>0.196 **</b>	<b>0.132</b>	<b>0.051</b>

Notes: + p<0.1, \* p<0.05, \*\* p<0.01.