E C O N O M I C S B U L L E T I N

Returns to Education and Wage Differentials in Brazil: A Quantile Approach

Patricia Stefani
Ibmec − SP

Ciro Biderman
FGV − SP

Abstract

This paper uses quantile regression techniques to analyze the returns to education across the conditional distribution of wages from individuals, separated both by gender and skin color, while accounting for the endogeneity of education decisions. Are the returns to education heterogeneous across the conditional distribution of earnings? Using data from the 1996 PNAD, the results indicate that the returns to education are, indeed, significantly heterogeneous across the distribution of earnings, as well as a considerable wage gap between the groups, according to gender and skin color.

The first author would like to thank CNPq for research assistant scholarship. The second author is also a researcher from CEM−Lume, related to CEPID, with financial support from FAPESP. The usual disclaim applies.
Citation: Stefani, Patricia and Ciro Biderman, (2006) "Returns to Education and Wage Differentials in Brazil: A Quantile Approach." *Economics Bulletin*, Vol. 9, No. 1 pp. 1−6
Submitted: March 31, 2005. **Accepted:** March 7, 2006.
URL: http://www.economicsbulletin.com/2006/volume9/EB−05I20004A.pdf

1 Introduction

The main goal of this paper is to analyze the heterogeneity in returns to schooling in Brazil, while accounting for the endogeneity of education decisions, for individuals separated both by gender and by skin color. Its central contribution rests in the attempt to provide a more realistic picture not only of the wage differentials among the groups being considered, but of a possibly heterogeneous pattern of the returns to education across the conditional quantiles of the wages distribution.

One's decision about how much education to acquire depends on factors such as parent's educational level and the availability of resources to be designated to this end. If we take the educational decision as being exogenous, we might have a heterogeneous pattern of the returns to education across the quantiles that are only reflecting non-observable and heterogeneous individual abilities, which translate in differences in initial endowment that, in some way, are being translated into higher earnings.

In addition, we attempt to quantify the wages disparities among individuals, separated by gender and skin color, while controlling for the main factors that contributes to determine one's earnings.

The wage and income inequality, the role of education and the disparities among individuals, especially between black and white, are a major issue concerning the future of Brazil. As black individuals are concentrated in the bottom of the earnings distribution and as poor individuals tend to acquire less education due to factors such as the high marginal cost that this activity represents and due to credit restrictions, this combined with the well known poor quality of the brazilian public educational system, may lead to an unbreakable cycle.

The results indicate that the returns to education are significantly heterogeneous across the distribution of earnings, as well as the wage disparities between the groups separated according to gender and skin color.

2 Methodology

As conditional mean regressions may not give the best portrait of the explanatory variables over the distribution of wages, to better represent the issues being addressed we have opted for a quantile regression technique, which allows us to infer the effect of the exogenous variables at any arbitrary point of the conditional wage distribution, using the whole sample to estimate the parameters at each quantile.

2.1 Wage Differentials

As proposed by Koenker and Bassett (1978) the quantile regression model can be represented as:

$$y_i = x_i^{'}\beta_\theta + u_{\theta i} \qquad , \qquad Quant_\theta\left(y_i|x_i\right) = x_i^{'}\beta_\theta \tag{1}$$

where $Quant_\theta\left(y_i|x_i\right) = x_i^{'}\beta_\theta$ is the θth conditional quantile of $y_i, 0 < \theta < 1$ given a vector of regressors x_i, and $\tilde{\beta}_\theta$ is the estimator of β_θ in (1), which solves the following minimization problem:

$$\min_\beta \frac{1}{n}\left\{\sum_{i:y_i \geq x_i^{'}\beta} \theta\left|y_i - x_i^{'}\beta\right| + \sum_{i:y_i < x_i^{'}\beta} (1-\theta)\left|y_i - x_i^{'}\beta\right|\right\} \tag{2}$$

The distribution of the disturbance term, $u_{\theta i}$,in (1), need not to be known, as long as the property $Quant(u_{\theta i}|x_i) = 0$ is being satisfied. Not only are the parameters of this type of regression extremely robust with respect to outliers, but thy can also vary across the conditional distribution of the dependent variable.

To account for the wage differentials among the individuals separated by gender and skin color, in white males – the control group – black males, white females and black females, we follow Oaxaca (1973), starting from a mincerian wage equation, for each group, such as

$$w_i = \alpha_i + \sum_k \beta_{ki} x_{ki} \tag{3}$$

where w_i is the natural logarithm of the hourly wage rate of the ith individual and x_{ki} is the vector composed by the k individual characteristics, which, in a quantile regression framework, would give us

$$\overline{w}_\theta^c - \overline{w}_\theta^j = \sum_k \beta_{k\theta}^c \overline{x}_k^c - \sum_k \beta_{k\theta}^j \overline{x}_k^j \tag{4}$$

with c =white male (the control group) and j = black males, white female, black female.

Omitting, for sake if simplicity, the k subscript, and adding and subtracting $\sum \beta_{k\theta}^c \overline{x}^j$, in order to preserve the inequality, we can write (4) as

$$\overline{w}_\theta^c - \overline{w}_\theta^j = \sum_i \overline{x}^c\left(\beta_\theta^c - \beta_\theta^j\right) + \sum_k \beta_\theta^c\left(\overline{x}^c - \overline{x}^j\right) \tag{5}$$

where $\overline{w}_\theta^c$ and $\overline{w}_\theta^j$ are, respectively, the hourly wage rate, at the quantile of interest, of a white male and an individual from one of the other groups in j;

$\overline{x}^c$ and $\overline{x}^j$ are the vectors with the mean of the regressors for this groups, and β^c_θ and β^j_θ are the vectors with the estimated coefficients at each quantile.

The first term represents the part of the wage differential that is due to, in Oaxaca terms, discrimination. If this difference is positive, then there exists a different remuneration for each group when compared with the control group, that is, white male.

The idea is that, if the individuals of both groups were equally productive, then this difference would be zero, as they would face the same wage structure.

The last term accounts for the difference in endowments. In the absence of discrimination, all the wage disparities would be accounted for differences in productive factors.

2.2 Instrumental variable Quantile Regression

As mentioned before, the endogeneity of the individual decision about how much education to acquire must be taken into account when we are trying to characterize the pattern of a family of returns to education.

The relative position that an individual occupies in the society is strongly related to his initial endowments and a strong manner trough where this relation becomes explicit is in the influence of parent background in determining, in some way, the potential income of its child. The child not only inherits physical and genetic endowments, but they are also subjected to their parent's education investment decisions.

Under these conditions, those who were born in more educated and wealthier families face a lower marginal cost of education, tending to acquire more education then those who come from poorer families.

Therefore, in trying to understand the nature of a possibly heterogeneous pattern of the returns to education across the conditional wage distribution, we need to take the endogeneity bias, or ability bias, into account. The underlying hypothesis is the treatment of ability and education as two distinct factors in the construction of human capital and, if we believe that there exist some sort of complementarity between these factors, then there exists an additional indirect effect that would make the returns to education heterogeneous across the distribution of wages.[1]

The instrumental variable quantile regression technique aims to take the latter problem into account. As in Arias et al. (2001), consider the following structural equation:

$$Y = Y_1\gamma + X_1\beta + u \tag{6}$$

where Y is the dependent variable, Y_1 is a $n \times g$ matrix with the endogenous variables being simultaneously determined with Y, γ is the vector of the associated coefficients and X_1 is a $n \times k_1$ matrix with the exogenous variables. If we

[1]See Arias et Al (2001) for more detailed discussion.

have a matrix X_2, containing the set of k_2 instruments, we can perform a two stage estimation, where, on a first stage, the endogenous explanatory variable is spanned in the space generated by the instruments, that are assumed to be uncorrelated with the disturbance term, as follows

$$Y_1 = X\Pi + v \tag{7}$$

where $X = [X_1, X_2]$ is a $n \times (k_1 + k_2)$, v is a vector of iid's disturbances and, as mentioned, X_1being the matrix of the exogenous variables and X_2 a matrix containing the instruments. The second stage consists in a quantile regression of the dependent variable, Y, on the projections obtained in the first stage:

$$Y = \widehat{Y}_1\gamma + X_1\beta + u \tag{8}$$

Hence, based on the OLS regression from the first stage, the individual's education, is regressed on the instruments – father education, mother education, family size – and on the exogenous variables – experience, formality control[2] and regional controls (Southeast, Northeast, North and South). On the second stage, the wage equation model described in section 2.1 is estimates, and with containing not only the predicted values for education on the first stage, but the exogenous variables as well.[3]

3 Data

The data come form the 1996 PNAD - Pesquisa Nacional por Amostra de Domicílios, a national household survey, and the particular year was chosen because it contains data regarding one's family background .The sample was separated in four groups: white male, black male, white female and black female and restricted to those who reported positive earnings and were between the age of 15 to 60.

Table 1 presents the means, by gender and skin color, of the selected variables. Its main feature to be highlighted is the low parent's educational level, for every group, especially for blacks, as well as the higher educational level of white individuals, specially white women. Finally it is worth noticing how low is the percentage of individuals who are working in the formal sector.

[2] One of the central issues concerning the brazilian labor market is the fact that a great part of the working force, due to high costs associated to working at the formal sector, choose to work at the "informal" sector. This control assumes a value of 1, if the individual works at the formal sector, and 0 otherwise.

[3] See Arias et Al. (2001) for details about the properties of this estimator. The covariance matrix was obtained via bootstrapping with 400 replications.

In figure 1 the wage of each group appears as a percentage of the white males wage, at each quantile. One can note that, as we ascend in the distribution, this percentage is descending. The behavior is clearly heterogeneous across the distribution; a black male located between the poorest 10%, at the median and between the richest 10%, earns, respectively, about 94%, 76% and 64%. The black female group is the one earning proportionately less, in every quantile, than a white male.

At last, in table 2, the sample correlation between an individual's education and his father and mother education is quite strong, which reinforces the latter discussion about the importance of the family background to the individual's educational decisions and his structure of future earnings.

4 Empirical Findings

One can see from figure 2, that returns to education are higher at the upper quantiles. Even when we take the endogeneity bias into account, the returns to education show a heterogeneous pattern across the conditional distribution of wages, what is confirmed by the tests of the interquantile differences.[4] This reinforces the idea of a complementarity degree between education and abilities, what gives an advantage to those located at the top of the distribution of wages, also enhancing the potential earnings of those located at the bottom, what would be consistent with the existence of a negative correlation between marginal costs and marginal benefits of education across the abilities. In this way, more able individuals tend to face higher returns to education distribution.

As for the groups being considered, black, both male and female, face lower returns to education than whites. Looking to the wage differentials due to disparities in observable characteristics, in figure 3, one can notice that black female loose about 13% to 16% of their wages in comparison to the wage of a white male, what indicates that the difficulties to have access to quality education, a problem faced by low income families, is an important issue for this group.

Concerning the group of the black males, the larger part of the wage differentials is due to disparities in observable characteristics; the group looses between 12 to 20% of its wages in relation to a white male wage because of this matter, what is the opposite of the results obtained for white female, whose wage disparity is mainly accounted by discrimination factors.

It is important to draw attention to the low presence in the formal sector of women in comparison to white male, especially at the bottom of the wage distribution, in determining the wage differentials. The informal sector ends up attracting those individuals with lower educational levels, whose opportunity

[4] Following Buchinsky (1995), this test is performed after an interquantile regression, which reestimates the model taking the difference between the coeficients across the wages distribution $\beta_{k\theta_1} - \beta_{k\theta_2} = 0$, where θ_1 and θ_2 are two distinct quantiles, say, .10 and .50 and the β_k refer to regressor k.

cost of searching for a job at the formal sector and being kept at it, paying the taxes that come with it, does not compensate an eventual benefit of contributing, for example, to public social security. Therefore, these individuals prefer to work for a lower wage at the informal sector

An under skilled worker, with low educational level has little benefits to extract from working at the formal sector of the economy. Given the incertitude of being kept on such jobs for a longer period of time, these individuals would not benefit from the legal contributions made during the working period. Hence, those compulsory contributions end up acting as a tax, making less costly to work at the informal sector.

5 Conclusions

The most important feature presented in this paper is the heterogeneity in returns to education for individuals separated according to gender and skin color. Not only the returns to education are heterogeneous across the conditional distribution of wages, but they are quite high as well. Our results corroborate the idea that policies of investment in education contribute to straighten wage dispersion. When the impact of education on an individual's earnings is expressive, the low educational level translates, inevitably, in income disparities.

Investments in education will be as low the poorer the family is, not only because of lack of resources and access to credit, but also due to the large opportunity cost implied by this investment. Therefore, it is natural to expect that less educated individual today will earn lower wages, have fewer resources to invest in their children, determining in a large way, therefore, the starting point of theses individuals.

The extent of the intergenerational mobility is determined by market and institutional factors, and by genetic ones. If the educational system is successful in distributing public expenditure in a manner that aims to reduce the educational disparities, it would be affecting the relative position that an individual occupies within a society.

This problem is especially important for black individuals, as they are largely represented at the bottom of the income distribution. The inequality of opportunities and the sub investment in human capital results in the relatively lower productivity of blacks to whites, being that the main factor contributing to the wage disparities between these groups[5].

[5] The results are also being worked on for the year of 1988, which also included data on family backgroud. From that, one can compare the evolution of the educational picture and the wage disparities in a interval of almost ten years.

References

Arias, O., Hallock, K. and Sosa-Escudero, W. (2001). Individual Heterogeneity in the Returns to Schooling: instrumental variable quantile regression using twin data. *Empirical Economics*, 26:7-40.

Buchinsky, M. (1995). Estimating the Asymptotic Covariance Matrix for Quantile Regression Models: A Monte Carlo Study. *Journal of Econometrics*, 68: 303-338.

Koenker, R. and Bassett, G. (1978). Regression Quantiles. *Econometrica*, vol. 46, Issue 1, 33 - 50.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, vol. 14, Issue 3, 693-709.

Table 1 - Descriptive Statistics

Variable	White Male	Black Male	White Female	Black Female
Years of Shooling	7.0	5.8	7.3	6.0
Experience	21.2	22.0	19.1	20.7
Age entering labor market	13.0	12.9	14.9	14.3
Father's years of schooling	2.4	1.9	2.4	1.9
Mother's yers of schooling	2.2	1.7	2.2	1.7
# of people in the family	4.3	4.5	4.1	4.3
Working at formal sector	33%	36%	17%	20%
Southeast region	34%	45%	34%	46%
Center-West region	11%	8%	11%	8%
Northeast region	30%	32%	31%	31%
Southern region	18%	10%	17%	11%
Urban Region	82%	82%	84%	85%

Table 2 - Sample correlation between an individual years of schooling and his parent's years of schooling

	Father	Mother
White Male	0.57	0.56
Black Male	0.51	0.46
White Female	0.54	0.55
Black Female	0.48	0.47

Table 3 - Estimated Coefficients at quantiles $\theta = 0.10$ and $\theta = 0.90$

	White Male		Black Male	
	Coef.	Std. Dev.	Coef.	Std. Dev.
$\theta = .10$				
years of schooling	0.14	0.004	0.10	0.02
experience	0.007	0.001	0.002*	0.003
formality control	0.17	0.01	0.16	0.06
Southeast	0.20	0.04	0.39*	0.23
Center-West	0.14	0.04	0.43*	0.27
Northeast	-0.1	0.04	0.19*	0.24
South	0.18	0.04	0.4*	0.25
Urban region	0.32	0.02	0.31	0.09
$\theta = .90$				
years of schooling	0.20	0.004	0.19	0.2
experience	0.02	0.001	0.02	0.02
formality control	-0.22	0.02	-0.38	-0.3
Southeast	0.05*	0.03	0.14	0.09
Center-West	0.08	0.04	0.4	0.18
Northeast	-0.06*	0.03	-0.01	-0.14
South	-0.008*	0.03	0.14	0.05
Urban region	0.29	0.02	0.11*	-0.03
	White Female		Black Female	
	Coef.	Desv. Pad	Coef.	Desv. Pad
$\theta = .10$				
years of schooling	0.15	0.005	0.06	0.02
experience	0.009	0.001	-0.002*	0.005
formality control	0.08	0.02	-0.002*	0.08
Southeast	0.15	0.04	0.17*	0.21
Center-West	0.09*	0.05	0.27*	0.26
Northeast	-0.24	0.04	-0.09*	0.21
South	0.17	0.04	0.33*	0.23
Urban Region	0.12	0.04	0.46	0.17
$\theta = .90$				
years of schooling	0.006	0.01	0.28	0.04
experience	0.002	0.003	0.029	0.007
formality control	0.02	0.05	-0.14*	0.14
Southeast	0.04*	0.09	-0.33*	0.54
Center-West	0.06	0.10	-0.007*	0.57
Northeast	0.05	0.09	-0.54*	0.53
South	0.05*	0.08	-0.25*	0.54
Urban Region	0.05*	0.08	-0.55*	0.31

* not significant at 5%

Table 4 - Interquantile Differences in the Returns to Educarion

Quantiles		White Male	Black Male	White Female	Black Female
0.10	0.25	0.028	0.039	0.025	0.080
0.10	0.50	0.052	0.068	0.051	0.126
0.10	0.75	0.070	0.093	0.070	0.147
0.10	0.95	0.067	0.089	0.066	0.219
0.25	0.50	0.024	0.028	0.025	0.046
0.25	0.75	0.042	0.054	0.045	0.067
0.25	0.90	0.039	0.050	0.041	0.139
0.50	0.75	0.018	0.025*	0.019	0.021*
0.50	0.90	0.015	0.021*	0.016	0.093
0.75	0.90	-0.002*	-0.003*	-0.003*	-0.001*

* not significant at 5%

Figure 1 - Wage of each group as a percentage of the wage of a white male at each quantile

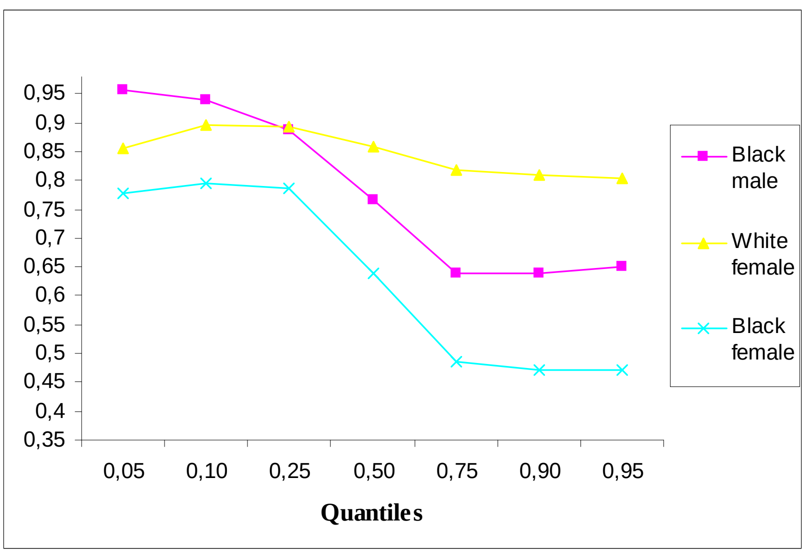

Figure 2 - Returns to Education

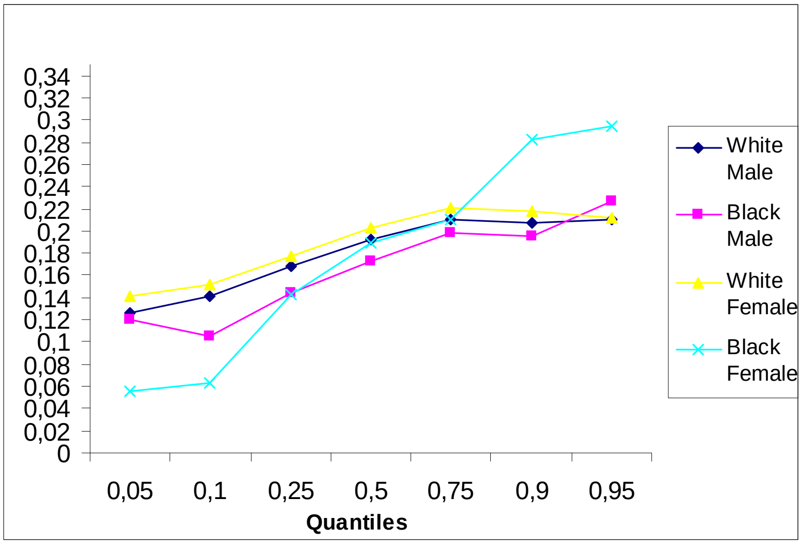

Figure 3 - Wage differentials due to differences in observable characteristics

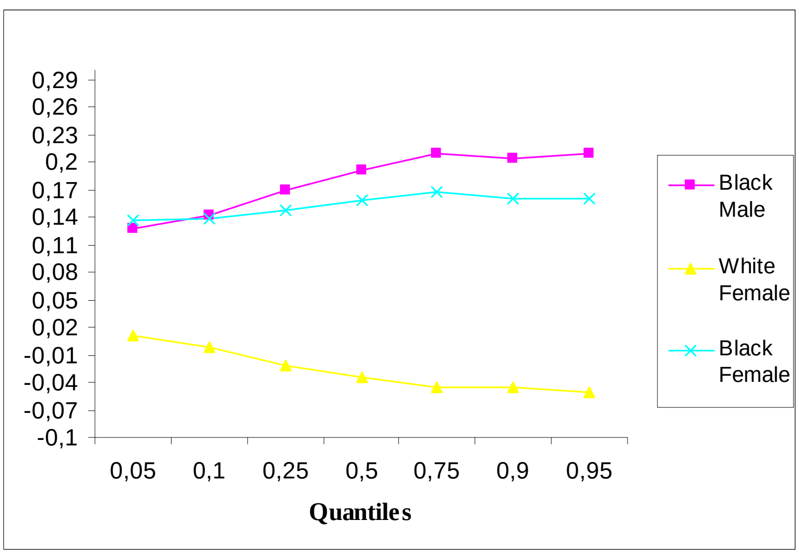

Figure 4 - Wage differentials due to discrimination