Accounting for Labor Income Differences in Brazil: the role of Human Capital

Juliana Ferraz Guimarães

Universidade Federal de Pernambuco, Departamento de Economia *

Tiago Cavalcanti

Universidade Federal de Pernambuco, Departamento de Economia

Raul Silveira Neto

Universidade Federal de Pernambuco, Departamento de Economia

July 2006

Abstract

This paper investigates the sources of wage inequality across Brazilian regions. Duarte, Ferreira and Salvato (2003), using a semi-parametric approach based on DiNardo, Fortin and Lemieux(1996), show that income differences across the Southeast and Northeast can be explained by educational disparities across their population. This result implies that any policy recomendation should prescribe direct investments in education in order to narrow the regional income gap. In this paper we apply the methodology developed by Machado and Mata (2004), based on quantile regression and resampling techniques, for the metropolitan areas of Brazilian Southeast and Northeast. The technique allows us to decompose the wage inequality across regions into a part that can be attributed to changes in workers' observable characteristics and another to changes in the returns to these characteristics. First we show that the difference between NE and SE wage distribution has a higher mass on the lower tail, and lower density on the upper tail, resulting from the well known fact that Northeast have a higher number of people living on poverty and lower number of people on high paying occupations compared to Southeast. Both returns to skills and workers' observed characteristics have an impact on this result. However, different from what Duarte *et al* (2003) show on their paper, schooling did not respond for this wage inequality across the two regions. Returns to schooling did have a higher impact on wage inequality than the distribution of education *per se*. These results imply that investments on education alone may not be enough to narrow regional income inequality since the availability of capital and quality of education may be driving those results.

Key words: Regional wage inequality, education, quantile regression. JEL Classification: J30, J31, J38, C01, I20

*Corresponding author: Juliana Guimarães, Universidade Federal de Pernambuco, Av.dos Economistas, S/N, Cidade Universitária, Recife-PE -Brazil, 50740-590, guimaras@fe.unl.pt.

Abstract

Este artigo investiga as causas da diferença salarial entre as regiões Brasileiras. Usando uma abordagem semi-paramétrica baseada em DiNardo, Fortin e Lemieux (1986), Duarte, Ferreira e Salvato (2003) mostram que as diferenças de renda entre o Nordeste e o Sudeste podem ser explicadas por disparidades regionais no nível educacional. Este resultado implica que políticas de combate a desigualdade regional de renda devem focar-se no investimento a educação para solução do problema. Neste artigo usamos a metodologia de Machado e Mata (2004), baseada em regressão quantílica e técnicas de reamostragem para dados do Nordeste e Sudeste do Brasil. Esta técnica nos permite decompor a desigualdade regional de salário em uma parte que é atribuída a diferenças de características dos trabalhadores e outra devida ao preço dessas características no mercado de trabalho. Primeiro mostramos que a subtração da distribuição de salário da região Nordeste menos a do Sudeste, apresenta uma maior massa nos quantis inferiores da distribuiçã e uma menor densidade nos quantis superiores. Isto resultado direto do fato de que o Nordeste têm um número maior de pessoas vivendo em situaçã de pobreza e menos pessoas em ocupações com altos salários, comparado ao Sudeste. Os retornos das características e as próprias variáveis demográficas são determinantes têm um impacto direto neste resultado. Quando se trata do nível educacional entretanto, diferente do resultado de Duarte, Ferreira e Salvato (2003), achamos que o retorno para os anos de escolaridade no mercado de trabalho têm um impacto mais forte do que as diferenças de níveis educacionais *per se* . Este é um resultado muito importante desde que o acesso ao capital assim como diferenças de qualidade de ensino podem estar levando a diferenças em retorno. Desta forma políticas públicas de combate a diferenças regionais baseadas apenas em investimento a educação podem na realidade não atacar o problema.

Palavras-chave: desigualdade reional de salário, educação, regressão quantílica . Clasificaçâo JEL: J30, J31, J38, C01, I20

1 Introduction

Recent research has shown that half of income inequality can be attributed to differences in human capital and occupational characteristics across regions [1]. This is an interesting result, especially when it comes to policies designed to narrow regional inequality and fight poverty.

Using a semi-parametric approach based on DiNardo, Fortin and Lemieux(1996), Duarte *et al* (2003) in particular, show that income differences across the Southeast and Northeast in Brazil, can be explained by educational disparities across the population. This result is extremely important when it comes to policy recomendation since it implies that any policy deseigned to narrow regional inequality should focus basicaly on investiments in education.

Using the methodology developed by Machado and Mata (2004), from now on MM, to Brazilian data, this paper seeks to deepen the empirical analysis on regional income inequality in metropolitan areas in Brazil. In order to do that, we decompose the observed income inequality across Northeast and Southeast, into a part that can be attributed to differences in workers characteristics (not only education) and another part that can be attributed to differences in labor market returns to theses characteristics. Moreover, we can obtain this decomposition for different quantiles of the conditional wage distribution.

Our results show that both regional differences in workers attributes as well as the difference in returns to these attributes explains the observed income inequality between SE and NE. However, the returns to workers characteristics had a stronger impact on this inequality, compared to the characteristics *per se* . This is true for all quantiles of the conditional wage distribution. Moreover, when we focus on the role of education on the regional wage inequality, we observe that differences in the distribution of education across regions had a smaller impact compared to the difference in returns to education between regions. This results go deeper on the discussion of the role of education on regional inequality, since the importance of returns to education tell us that there is more than simply human capital inequality across the two regions. Therefore policy recommendations that prescribe investments on education alone may not be enough to narrow the regional gap.

The present paper has the following structure: the next section will present some important facts underlying the evolution and persistence of regional income inequality in Brazil. The third section describes the methodology, the fourth section the analysis of the data and the remaining sections the analysis of the results and concluding remarks.

2 Regional Inequality in Brazil

The study of regional income inequality in Brazil may be summarized in few words : extent, severe consequences to welfare and persistence over decades. [2]. Silveira Neto and Campelo (2003) show that Northeast's *per capita* income is little more than half (56%) of the Brazilian *per capita* income. Moreover, it is not even half of Southeast's *per capita* income (46%). When it comes to the state level, data from *Pesquisa Nacional por Amostra de Domicílio (PNAD) 2002* show that Maranhão's *per capita* income is about 34% of the Brazilian's income and 25% of São Paulo's income.

[1] See for instance Servo (1999), Azzoni and Servo (2001), Barros (2002), Miranda *et al* (2002), Silveira Neto and Campelo (2003), Duarte, Ferreira and Salvato (2003).

[2] For a detailed analysis concerning reginal inequality since beginning of the century, its evolution on the 1990's and its long term evolution see Silveira Neto and Campelo (2003), Azzoni and Servo (2002) and Azzoni (1997)

International data show that the extent of regional inequality in Brazil is above the ones observed in other countries. Using the ratio between *per capita* income of the richest and the poorest state, Shankar and Shah (2003) show that for a pool of 8 countries, the Brazilian income inequality is one of the highest, loosing position only to Russia and China. The persistence over time and extent is certainly the major characteristics of regional income inequality in Brazil. Data from PNUD and Brazilian 2000 Census, show that there is a close relationship between regional income inequality and the percentage of people living under poverty. Not only all of the Northeastern states have more than 50% of their population living on poverty, a percentage that is way above the one in São Paulo, but also, poverty is much more intense in the Northeast. People living on poverty in the Northeast are bellow poverty line compared to other regions [3].

Here we will access the role of workers characteristics and their price on the labor market on income inequality across the Brazilian Norhteast and Southeast. Next section describes the methodology.

3 Methodology

The model estimated here directly follows the methodology introduced by MM (2004). They proposed a method based on the estimation of the marginal wage density in a given year implied by counterfactual distributions of some or all of the workers' measured characteristics. Here we apply this same methodology to Brazilian data, in order to have and idea on the impact of both the differences in the distribution of skills and the returns to this skills on the regional wage disparities.

The MM approach can be seen as an extension of the DiNardo, Fortin and Lemieux (1996) methodology. Their approach consists of the estimation of the wage distribution conditional on the variables of interest, which is accomplished using quantile regression, i.e. through the estimation of models for the quantiles of the conditional wage distribution. After obtaining estimates for each quantile of the wage density we can marginalize the conditional wage distribution using different scenarios for the distribution of worker's attributes. By comparison of this counterfactual with the marginal density observed, we have an idea on the impact of each attribute and their returns on the overall change in the wage schedule.

3.1 The details of the technique

The counterfactual decomposition of changes in the wage density is obtained by means of quantile regression and resampling methods. It can be divided into 4 steps: (1) the estimation of the conditional wage distribution; (2) the estimation of the marginal densities implied by the model for both regions; (3) the counterfactual densities; (4) the decomposition of changes into the contribution of coefficients, covariates and a residual. The following subsections explain each one in detail.

3.1.1 First Step: Conditional Wage Distribution

The starting point is the estimation of the wage distribution conditional on the covariates of interest which is obtained using the quantile regression technique developed by Koenker and Basset (1978).[4]

[3]Notice that a person is considered to be under the poverty line if he/she has income bellow R$ 75.00

[4]Application of quantile regression methods to Mincerian wage equation can be traced back to Buchinsky (1994), Chamberlain (1994), Mwabu and Schultz (1996), Fitzenberger and Kurz (1997) and recently to Arias, Hallock and

The first step is to specify the τ^{th} conditional quantile of the distribution of (log) hourly wage ω on a vector X of workers attributes as $Q_\omega(\tau|x)$ for $\tau\epsilon(0,1)$. The set of individual attributes X used in the empirical analysis consist of information about schooling, job experience and indicator variables for gender, race, immigration, marital status, union status, formal labor contract[5], and economic sectors (service,commerce, agriculture and public administration, where industry was the one left out). Based on that we specify the conditional quantile function as:

$$Q_\omega(\tau|x) = x'\beta(\tau) \tag{1}$$

where $\beta(\tau)$ is the quantile regression coefficient which will be estimated by solving:

$$\min_{\beta\epsilon\Re} n^{-1} \sum_{i=1}^{n} \rho_\tau(\omega_i - x_i'\beta)$$

where,

$$\begin{cases} \tau u & for \quad u \geq 0 \\ (\tau - 1)u & for \quad u < 0 \end{cases}$$

Note that just as $\hat{\beta}_{OLS}$ gives us the mean returns to individual characteristics in a Mincerian equation, quantile regression estimate $\beta(\tau)$ presents the return to these characteristics for different quantile of the wage distribution. Also, each conditional quantile $Q_\omega(\tau|x)$ will return different points of the conditional wage density and thus provide a full picture of the wage distribution conditional on the covariates.

3.1.2 Second Step: Marginal Density of Wages

This part of the estimation is designed to obtain the marginal density [6] of wages based on the conditional distribution estimated above. To better understand this process notice that, as observed by MM, the marginal density could be easily obtained from the data, but it would not be conditioned on the covariates specified above and therefore would not allow us to proceed with the counterfactual exercise.

Recall that if U is a uniform random variable $U \ \epsilon \ [0,1]$, then $F^{-1}(U)$ has distribution F. Thus given a sample of $\tau_1, \tau_2, ..., \tau_m$ from a Uniform (0,1) such that the conditional quantile of wages given x, $\{x'\beta(\tau_i)\}_{i=1}^m$ consisted of a random sample from the conditional distribution of wages given $X = x$, estimated in step 1. Here instead of arbitrarily choosing a specific X we resorted to resampling methods and bootstraped a sample of covariates from an appropriate distribution. The algorithmic below summarizes this procedure:

Consider two different regions, Northeast and Southeast such that r=NE, SE, and $\omega(r)$ and $x(r)$ denote wages and p covariates at an specific region. Moreover, let $g(x;r)$ be the joint distribution of covariates at region r such that we are able to obtain a random sample from the wage density that would prevail at region r if model (1) was true and covariates were distributed as $g(x;r)$. After that we may:

Sosa-Escudero (2001).

[5]The variable used as a proxy for formal labor contracts was contribution to social security system

[6]The density estimations were performed using R **density** function, which follows a kernel-density smoother procedure with a Gaussian kernel and bandwidth $4.24n^{\frac{-1}{5}}min(\sigma, \frac{R}{1.34})$, where n denotes the sample size, σ the standard deviation and R the interquartile range (Silverman, 1986, and Venables *et al*, 1997).

1. Generate a random sample of size m from a Uniform (0,1): $u_1, ..., u_m$.
2. For data at region r, $X(r)$ being the n_{rxp} covariates matrix and u_i obtained above, compute:

$$Q_\omega(u_i|x;r)$$

   that gives us m estimates of the QR coefficients $\hat{\beta}_r(u_i)$.
3. Generate a random sample with replacement of size m from the rows of the covariate matrix $X(r)$: $\{x_i^*(r)\}$, $i = 1, ..., m$.
4. Now estimate

$$\{\omega_i^*(r) \equiv x_i^*(r)'\hat{\beta}_r(u_i)\}_{i=1}^m$$

   which is a random sample of size m from the desired distribution.

3.1.3 Third Step: Counterfactual Densities

The first counterfactual exercise corresponds to the estimation of the density function in the Northeast if all covariates had been distributed as in the Southeast, and workers had been paid according to their northeast wage structure. In order to do this we may follow the algorithm above and in the third step, instead of drawing the sample of covariates from the rows of the Northeast matrix, $X(NE)$, we bootstrap it from the rows of the Southeast matrix, $X(SE)$.

The second kind of exercise is the estimation of the density in Northeast if only one covariate had been distributed as in Southeast. This exercise is not as straightforward as the first one. Consider one specific covariate say $z(r)$, lets say union. We want to estimate the wage distribution that would have prevailed in the Northeast region if z had been distributed as in the Southeast and all other covariates as in the Northeast. By suggestion of MM (2004), it is convenient to partition the space of the individual covariates into classes. The algorithm for the second kind of counterfactual exercise goes as follows:

1. Follow step 1 to 4 of the previous algorithm to generate a sample of size m of wage densities at $r = NE$: $\{\omega_i^*(NE)\}_{i=1}^m$.
2. Take one class $C_1(SE)$ (*say formal)*

   (a) Select a subset of the random sample of wage generated on step 1 corresponding to the space $I_1 = \{i = 1, ..., m|z_i(NE)\epsilon C_i(SE)\}$ that is $\{\omega_i^*(NE)\}_{i\epsilon I_1}$.*(Select a sub-sample of formal workers from the marginal densities of wages estimated on step(1)).*

   (b)Generate a random sample of size m x $f_1(SE)$ [7] with replacement from $\{w_i^*(NE)\}_{i\epsilon I_i}$. (*That is, generate a random sample from the sub-sample of formal workers' wage in the Northeast obtained on last step, equal to the number of formal workers in the Southeast.)* [8].

[7]These will gives us the proportion of formal workers in the Southeast. Notice that $f_j(t)$ are the frequencies for each class and clearly $f_1(SE) + f_2(SE) + ... + f_J(SE) = 1$.

[8]This sample gives us the counterfactual wage distribution if the level of formal had remained as in the Southeast and workers have been paid as in the Northeast. For more on this kind of counterfactual exercise see DiNardo, Fortin and Lemieux (1996).

3. Repeat step 2 for $j = 2, ..., J$. (*Repeat step 2 for informal workers*)
4. Assemble all wage samples together.

Formally the density of wages that would have prevailed in the Northeast if, for instance, formal workers had been distributed as in the Southeast is given by:

$$\int f(\omega|z, r_\omega = NE) dF(z|r_z = SE)$$

where $f(\omega|z; r_\omega = NE)$ denotes the distribution of wages in the Northeast for formal workers z, and $F(z|r_z = SE)$ is the distribution of formal workers in Southeast. The Northeast marginal distribution of wages can therefore be written as a composition of the wage distribution for formal and informal workers with weights equal to the proportion of formal and informal workers in Southeast. Just as in DiNardo, Fortin and Lemieux (1986), we can manipulate those weights, based on the frequencies, in order to estimate different counterfactual densities.

3.1.4 Fourth Step: Decomposition of Wage Density

Finally, after obtaining the desired counterfactual densities we are able to decompose the overall wage difference across regions into a part attributable to coefficients, (βs), another to the covariates (Xs) and a last one to the residual. Moreover, the impact of covariates can be decomposed for each one of the individual attributes considered in the model.

Let $f(\omega(r))$ denote the estimator of the marginal density of ω (log of hourly wages) at region r obtained by the sample, i.e. the empirical density, and $f^*(\omega(r))$ an estimator of the density of ω at r based on the generated sample $\{\omega_i^*(r)\}$, i.e. the marginal implied by the model. Extending this notation to the counterfactual distributions we may have: $f^*(\omega(NE); x(SE))$ as the density that would have prevailed in the Northeast if all covariates have been distributed as in the Southeast and the workers have been paid as in Northeast. Also, $f^*(\omega(NE); z(SE))$ the wage density that would prevailed in the Northeast if only one covariate, z, has been distributed as in the Southeast.

Let $\alpha(.)$ be an usual summary statistics such as quantile or scale measure, we may decompose the changes from $f(\omega(SE))$ to $f(\omega(NE))$ into:

$$\alpha(f(\omega(NE)) - f(\omega(SE))) =$$

$$\alpha[f(\omega(NE) + f^*(\omega(NE); x(SE)) - f^*(\omega(NE); x(SE)) - f(\omega(SE))] =$$

$$\underbrace{\alpha(f^*(\omega(NE); x(SE))) - \alpha(f^*(\omega(SE)))}_{coefficients} + \underbrace{\alpha(f^*(\omega(NE))) - \alpha(f^*(\omega(NE); x(SE))}_{covariates} + Residual$$

This decomposition will then give us different measures for the contribution of the coefficients, the covariates and an unexplained part (residual) on the differences in the conditional wage distribution observed across the regions. After that we will be able to study the sources of disparities across the conditional wage distribution in Brazil related to changes in the distribution of human capital factors and changes in the distribution of their returns (QR coefficients) for different quantiles of the conditional wage density.

4 Data

This paper uses data from the 2002 Brazilian household survey called *Pesquisa Nacional por Amostra de Domicílio (PNAD)* , conducted by the *Instituto Brasileiro de Geografia e Estatística - IBGE* in September of each year. Data are collected from a representative national sample of households and contain information on personal characteristics such as age, gender, education, marital status, monthly wage, hours worked, union status, job sector, contribution to social security, actual labor market experience and others.

Our sample consists of workers on the metropolitan areas of both Brazilian Northeast and Southeast regions (17.000 for the NE and 21.000 for the SE), between 10 and 65 years old that reported a positive wage in the week preceding the survey. All nominal data were deflated using the Living Cost Index, estimated in Azzoni *et al* (2000) [9].

From the basic statistics reported on Table 1 [10] we notice that workers on the NE earn on average 40% less than those on the SE and this reflects the fact that the Northeast conditional wage distribution (solid line) remains to the left of the Southeast distribution as we can see form the first plot of Figure 5. This difference is even stronger when we compare hourly wages, besides the fact that there is no major differences between numbers of hours worked for the NE and SE. There is also slightly differences between the numbers of years of labor market experience where NE stays a little behind with approximately one year less of experience per worker on average. Also, average education differs very little (less than one year) from one region to another, and this is due to the fact that the NE has a higher percentage of people with few years of education compared to the SE (bottom of Table 1). We can also analyze the education distribution on both regions trough the plots on Figure 2. Notice that both distributions do not differ widely except from the lower tail where the NE present a higher mass (NE dotted line).

Racial composition is widely different from both regions since the SE has almost twice the number of white people compared to the NE. When it comes to immigration, we notice that while the SE had on average 27% of immigrants on 2002 the NE had no more than 8% that year. Notice also that informality seems to be higher in the NE and so is the presence of agriculture, considering we are dealing with urban areas only.

5 Results

We will start this section analyzing the quantile regression estimates for our model for some specific covariates. In Figure 1 in the appendix the plots on the left show the quantile regression (QR) estimates as well as the 95% confidence intervals for the Southeast Region. On the center we have the same kind of plots for the Northeast and on the right the difference in coefficients for the NE and SE. The least square estimates are presented as the dotted horizontal line in each plot.

As we can see the returns to schooling in both regions follow the international pattern verified in Buchinsky (1994), Chamberlain (1992), Machado and Mata (2004), increasing with the quantiles of the conditional wage distribution. Notice also that these returns are more dispersed for the Northeast region, being lower than the Southeast returns in the lower tail and higher in the upper tail of the conditional wage distribution. This is an interesting result since we cannot really see

[9]We would like to thank Tatianne Menezes who kindly gave us those indexes.

[10]All Figures and Tables are on the Appendix of the paper.

huge differences between the distribution of education in both regions (Figure 4). White workers are receiving higher returns in high paying jobs compared to no whites and this is true for both NE and SE. Experience follows the same pattern as education, increasing with the quantiles, which shows that more experienced workers are receiving higher returns on high paying jobs compared to the ones on low paying jobs.

Workers on low-paying jobs receive higher returns for having a formal contract than the ones on high-paying occupations (Figure 1 -cont.). This is probably due to the lack of opportunities that workers in the bottom of the conditional wage distribution face, relative to those in the upper tail. High paid workers are more likely to be high educated, high ability workers for whom informality is more of an option, especially for those willing to evade mandated benefits. This is not true for low-paid, low-skilled workers, for whom a formal contract may be the best alternative, given that their wage will be complimented by benefits such as , retirement plans, health insurance, maternity leave, and others. Notice that, one more, the only difference between regions relays on the fact that on the NE, workers on the lower tail receive higher returns compared to the ones on the SE. Immigrants receive lower returns on low paying jobs on the NE, and also, compared to the SE, immigrants on the upper tail receive higher returns.

After analyzing the QR plots we present the results for the decomposition we described in the methodology. The plots on Figure 3 show this decomposition graphically. On the first plot we have the difference between the Northeast and Southeast empirical densities $(f^*(\omega(NE)) - f^*(\omega(SE)))$. As we notice the difference is positive for lower incomes and negative for higher ones, resulting from the fact that Northeast have more people leaving with low income compared to the Southeast and way less rich people than on the Southeast. The question that we are trying to answer here is what drives this observed results for our estimated conditional wage distribution? Based on the decomposition developed by Machado and Mata (2004) we can see the impact of both differences in workers characteristics across regions and differences in the job market returns to these observed characteristics.

The results show (second and third plot in Figure 3) that both the returns and the observed workers characteristics (Xs) had some influence on the regional differences on wage distribution showed on the first plot (notice the same patten of change). However, the influence of the job market returns (the β s), is stronger. This is clear when we analyze the difference from the numbers on Table 2, where we notice that the contribution of the coefficients for most of the quantiles is at least twice the contribution of workers characteristics (covariates). When analyze the impact of the workers characteristics, taken individually, we notice that most of them have little or no impact on the regional labor income differences, except for the racial composition, marital status and immigrantion. The fact that in the SE around 60% of the sample is white whereas on the NE this is only 31%, and that for most of the quantiles white workers on NE earn lower returns for being white compared to the SE (except on the upper quantiles), may be driving the fact that racial composition influences the wage inequality observed between the two regions.

Also, the SE have a higher percentage of immigrants compared to the NE. However from some of the basic statistics for the immigrants on both regions we cannot tell what is driving this results. On average immigrants in the SE have lower education, and are mostly low skilled workers compared to the NE. Therefore here we conclude that the average statistics are uncovering a richer pattern of immigrants distribution in the SE which shifts the wage distribution towards the right of the NE conditional wage density.

The impact of marital status distribution across regions is also an intriguing fact. We may claim that on the NE people get married earlier on average than in the SE and that this fact alone has all sorts of consequences on their job market outcome. Also, from basic statistics we notice that on the NE married people are not only younger, but less educated and have lower labor market outcome, but this is true for the whole sample, not only for married. Therefore when we come to analyze the influence of marital status on labor income inequality across regions, we need to go further and study married women participation on both regions and also number of children per couple and see if those pattern really differ across regions.

Since the impact of the returns to workerss characteristics was stronger than the impact of the workers attributes *per se* we decided to plot the influence of the returns individually (Figure 4). It turns out that all of the returns contributed to the wage differences observed between the two regions. Here we would like to focus on one result specifically. Besides the fact that the differences in the distribution of education had no impact on the wage inequality observed between the SE and NE, the returns to education, has a strong influence on the regional wage inequality studied here. We saw on Figure 2 that the schooling distribution differ very little between the NE and the SE. However, the fact that on the NE workers are earning lower returns compared to the ones in the SE, for most of the quantiles, is driving the wage inequality between this two regions. This can also be verified on Table 2 where the returns to education taken individually responds to 70 to 80% of the labor income inequality across regions and the distribution of education alone responds for the rest 30 to 20%. Returns to experience also, contributed to the difference in wage distribution. Therefore, the lower returns to human capital in the NE is driving the regions wage distribution to the left.

This qualitative results are extremely important in terms of policy recomendations. Notice that the distribution of education does not differ greatly among regions. However, nothing can be said with the information we have here, about the access to capital for every worker or even the quality of education.

Our results show a completely different picture than the results presented by Duarte, Ferreira and Salvato (2003). However, we found that what the authors claim to be the consequence of differences in educational distribution across both regions, is indeed is due to the differences in returns to years of education. In order to study more deeply the differences in both results, ours and Duarte et all (2003), we decided to follow closer their decomposition and we end up finding that when they are comparing CE/SP x SP and also, NE/SE1 x SE1 (or even NE/SE2 x SE2) what they are doing is:

$$\alpha(f^*(\omega(CE); z(SP))) - \alpha(f^*(\omega(SP)))$$
$$\alpha(f^*(\omega(NE); z(SE))) - \alpha(f^*(\omega(SE)))$$

This would be basically accounting for the impact of the returns to workers characteristics, that is the βs, in their case, specifically returns to schooling. Most of their results show impacts that ranges from 50% to 80% across quantiles, which is basically what our results show here. Therefore we found that what they claim to be due to differences in educational distribution is indeed a consequence of the returns to education. We go further and claim that most of the wage inequality observed across both regions is a consequence of the differences in the job market returns paid on both NE and SE. Lower returns to human capital and job characteristics in the NE which is certainly a matter of production process, is contributing for the weak performance of the Northeast Brazil. Here verified these facts. In order to test for the reasons for this lower labor market returns in the NE we will need to have other controls such as amount of physical capital by workers in each regions and also on the quality of education.

The importance of the results found in our paper relies on the fact that policy recommendations that focus solely on investments on education may have small impacts on regional income inequality since it desregards the fact that what is driving the wage inequality across region is not the distribution of education alone, but workers labor market outcome, regardless their level of schooling.

6 Conclusion

This paper analyzed the regional income differences in Brazil under the lenses of labor market outcomes. We verified that the poor performance of the Northeast region can be attributed mostly to the lower labor market returns received by workers compared to the Southeast region. Workers characteristics altogether did have some influence on the regional labor income differences observed. However, few of the workers characteristics taken separately, can explain the fact that Northeast wage distribution remains to the left of the Southeast distribution.

At first it seems that these results would go completely against the results found recently by Duarte, Ferreira and Salvato (2003). However analyzing their results through the lenses of Machado and Mata (2004) decomposition, it is clear that what they claim to be a consequence of educational regional differences, is more of a matter of differences in returns. We went further and verified that not only returns to education are contributing to this outcomes, but that returns to human capital and job characteristics (sectors, type of contracts, etc.) as a whole, are greatly influencing regional wage inequalities.

Notice that policies recommendations regarding regional inequality may be completely different whether we claim the difference in educational level to be responsible for the regional wage inequality, or we consider the returns to be driving those results. Future work should focus on the reasons for this differences in returns, and therefore should use controls that would account for differences in productive structure on both regions.

7 References

Arias, O., Kevin Hallock and Walter Sosa-Escudero (2001), "Individual Heterogeneity in the Returns to Schooling: Instrumental Variables Quantile Regression using twins data". *Empirical Economics.*20, 7-40.

Azzoni, C.R and Servo, L. (2002). *Education, cost living and regional wage inequality in Brazil.* Universidade de São Paulo - USP, mimeo.

Azzoni, C.R , Do Carmo, H. e Menezes, T. (2000). "Índice de Custo de Vida Comparativo para as Principais Regiões Metropolitanas Brasileiras: 1981-1999", *Estudos Econômicos,*vol. 30, n.1, jan/mar, p. 1-22.

Barros, A.H. (2002). *Is there a regional problem in Brazil?* Universidade Federal de Pernambuco. mimeo.

Chamberlain, Gary, (1994). “Quantile Regressions, Censoring and the Structure of Wages”. *Advanced Econometrics,* C.Sim ed. 6^{th} World congress, Cambridge University Press,pp.171-209.

Buchinsky, Moshe (1994). “Changes in the U.S. Wage Structure 1963-1987: An Application of Quantile Regression”. *Econometrica,* March, 62(2). pp.405-58.

DiNardo, John, Nicole M. Fortin and Thomas Lemieux, (1996).“Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach”. *Econometrica.* 64(5). September, pp.1001-44.

Duarte, Angelo José, Pedro C. Ferreira and Márcio Salvato, (2003). *Regional or Educational Disparities? A Counterfactual Exercise.* Ensaios Econômicos n.552, Fundação Getúlio Vargas - EPGE. Brazil.

Koenker, Roger and G. Bassett, (1978). “Regression Quantiles” *Econometrica.* 46(1), pp.33-50.

Machado, José and José da Mata,(2000). “ Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression”. *Journal of Applied Econometrics.* (forthcoming).

Servo, L. (1999). “Diferenciais regionais de salários no Brasil”. *Master Dissertation,* Depto. de Economia, Universidade de So Paulo.

Shankar, R. and A. Sha (2003).”Bridging the economic divide within countries: a scored on the performance of regional income disparities”, *World Development,* vol.31, Iss. 8, august, p.1421-1441.

Silveira Neto, R.M. and A. K. Campelo (2003). “Radiografando as disparidades regionais de renda no Brasil: evidências a partir de regressões quantílicas”, *Revista Econômica do Nordeste,* v.34 n. 3, jul.-set., p.359-378.

8 Appendix

	NE	**SE**
Observations	17195	21390
Monthly Wages	509.79	833.15
	(954.45)	(1296.72)
weekly hours worked	42.329	42.58
	(15.72)	(13.08)
(log) Hourly Wages	0.4664	1.022
	(1.025)	(0.9442)
experience	19.29	20.53
	(12.74)	(12.89)
education	8.93	9.63
	(4.29)	(4.10)
gender(women=1)	0.441	0.427
	(0.496)	(0.494)
married	0.7060	0.7317
	(0.4557)	(0.4947)
white	0.3143	0.5930
	(4642)	(0.4912)
union	0.1720	0.1675
	(0.3774)	(0.3726)
formal	0.5048	0.6224
	(0.499)	(0.484)
immigrants	0.085	0.269
	(0.28)	(0.44)
commerce	0.2152	0.1910
	(0.4110)	(0.3931)
services	0.390	0.374
	(0.487)	(0.484)
pub. administration	0.057	0.050
	(0.232)	(0.218)
agriculture	0.017	0.006
	(0.13)	(0.08)
Years of Education(%)		
0	7	3
1	8	6
5	10	12
8	26	25
15 or more	9	1

TABLE 1 -Summary Statistics - sample averages; sample standard errors in parenthesis

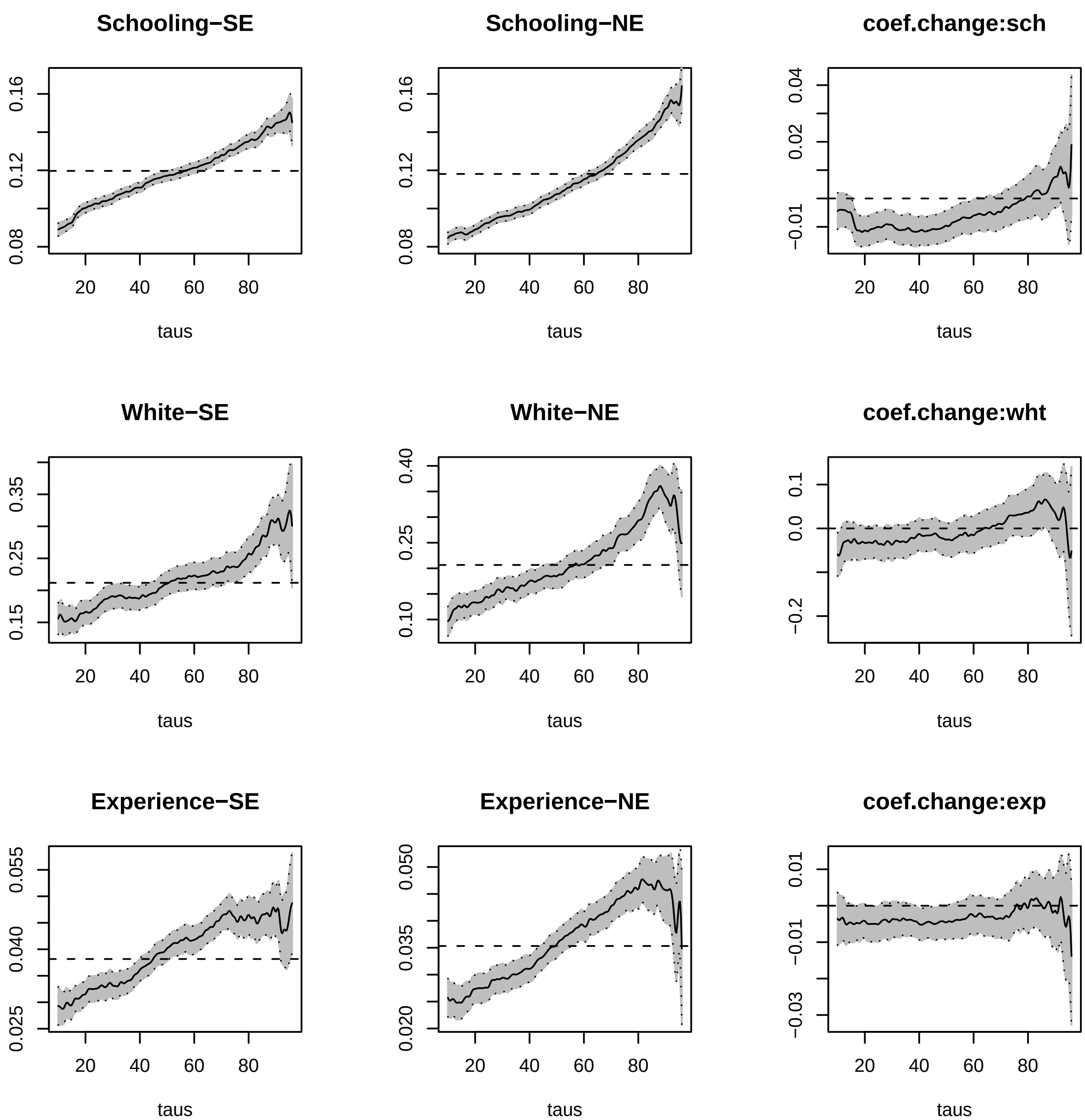

Figure 1:The above plots present the quantile regression estimates for each individual covariate indicated. The solid lines are the 95% confidence intervals and the dotted line presents the Ordinary Least Square estimates. The plots on the left bring the estimates for the Northeast the ones at the center the Southeast estimates and the plot on the right the difference between NE and SE estimates.

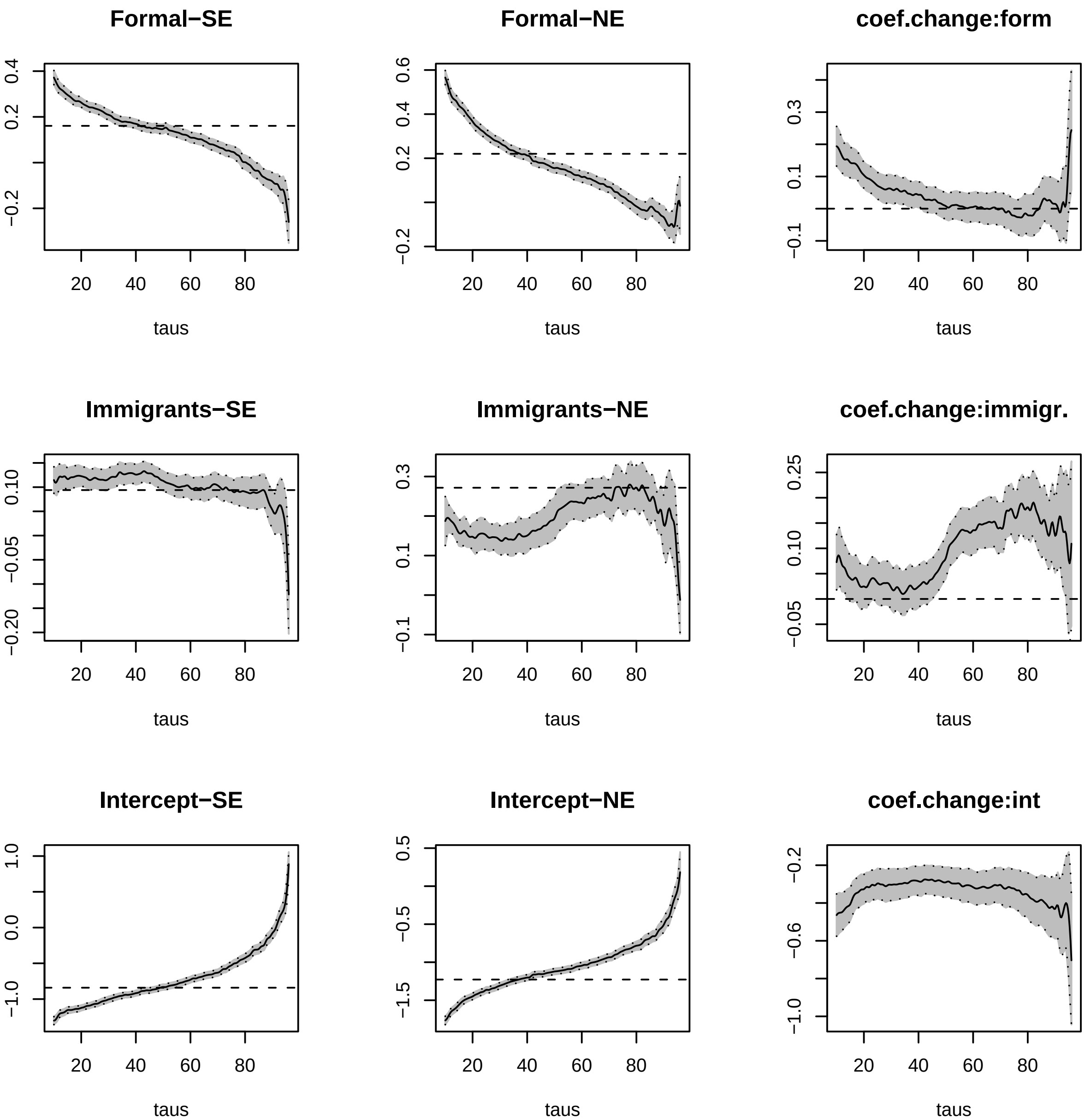

Figure 1(Cont.):The above plots present the quantile regression estimates for each individual covariate indicated. The solid lines are the 95% confidence intervals and the dotted line presents the Ordinary Least Square estimates. The plots on the left bring the estimates for the Northeast the ones at the center the Southeast estimates and the plot on the right the difference between NE and SE estimates.

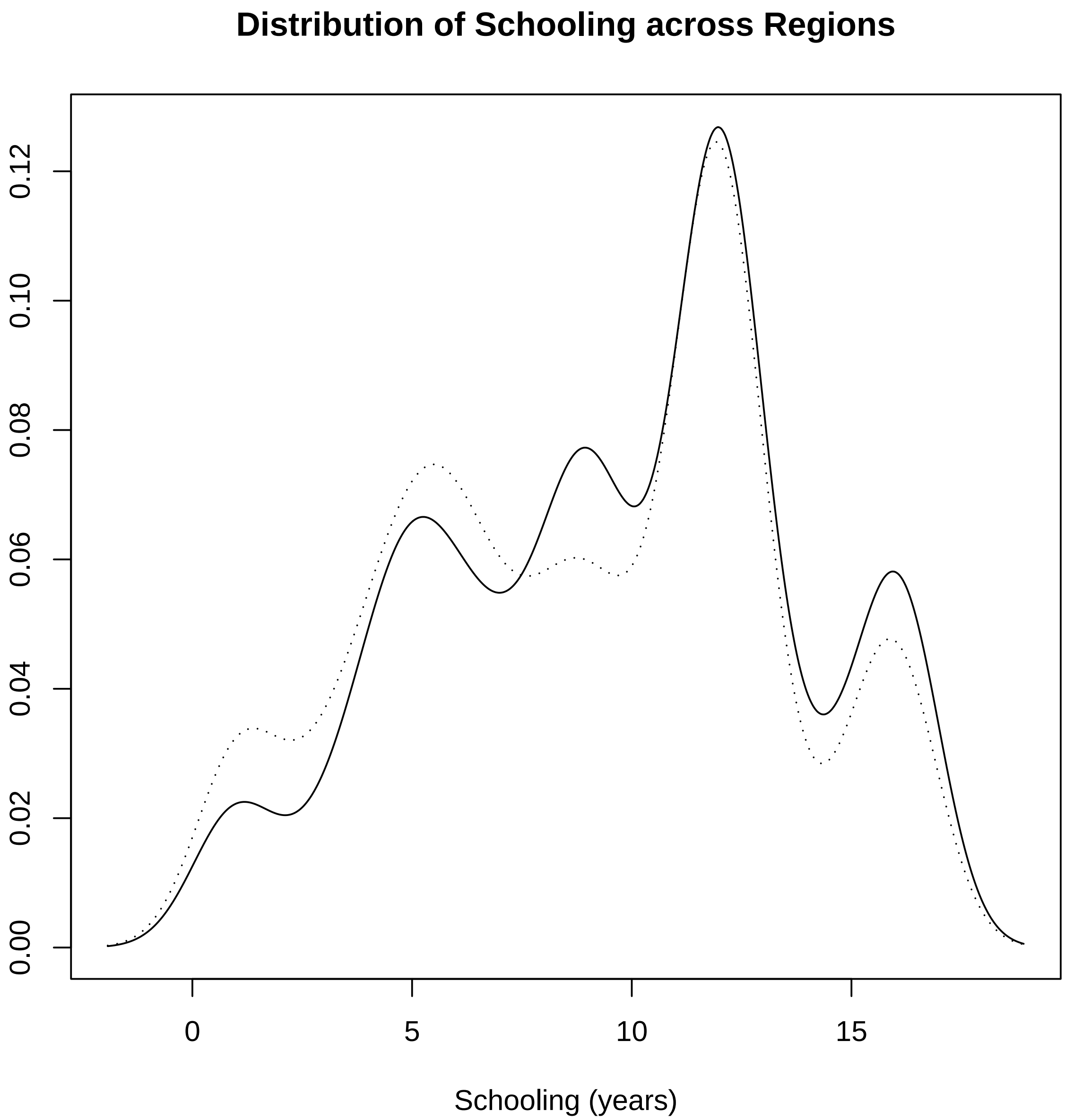

Figure 2: This plots present the Distribution of Schooling in both regions, being NE the dotted line.

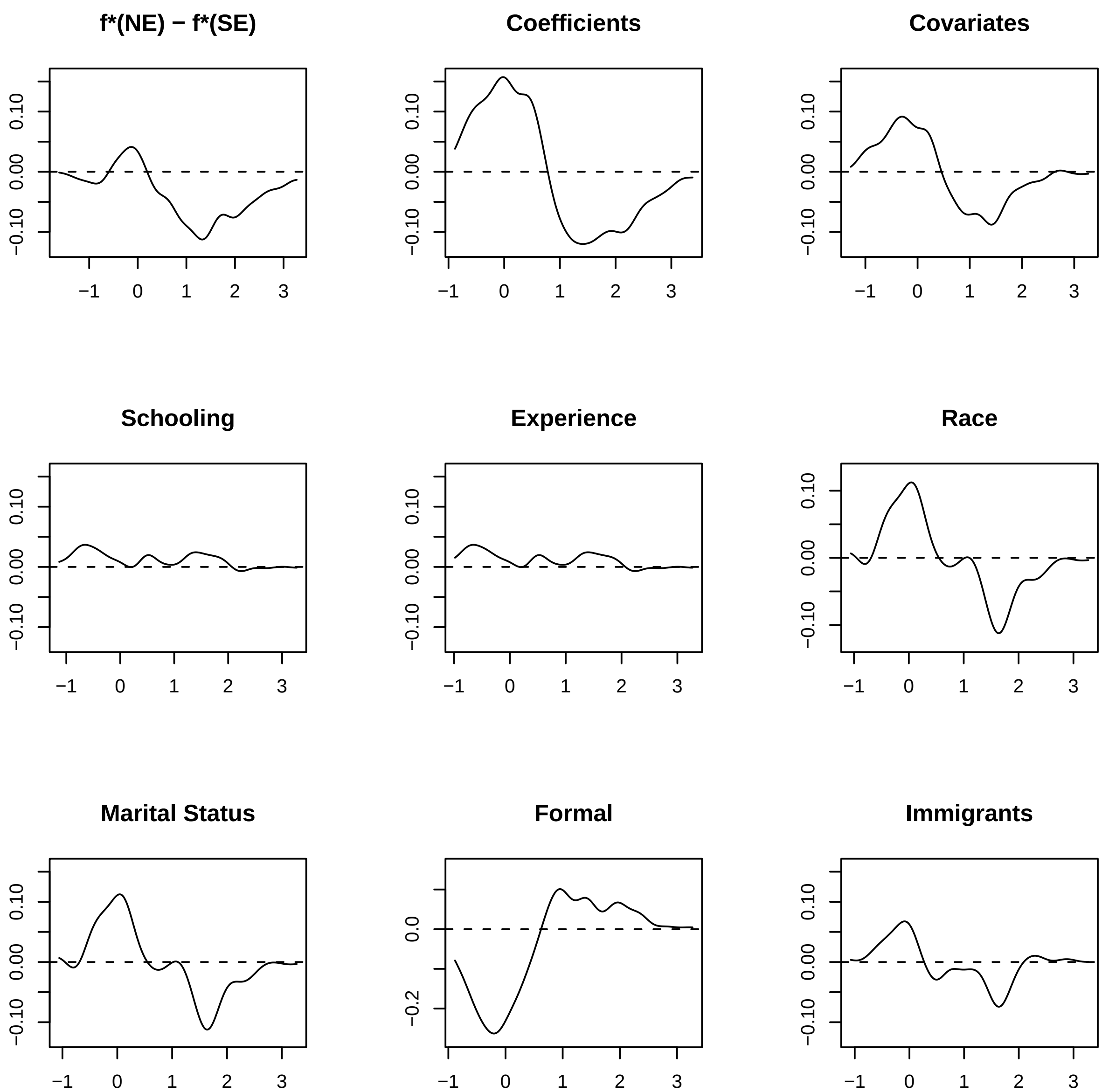

Figure 3: The graphics above describe the differences in densities. The first plot shows the difference in empirical densities from the NE to the SE for each quantile $\alpha(f(\omega(NE)) - f(\omega(SE)))$. The second plot presents the influence of the coefficients which is estimated by the difference between the counterfactual with all covariates in SE and the estimated SE density $\alpha(f^*(\omega(NE), x(SE) - f^*(\omega(SE)))$. The third plot brings the influence of the covariates in the overall difference. This is estimated by the difference between the estimated density in the Northeast and the counterfactual with all covariates in SE $\alpha(f^*(\omega(NE)) - f^*(\omega(NE), x(SE)))$. The remaining plots repeat the analysis on plot 3 for specific covariates $\alpha(f^*(\omega(NE)) - f^*(\omega(NE), z(SE)))$.

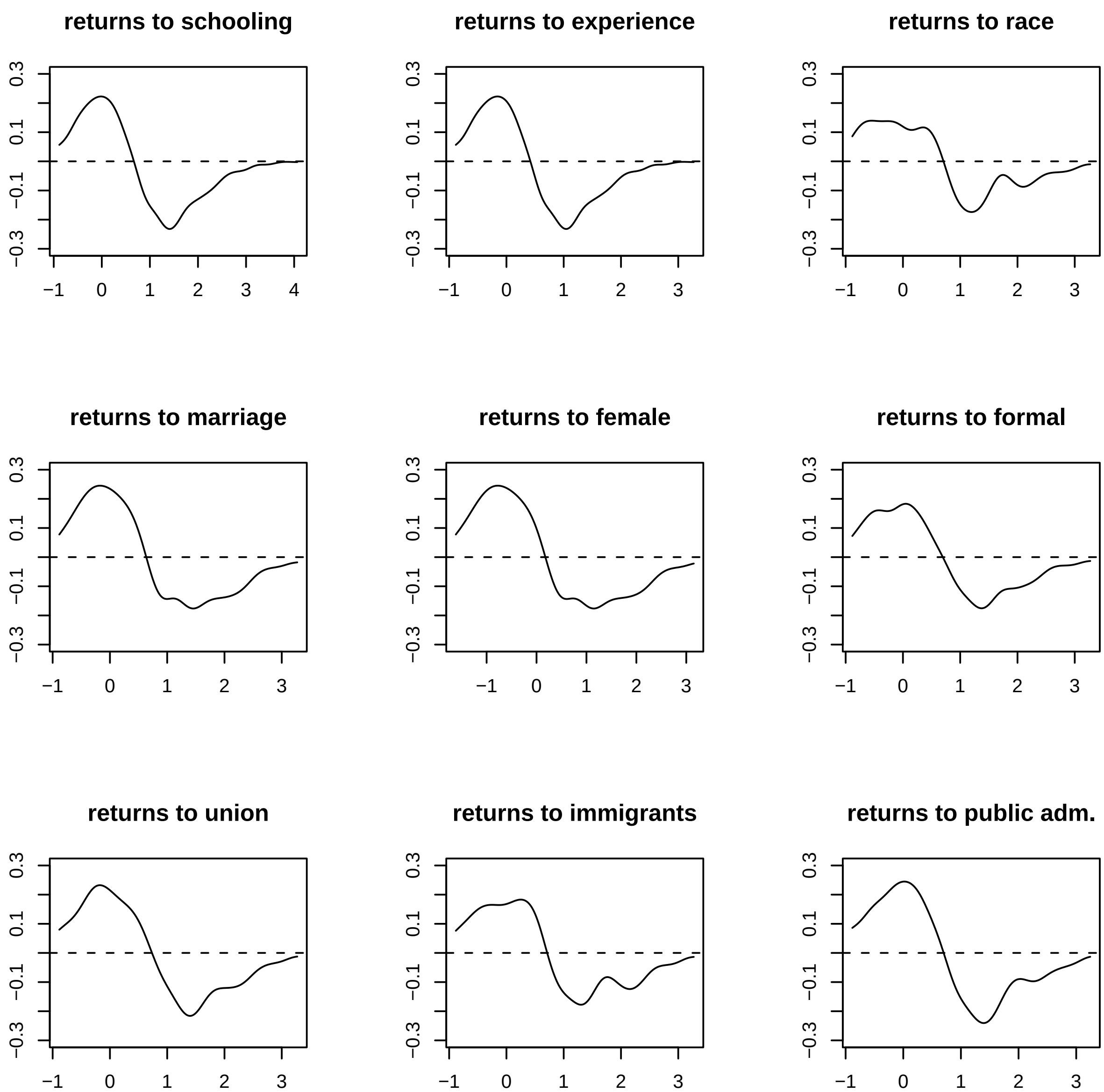

Figure 4: The graphics above presents the influence of the individual coefficients which is estimated by the difference between the counterfactual with one covariates in SE and the estimated SE density $\alpha(f^*(\omega(NE), z(SE) - f^*(\omega(SE)))$.

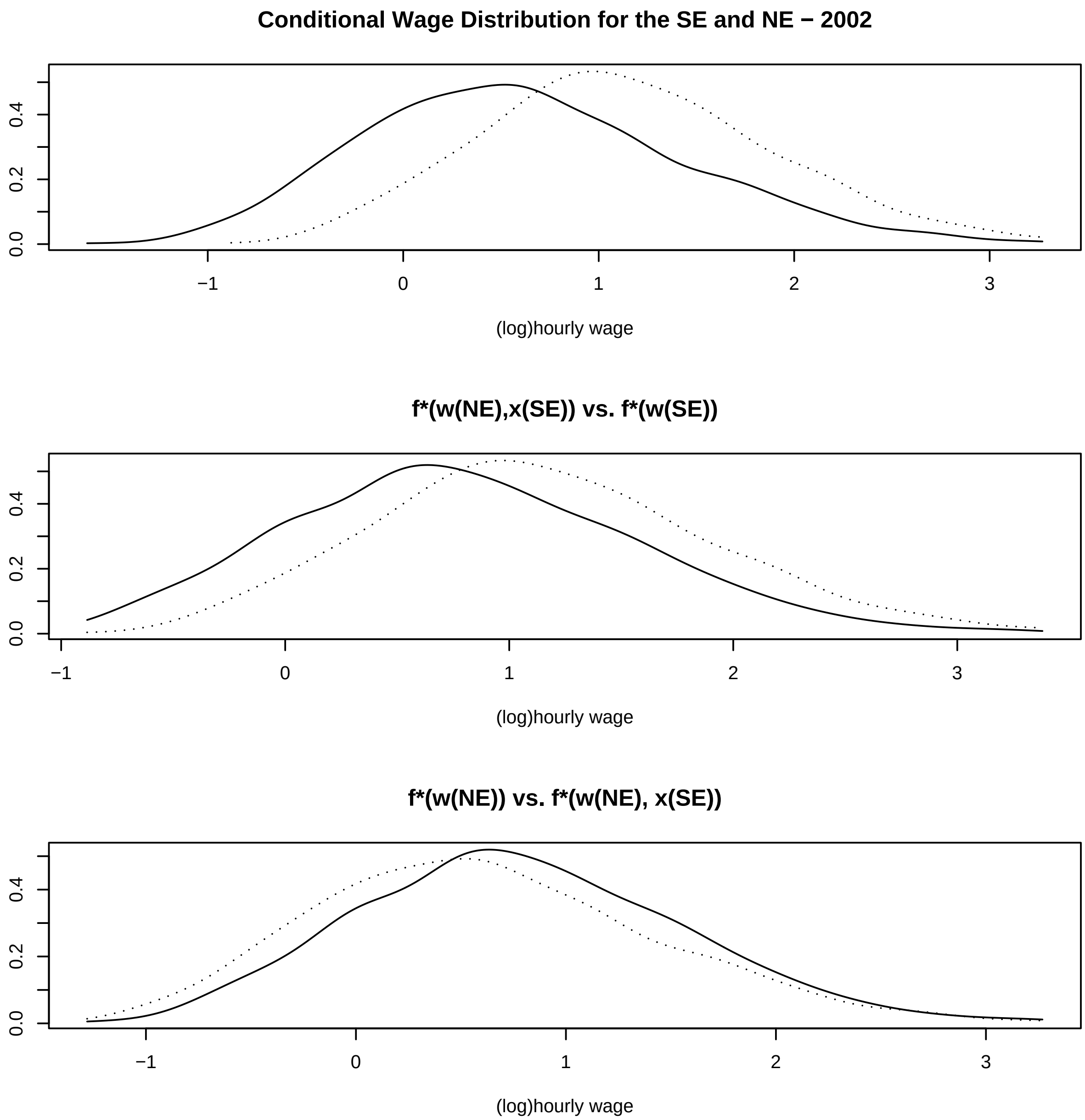

Figure 5: The first plot presents the estimated densities for NE (solid line) and the SE. The second plot brings a comparison between the counterfactual with all covariates as in the SE (solid line) and the estimated SE density, which gives us the influence of the coefficients on the overall regional income difference. The third plot compares the NE density with the counterfactual with all covariates as in the SE (solid line), from which the influence of covariates derives.

	Marginals			Contributions				
	NE	SE	NExSE	Covariates	Coefficients	Residual	Education	Returns to Educ.
10th Quant.	-0.379	-0.189	-0.568	35%	64%	1%	16%	83%
25th Quant.	0.0362	0.627	-0.591	32%	67%	1%	12%	87%
Median	0.558	1.098	-0.540	34%	65%	1%	13%	86%
75th Quant.	1.124	1.628	-0.503	36%	63%	1%	23%	77%
90th Quant.	1.733	2.159	-0.426	15%	85%	0%	83%	11%

TABLE 2 -Decomposition of Wage Densities: The entries in the first three columns corresponds to the quantile of the estimated difference. The entry on the remaining columns shows the proportion of the total change explained by the factor indicated.