Educational upgrading and wage inequality in Colombia: A quantile regression approach

John Fredy Ariza Bulla[1]

Universitat Autònoma de Barcelona

March 1, 2012

Abstract

In this paper I study the sources of the fall in the overall male wage inequality in Colombia between 2001 and 2010 and test whether the increase in the supply of more educated workers was offset by increased demand. Using Quantile Regressions I conclude that over the sample period schooling had a positive impact upon within-levels wage inequality and that the returns to most of the workers´ characteristics fall because of a strong supply effect in the labor market. The results from the decomposition exercise suggest that the changes in observable prices account for the 42% of the total decrease in inequality while the changes in observable characteristics explain about 33%.

JEL Classification: J31, C14

Key words: wage inequality, quantile regression, decomposition methods

[1] PhD student in Applied Economics. Department d'Economia Aplicada, Universitat Autònoma de Barcelona, Edifici B, Campus UAB. 08193 Bellaterra, Spain; tel: +34-696406364; Fax: +34-935812292; e-mail: JohnFredy.Ariza@e-campus.uab.cat. I wish to thank Josep Lluis Raymond and Xavier Ramos for their suggestions on earlier versions of this document.

1. Introduction

In the economics literature different arguments have been proposed and assessed in order to explain changes in the wage inequality as result of supply, demand, and institutional factors. Hypotheses related to changes in labor force composition, the influence of technology in labor demand, the impact of de-unionization, trade reforms and minimum wages have been tested. In developing economies most of the evidence mainly focuses on the impact of trade reforms and technological change. In the case of Colombia, the trade reform implemented in the beginning of the nineties and its effect upon the demand for skills had a little impact upon the wage distribution (Attanasio *et al.*, 2004). In fact, the evidence for Colombia suggests both the increasing share in total employment of more educated workers and the rising in their within-group wage inequality as the main features in the wage inequality in the nineties (Attanasio *et al.,* 2004; and Santamaria, 2001).

Over the last ten years however, a larger share of male workers with higher educational attainment (some university or more) was accompanied by a reduction in overall wage inequality[2]. In Table 1, we display the main changes in the male labor force composition as well as the changes in male wage inequality during 2001-2010. According to that, the changes in the composition of the labor force were characterized by a larger share of older workers with less experience and with higher level of schooling. We also see that for most of the defined groups the wage inequality fall. This decrease in inequality was greater for the groups with increasing share in total employment.

In this paper we study the sources of the changes in the male wage inequality in Colombia between 2001 and 2010 and test whether the supply of more educated workers was offset by increased demand. Our theoretical approach departs from a simple framework where under the assumption of heterogeneity between groups of workers an increase of more educated workers (*ceteris paribus*) leads to changes in overall wage inequality via two effects: first, through a price effect which shrinks the relative wage of skilled workers because they become relatively more abundant and second, through a composition effect which increases inequality due to an increase in the group of workers earning the highest wage.

[2] We only consider the men for two reasons. First, the changes in overall wage inequality were mainly driven by the changes in male wage inequality. Second, for the case of women we would need to correct the sample selection in wage equation which is beyond of the goal of this paper.

Empirically we evaluate these two effects by using wage and employment data from Colombian household surveys *Encuesta Continua de Hogares* (ECH) and *Gran Encuesta Integarda de Hogares* (GEIH). By means of Quantile Regressions (QRs hereafter) we study the returns to the workers´ characteristics, the returns to the characteristics of the industry and the returns to the characteristics of the firm at several percentiles of the wage distribution. We also evaluate the effect of each characteristic upon wage inequality. Finally we decompose changes in wage inequality into quantities, prices and residuals as in Juhn, Murphy and Pierce (1993).

The results suggest that the decrease in wage inequality is due to changes in both the middle and the lower part of the distribution. The changes in the 75^{th} - 50^{th} wage gap explain about 53 percent of the total fall while the changes in the 25^{th}-10^{th} wage gap accounts by the 31 percent of the total decrease. As we expected, wages increase with age, tenure and education across the whole distribution. In short, most of the estimates from the quantile regressions are statistically significant and have the expected signs. One of the most important results is that over the period most of returns to characteristic fell. It suggests a supply net effect upon inequality. The results from the decomposition exercise indicate that changes in the characteristics of the labor force explain about 33% of the total change of wage inequality, while changes in the prices explain about 42% of this total change.

The paper is organized as follows. Section 2 presents a literature review of the sources of changes in wage inequality. Section 3 describes our theoretical framework. Section 4 explains our empirical strategy and some methodological issues. The data are described in section 5. The results of the quantile regressions and the decomposition exercise are presented in section 6. Finally, section 7 provides some concluding remarks.

2. Literature review

In the economics literature different arguments have been proposed and assessed in order to explain changes in the wage inequality as result of supply, demand, and institutional factors. Juhn, Murphy and Pierce in 1993 suggest that the increased demand for skills was the main explanation for the rise in wage inequality in US during the 1970s and the 1980s. They conclude that the increase in returns to unobserved skills was a consequence of a greater demand for skills. In 2006 however, Lemieux states that the changes in wage inequality during these decades were driven by

composition effects (increases in experience and education of the groups with the highest variance) rather than factors of demand as previous literature considered. An approach developed by DiNardo, Fortin and Lemieux in 1996 highlights the role of minimum wages and de-unionization in the structure of the wages. They point out that de-unionization of the workforce was as important as supply and demand shocks in explaining the rise inequality in that period. Nowadays, the discussion about what factor has had the mayor role explaining wage inequality is still open.

In developing economies most of the evidence mainly focuses on the impact of trade reforms and technological change upon wage inequality between skilled and unskilled workers. In the case of México, Esquivel *et al*, (2003) assert that trade liberalization would have to reduction the wage (skill) gap, while technological progress would impact negatively the real wage of unskilled workers. Since the effect of trade liberalization on wage gap was nil the slight increase in wage inequality would have been driven by technological progress. For Brazil, Ferrerira *et al*, (2007) suggest that trade liberalization made a significant contribution towards a reduction in wage inequality. In this country, the effects occurred through the employment flows across sector and formality categories. They also conclude that the changes in the economy-wide skill premium were important. For Colombia the evidence suggests a little impact of trade reforms upon wage distribution. In 2004, Attanasio *et al.* point out that the overall effect of trade reforms on the wage distribution could have been small. Increasing returns to education, changes in industry premium and informality alone cannot fully explain the increase in wage inequality over 1984-1998. Santamaria (2001) suggest that after 1988 the rise in earnings of the highly educated and the more experience workers explain the changes in wage inequality.

3. Theoretical framework

Our theoretical approach uses a simple market framework as the suggested by Machado and Mata (2005) where there are two types of workers in the economy: low-skilled and high-skilled. So, when the workers are homogeneous, an increase of highly educated workers leads to a reduction in wage inequality via two effects: first, through a price effect that shrinks the relative wage of skilled workers (because they become relatively more abundant), and second, through a composition effect (since a greater proportion of workers are now in the skill group that have the highest wage). These two effects combined lead to the prediction that shifts to the right in both demand and supply of skilled workers that would leave price unchanged would result in less wage inequality.

However, when these workers are not homogeneous or the degree of heterogeneity within type is not the same, an increase in the level of schooling reduces the weight of the low spread group, contributing to both an increased overall wage dispersion and greater inequality.

Most of the wage inequality literature focusses on decomposing changes in inequality into changes in characteristics, prices and residuals. In this way researchers are able to identify and quantify what factors are behind the changes in inequality. Here the distinction between observable and unobservable quantities and prices is crucial. For example, the hypothesis related to SBTC relies on the assumption that the increase in demand for skills explains most of the increase in unobservable prices and therefore would explain the trend in the residual term of a wage equation. This is the approach followed by JMP in 1993. Other approaches however, consider additional functional forms for residuals rather than those based mainly on unobservable prices (Lemieux, 2006). So, we perform a decomposition exercise under this framework as a basic complementary technique to evaluate our hypothesis.

Choosing the decomposition technique however is not easy. We want the most appropriate approach but face software limitations, and writing the code for uncanned decompositions is beyond the scope of this first paper. In principle and since we are working with semiparametric techinques we should try to decompose differences in quantiles, as Machado and Mata (2005) and Melly (2005). Nevertheless, these kinds of procedures have important drawbacks. According to Fortin *et al.* (2001), there has been only a few (and not always satisfactory) decomposition techniques for general distributional measures (quantiles in particular). The basic model to decompose changes in inequality (that decomposes changes at the mean) is the one suggested by Oaxaca in 1973. Decomposing at several quantiles is however difficult to implement, and we will not pursue it here. So despite most of the analysis in the paper is based on QRs to decompose the changes in wage inequality we use OLS regressions as JMP (1993). Since we relate the fall in overall wage inequality mainly to the changes in the composition of the labor force, both observed quantities and prices in the decomposition exercise should explain most of the changes in inequality.

4. Empirical strategy

Traditionally the changes in the structure of wages have been analyzed by considering the average wage into the usual supply and demand framework. However, in order to study the whole wage distribution as well as differences between distributions in a period of time or over time, statistical methods as QRs have been implemented.[3] According to Koenker (2005), QR constitutes a parsimonious way of describing the whole distribution and provides a way to explore sources of heterogenity in the response that are associated with the covariates.

The conditional quantile regression $Q(p/x)$ viewed as a funtion of x for a given p, may be characterized as the unique solution to the problem

$$\underset{g\in\Xi}{Min}\ E\ell_P\left(Y_i - \mathrm{g}(X_i)\right), \qquad 0 < p < 1 \qquad (1)$$

Where Ξ is the class of real valued functions defined on $\mathbb{R}^k$ and ℓ_P denotes the asymmetric absolute loss function. This suggests estimating $Q(p/x)$ by choosing a function of x, given p, out of a suitable of family $\varphi \subseteq \Xi$. The various approaches differ in the choice of the family functions φ. In Koenker and Bassett (1978), φ is the class of linear functions of x such that $Q(p/x) = x'\beta(p)$. The problem (1) is reduced to estimating k-dimentional parameter $\beta(p)$ by any solution $\hat{\beta}(p)$ to the problem

$$\underset{\beta\in\mathbb{R}^k}{Min}\ n^{-1}\sum_{i=1}^{n}\ell_P\left(Y_i - \beta' X_i\right), \qquad 0 < p < 1 \qquad (2)$$

Therefore, the estimate of $Q(p/x)$ is $\hat{Q}(p/x) = x'\beta(p)$. The problem as a linear programming problem has simple and fast computational algorithms available which allows us to efficiently compute the whole set of conditional quantiles for $p \in (0,1)$. Linear quantile regression estimators can be $\sqrt{n}$ - consistent and asymptotically normal under general conditions. We use the Stata software to estimate these regressions. We specify 400 replicates to ensure a large enough number of bootstrap samples for stable estimates of the standard errors and 95% confidence intervals.

We also decompose changes in wage inequality into quantities and prices as JMP (1993). The method starts from a simple wage equation where the cumulative distribution function of the

[3] QR methods also have advantages beyond providing a richer characterization of the data. Median regression is more robust to outliers than OLS regression. Moreover, QR estimators can be consistent under weaker stochastic assumptions than possible with least-squares estimation.

residuals is given by $\mu_{it} = F_t^{-1}(\theta_{it}|\,\boldsymbol{x}_{it})$ with $F_t^{-1}(\cdot\,|\,\boldsymbol{x}_{it})$ as the inverse cumulative residual distribution for workers with characteristics x_{it}. By defining $\overline{\beta}$ to be the average prices for observables over the whole period and $\overline{F}^{-1}(\cdot\,|\,x_{it})$ to be the average cumulative distribution, they decomposes the level of inequality as $y_{it} = x_{it}\,\overline{\beta} + x_{it}\left(\beta_t - \overline{\beta}\right) + \overline{F}^{-1}(\theta_{it}|\,x_{it}) + \left[F_t^{-1}(\theta_{it}|\,x_{it}) - \overline{F}^{-1}(\theta_{it}|\,x_{it})\right]$. By allowing the variation of observable prices, quantities and distribution of residuals, the three equations that reflect counterfactual scenarios of the wage distribution are

$$y_{it}{}^1 = x_{it}\,\overline{\beta} + \overline{F}^{-1}(\theta_{it}|\,x_{it})\,; \quad y_{it}{}^2 = x_{it}\,\beta_t + \overline{F}^{-1}(\theta_{it}|\,x_{it})\,; \quad y_{it}{}^3 = x_{it}\,\beta_t + F_t^{-1}(\theta_{it}|\,x_{it})$$

The basic technique proposed is to calculate the distribution of $y_{it}{}^1$, $y_{it}{}^2, y_{it}{}^3$ for each year and attribute the change through time in inequality in the $y_{it}{}^1$ distribution to changes in observable quantities. Then, attribute any additional change in inequality in $y_{it}{}^2$ to the changes in observable prices, and finally attribute any additional change in inequality for $y_{it}{}^3$ beyond those found for $y_{it}{}^2$ to changes in distribution of unobservables (i.e. changes in unmeasured prices and quantities).

5. Data

The results in this paper are based on wage data for Colombian household surveys ECH for 2001 and GEIH for 2010 administered by *Departamento Administrativo Nacional de Estadistica* (DANE). They provide information on demographic characteristics (gender, age, marital status, family background, educational level), and labor market (employment, unemployment and inactivity) by regions and cities. Throughout the paper we focus on log hourly wages for male workers. We deflate wages by *Indice de Precios al Consumidor* (IPC, base 2008) and define the log hourly wage as the logarithm of the deflated wage divided by the weekly hours worked. For purposes of analysis we select a representative sample of workers with a reasonably strong labor force attachment. Our sample selection criteria is that the workers be aged 18-65, work at least 20 hours weekly in urban areas, not be self-employed, not work in agricultural jobs, and earn a minimum of one dollar per hour. We discard observations of employed people whose information of key variables are reported as missing or outside the coding provided by DANE.

We use these household surveys to create several variables. From the composition of the labor force side, we consider age, tenure and educational attainment.[4] As regards with employer characteristics, we consider indicators as the type of industry of employment, the type of firm, the size of the firm, the type of contract and the type of occupation. We also divide workers between those who are without contract of work and those who have one either to a fixed term or a long term. We also distinguish among six occupation categories: professional/technical, manual/manufacture, management, personnel, servant and sales. Finally, we reclassify the thirteen metropolitan areas defined in Colombia into five ones according to their importance in population and economic activity.

6. Empirical results

6.1 Main changes in the distribution of wages

In figure 1 we plot the empirical densities of log hourly wages as well as both the location and shape shifts of the distribution. During the period, the distribution moves to the right side increasing the real hourly wages. This wage increase however was very unevenly distributed since the main location shifts taken place in the lower part of the distribution. In Table 2 we display more detailed information about the main changes of real wages. For our purpose, we focus on the difference of the log wage between 90^{th} and 10^{th} quantile as a measure of wage inequality. The decrease in 0.19 of this wage gap is explained basically for both changes in the middle and in the lowest part of the distribution. The changes in the 75^{th} - 50^{th} wage gap explain about 53 percent of the total fall while the changes in the 25^{th}-10^{th} wage gap accounts by the 31 percent of the total decrease.

6.2 Quantile Regressions

As a first step in our empirical analysis, we consider the results from QRs showed in figure 2 and the figure 3. The plots show the coefficient estimates for $\beta_i(\theta),\ i = 1, \ldots . k$ for $\theta \in (0,1)$ by year and the associated confidence bands (represented by the shaded area). This shaded area indicates that

[4] Our data do not contain direct information of the individuals´ experience in the labor market. Instead experience, our data contains direct information on the tenure within a firm. We decide work with tenure rather than construct a proxy variable of this from the age and the number of years of schooling as usual, because at the end this proxy of experience would capture in general, the same effect of the age.

the effect of a covariate is significant for particular quantiles if the area does not cross zero. The coefficients estimated by mean regression (OLS) are reported as a dashed horizontal line with their confidence intervals (represented by the dots). The graph of the constant coefficient is a predicted quantile function for log hourly wage for the typical worker and serves as baseline (Hao, *et al.*, 2007).

According to the plots in figure 2, most of the returns graphed are statistically significant different from zero. The plots allow us to see how quantile regression is a straightforward way to capture heterogeneity in the returns in the entire distribution of wages. The first issue to highlight is that returns to all covariates at several quantiles are different from the mean (OLS) estimates. This full characterization of the returns has been so important in wage inequality empirical work. The second issue to comment is that while in some cases returns for categories of some covariates only change in the tails of the distribution, in other cases, there is a more defined pattern of these returns across distribution. Clearly, this is the case for age, educational attainment, type of firm, type of contract and management as type of occupation. The return to age and education increases while the returns to type of contract decrease along distribution. This implies that age and education have a larger impact wage inequality in the upper part of the distribution.

In Table 3 we present the estimated coefficients for the 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentile by year. We also present the OLS estimates in the first column of each year. So, according to the OLS results, on average in 2001, workers who are between 30 and 45 years old earn 12% more than younger workers. For the oldest group this gap is about 22%. That is, the older the worker, the higher the returns for age.[5] As we discuss previously, the results from QRs describe the heterogeneity of the returns across different quantiles. It is clear that wages increase with education across the whole distribution. The effect is more important at the highest part of the distribution than at the lowest, which means that education increases wage dispersion. That is, samples with more educated workers show higher wage dispersion than samples of less educated workers. This fact is qualitatively similar to comparable findings in literature for other countries (Machado and Mata, 2005). By characteristics of industry, we found that the premium for manufactures and other industries is greater than for commerce-hotels-restaurants sector. High

[5] As Buchinsky (1994) and Koenker (2005), the term "return to education" (and likewise "return to experience") refers merely to the effect of experience on the conditional quantile of log wage. The causal effect interpretation is beyond the scope of this study.

levels of productivity and important developments in sector such as mining, energy and water and construction could explain these differences. Higher returns for bigger firms are also associated with greater levels of productivity per work. Having a long term written contract represents more earnings especially for workers at the lower part of the distribution.

6.2 Changes in the returns to the covariates in labor market

During the period most of the returns to the characteristics fell (Table 3). The magnitude of these changes was however different across quantiles. For instance, a worker between 30 and 45 years old earned on average 12 % more than one in the younger category in 2001 but only 11% in 2010. At the upper decile the falling in these returns went to 19% to 10%. Only the returns for the most educated workers and the returns to sector of other industries increase. For example, in the case of the most educated workers, the increase on average return (in regards to the primary or less group) went to the 100% to the 104% while at the upper tail was from 64% to 66%.

According to our framework, these changes in returns are consequence mainly of both the changes composition of the labor force and the changes in the structure of employment. The falling in returns for the most of the characteristics is associated to the supply net effect. That is, the demand for this kind of workers did not growth enough to compensate the increase of supply. This effect is strong and also leaves to fall in overall wage inequality. In the case of the positive changes in returns we could associate it with a more dynamic role of demand in some sectors. For example, sectors such as energy, water and construction would have been important in economic growth in last decade in Colombia. This hypothesis however should be study further.

6.3 Impacts of covariates upon wage inequality

The information provided by QR estimates allows us study the effect of each covariate upon wage inequality. As a measure of inequality we consider the difference in log wages of some percentiles for 2001 and for 2010 and we present them in table 4. If the difference between the 90^{th} and 10^{th} percentile coefficient on a covariate is positive, a higher value of this variable increases within group inequality. In section 6.2 we suggest that education impacts more the wage dispersion in the upper deciles, now we test it statistically. The results indicate that for most of the covariates the interdecile range is significantly different from zero. This implies that heteroscedastic inconsistent

methods will yield biased results (Melly, 2005). So we confirm that within group inequality grows as function of age and schooling for 2001, and furthermore as a function of tenure in 2010. [6] At the beginning of the decade the statistically significant impact on the dispersion of wages is higher for service sector with some type of contract working in a small firm. For 2010 the impact of each previous variable increases. Only the effect of the fixed-term contract becomes smaller.

6.4 A basic decomposition exercise of changes in wage inequality

As JMP (1993) we decompose the contribution to changes in inequality due to changes in quantities, prices and residuals. Table 5 shows the results of this wage gap decomposition. According to these results, most of the changes in overall inequality (90^{th}-10^{th} wage gap) are explained by changes in prices (42%). Changes in observed quantities account only for 32% of the total decrease while changes in the residuals commonly related to skill biased technological change SBTC accounts for the 25%. In fact we see that the residual component becomes more important in the lower wage gap. As we commented previously the discussion in the literature about the impact of changes in both observable and unobservable quantities and prices on wage inequality is addressed here throughout this basic exercise. So, one possible implication is that the SBTC had an important impact on the changes in wage inequality in the lower part of the distribution. On the other hand, we need to make sure that we control for all of the covariates in the regression models. Thus these hypotheses should be study further.

7. Concluding remarks

Unlike the 1990s, in the last ten years in Colombia wage inequality fell while the share of male workers with high educational attainment increased. Such increasing sharing of workers with more schooling (some college and college or more) went to 25% to 35% whereas the difference of the log of hourly wages between the 90^{th} and the 10^{th} quantile decreased from 1.84 to 1.65. In this paper,

[6] Martins and Pereira (2004) suggest three possible factors that explain why the spread of the returns increases for higher educational levels. The first is the over-education since workers are taking jobs less according to their skills. Another possibility is the interaction between schooling and ability, in which the most able can benefit more from their schooling and the pay gap between the more and less able deepens from higher educational levels. Finally, they also consider the differences in school quality or fields of study as sources of higher dispersion.

we study the sources of this change in wage inequality and test whether the supply of more educated workers was offset by increases in demand.

By considering a model with two types of workers, as in Machado and Matta (2005), we conclude that the fall in overall male wage inequality over the sample period is explained by the decrease in the returns for most of the characteristics, which in turn is due to a net supply effect. That is, the evidence suggests that although the variance of wages grows with age, tenure and education, the net impact of the observed increase in age and education of the workforce upon within-wage inequality reduced the overall inequality.

The decomposition exercise corroborates the main findings in the QRs. The changes in overall inequality are explained mainly by changes in quantities and prices. Changes in quantities explain about 33% of the overall variation of inequality while changes in prices account for 42%. One important result is the role of the residual component in the wage inequality in the lower part of the distribution (It explains the 21% of the total variation in overall inequality). Since it has been related to the SBTC hypothesis, further tests should be done. At this point, the discussion about the relationship between education and ability and the rewards in the labor market to observed and unobserved characteristics become crucial issues to be considered.

References

[1] Attanasio, O. Goldberg, P. and Pavcnik, N. (2004) "Trade Reforms and Wage Inequality in Colombia". *Journal of Development Economics*, Vol. 74, No, 2, pp. 331-366.

[2] Buchinsky, M. (1994). "Changes in the US wage structure 1963-1987: Application of quantile regression". *Econometrica,* Vol 62, No 2. Pp. 405-458.

[3] Dinardo, J., Nicole, F. and Lemieux, T. (1996). "Labor market institutions and the distribution of wages, 1973-1992. A semiparametric approach", *Econometrica*, Vol. 64, No, 5, pp. 1001-1044.

[4] Esquivel, G. and Rodriguez-López, J. (2003). "Technology, trade, and wage inequality in Mexico before and after NAFTA", *Journal of Development Economics*, 72 (2), Pp. 543-565.

[5] Ferreira, F. Leite Leite, P. and Wai-Poi, M. (2007). “Trade liberalization, employment flows and wage inequality in Brazil”, Working Papers UNU-WIDER Research Paper, *World Institute for Development Economic Research* (UNU-WIDER).

[6] Hao, L. and Naiman, D. Q. (2007). “Quantile Regression". *Series Quantitative Applications in the Social Sciences*, No. 149 Sage Publication, Thousand Oaks.

[7] Juhn, C. Murphy, K. and Pierce, B. (1993). “Wage Inequality and the rise in return to skill”, *Journal of Political Economy*, 101 (3), Pp. 410-442.

[8] Katz, L. and Autor, D. (2000). “Changes in the wage structure and earnings inequality", *Handbook of Labor Economics*, Vol. 3, Part A. pp. 1463-1555.

[9] Koenker, R. and Bassett, G. (1978). “Regression Quantiles". *Econometrica*, No. 46, pp. 3350.

[10] Koenker, R. (2005). *Quantile Regression*, Cambridge University Press, New York.

[11] Lemieux, T. (2004). “Increasing residual wage inequality: compositions effects, noisy data or rising demand for skill?”, *The American Economic Review*, Vol. 96, No. 3, pp. 462-500.

[12] Lemieux, T. (2006). “Postsecondary education and increasing wage inequality", *The American Economic Review*, Vol. 96, No. 2, pp. 195-199.

[13] Machado, M. and Mata, M. (2005). “Counterfactual decomposition of changes in wage distribution using quantile regression", *Journal of Applied Econometrics*, John Wiley & Sons, Ltd., Vol. 20(4), pp. 445-465

[14] Martins, P. and Pereira, P. (2003). “Does education reduce wage inequality? Quantile regression evidence from 16 countries", *Labour Economics*, No. 11, pp.355-371.

[15] Melly, B.(2005). “Decomposition differences in distribution using quantile regression", *Labour Economics*, No. 12, pp. 577-590.

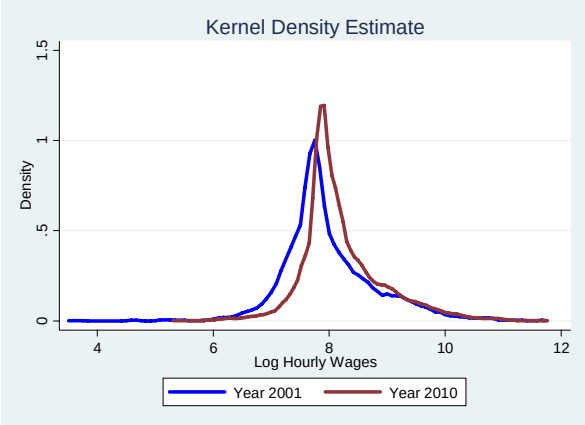
[16] Oaxaca, R. (1973). “Male-female wage differentials in urban labor markets”. *International Economic Review*. No. 14, pp. 693-709.

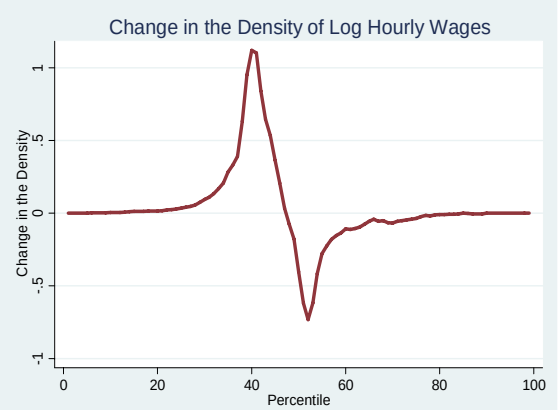
[17] Santamaria, M. (2001). “External trade, skill, technology and the recent increase of income inequality in Colombia", *Archivos de Economía* No. 171, DNP.

Table 1. Percentage share and wage inequality measures of workers by some covariates

	%		% Change	Gini	
	2001	2010		2001	2010
Real wage				0.48	0.44
				(0.01)	(0.01)
Age					
18-29	39.78	36.76	-3.02	0.32	0.31
				(0.01)	(0.01)
30-44	43.44	41.56	-1.88	0.48	0.44
				(0.01)	(0.01)
45 or more	16.78	21.68	4.9	0.56	0.51
				(0.02)	(0.02)
Tenure					
0-3	57.01	63.38	6.37	0.44	0.38
				(0.02)	(0.01)
4.-6	15.62	14.4	-1.22	0.42	0.44
				(0.02)	(0.02)
7 or more	27.37	22.22	-5.15	0.50	0.48
				(0.02)	(0.01)
Educational attainment					
Primary or less	19.45	15.21	-4.24	0.27	0.23
				(0.01)	(0.01)
High School Dropouts	22.88	15.84	-7.04	0.26	0.23
				(0.01)	(0.01)
High school	32.02	33.57	1.55	0.32	0.25
				(0.01)	(0.01)
Some college	10.93	19.47	8.54	0.34	0.30
				(0.01)	(0.01)
College or more	14.73	15.9	1.17	0.42	0.39
				(0.02)	(0.01)

Source: ECH, GEIH - DANE.
Note: All calculations use sample weights. The standard errors are estimated by bootstrapping methods with 100 replicates.





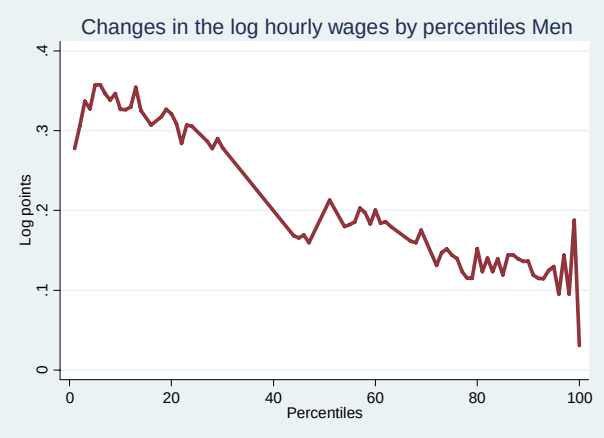


Figure 1. Main changes in the male wage distribution.

Table 2. Descriptive statistics of male real wages

	2001	2010		2001	2010	
Real wage (COP)			**Lnrealwage**			
Mean	4,381	5,054	Mean	8.02	8.22	
Std. Dev.	6,405	6,382	Median	7.82	8.03	
Min	36.3	216.1				
Max	116,059	119,688	**Dispersion**			**Change**
sk	7.1	5.9	Log(90)-Log(10)	1.84	1.65	-0.19
k	84.9	60.0	Log(90)-Log(50)	1.29	1.17	-0.12
			Log(50)-Log(10)	0.55	0.48	-0.07
Quantiles						
p10	1,347	1,926	Log(75)-Log(25)	0.79	0.68	-0.11
p25	1,889	2,568	Log(75)-Log(50)	0.60	0.50	-0.10
Median	2,418	3,078	Log(50)-Log(25)	0.19	0.18	-0.01
p75	4,241	4,987				
p90	9,067	9,974	Log(25)-Log(10)	0.36	0.29	-0.06

Source: ECH, GEIH - DANE. Note: All calculations use sample weights

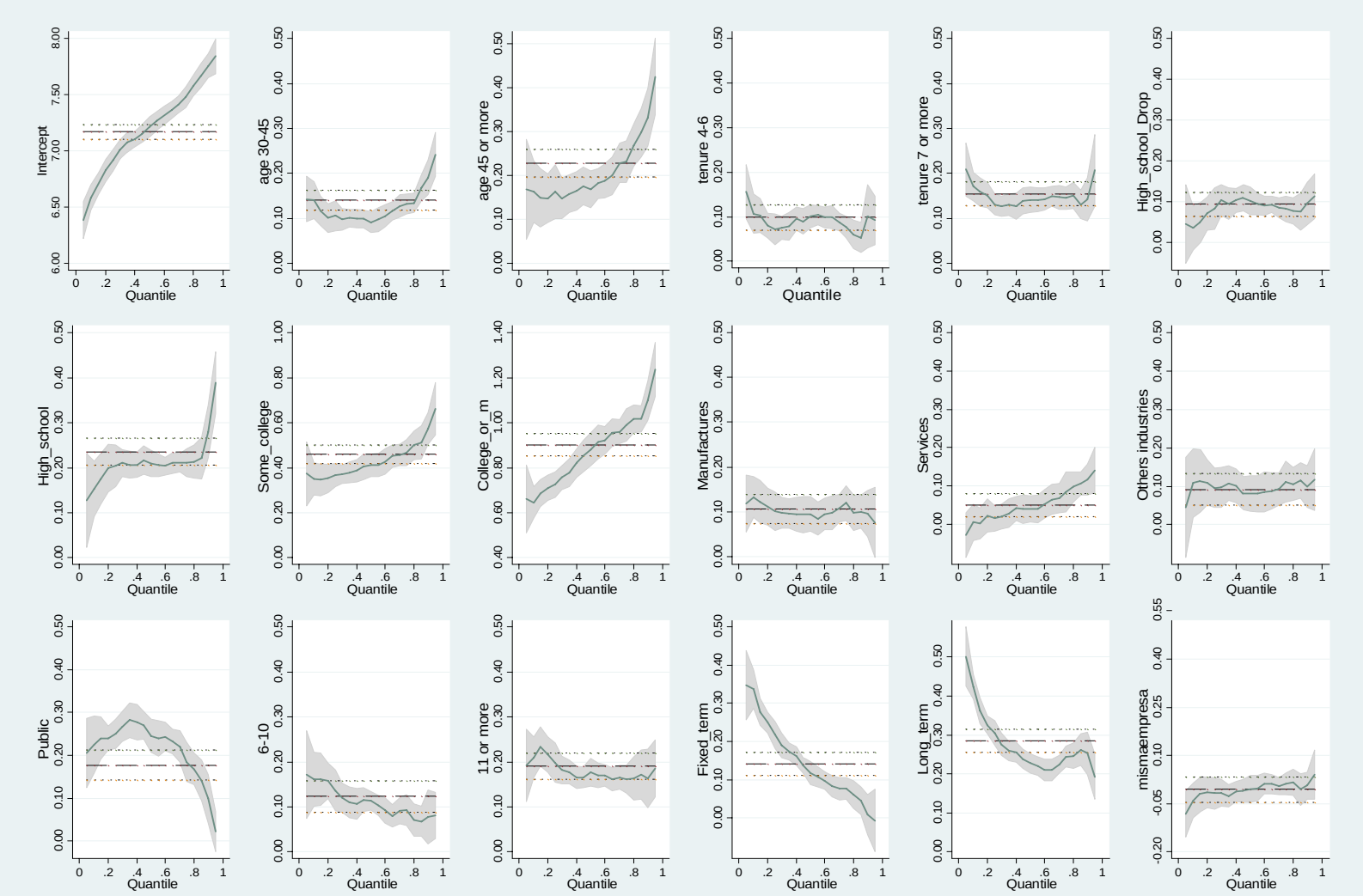


Figure 2. Quantile Regression coefficients for 2001. The shaded area represents 95% confidence intervals for the deciles, the dashed horizontal line represents the least squares estimate.

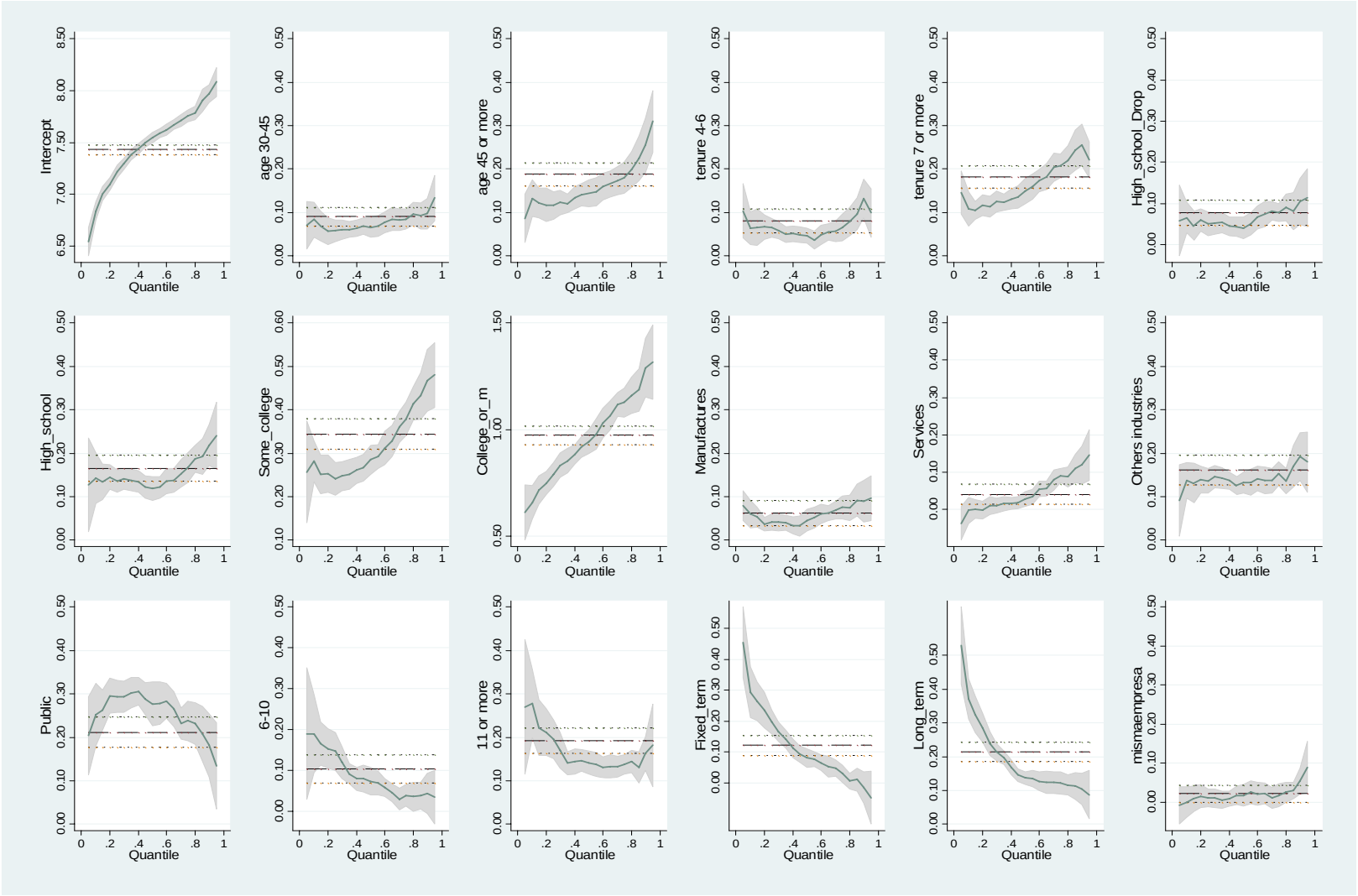


Figure 3. Quantile Regression coefficients for 2010. The shaded area represents 95% confidence intervals for the deciles, the dashed horizontal line represents the least squares estimate.

Table 3. Quantile Regressions of Log hourly wage on a set of worker and firm characteristics

	2001						2010					
	OLS	**10th**	**25th**	**Median**	**75th**	**90th**	**OLS**	**10th**	**25th**	**Median**	**75th**	**90th**
Age 30-45	0.125***	0.141***	0.106***	0.092***	0.133***	0.191***	0.113***	0.084***	0.058***	0.066***	0.085***	0.097***
Age 45 or more	0.221***	0.164***	0.163***	0.168***	0.231***	0.332***	0.208***	0.133***	0.116***	0.144***	0.180***	0.255***
Tenure 4-6	0.070**	0.107***	0.071***	0.102***	0.077***	0.102***	0.053**	0.064***	0.065***	0.046***	0.066***	0.132***
Tenure 7 or more	0.199***	0.171***	0.131***	0.141***	0.145***	0.142***	0.170***	0.108***	0.114***	0.149***	0.208***	0.255***
High_school_Drop	0.092***	0.035	0.083***	0.102***	0.083***	0.096***	0.061***	0.065**	0.050***	0.040***	0.077***	0.104***
High_school	0.250***	0.153***	0.204***	0.209***	0.211***	0.280***	0.149***	0.143***	0.136***	0.119***	0.167***	0.219***
Some_college	0.493***	0.352***	0.367***	0.411***	0.467***	0.575***	0.341***	0.282***	0.241***	0.285***	0.379***	0.468***
College_or_m	1.001***	0.646***	0.725***	0.882***	0.990***	1.098***	1.038***	0.660***	0.791***	0.944***	1.127***	1.289***
Manufactures	0.104***	0.132***	0.101***	0.095***	0.119***	0.096***	0.045**	0.060***	0.042***	0.046***	0.075***	0.089***
Services	0.029	0.006	0.016	0.041**	0.084***	0.116***	0.045*	-0.002	0.011	0.029**	0.091***	0.119***
Others industries	0.097***	0.108***	0.096***	0.081***	0.112***	0.100***	0.125***	0.138***	0.136***	0.132***	0.153***	0.192***
6-10	0.121***	0.162***	0.137***	0.113***	0.093***	0.078**	0.090***	0.189***	0.146***	0.073***	0.038**	0.043
11 or more	0.192***	0.210***	0.198***	0.177***	0.161***	0.164***	0.172***	0.279***	0.195***	0.142***	0.138***	0.163***
Fixed_term	0.105***	0.337***	0.217***	0.117***	0.078***	0.007	0.096***	0.293***	0.196***	0.080***	0.030*	-0.015
Long_term	0.245***	0.424***	0.305***	0.227***	0.245***	0.253***	0.187***	0.371***	0.239***	0.136***	0.122***	0.104***
Cons	7.250***	6.586***	6.920***	7.214***	7.478***	7.757***	7.423***	6.840***	7.218***	7.548***	7.757***	7.970***

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are based on bootstrapping with 400 replicates. We also control by type of occupation, metropolitan areas and ownership of the firm. Constant: Age 18-29, Tenure 1-3, Primary or less, Com-hotels-rest, Private sector, Firm with 2-5 employees, without contract, working as a seller and living in Bogota

Table 4. Interquantile regressions of Log hourly wage on a set of worker and firm characteristics

	2001			2010		
	90th-10th	75th-50th	25th-10th	90th-10th	75th-50th	25th-10th
Age 30-45	0.050*	0.041***	-0.035**	0.013	0.019*	-0.026*
Age 45 or more	0.169***	0.063***	-0.000	0.122***	0.036**	-0.017
Tenure 4-6	-0.005	-0.025	-0.035*	0.069*	0.020	0.002
Tenure 7 or more	-0.029	0.004	-0.040***	0.148***	0.059***	0.006
High_school_Drop	0.060*	-0.019	0.048**	0.039	0.037***	-0.015
High_school	0.127***	0.002	0.051***	0.076***	0.048***	-0.007
Some_college	0.223***	0.056**	0.016	0.186***	0.093***	-0.041*
College_or_m	0.453***	0.108***	0.079**	0.629***	0.183***	0.131***
Manufactures	-0.037	0.024	-0.032	0.029	0.030**	-0.018
Services	0.111***	0.044***	0.010	0.121***	0.062***	0.013
Others industries	-0.009	0.031	-0.012	0.055	0.021	-0.002
6-10	-0.084*	-0.021	-0.025	-0.147***	-0.035*	-0.043
11 or more	-0.046	-0.016	-0.012	-0.116***	-0.004	-0.083***
Fixed_term	-0.329***	-0.040***	-0.120***	-0.308***	-0.050***	-0.097***
Long_term	-0.171***	0.018	-0.119***	-0.267***	-0.015	-0.132***
Cons	1.170***	0.264***	0.334***	1.131***	0.209***	0.379***

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are based on bootstrapping with 400 replicates. We also control by type of occupation, metropolitan areas and ownership of the firm. Constant: Age 18-29, Tenure 1-3, Primary or less, Com-hotels-rest, Private sector, Firm with 2-5 employees, without contract, working as a seller and living in Bogota

Table 5. Juhn, Murphy and Pierce Decomposition for Log real hourly wage gap

Wage Gap	Total Change	Observed Quantities	Observed Prices	Unobserved Prices and Quantities
90th -10th percentile	-0.190	-0.062	-0.080	-0.048
	100%	32.6%	42.1%	25.2%
90th -50th percentile	-0.073	-0.027	-0.038	-0.008
		14.2%	20%	4.2%
50th -10th percentile	-0.117	-0.035	-0.042	-0.040
		18,4%	22,2%	21.0%

Source: *ECH, GEIH- DANE.*