

Regional Wage Gaps, Education and Informality in an Emerging Country: The Case of Colombia

PAULA HERRERA-IDÁRRAGA, ENRIQUE LÓPEZ-BAZO and
ELISABET MOTELLÓN

(Received February 2015; accepted May 2016)

ABSTRACT *This paper uses Colombian micro-data to analyze the role of education and informality in regional wage differentials. The hypothesis is that apart from differences in the endowment of human capital, regional heterogeneity in the incidence of informality is another important source of regional wage inequality in emerging countries. This is confirmed by the evidence from Colombia, which also reveals remarkable spatial heterogeneity 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 at the bottom.*

L'écart des salaires régionaux, l'éducation et l'informalité dans un pays émergent: étude de cas de la Colombie

RÉSUMÉ *Cet article emploie des données microéconomiques auprès de la Colombie afin d'analyser le rôle de l'éducation et de l'informalité en matière de l'écart des salaires régionaux. L'hypothèse est que, hormis les différences de la dotation en capital humain, l'hétérogénéité régionale de la fréquence de l'informalité constitue un autre facteur explicatif de l'inégalité régionale des salaires dans les pays émergents. Cette hypothèse est confirmée par des résultats auprès de la Colombie, ce qui met en lumière aussi l'hétérogénéité spatiale exceptionnelle des salaires en fonction des caractéristiques de l'individu. Il s'avère que l'hétérogénéité régionale du rendement de l'éducation est particulièrement importante au niveau supérieur de la distribution des salaires. À son tour, l'hétérogénéité des pertes de salaires informelles est plus pertinente au niveau inférieur.*

Diferencias salariales por regiones, educación e informalidad en un país emergente: el caso de Colombia

RESUMEN *En este artículo utilizamos datos de Colombia a nivel micro para analizar el papel de la educación y la informalidad en las diferencias salariales por regiones. La hipótesis es que aparte de las*

Paula Herrera-Idárraga, Department of Economics, Pontificia Universidad Javeriana, Calle 40 N 6-23 Piso 7 Edificio Gabriel Giraldo, S.J., Bogota, Colombia. Email: pherrera@javeriana.edu.co (to whom correspondence should be sent). Enrique López-Bazo, AQR Research Group – University of Barcelona, Barcelona, Spain. Email: elopez@ub.edu. Elisabet Motellón, Universitat Oberta de Catalunya, Barcelona, Spain and AQR Research Group – University of Barcelona, Barcelona, Spain. Email: emotellon@uoc.edu.

diferencias en la dotación de capital humano, la heterogeneidad regional en la incidencia de la informalidad es otra fuente importante de desigualdad salarial por regiones en países emergentes. Esto se confirma con el ejemplo de Colombia, que también muestra una marcada heterogeneidad espacial de las características de la rentabilidad salarial para las personas. La heterogeneidad regional en la rentabilidad de la educación es especialmente intensa en la parte superior de la distribución de salarios. A su vez, la heterogeneidad en la penalización de pagos informales es más relevante en la parte inferior.

新兴国家的区域薪资落差, 教育与非正式性: 哥伦比亚的案例

摘要 本文运用哥伦比亚的微观数据, 分析教育和非正式性在区域薪资差异中所扮演的角色。本研究的假设是, 除了人力资本的天赋差异之外, 非正式性影响范围内的区域异质性, 是新兴国家区域薪资不均的另一个重要来源。哥伦比亚的证据证实了上述假说, 并同时揭露了个人特征的薪资回报中的巨大空间异质性。教育回报的区域异质性, 在薪资分佈的上层特别强烈。反之, 底层中非正式支薪损失的异质性则更具相关性。

KEYWORDS: *Wage differentials; quantile-based decompositions; informal jobs; economic development*

JEL CLASSIFICATION: *C21; J31; O17; R23*

1. Introduction

Over the past decade several studies have registered the decline in income inequality for Latin American countries (López-Calva & Lustig, 2010; Gasparini et al., 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 this part of the world. Analyses focusing on the regional dimension are of great relevance because even in the presence of declining income inequality at the national level, important interregional 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 American countries and faces large geographical inequalities.¹ 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 & Meisel, 2008; Joumard & Londoño, 2013). In particular, differences in wages deserve attention from a regional perspective as, for example, in 2010 the average gross hourly wage in a small city, such as Cucuta, was only 66% of that paid in Bogota, the country's capital.

This study pays special attention to spatial imbalances in the endowment of human capital, and to the extent that 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 also focuses on one important feature of almost all developing

and emerging countries: the large proportion of workers employed in informal jobs. In this regard, our hypothesis is that regional differences in the availability of good jobs that generate higher wages explain a large part of regional wage gaps in a developing country such as Colombia. This means that, apart from differences in the endowment of and return to human capital, 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. We estimate the return to education and the pay penalty of informal work for each Colombian region, controlling for the effect of a large set of other observed characteristics by using mean and quantile regression models. Afterwards, we use these estimates to decompose the regional wage gaps into the contribution of differences in the regional distribution of characteristics, and into the contribution of differences in wage structures. 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 et al. (2009) and Fortin et al. (2011). In contrast with other procedures (Machado & Mata, 2005; Melly, 2005), the latter approach allows us to isolate the particular contribution of education and informality to the regional wage gap at different quantiles. Galego & Pereira (2014) applied this method in their study of regional wage differentials in Portugal. As far as we know, this is the first application for the analysis of regional wage differentials of a developing country.

A comprehensive micro-level dataset for Colombia is exploited to obtain the empirical evidence in this paper. Therefore, the approach differs from that in other recent studies that have analysed variation in wages across the space using aggregate data for a set of spatial units (e.g., Fingleton & Palombi, 2016). In contrast with the focus of these aggregate studies on the regional determinants of wages (including spatial spillovers), the analysis in this paper concentrates on the role paid by individual-level determinants, particularly schooling and informal job. The results confirm previous evidence on the existence of significant regional wage differentials between the Golden Triangle region, demarcated by the cities of Cali, Medellín and Bogotá, and other regions in the country. Interestingly, they show that Colombian regions not only differ in earning-relevant characteristics, but also display sizeable regional variability in the returns to these characteristics. In particular, the return to education is far from regionally uniform and workers face rather variable informal pay penalties throughout the territory. Altogether, the evidence in this study points to the conclusion that some public policies aimed at reducing human capital differences among regions will help to decrease regional wage gaps, especially at the higher part of the wage distribution. However, equalizing years of education of workers across regions would not be enough to reduce regional wage disparities due to the sizeable differences in returns to education at higher quantiles. Meanwhile, policies that point towards the reduction of informality will help to reduce regional wage gaps at the lower end of the wage distribution, particularly for those regions with sizeable informality.

The paper is organized as follows. The next section briefly summarizes the previous related literature, and the third section presents a description of the data used. The methodology applied in this study is outlined in the fourth section, and the fifth section reports and discusses the results. Finally, conclusions are presented in the sixth section.

2. Brief Literature Review

To explain large spatial wage disparities, a number of arguments have been proposed. One emphasizes that wage disparities across areas are caused by differences in amenities. For instance, certain areas may have a favourable 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 et al., 2008). Workers with better labour market characteristics tend to sort into regions that concentrate industries with high skill requirements, where wages tend to be higher. Related to this last explanation, the third is based on agglomeration economies. A larger pool of highly skilled workers in an area may provide a source of important knowledge spillovers that can lead to productivity gains (Glaeser et al., 1992). Also, labour pooling improves the matching of firms and workers, which could also increase economic efficiency and lead to higher wages (Andersson et al., 2007).

A number of studies have been devoted to measuring the degree of regional wage gaps and identifying their origin. For instance, Blackaby & Murphy (1995), Duranton & Monastiriotis (2002), and Dickey (2007) analyse the case of Great Britain; García & Molina (2002), Motellón et al. (2011), and López-Bazo & Motellón (2012) that of Spain; and Galego & Pereira (2014) and Pereira & Galego (2014) that of Portugal.² These studies centre 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 differently these characteristics are rewarded (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 differs considerably across and within countries. Some studies conclude that regional wage differentials are mostly due to differences between regions in individual characteristics (Blackaby & Murphy, 1995). Other studies found that a significant part of wage differentials is explained by differences in returns (Motellón et al., 2011; Galego & Pereira, 2014; Pereira & Galego, 2014), while some others point out that both components play an important role (García & Molina, 2002). Less evidence is available on the amount and origin of regional disparities using individual-level data in the Latin American countries. Azzoni & 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. They found education as the most important variable for explaining regional wage disparities.

A distinctive feature of almost all developing and emerging countries is the large proportion of workers employed in informal jobs. Half of the employed population in almost all Latin American labour markets works in informal jobs (e.g., Bacchetta et al., 2009). Informal workers generally earn lower wages compared with similar formal workers, according to the traditional segmentation hypothesis (Fields, 1975), are excluded from the social security system and are

trapped in unproductive activities (ILO, 2011). 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%. Ortiz et al. (2008) suggested that the Colombian labour market is segmented in two dimensions. The first is 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 interregional segmentation, which is mainly due to the barriers of mobility of labour and other factors between regions. This double segmentation may be responsible for a great deal of the lower wage levels in regions more prone to generate informal jobs. As we mentioned above, this feature of the labour market of a developing country such as Colombia has been ignored so far in studies analysing regional wage disparities. Our analysis in the following section aims at filling this gap.

3. 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 Departamento Administrativo Nacional de Estadística (DANE), which is the governmental institute in charge of official statistics in Colombia.³ The survey gathers information about employment conditions for the population aged 12 years and above including income, occupation and industry sector at a two-digit level, in addition to the general population characteristics such as sex, age, marital status and educational attainment. The CHS is representative of the 13 major metropolitan areas in Colombia.

The analysis is restricted to salary workers not carrying formal studies, aged between 15 and 60 years, and who reported working more than 16 hours per week. We do not include self-employed workers and employers in the analysis because their source of income is a combination of labour and physical capital and, therefore, may not be compared with earnings of other employees. We also exclude public employees from the sample as their wages are fixed at the national level. After excluding observations with missing values or inconsistencies, 13,796 individuals remain in our sample.

As for the measure of wages, we combine 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 13 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. Averages of the *adjusted* gross hourly wages are shown in the third column of Table 1. It is observed that the position in the regional

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 (%)
Barranquilla	1,037	3,663.16 (2,947.25)	3,510.73 (2,824.61)	11.31 (3.45)	35.29 (0.015)
Cartagena	809	3,760.54 (2,518.59)	3,605.99 (2,415.08)	11.74 (3.44)	22.00 (0.015)
Monteria	759	3,650.30 (3,218.13)	3,493.12 (3,079.56)	11.26 (3.59)	36.89 (0.018)
Cucuta	754	2,825.23 (1,837.99)	2,634.22 (1,713.73)	9.39 (4.07)	59.15 (0.018)
Bucaramanga	988	3,662.94 (2,562.04)	3,442.25 (2,407.68)	10.65 (3.87)	31.88 (0.015)
Villavicencio	862	3,306.05 (2,464.41)	3,141.81 (2,341.98)	10.11 (3.48)	43.85 (0.017)
Manizales	1,109	3,506.84 (2,680.53)	3,402.62 (2,600.87)	11.19 (3.74)	20.83 (0.012)
Pereira	1,014	3,351.98 (2,547.55)	3,230.37 (2,455.12)	10.24 (3.89)	28.60 (0.014)
Ibague	869	3,678.27 (2,913.20)	3,501.31 (2,773.05)	11.06 (3.73)	36.02 (0.016)
Pasto	733	2,981.61 (2,668.21)	2,885.20 (2,581.93)	10.53 (4.14)	49.39 (0.018)
Medellin	1,913	3,903.84 (2,904.72)	3,718.43 (2,766.76)	10.96 (3.76)	18.98 (0.009)
Bogota	1,754	4,305.70 (3,566.44)	4,132.05 (3,422.61)	11.33 (3.96)	23.95 (0.010)
Cali	1,195	3,872.52 (3,147.60)	3,745.43 (3,044.30)	10.68 (3.83)	28.62 (0.013)
Colombia	13,796	3,662.54 (2,894.79)	3,504.48 (2,773.67)	10.86 (3.82)	31.05 (0.004)

Note: Sample means. Standard deviation is shown in parenthesis.

ranking of wages is roughly 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 fairly recent base year, 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.⁴

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 can be seen, there are notable differences in education. On average, workers in Cartagena have more than two years of education over 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 proportion of informal workers. Moreover, informal workers tend to be concentrated in less-productive occupations. They are more likely to be salesmen, shop assistants, cooks, waiters, bartenders, bricklayers, carpenters or other construction workers. These occupations account for almost 40% of all informal workers. In this paper we define workers as formal if they contribute both to health and old-age insurance. According to this legal definition, an informal job is an activity that is unregulated by the formal institutions and regulations of a country.⁵ Importantly, since data come 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 to 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. It is observed that 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. We grouped Bogota, Medellin and Cali into one region which we will refer to as the Golden Triangle.⁶ These metropolitan areas are the most productive and dynamic of the country. The most productive firms, most of research and development (R&D) investments and the most highly skilled workers are concentrated in these three areas. Besides the 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 Ibagu ), and Pacific, comprising only 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

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

	Average	SD of logs	Gini	Quartiles		
				25th	50th	75th
Atlantic	3,535.18	0.57	0.33	2,395.67	2,617.42	3,727.07
Oriental	3,108.82	0.54	0.31	2,000.76	2,489.83	3,321.36
Central	3,372.9	0.54	0.32	2,144.57	2,467.86	3,489.06
Pacific	2,885.19	0.69	0.39	1,458.48	2,325.62	3,010.51
Golden Triangle	3,874.31	0.57	0.34	2,384.57	2,778.14	4,167.22

differences is lower than that found for the 13 metropolitan areas. The average hourly wage of the region with the lowest level, Pacific, is 74% of that in the region with the highest level, the Golden Triangle. So, by grouping metropolitan areas into regions disparities are attenuated, but they still remain sizeable. 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 (standard deviation of the logarithm of wages and the Gini index). Interestingly, from the value of wages at the quartiles of the distribution (25th, 50th and 75th percentiles),⁷ 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 behaviour. 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 mentioned above, the spatial distribution of human capital, informality and other earning determinants are supposed to contribute to regional disparities in wages. A simple description of the amount of regional variability in 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. Interestingly, informality also differs considerably between regions, ranging from 49% in Pacific to 23% in the Golden Triangle. 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 with failures in

Table 3. Descriptive of worker and firm characteristics

	Atlantic	Oriental	Central	Pacific	Golden Triangle
<i>Adjusted Hourly Wage (pesos)</i>	3,535.18	3,108.82	3,372.9	2,885.19	3,874.28
Informal workers	0.32	0.44	0.28	0.49	0.23
<i>Workers' characteristics</i>					
Schooling (years)	11.43	10.10	10.83	10.53	11.03
Experience (years)	18.02	17.09	18.55	17.99	18.05
Tenure (months)	53.91	36.92	48.57	44.74	50.21
Women	0.39	0.43	0.42	0.43	0.45
Married	0.60	0.51	0.49	0.52	0.51
Head of household	0.43	0.40	0.43	0.43	0.44
<i>Type of contract</i>					
No contract	0.25	0.44	0.26	0.43	0.23
Temporary	0.21	0.21	0.24	0.28	0.24
Permanent	0.54	0.36	0.50	0.29	0.52
<i>Firm size</i>					
Micro	0.27	0.44	0.32	0.50	0.28
Small	0.21	0.21	0.18	0.16	0.20
Medium	0.06	0.05	0.05	0.02	0.07
Large	0.46	0.31	0.44	0.32	0.45
<i>Sector</i>					
Mining, electricity, gas and water	0.04	0.04	0.03	0.02	0.02
Industry	0.21	0.19	0.24	0.16	0.26
Construction	0.05	0.11	0.08	0.07	0.06
Sales, hotels and restaurants	0.29	0.34	0.26	0.38	0.27
Transportation	0.10	0.08	0.10	0.08	0.07
Financial intermediation	0.11	0.09	0.11	0.07	0.15
Social services	0.20	0.16	0.18	0.22	0.17
Observations	2,605	2,604	2,992	733	4,862

Note: Sample means. All figures are in percentages, excepting those indicated in parenthesis.

regional labour markets, as similar workers in comparable firms but 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.

4. Empirical Strategy

4.1. 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:

$$W_{ir} = X_{ir}\beta_r + \varepsilon_{ir} \quad (1)$$

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; and β_r is the vector of prices or returns at region r associated with the characteristics in X_{ir} . Equation (1) is estimated for each region, so that it is consistent with interregional segmentation, i.e., 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 effect 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 & Bassett (1978). If the conditional effect of a specific variable in X_r varies over the levels of other covariates in X_r , the estimator based on the CQR may be a consistent estimator of its conditional effect at the mean values of the other $k - 1$ remaining covariates, but it is not a consistent estimator of the effect on the unconditional wage distribution (e.g., Borah & Basu, 2013). It is possible to estimate the unconditional quantile effect of each variable in X_r using the approach proposed by Firpo et al. (2009) based on the influence function (IF) and recentred influence function (RIF). In the context of wages, the IF is:

$$\text{IF}(W_r; q_\tau) = \frac{(\tau - I\{W_r \leq q_\tau\})}{f_{W_r}(q_\tau)} \quad (2)$$

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

$$\text{RIF}(W_r; q_\tau) = q_\tau + \text{IF}(W_r; q_\tau) \quad (3)$$

Firpo et al. (2009) demonstrate that the implementation of the UQR is straightforward and similar to the ordinary least squares (OLS) regression. For a specific quantile τ , the first step is to estimate the RIF of the τ th quantile of W_r following equations (2) and (3). The second step is to run OLS regression of $\text{RIF}(W_{ir}; q_\tau)$ on the observed covariates, X_{ir} :

$$E[\text{RIF}(W_{ir}; q_\tau | X_{ir})] = X_{ir} \beta_\tau \quad (4)$$

where coefficients β_τ represent the approximate marginal effects of the explanatory variables on the unconditional quantile q_τ of wages for workers in region r .

It can be reasonably argued that two sources of bias may affect the estimation of the wage equations. One is related to sample selection due to the employment status. It arises because some unobserved characteristics could be correlated with the likelihood of both employment and wages. Another source of sample selection comes from the probability of being a migrant as long as unobservable factors that affect the probability to migrate correlate with wages. Although both sources of selection may lead to biased results, there are two reasons why they are not controlled for in the results reported in this study. The first is that previous analyses that have controlled for employment selection in Colombia have found that results are not strongly affected (Quiñones & Rodríguez, 2011). We obtain the same conclusion when the wage equations for the

Colombian macro-regions were estimated controlling for selection (results are available from the authors upon request). 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 et al., 2008).⁸ Also, the aim of this study is to analyse wage differences between workers in different regions rather than between workers in the formal and informal sectors in each region. That is the reason why the wage equation for each region is estimated pooling observations corresponding to both formal and informal workers, in a specification that includes a dummy variable to control for the type of job (formal or informal). This means that the strategy is different to that by, for instance, Christopoulou and Monastiriotis (2014), as they estimate separate wage equations for the two groups of workers under analysis: private and public sectors workers. It is the distinction between the two groups of workers that forces them to control for endogenous classification in each sector by means of an endogenous switching regression model. In contrast, in our case it makes no sense to control for selection (into the formal or informal sectors) by estimating an endogenous switching model, because this will force us to analyse separately regional wage disparities of formal and informal workers. Results on the contribution of regional differences in the incidence of informality could not be obtained in that case.⁹

Another 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 not 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 estimate of differences between regions.

Finally, it should be noted that the identification of the individual effect on wages of education and informality requires that they be not highly correlated. We explored this issue and realized that the degree of association between our measures of education and informality is moderate in the sample of workers in each region. The correlation coefficients are not high; they range from 0.31 to 0.43. In addition, we observed that the mapping between the endowment of education and the incidence of informality at the regional level is far from perfect. To do so, we compared the two aggregate magnitudes for the Colombian cities, obtaining that informality was high in cities with a small endowment of education and also in cities in which the endowment of education was large. Therefore, it can be concluded that the amount of overlapping in the information contained in the two magnitudes is not very high. This evidence, combined with the fact that we use large sample sizes (which means high sample variability) to estimate the wage equations, lead us to conclude that our estimates should be providing precise enough estimates of the effects of education and informality.¹⁰

4.2. Decomposition of Regional Wage Gaps

The Blinder–Oaxaca method uses the estimation of the wage equation in equation (1) to decompose the mean wage gap between a high- ($r = h$) and a low-wage region ($r = l$) as follows:

$$\overline{W}_h - \overline{W}_l = (\bar{X}_h - \bar{X}_l)\hat{\beta}_h - \bar{X}_l(\hat{\beta}_h - \hat{\beta}_l) \quad (5)$$

The first term in the right-hand side 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.¹¹

Similarly, Firpo et al. (2009) proposed a decomposition of the wage differential at quantile τ using the RIF regression estimates. Any distributional parameter, such as 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 τ between a high-wage region and a low-wage region, Δ^{q_τ} , can be written as:

$$\Delta^{q_\tau} = q_\tau(F_{w_h|r=h}) - q_\tau(F_{w_l|r=l})$$

$$\Delta^{q_\tau} = [q_\tau(F_{w_h|r=h}) - q_\tau(F_{w_l|r=h})] + [q_\tau(F_{w_l|r=h}) - q_\tau(F_{w_l|r=l})] \quad (6)$$

$$\Delta^{q_\tau} = \Delta_S^{q_\tau} + \Delta_X^{q_\tau}$$

where $q_\tau(F_{w_h|r=h})$ indicates the actual wage quantile of workers belonging and rewarded under the wage structure of region $r = h$; and $q_\tau(F_{w_l|r=h})$ represents the counterfactual wage quantile, that is the wage quantile that would prevail 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, Δ^{q_τ} , 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, 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 et al., 2011). First, a reweighting factor is calculated as:

$$\Psi(X) = \frac{\Pr(r = h|X)/\Pr(r = h)}{\Pr(r = l|X)/\Pr(r = l)} \quad (7)$$

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

$$\hat{\Delta}^{q_\tau} = (\bar{X}_h\hat{\beta}_{\pi h} - \bar{X}_l^c\hat{\beta}_{\pi l^c}) + (\bar{X}_l^c\hat{\beta}_{\pi l^c} - \bar{X}_l\hat{\beta}_{\pi l})$$

$$\hat{\Delta}^{q_\tau} = \hat{\Delta}_S^{q_\tau} + \hat{\Delta}_X^{q_\tau} \quad (8)$$

where \bar{X}_r denotes the vector of means in region r ($= l$ and h); and \bar{X}_l^c are the counterfactual means for region l using the reweighting factor in (7) to make the distribution of the characteristics, X , in the region with low wages similar to that of the region with high wages.

The wage structure effect can be divided into a pure wage structure effect and a reweighting error:

$$\begin{aligned}\hat{\Delta}_S^{q\tau} &= \bar{X}_h(\hat{\beta}_{\pi h} - \hat{\beta}_{\pi}^c) + (\bar{X}_h - \bar{X}_l^c)\hat{\beta}_{\pi}^c \\ \hat{\Delta}_S^{q\tau} &= \hat{\Delta}_{S,p}^{q\tau} + \hat{\Delta}_{S,e}^{q\tau}\end{aligned}\quad (9)$$

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 specification error (which should be zero as long as the model is linear):

$$\begin{aligned}\hat{\Delta}_X^{q\tau} &= (\bar{X}_l^c - \bar{X}_l)\hat{\beta}_{\pi}^c + \bar{X}_l(\hat{\beta}_{\pi}^c - \hat{\beta}_{\pi}) \\ \hat{\Delta}_X^{q\tau} &= \hat{\Delta}_{X,p}^{q\tau} + \hat{\Delta}_{X,e}^{q\tau}\end{aligned}\quad (10)$$

The pure wage structure and composition effects in equations (9) and (10) can be decomposed into the contribution corresponding to each characteristic. In particular, our empirical exercise isolates the specific contribution of education and informality.

The procedure for decomposing the regional wage gaps entails two assumptions. First, the similarity of the conditional distribution of unobservable characteristics affecting wages in the corresponding regions and, therefore, that neither the wage structure nor the composition component is confounded by differences between regions in the conditional distributions of unobservables. Second, the overlapping of the distribution of the observable characteristics for the two regions for which the gap is decomposed (i.e., existence of a common support). As indicated by Nicolici et al. (2014), the main shortcoming of the approach is that it provides only local linear approximations to the effects of (possibly large) changes in characteristics and returns.

5. Results

5.1. Ordinary Least Squares (OLS) and Quantile Regression Estimates

Table 4 reports estimates of the effect of education and informal work at the mean (OLS) and at the quartiles for the five regions and Colombia as a whole.¹² All the variables described in Table 3 were included as control variables and the corresponding estimated coefficients are in Table A1 in the supplemental data online. The OLS estimates of the return to schooling for each region and the entire country are displayed in the first column of the upper panel. Results confirm that for Colombia as a whole, investments in education are quite profitable, since the estimated return is 7.42%, and highly significant. This is also so in the five regions under analysis, although, as expected, there are significant differences across them in the return to years of education. A higher return to schooling is observed in those regions with the highest level of wages: 8.14% and 8.26% in Atlantic and the Golden Triangle, respectively. In turn, regions with the lowest levels of hourly wages display the lowest returns to schooling: 5.57% in Oriental and 6.82% in Pacific. Differences across regions in the return to schooling are confirmed by the result of a Wald test. The value of the statistic, displayed in Table 4, leads to the clear rejection of the null hypothesis of equality of returns. Thus, in addition to differences in the endowment of education, returns to schooling may be thought to be an important factor in explaining regional wage gaps.

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

		UQR		
	OLS	25th	50th	75th
<i>Years of education</i>				
Atlantic	0.0826** [0.0028]	0.0087** [0.0012]	0.0435** [0.0025]	0.1319** [0.0056]
Oriental	0.0557** [0.0027]	0.0215** [0.0036]	0.0253** [0.0022]	0.0740** [0.0046]
Central	0.0752** [0.0024]	0.0214** [0.0024]	0.0306** [0.0016]	0.1148** [0.0048]
Pacific	0.0682** [0.0050]	0.0419** [0.0099]	0.0288** [0.0051]	0.0899** [0.0079]
Golden Tr.	0.0814** [0.0020]	0.0118** [0.0011]	0.0519** [0.0019]	0.1617** [0.0047]
Colombia	0.0742** [0.0012]	0.0139** [0.0009]	0.0374** [0.0010]	0.1254** [0.0024]
Test of equal effects	16.87** (0.0000)	6.27** (0.0000)	5.18** (0.0004)	18.34** (0.0000)
<i>Informal</i>				
Atlantic	-0.1023** [0.0257]	-0.1137** [0.0138]	-0.0874** [0.0258]	-0.0472 [0.0525]
Oriental	-0.0991** [0.0257]	-0.2710** [0.0355]	-0.0810** [0.0231]	0.0123 [0.0445]
Central	-0.0951** [0.0274]	-0.2389** [0.0326]	-0.0572** [0.0215]	0.0414 [0.0493]
Pacific	-0.2680** [0.0558]	-0.3085* [0.1200]	-0.3499** [0.0642]	-0.2939** [0.0868]
Golden Tr.	-0.1356** [0.0227]	-0.1473** [0.0147]	-0.0470 ⁺ [0.0249]	-0.0215 [0.0487]
Colombia	-0.1430** [0.0125]	-0.1881** [0.0109]	-0.0917** [0.0118]	-0.0471 ⁺ [0.0242]
Test of equal effects	2.35 ⁺ (0.0522)	2.42* (0.0461)	1.39 (0.2346)	2.55* (0.0373)

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 are shown in brackets. OLS refers to the ordinary least square estimates of the wage equation; UQR denotes the results of the unconditional quantile regressions. Test of equal effects is the result of a Wald test for the hypothesis that returns to years of education or pay penalty of informality is the same in all Colombian regions. The corresponding *p*-value is shown in parentheses.

⁺*p* < 0.1; **p* < 0.05; ***p* < 0.01.

The OLS estimates of the informal pay penalty, reported in the first column of the lower panel of Table 4, show a more complex pattern. The Pacific region, which has the lowest wage level, has 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, the 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 with the region with the highest, there seems to be no 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 but also in the mean wages earned by otherwise similar formal and

informal workers. This is confirmed by the result of the Wald test reported in [Table 4](#). Even though the null hypothesis of equality in the effect of informality is rejected with a p -value slightly above 5%, pairwise tests reveal that the difference is highly significant between the effect in Pacific and the other Colombian regions.

The UQR estimates of the returns to schooling and the informality pay penalty are summarized in the last three columns in [Table 4](#). They reveal a heterogeneous pattern in the return to schooling along the unconditional wage 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 10-fold higher at the upper part of the wage distribution than at the bottom. In other words, education clearly seems to contribute to increasing wage inequality. A similar pattern is observed in all regions, although the increase in the return is more pronounced in some and less in others. Another interesting feature derived from the UQR results is that 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, in regions with low wage levels. Therefore, increasing education does not raise intra-regional inequality at the middle-bottom part of the distribution in low-wage regions. This feature is not observed in regions with high wages, in which the return increases monotonically along the three quartiles.

As for the effect of informality along the unconditional distribution, the UQR estimates 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 about 9% for workers with median wages. In turn, the pay penalty is only marginally significant for workers at the upper end 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. This means that in Pacific, informality reduces wages of workers with low, medium and high wage levels by about 30%. Note too that the pay penalty in that region is higher than in any other region all over the distribution. Altogether, the UQR estimates suggest that reducing informality would contribute to decreased 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 somewhat 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.

The effect of education and informality varies significantly across regions in each of the quartiles. The results of Wald tests for each quartile based on the UQR regression estimates clearly lead to the rejection of the null hypothesis of equality of the returns to schooling. The same applies to the effect of informality at the highest and lowest quartiles, but not at the median. Nonetheless, results of the Wald test reported in [Table 4](#) correspond to the UQR estimation that pools observations for all regions, which means that they correspond to the effects on the unconditional wage distribution in the whole country rather than in each particular region. In other words, the result of the Wald statistic in each quartile may be seen as a lower bound of a test of equality of the effects based on the UQR estimates.

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 corroborate that workers face different informal pay penalties throughout the territory, which affects mostly individuals with low wages (which is consistent with the fact that informal workers concentrate at the bottom of the wage distribution).¹³ 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.

5.2. Decomposition of Regional Wage Gaps

Following the method sketched in section 4.2, the decomposition of regional wage differentials in Colombia is analysed by considering the difference between the Golden Triangle, the region with the highest level of wages, and the other regions. The regional wage differentials relative to the Golden Triangle for the mean and the quartiles are reported in the first row of information for each region in Table 5, followed by the results of the global and detailed decompositions. Given our goal in this paper, the details of the specific contribution of education and informality are presented in Table 5, while the contributions of the other observable characteristics have been grouped in the term labelled *rest*.

Wage differentials between the Golden Triangle and each of the other four regions, calculated at the mean, are all statistically significant. The highest wage gap (measured in logarithmic points) is found in the Pacific region, while the lowest is that of Atlantic. 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 the 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.¹⁴ The decomposition shows that these two groups also differ in the origin of the gap. Results from the global decomposition reveal that 0.207 logarithmic points (l.p.) out of the 0.362 l.p. of the mean gap for Pacific are attributable to differences in observed characteristics between this region and the Golden Triangle. The contribution of this component is even larger in the case of Oriental, where 0.171 l.p. out of 0.19 l.p. 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, and the Golden Triangle cannot be explained by differences in observed characteristics. Only 0.036 l.p. out of 0.119 l.p. of the wage gap in the Central region are explained by characteristics. In the case of Atlantic, results even suggest that the average wage would have been higher than in the Golden Triangle (by 0.021 l.p.) 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 (0.004 l.p. in the first versus 0.144 l.p. in the last quartile in Atlantic, and 0.069 l.p. and 0.138 l.p. respectively in Central). Therefore, the global decomposition reveals

Table 5. Regional wage gap decomposition

	Atlantic				Oriental			
	Mean	Quartiles			Mean	Quartiles		
		25th	50th	75th		25th	50th	75th
Overall wage gap	0.087**	0.006	0.068**	0.114**	0.190**	0.187**	0.118**	0.238**
<i>Composition effect attributable to ...</i>								
Education	-0.020**	-0.002*	-0.009*	-0.033*	0.067**	0.014**	0.045**	0.113**
Informal	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
Total explained by characteristics	-0.021**	0.002	-0.022*	-0.030	0.171**	0.165**	0.115**	0.221**
<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
Total wage structure	0.109**	0.004	0.091**	0.144**	0.019	0.022	0.003	0.017
	Central				Pacific			
	Mean	Quartiles			Mean	Quartiles		
		25th	50th	75th		25th	50th	75th
Overall wage gap	0.119**	0.111**	0.127**	0.189**	0.362**	0.499**	0.180**	0.334**
<i>Composition effect attributable to ...</i>								
Education	0.021**	0.005**	0.010**	0.035**	0.065**	0.023**	0.040**	0.105**
Informal	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
Total explained by characteristics	0.036**	0.043**	0.040**	0.051**	0.207**	0.303**	0.047 ⁺	0.283**
<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 ⁺
Total wage structure	0.083**	0.069**	0.087**	0.138**	0.155**	0.196**	0.132	0.051

Note: ⁺ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$.

that the origin of the much lower wages in Pacific and Oriental is essentially on their lower endowment of characteristics that favour high wages, whereas wage differentials in regions with wage levels closer to those in the Golden Triangle can be explained almost completely by differences in returns to characteristics (wage structure), which are higher in the benchmark region.

From the detailed decomposition, it is observed that differences in years of schooling and in the incidence of informality greatly contributed to widen the gap in Pacific. To be sure, 0.065 l.p. of the mean wage gap between this region and the Golden Triangle corresponds to the higher level of education of the working population in the latter region, whereas differences in the share of informal work account for 0.066 l.p. A similar portion (0.067 l.p.) is attributable to education in the case of Oriental, though the contribution of informality is lower in this case (0.017 l.p.). As for the regions in which the gap is narrower, the contribution of differences in education and informality is less intense. Actually, the better endowment of education in Atlantic with respect to the Golden Triangle reduces the magnitude of the wage gap by 0.021 l.p.

The detailed decomposition at the different quartiles supports one of the major hypotheses in this paper, which is that differences across regions in the level of education provoke regional disparities at the upper end of the distribution, whereas differences in informality explain a large part of the gap at the bottom. This feature is particularly intense in the regions with the widest wage gaps. In Pacific, differences in the endowment of education with respect to the Golden Triangle account for 0.105 l.p. of the gap at the third quartile and only 0.023 l.p. at the first. The same applies to Oriental (0.113 l.p. in the third and 0.014 l.p. in the first quartile). Remarkably, differences in returns to schooling, reported in section 5.1, also contribute greatly to the wage gap at the upper part of the distribution in these two regions. The joint effect attributable to differences in the endowment and in the return to education (0.548 l.p. in Pacific and 0.563 l.p. in Oriental) exceeds by a large amount 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 end of the distribution in all regions, with almost no effect for median and top wages. In this respect, the results for Pacific are of particular interest, since they show 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 the Golden Triangle (the part explained by differences in the share of informal workers is 0.165, which corresponds to 33% of 0.499). In turn, the contribution of this component is a bit less than 0.05 l.p. 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 the Golden Triangle, increases the wage gap by 0.117 l.p. at the first quartile, but only 0.034 l.p. at the median and by a non-significant amount at the third quartile. Overall, the total effect linked to informality at the bottom quartile in Pacific amounts to as much as 0.282 l.p., which represents more than 56% of the gap for workers earning the lower wages.

Finally, Table 5 also reports the contribution of the reweighting and specification errors. The former is small in magnitude and not significant in most cases. As for the latter, a significant contribution of the departure from linearity is only observed

in a few cases: the lowest quartile in Oriental, and the median in Central and Pacific. In these cases, the contribution of the specification error is intense in magnitude, exceeding even that of education and informality. Therefore, although there seems not to be a systematic pattern in this error term, one should accept with some caution the specific results for the above-mentioned cases.

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 that of informality basically affects workers at the bottom.

5.3. Results by Gender

The analysis so far did not distinguish by gender. However, it is known that women and men behave differently in the labour market. In particular, the international evidence suggests that the return to education is higher for female than for male workers (e.g., Psacharopoulos & Patrinos, 2004), while it is also likely that the incidence of informality varies between women and men.¹⁵ Therefore, it is worthwhile checking if the evidence derived so far for the whole sample of workers also applies for both genders. Note that we are interested in ascertaining whether the features described in the former case also apply to the latter, rather than in testing for gender differences within each region. To save space, we only report results of the wage gap decomposition for women and men. Those corresponding to the estimates of the returns to schooling and the informality pay penalty are available from the authors upon request. In any case, consistent with the previous literature, a higher return to education is estimated for females in all regions—in the mean and in all the quartiles. Interestingly, the pay penalty of informality is, in general, also higher for females.

Results in Table 6 show that the main features detected for the entire sample are also observed when the regional wage gap decompositions are computed separately for female and male workers. In the first place, the prominent role of the contribution of differences in characteristics in Pacific and Oriental is common to both genders. However, this component cannot explain the bulk of the wage gap in Atlantic and Central, for neither females nor males. Therefore, similar to that described in the previous section for the entire sample, the evidence suggests that, for both females and males, the gap in the regions with the lowest wage levels, Pacific and Oriental, is the consequence of their poorer endowment of characteristics that favour higher wages. By contrast, the difference in the wage structure accounts for most of the gap of women and men in the regions with wage levels closer to those in the Golden Triangle (Atlantic and Central).

As for the particular contribution of education and informality, beyond some specific differences between the two genders, results in Table 6 are also consistent with those commented for the entire sample. The contribution of the endowment of education is stronger in the upper part of the distribution for both genders, though particularly so in the case of men. In turn, the contribution of informality is concentrated at the bottom for females and males. Finally, differences in the return to education for males exert a large effect at the top of their wage distribution in Pacific and Oriental. The contribution of this component is even stronger for

Table 6. Regional wage gap decomposition by gender

	Men				Women			
	Mean	Quartiles			Mean	Quartiles		
		25th	50th	75th		25th	50th	75th
Atlantic								
Overall wage gap	0.089**	0.027**	0.092**	0.108**	0.083**	−0.003	0.032 ⁺	0.137**
<i>Composition effect attributable to ...</i>								
Education	−0.030**	−0.003**	−0.010**	−0.038**	−0.033*	−0.004*	−0.014*	−0.059*
Informality	0.014**	0.014**	0.007**	0.000	0.009*	0.012**	0.005*	−0.004
Rest	−0.023*	−0.013**	−0.010**	−0.006	−0.013	−0.007	−0.015**	−0.018
Error	0.006	−0.004	−0.029**	−0.027	0.008	−0.001	−0.029*	0.074*
Total explained by characteristics	−0.033**	−0.006	−0.042**	−0.070**	−0.029**	−0.001	−0.053**	−0.006**
<i>Wage structure effects attributable to ...</i>								
Education	0.042	0.042 ⁺	0.220**	0.513**	0.039	0.039	0.238**	0.299*
Informality	0.013	0.007	0.016 ⁺	0.012	−0.007	−0.004	0.001	−0.037
Rest	0.084	0.037	−0.038	−0.278	0.356*	0.040	0.064	0.283
Constant	−0.008	−0.051	−0.059	−0.052	−0.267	−0.075	−0.214	−0.386
Error	−0.008	−0.003	−0.005	−0.017	−0.009	−0.003	−0.006	−0.016
Total wage structure	0.123**	0.033**	0.134**	0.178**	0.112**	−0.002	0.084**	0.143**
Oriental								
Overall wage gap	0.179**	0.148**	0.119**	0.219**	0.204**	0.260**	0.123**	0.240**
<i>Composition effect attributable to ...</i>								
Education	0.068**	0.018**	0.044**	0.123**	0.053**	0.010**	0.035**	0.087**
Informality	0.019*	0.028**	0.004	−0.011	0.015*	0.046**	0.007	−0.024 ⁺
Rest	0.095**	0.057**	0.080**	0.139**	0.091**	0.063**	0.065**	0.100**
Error	−0.004	0.016	−0.015	−0.001	−0.003	0.122**	0.006	0.013
Total explained by characteristics	0.178**	0.119**	0.113**	0.250**	0.156**	0.241**	0.113**	0.176**
<i>Wage structure effects attributable to ...</i>								
Education	0.139**	−0.047	0.057	0.284**	0.013	−0.032	−0.018	0.425**
Informality	−0.005	−0.003	0.001	0.000	−0.017	0.022*	−0.007	−0.058*
Rest	0.164	0.212*	0.208	−0.280	0.039	0.057	−0.051	0.029
Constant	−0.301*	−0.133	−0.263 ⁺	−0.044	0.023	−0.024	0.093	−0.315
Error	0.004	0.000	0.002	0.008	−0.009	−0.004	−0.008	−0.017
Total wage structure	0.001	0.029**	0.005	−0.031	0.049**	0.019**	0.010**	0.064**

(Continued)

Table 6. Continued.

	Men				Women			
	Mean	Quartiles			Mean	Quartiles		
		25th	50th	75th		25th	50th	75th
Central								
Overall wage gap	0.123**	0.114**	0.092**	0.176**	0.114**	0.111**	0.093**	0.170**
<i>Composition effect attributable to ...</i>								
Education	0.028**	0.005*	0.011**	0.037**	-0.007	-0.002	-0.003	-0.013
Informality	0.006*	0.010**	0.004*	-0.005	0.002	0.005	0.001	-0.002
Rest	0.005	0.018**	0.006	0.005	0.011	0.004	0.002	0.018
Error	0.002	0.017	-0.029**	0.007	0.004	0.007	0.002	0.010
Total explained by characteristics	0.041**	0.050**	-0.007**	0.044*	0.010	0.015	0.002	0.013
<i>Wage structure effects attributable to ...</i>								
Education	-0.024	-0.031	0.144**	0.368**	0.094	-0.114*	0.177**	0.287*
Informality	-0.003	0.007	0.012	-0.008	-0.010	0.019 ⁺	-0.011	-0.050
Rest	0.045	0.205**	0.088	-0.075	0.045	0.237*	-0.143	0.163
Constant	0.062	-0.119	-0.146	-0.154	-0.023	-0.040	0.069	-0.236
Error	0.002	0.001	0.002	0.001	-0.002	0.001	-0.001	-0.008
Total wage structure	0.082**	0.064**	0.099**	0.132**	0.104**	0.103**	0.092**	0.157**
Pacific								
Overall wage gap	0.386**	0.507**	0.218**	0.346**	0.329**	0.461**	0.138**	0.269**
<i>Composition effect attributable to ...</i>								
Education	0.057*	0.008	0.020*	0.088*	0.042 ⁺	0.010	0.028 ⁺	0.062 ⁺
Informality	0.091**	0.180**	0.076**	0.082	0.048*	0.124**	0.053*	-0.002
Rest	0.054 ⁺	0.042	0.061*	0.023	0.095**	0.081**	0.032	0.080 ⁺
Error	-0.008	0.051	0.026	0.038	0.004	0.098 ⁺	-0.102**	0.106
Total explained by characteristics	0.194*	0.281**	0.182**	0.231**	0.189**	0.313**	0.011*	0.246
<i>Wage structure effects attributable to ...</i>								
Education	-0.043	-0.001	0.191**	0.169	0.212*	-0.032	0.107	0.904**
Informality	0.061*	0.139**	0.069**	0.091 ⁺	0.012	0.085**	0.037 ⁺	-0.025
Rest	-0.143	0.541*	0.522**	-0.183	0.208	0.076	-0.361	0.311
Constant	0.316	-0.456	-0.745**	0.039	-0.265	0.031	0.362	-1.124*
Error	0.002	0.002	0.001	0.000	-0.028	-0.013	-0.019	-0.043
Total wage structure	0.193**	0.226**	0.036**	0.116**	0.139**	0.147**	0.127**	0.023**

Note: ⁺ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$.

female workers in these poor regions. The same applies to the case of Atlantic and Central, although in this case the contribution is stronger for males.

All in all, the analysis of the regional wage gaps separately for women and men supports the role played by differences in education and informality. It also confirms that the contribution of these two characteristics is concentrated in different parts of the wage distribution.

6. Conclusions

We have reported evidence that confirms that spatial differences in education and informal work explain a large part of regional wage disparities in an emerging country such as Colombia. The analysis of the entire wage distribution permitted us to probe that differences in education across Colombian regions account for gaps at the upper end of the wage distribution. Meanwhile, the evidence supports that the effect of differences in informal work is limited to workers with medium and low wages. We must admit that results 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 to think that the most productive and prosperous territories offer greater 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 into the estimated return to schooling, one would expect higher estimated returns in the most developed regions, which is what our results reveal. 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.

The evidence from Colombia leads 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 large 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 end of the wage distribution, particularly for those regions with sizeable informality. In addition, evidence has been obtained suggesting that improvements in 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 helping to decrease within-region inequality. This is so since the wage effect of decreasing informality is stronger for low than for high wage levels.

Funding

The authors gratefully acknowledge financial support from the Ministerio de Ciencia e Innovación, Plan Nacional de I+D+I [project number ECO2011-30260-C03-03], and the Seventh Framework Programme [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 number FI-DGR 2012] and from Pontificia Universidad Javeriana.

Disclosure statement

No potential conflict of interest was reported by the authors.

Supplemental Data

Supplemental data for this article can be accessed at [10.1080/17421772.2016.1190462](https://doi.org/10.1080/17421772.2016.1190462).

Notes

1. The strong performance of Colombia's economy in the last decade has contributed to a decline in unemployment, informality, inequality and poverty. During most of the 2000s, Colombia benefited from growing external demand, high commodity prices and abundant foreign direct investment (FDI). Confidence in the government's macroeconomic policies and the improvement in the security situation led to an impressive rise in private investment in general, and in FDI in particular. However, as indicated in World Bank (2015), 'the country faces considerable challenges to boost and sustain growth: (i) access to finance for firms and infrastructure investment is low and narrow compared to regional peers, limiting effective capital accumulation, and infrastructure provision; (ii) lack of adequate education and professional skills and skill matching mechanisms contribute to low levels of labor productivity, and weak labor market outcomes; (iii) the country lacks a coherent and integrated innovation financing system that will encourage productivity gains and sophistication of firms over time; and (iv) regulations and procedures increase business costs, limiting trade activities' (p. 2).
2. Pereira & Galego (2015) perform a similar analysis for the Portuguese regions, although they focus on intra-regional wage inequality.
3. We used the last wave of the CHS that was available when this study was initiated. An inspection of the data available for other years suggests that using this information rather than that for 2010 had led to results similar to those reported in this paper. The same applies with respect to the data for the other quarters of the year used in the study. We decided not to pool the data for different quarters to prevent further complications with the decomposition in case of including controls for each quarter. We also wanted to avoid the problem caused by the inclusion in the sample of different observations corresponding to the same individual. Since it is not possible to identify individuals in the different quarters, the treatment of observations corresponding to the same individuals is not feasible. In any case, there is no reason to expect differences in the results when pooling the sample for the four quarters, whereas it should be stressed that the number of observations available for each quarter is large enough to guarantee the quality of the estimates. On the other hand, as pointed out above in the introduction, the fall in inequality happened during the first decade of the century in many Latin American countries. In the Colombian case the decline in income inequality was not stable until 2007 and subsequent years. However, in 2008 the Colombian economy was hit by the economic crisis, and after 2010 it experienced an economic boom. Therefore, in our view 2010 is a year that may reflect the *standard* performance of the Colombian economy and in which there was also a fall in inequality.
4. A similar procedure has been applied in, for instance, Pereira & Galego (2014).
5. The definition of the Seventeenth International Conference of Labour Statisticians of informal employment is 'based on the characteristics of the individual's employment, job or position. A worker has an informal job if the employment relationship is, in law or in practice, not subject to national labour or social legislation. This condition of informal employment is observed in persons employed in both formal and informal enterprises, as well as in those employed in domestic service by households' (ILO, 2011, p. 64).
6. Colombia's Golden Triangle refers to an urban region limited by a triangle whose vertexes are defined by the three largest cities: Bogotá, Medellín and Cali. In our particular case, we are not referring to the region but only to the three cities that demarcate the triangle.
7. Due to space, we do not reproduce the results for other percentiles here, although they are available from the authors 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.
8. See Pereira & Galego (2014) for similar arguments in the case of Portugal.

9. In any case, it should be acknowledged that endogeneity of informality is likely to be an issue. We explored the possibility of instrumenting this variable using some household characteristics (similar to those used in the previous literature dealing with selection bias due to endogenous classification). Unfortunately, the test statistics revealed that instruments were weak in all cases, resulting in unreliable estimates.
10. A similar concern could be formulated as regards any estimate reported in the literature on the effect of the set of characteristics included in a wage equation (public/private sector, education, firm size, etc.). We thank an anonymous referee for raising this point.
11. The Blinder–Oaxaca decomposition has been applied in several studies analysing wage gaps by gender, race, public and private sector, etc. It has also been applied recently for understanding regional wage gaps (García & Molina, 2002; Motellón et al., 2011; Galego & Pereira, 2014; Pereira & Galego, 2014) and regional differences in wage inequality (e.g., Pereira & Galego, 2015).
12. Estimates were also obtained for other percentiles. They are not reported here but are available from the authors upon request.
13. The share of informal employment in the lowest quartile of wage distribution ranges from about 40% in the Golden Triangle to 70% in Oriental. In turn, these shares are about 10% and 30% respectively in the upper quartile. An anonymous referee kindly suggested this explanation.
14. Motellón et al. (2011) found an increasing wage differential across the distribution for Spain, and Pereira & 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.
15. We thank an anonymous reviewer for suggesting the separate analysis by gender. As a side effect, it may be assumed that any concern about sample selection is mitigated when analysing women and men separately.

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, 157–175.
- Bacchetta, M., Ernst, E. & Bustamante, J. P. (2009) *Globalization and informal jobs in developing countries*, Geneva, International Labour Office and World Trade Organization.
- 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. & Basu, A. (2013) Highlighting differences between conditional and unconditional quantile regression approaches through an application to assess medication adherence, *Health Economics*, 22, 1052–1070.
- Christopoulou, R. & Monastiriotis, V. (2014) The Greek public sector wage premium before the crisis: size, selection and relative valuation of characteristics, *British Journal of Industrial Relations*, 52, 579–602.
- Combes, P., Duranton, G. & Gobillon, L. (2008) Spatial wage disparities: sorting matters! *Journal of Urban Economics*, 63, 723–742.
- Dickey, H. (2007) Regional earnings inequality in great Britain: evidence from quantile regressions, *Journal of Regional Science*, 47, 775–806.
- Duranton, G. & Monastiriotis, V. (2002) Mind the gaps: the evolution of regional earnings inequalities in the U.K., 1982–1997, *Journal of Regional Science*, 42, 219–256.
- Fields, G. S. (1975) Rural–urban migration, urban unemployment and underemployment, and job–search activity in LDCs, *Journal of Development Economics*, 2, 165–187.
- Fingleton, B. & Palombi, S. (2016) Bootstrap *J*-test for panel data models with spatially dependent error components, a spatial lag and additional endogenous variables, *Spatial Economic Analysis*, 11, 7–26.
- 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 & D. Card (eds) *Handbook of Labor Economics*, Vol. IV, Part A, pp. 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 Economics Letters*, 9, 209–215.
- Gasparini, L., Cruces, G. & Tornarolli, L. (2011) Recent trends in income inequality in Latin America, *Economía*, 11, 147–190.

- 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.
- International Labour Organization. (2011) *2011 Labour overview: Latin America and the Caribbean*, Lima, ILO / Regional Office for Latin America and the Caribbean.
- Journard, I. & Londoño, 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. & Bassett, 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. (2010) Explaining the decline in inequality in Latin America: technological change, educational upgrading, and democracy, in: L. F. López-Calva & N. Lustig (eds) *Declining inequality in Latin America: a decade of progress?*, pp. 1–24, 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.
- Nicolic, J., Rubil, I. & Tomić, I. (2014) Changes in public and private sector pay structures in two emerging market economies during the crisis, *EIZ Working Papers*, EIZ-WP-1403, Zagreb, Croatia.
- 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.
- Pereira, J. & Galego, A. (2014) Inter-regional wage differentials in Portugal: an analysis across the wage distribution, *Regional Studies*, 48, 1529–1546.
- Pereira, J. & Galego, A. (2015) Intra-regional wage inequality in Portugal, *Spatial Economic Analysis*, 10, 79–101.
- Psacharopoulos, G. & Patrinos, H. A. (2004) Returns to investment in education: a further update, *Education Economics*, 12, 111–134.
- 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.
- World Bank. (2015). *Colombia - Second Programmatic Sustained Growth and Income Convergence Development Policy Loan Project*, Washington, DC, World Bank Group.

Copyright of Spatial Economic Analysis is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.