Labor Market Discrimination: A Quantile Regression Approach*

Tiago V. de V. Cavalcanti
University of Illinois at Urbana-Champaign
Department of Economics, cavalcan@uiuc.edu
and
Juliana F. Guimarães
University of Illinois at Urbana-Champaign
Department of Economics, guimaras@uiuc.edu

Revised in May 2001

*We would like to thank without implicating Roger Koenker, Kevin Hallock, Wallace Hendricks and participants of the 1999 Brazilian Econometric Association (SBE) meeting for their useful comments on an earlier draft of this paper. We are responsible for the remaining errors.

Abstract

In this study we apply quantile regression techniques to the well-known Oaxaca coefficient of discrimination. Our methodology provides different coefficients for different quantiles of the conditional wage distribution and is more informative than the technique based on OLS regression, which concentrates only on the conditional mean of the wage distribution and provides only an average estimate. Results, using data from the 1999 Current Population Survey, show that the part of the wage gap not explained by differences in skills between the genders is higher at the upper quantiles of the conditional wage distribution, suggesting that "discrimination" against women increases as we move from low to high quantiles. That is, more than twenty years after Oaxaca's (1973) results, we notice that women and men are much more alike in terms of labor market experience, education and attachment to the labor force. Moreover, the data show that in the last twenty years female workers increased their share in predominantly male occupations and decreased their participation in female occupations. During this period, the wage gap actually decreased, however from the little wage gap remaining, almost nothing can be explained by differences in observable skills, and women at higher paying jobs are the most affected by this unexplained wage differential.
Key Words: Discrimination, Oaxaca's coefficient of discrimination, quantile regression.

1 Introduction

Market Discrimination is generally understood to exist when workers who have identical productive characteristics are treated differently because of the demographic group to which they belong . This is the basic idea developed by Gary Becker in his Ph.D. dissertation, subsequently published in 1957 as *The Economics of Discrimination.* [1] Labor market discrimination can take two prominent forms. First, discrimination can exist when wage gaps among different groups, say women and men, are not accounted for by productivity differences; this is so called *wage discrimination.* In addition, employers may discriminate against women in access to training programs or in hiring for particular occupations. This later example is labeled *occupational discrimination.*

A reason why economists and other social scientists have been studying market discrimination is based on an efficiency argument. According to D'Amico (1987), a consequence of market discrimination is that it generates a clear loss of efficiency, since scarce resources are overallocated to relatively unproductive members of the "favored" group (men) and underallocated to more-productive members of the discriminated group, or "target" population (women). Thus society's aggregate output will fall below its potential size and the size of this shortfall will depend, among other things, on the size of the "target" population and how effective is the discrimination. In addition to society's loss of efficiency, market discrimination also imposes costs on the members of the discriminated group. These personal losses include a lower per capita real income, poorer living conditions and lower social status relative to the situation that would have existed in the absence of discrimination.

Additional evidence on the importance of studying labor market discrimination is provided in a number of U.S. Court cases related to sex and race discrimination, since the Equal Pay Act of 1963 and the Civil Rights Act of 1964 (Kuhn, 1987). Ashenfelter and Oaxaca (1987) also show that the statistical techniques used to support legal charges of discrimination are very similar to those empirical methods developed by economists to measure labor

[1] For additional theoretical work see Arrow (1972) and Aigner and Cain (1977). For empirical work see Oaxaca (1973), Cain (1986), Ashenfelter and Oaxaca (1987), Blau and Ferber (1987) and Garcia at al (2001), among others.

market discrimination in the late 1960s and early 1970s, demonstrating the large impact of these methods.

Among the diverse methods of statistical measures of sex and race discrimination, the most often used technique was developed by Oaxaca (1973) and Blinder(1973). Oaxaca's approach follows Becker (1957) and is based on the estimation of a wage function for each group. It attempts to decompose the wage rates of the two groups into a part that controls for differences in the vector of productive characteristics and a part attributable to discrimination. Up until now, only traditional regression methods, which concentrate on the conditional mean wage, have been applied to this problem.[2] Nevertheless, "merely to use some measure of the average discrimination coefficient does not suffice" (Becker 1957,p.18), since this coefficient will depend largely on idiosyncratic factors (as "individual's taste for discrimination"). Thus in order to take into account this hetereogeneity problem, the complete distribution of discrimination coefficients among individuals must be made explicit. The core of this paper is to provide a quantile regression technique where we can have different measures of labor discrimination for different quantiles of the wage distribution, giving, therefore, a more complete picture of the discrimination problem. In addition, as is well known, quantile techniques are more robust to non-Gaussian distributions and heterogeneous observations. Finally, it also allows us to deal explicitly with the heterogeneity problem, which is common in estimating wage functions (for more detail see, for example, Buchinsky (1994) and Chamberlain(1994)).

This paper is unique in the sense that it introduces a more robust and informative methodology to measure wage differentials across different demographic groups. In order to do this it uses a sample from the CPS Outgoing Rotation Group dataset for 1999, providing new empirical results regarding wage discrimination in the U.S.. Defining "discrimination" is obviously very difficult. For the purposes of this paper, we follow Oaxaca and define discrimination as the part of the wage gap that cannot be explained by differences in workers' observed characteristics between the two groups of interest, in our case, men and women.

This paper has four additional sections. Section 2 is divided in two parts;

[2] An exception is Garcia et al (2001) which introduce a quantile regression methodology using a Spanish data set.

in the first we briefly explain Oaxaca's (1973) methodology and some problems this approach presents. In the second part we extend the later approach to quantile regression and provide a solution to the problems discussed earlier. The third section describes our data set and shows the estimations using the original Oaxaca approach. The fourth section presents the results for the proposed estimation. And finally, section 5 discusses implications of this new methodology and offers concluding remarks.

2 An Extension of Oaxaca's Decomposition: The Quantile Regression Approach

2.1 Oaxaca's Technique

The measure of discrimination proposed by Oaxaca (1973) compares the wage rate[3] of two groups, male and female, as they are observed and as they would have prevailed in the absence of discrimination. Then the coefficient of discrimination against women is defined as

$$D = \frac{\frac{W_m}{W_f} - \frac{W_m^0}{W_f^0}}{\frac{W_m^0}{W_f^0}} \tag{1}$$

where $\frac{W_m}{W_f}$ is the observed male-female wage rate and $\frac{W_m^0}{W_f^0}$ is the male-female wage rate in the absence of discrimination.

As we can easily see, since $\frac{W_m}{W_f}$ is observed, to estimate D we must find $\frac{W_m^0}{W_f^0}$. Taking natural logarithms yields the following equation:

$$log(D+1) = log(\frac{W_m}{W_f}) - log(\frac{W_m^0}{W_f^0}) \tag{2}$$

[3] As noted by Ashenfelter and Oaxaca (1987), "it is clear that a similar definition of market discrimination is available for any market outcome, whether it be the number employed, hired, or discharged, or some other measure of compensation."

Oaxaca assumed that some matrix vector X of individual characteristics determines the wage function in the absence of discrimination. He also assumed the following Mincerian parametric relationship between wages and the matrix vector X

$$log(W_i) = X_i'\beta + u_i \text{ for } i = 1, ..., n \tag{3}$$

where β is an unknown vector of regression coefficients and u_i is the disturbance term. In some of the studies of Oaxaca (1973), Cain (1986), Kuhn (1987), Ashenfelter and Oaxaca (1987) and Blau and Ferber (1987) the authors do not assume any parametric form of the disturbance term. However, since they use Ordinary Least Squares (OLS) regression to estimate the coefficient vector β and OLS estimation yields inefficient results in the presence of heteroskedasticity and inconsistent estimates if the error term is non-normal[4], it is implicitly assumed in their models that the disturbance follows a normal distribution with 0 mean and constant variance. Otherwise, their results will be inconsistent and inefficient.

The main focus of these studies is on the characteristic matrix vector X. This issue arises from the discussion about wage discrimination and occupational discrimination. As we noted before employers may discriminate against women in access to training programs or in hiring for particular occupations. According to Cain (1986) the estimate of discrimination is considerably higher when a control for occupation is omitted. One of the main conclusions of Oaxaca (1973: 708) that "We are in agreement with other researchers that unequal pay for equal work does not account for very much of the female-male wage differential. Rather it is the concentration of women in lower paying jobs that produces large differentials", also calls attention for the problem.

From (3) we have that

$$\hat{l}og(\frac{W_m^0}{W_f^0}) = \hat{l}og(W_m^0) - \hat{l}og(W_f^0) = \bar{X}_m'\hat{\beta}_m - \bar{X}_f'\hat{\beta}_f \tag{4}$$

[4]See Greene (1997) chapters 11 and 12.

With the assumption that the current female wage structure would apply to both males and females in a nondiscriminatory labor market, the expected ratio is (Oaxaca, 1973: 696)

$$log(\widehat{\frac{W_m^0}{W_f^0}}) = \Delta\bar{X}'\hat{\beta}_f = (\bar{X_m}' - \bar{X_f}')\hat{\beta}_f \tag{5}$$

and

$$log(\widehat{D+1}) = -\bar{X}_m'\Delta\hat{\beta} = -\bar{X_m}'(\hat{\beta}_f - \hat{\beta}_m) \tag{6}$$

In an analogous way, if we assume that the male wage structure could apply to both males and females in a nondiscriminatory labor market, it can be shown that

$$log(\widehat{\frac{W_m^0}{W_f^0}}) = \Delta\bar{X}'\hat{\beta}_m = (\bar{X_m}' - \bar{X_f}')\hat{\beta}_m \tag{7}$$

and

$$log(\widehat{D+1}) = -\bar{X}_f'\Delta\hat{\beta} = -\bar{X_f}'(\hat{\beta}_f - \hat{\beta}_m) \tag{8}$$

Oaxaca used both approaches in order to determine a range of possible values. There are two econometric problems with this approach. First, we are interested in D, the coefficient of discrimination, instead of $log(D+1)$. We know that OLS regression estimates the expected value of the dependent variable, in this case $log(W)$, conditional on a vector of independent variables, X. Moreover, the fact that $log(\bullet)$ is a concave function implies by *Jensen's inequality* that $E(log(W|X))$ is smaller than $log(E(W|X))$, so we cannot have D by just taking the exponential value of $log(D+1)$, this is just an approximation.[5] This result is noted by Koenker and Portnoy (1996:

[5]Notice that $D+1 = \frac{\frac{W_m}{W_f}}{\frac{W_m^0}{W_f^0}}$ and $log(D+1) = log(\frac{W_m}{W_f}) - log(\frac{W_m^0}{W_f^0})$ so if we estimate $log(W^0)$ by OLS, we cannot use $exp(log(W^0))$ to find D.

31): "one often sees $h^{-1}(x'\beta)$ used in place of $E(y|x)$ in such circumstances, $exp(x'\beta)$ when the model has been specified as $log(y) = x'\beta$, for example, but this is difficult to justify formally".

In addition, Oaxaca's technique is based on least squares estimation of equation (3). Recall that the performance of the OLS regression to estimate the wage equation (3) depends on whether the distribution of the disturbance term u is Gaussian with constant variance over the population. Nevertheless, Bunchinsky's (1994) and Chamberlain's (1994) studies of the U.S. wage structure, using the Current Population Survey and the same parametric form of equation (3), show that the wage structures are different for different quantiles of the wage distribution. In particular, they found that the returns to education and experience are quite different at different quantiles, and they increase for higher quantiles. Following Koenker and Bassett (1982), these differences in the estimates for different quantiles imply that we should not believe in the Gaussian distribution with constant variance of the error term.

Finally, Oaxaca's approach is helpful if the object of interest is the mean, however, it is not very useful for measuring the distributional consequences of discrimination. Back to Becker (1957), we notice that the discrimination coefficient is highly dependent on individual's taste for discrimination, and thus it varies across different points of the wage distribution. In the next part of this session we introduce an extension of the Oaxaca technique, where we address the econometric problems mentioned above.

2.2 Quantile Approach

This subsection discusses the methodology proposed in this study and provides some interpretations. Following Oaxaca (1973) we define the proportional market discrimination against females in the τ-th quantile of the wage distribution by

$$D^{\tau} = \frac{\frac{W_m^{\tau}}{W_f^{\tau}} - \frac{(W_m^{\tau})^0}{(W_f^{\tau})^0}}{\frac{(W_m^{\tau})^0}{(W_f^{\tau})^0}} \tag{9}$$

where $\frac{W_m^\tau}{W_f^\tau}$ is the observed male-female wage ratio for the τ-th quantile and $\frac{(W_m^\tau)^0}{(W_f^\tau)^0}$ is the τ-th male-female wage ratio in a nondiscriminatory labor market, so that

$$log(D^\tau + 1) = log(\frac{W_m^\tau}{W_f^\tau}) - log(\frac{(W_m^\tau)^0}{(W_f^\tau)^0}) \tag{10}$$

Once more, in order to measure labor market discrimination we need to specify the male-female wage ratio for the τ-th quantile in the absence of discrimination. Although instead of estimating the wage function $log(W_i) = X_i'\beta + u_i$ for males and females using OLS techniques, we estimate this function by quantile regression. [6] In the same way that $\hat{\beta}_{\text{OLS}}$ minimizes the sum of the loss function $(log(W_i) - X_i'\beta)^2$, $\hat{\beta}(\tau)$ minimizes the sum of the following linear loss function $\rho_\tau(log(W_i) - X_i'\beta)$. [7] Thus the τ-th conditional quantile function is given by

$$Q_{\text{log(W}_\text{i})}(\tau|X) = X'\beta(\tau) \tag{11}$$

The conditional quantile function will give a family of functions, one for each τ, which provides a more complete characterization of the relationship between $\log(W_i)$ and X compared to the one given by OLS regression, which concentrates on the first conditional moments (Arias, Hallock and Sosa, 2001). In addition, Koenker and Portnoy (1996, pp. 36-42) show that the quantile functions have, in general, the same robustness properties to outlying observations as the ordinary τ-th sample quantiles. These robustness properties are very important when the distribution of the disturbance term deviates from the Gaussian distribution.

Finally, the slope parameters of the family of estimated quantile functions provides a way to test for the presence of heteroskedasticity in the model (Koenker and Bassett, 1982). For instance, if some slope coefficients

[6] We do not discuss quantile regression in detail. Instead we will just comment on some important properties of this approach, which are useful for this study. We suggest the works of Koenker and Portnoy (1996) and Koenker and Hallock (2001) as comprehensive sources of how to understand quantile regression.

[7] $\rho_\tau(u) = u\tau - uI(u \leq 0)$ where $I(u \leq 0)$ is an indicator function.

are changing with τ then this is indicative of some form of heteroskedasticity. Therefore, in the quantile regression technique we can address the heterogeneity of the unobservable effects in an informative and constructive way. Formally, Koenker and Bassett (1982) propose a Wald-type statistic to test if the slope parameters are equivalents for different quantiles.[8]

Therefore our estimation proceeds as follows, in order to estimate D^τ we define :

$$G^\tau = \frac{W_m^\tau - W_f^\tau}{W_f^\tau} \tag{12}$$

then

$$log(G^\tau + 1) = log(W_m^\tau) - log(W_f^\tau) \tag{13}$$

where W_m^τ and W_f^τ are the hourly wage for the τ-th quantile conditional on workers mean characteristics for male and female respectively, and are defined as:

$$l\hat{o}g(W_m^\tau) = \bar{X_m}'\hat{\beta}(\tau) \tag{14}$$

and

$$l\hat{o}g(W_f^\tau) = \bar{X_f}'\hat{\beta}(\tau) \tag{15}$$

where $\bar{X_m}$ and $\bar{X_f}$ are the vectors of the average covariates for male and female, respectively. Thus

$$l\hat{o}g(G^\tau + 1) = \bar{X_m}'\hat{\beta_m}(\tau) - \bar{X_f}'\hat{\beta_f}(\tau) \tag{16}$$

[8]See Koenker and Bassett (1978) and Koenker and Portnoy (1996) for more details.

Now let $\Delta\bar{X} = \bar{X_m} - \bar{X_f}$ and $\Delta\hat{\beta}(\tau) = \hat{\beta_f}(\tau) - \hat{\beta_m}(\tau)$ and as proposed by Oaxaca (1973) substitute $\hat{\beta_m} = \hat{\beta_f}(\tau) - \Delta\hat{\beta}(\tau)$ in (19) yielding

$$\hat{log}(G^{\tau} + 1) = \Delta\bar{X}'\hat{\beta_f}(\tau) - \bar{X_m}'\Delta\hat{\beta}(\tau) \tag{17}$$

If the female wage structure could apply for both male and female in a nondiscriminatory labor market we would have:

$$\hat{log}\frac{(W_m^{\tau})^0}{(W_m^{\tau})^0} = \Delta\bar{X}'\hat{\beta_f}(\tau) \tag{18}$$

and

$$\hat{log}(D^{\tau} + 1) = -\bar{X_m}'\Delta\hat{\beta}(\tau) \tag{19}$$

Alternatively, we can substitute $\hat{\beta_f} = \Delta\hat{\beta}(\tau) - \hat{\beta_m}(\tau)$ in (19) and assume that the wage structure of males apply for both males and females in a labor market with no discrimination. We should then have:

$$\hat{log}\frac{(W_m^{\tau})^0}{(W_m^{\tau})^0} = \Delta\bar{X}'\hat{\beta_m}(\tau) \tag{20}$$

and

$$\hat{log}(D^{\tau} + 1) = -\bar{X_f}'\Delta\hat{\beta}(\tau) \tag{21}$$

In contrast with the previous approach, we now can have the estimated D^{τ} by just taking the exponential of equations (19) and (21). This is true since quantile regression is equivariant to monotone transformations, i.e., $Q_{\mathrm{h(Y)}}(\tau|X) = h(Q_{\mathrm{Y}}(\tau|X))$ given that $h(\bullet)$ is a nondecreasing function on $\Re$. It follows from the fact that the quantiles of Y are based on the distribution function of Y and $P(Y \leq y) \equiv P(h(Y) \leq h(y))$, unless the function $h(\bullet)$ is decreasing in some range of $\Re$ [9].

Notice from equations (19) and (21) that now we have a family of D^τ to interpret, instead of just one approximated D. Thus, we can have a more complete picture of the discrimination in the labor market as suggested by Becker (1957) and not just an average value, as proposed by Oaxaca (1973).

3 Data

The estimation of this proposed quantile approach of Oaxaca's coefficient of discrimination is based on data on 146,796 individuals (71,238 women and 73,558 men) draw from the 1999 CPS Outgoing Rotation Group (ORG) files. The sample is the intersection of the following sets : (a) adults sixteen to sixty five years old; (b) those individuals who showed an hourly wage in the week preceding the survey and (c) workers who reported their race as either white or black.

The hourly wage, our dependent variable, is defined by the total weekly wage divided by the total numbers of hours worked during that week. The control variables we use in our model are: education, defined as years of schooling standardized using Jaeger's (1997) method to reconcile CPS's sample codes, [10] potential experience defined as age - schooling - 6 (linear and quadratic terms), and categorical variables for marital status, race, union status, public job, 1-digit occupation, urban area and region of the country.

The impact of additional education on earnings has been carefully studied in many theoretical and empirical studies.[11] One important aspect of the relationship between years of schooling and earnings is that unobserved (and unobservable) factors such as individuals ability and family background (e.g. parents' schooling) may affect the acquisition of human capital and therefore

[9]See Koenker and Portnoy (1996: pp.30-31) for more details.

[10]In 1990 the Bureau of the Census changed the question on education attainment from years of education to degree receipt. In order to reconcile both questions Jaeger (1997) used the sample median for the new variable code and got robust results for returns to education comparing old and new samples. We use Jaeger's (1997) methodology.

[11]See, for example Card (1995).

influence how education affects earnings. Disregarding the influence of these unobservable variables certainly will lead to a bias since education is not affecting earnings equally for each individual. Here we do not use any proxy for family effects and ability, however by using Quantile Regression we show the heterogeneity in the returns to schooling across quantiles of the wage distribution. We report how this outcome reflects on the wage gap and on the discrimination coefficient.

Potential work experience is a reasonable proxy for actual experience in the case of males.[12] Nevertheless, potential experience could overestimate the actual work experience for female workers since, on average, females are more likely to leave the labor force due to childbearing activities. Thus, the fact that the wage is positively correlated with experience implies that there would be a bias towards finding discrimination against women. Oaxaca (1973) overcomes this problem by controlling for numbers of children born to the female. We did not use this approach since the 1999 CPS Outgoing Rotation Group files do not contain such information. In part, the quantile approach deals with this issue, since it allows us to estimate the wage function conditional on the covariates for different quantiles of the wage distribution. Therefore, it estimates different wage functions for females with different wages whose years of experience are equivalent. It seems reasonable to estimate different wage functions for two women with the same years of potential experience and same characteristics but one of them was out the labor force for a year due to childbearing activities while the other was in the labor force without interruption. This can be accomplished using quantile regression, however it is not done here. Notice that we compare men and women in the same quantile of their wage distribution conditional on the covariates. In order to have an idea of the impact of measurement errors in the female experience variable we estimated our model using the March CPS file and controlling for a new proxy for actual experience following Buchinsky (1998). Our results can be found at the appendix.

Table 1 describes the basic statistics. The data indicate that women in our sample earn less than men, showing an unconditional wage gap of 80%. Men and women however, are about the same age, have the same level of education and potential labor market experience. Notice, however, that potential experience could overstate women's actual labor market experience

[12]See Oaxaca (1973) and Blau and Ferber (1987).

as it does not account for females absence due to childbearing related activities. Table 1 also shows that the female sample is, on average, less likely to be white, less likely to be married and has a higher divorce rate. Not surprisingly, women are less likely to be unionized and have a higher participation in public occupations. The majority of workers in our sample lives in metropolitan areas and most are evenly distributed across the country's different regions.

The second part of table 1 presents the summary statistics by occupation. Notice that women and men are equally distributed across executive positions. Women however have a higher representation in professional activities, sales, clerical, private households and services. While male workers account for a higher fraction in production, operatives, transport and labor occupations. Note that the sample for male and female are statistically different in most of the productive characteristics except for some occupations, such as executive, also for urban region and country regions.

4 Empirical Results

Before analyzing the quantile regression results it is reasonable to start with the empirical results for the standard Oaxaca coefficient estimates. Tables 2 and 3 present the estimations for the Oaxaca coefficient for a base-line model with personal characteristics and a full-scale wage regression that controls for 1-digit occupation.

Both tables are divided into two different panels. The first column in panel 1 reports the estimates from the base-line log wage regression for the male sample ($\hat{\beta_m}$). Column 2 presents the estimates using only the female sample ($\hat{\beta_f}$) and the third column shows the difference between the female and male estimates ($\Delta\beta = \hat{\beta_f} - \hat{\beta_m}$). In panel 2 we estimate the wage differential that cannot be accounted by differences in skills ($ln(\hat{D} + 1)$), using both female and male weights. This differential is performed for each individual covariate and the total $ln(\hat{D} + 1)$ is given at the bottom of each table. Notice that we calculate the percentages of this wage gap not accounted by productive differences in terms of the total conditional wage differential.

The results for the base-line model show that the conditional mean logarithm of hourly wages are 2.61 for men and 2.40 for women, so the wage difference in logarithmic terms , $ln(G+1)$, is 0.21. The effects of discrimination as described before are obtained by subtracting the effect due to differences in workers characteristics from the conditional wage gap. More formally, Oaxaca's coefficient of discrimination $ln(\hat{D}+1)$ is estimated using both female and male weights, for each individual covariates respectively. As the results show (table 2), in both cases gender differences in the distribution of experience contributed to widening the wage gap (columns 5 and 7).[13] This is also true for the race and marital status variables, i.e. gender differences in the distribution of whites and marital status also contributed to widening the wage gap between males and females. Schooling seems to exert a greater influence towards narrowing the wage differential. This is also true for the dummy variable urban, showing that the fact that the worker lives in a urban area contributes to the narrowing of the wage gap between men and women.

After accounting for the contribution of the personal characteristics we found that the wage gap not explained by differences in skills accounted for, on average, 98% (bottom of column 5 and 7, table 2). That is, most of the gender wage differential found in the base-line model cannot be accounted by productivity differences between men and women. This is slightly higher than the results found by Oaxaca (1973) when he found that 77% of the gap for whites and 93% of the gap for blacks, could not be explained by differences in personal characteristics between men and women. This suspects, in fact, that most of the gap is due to discrimination, which may seem suspicious. However, based on our definition of discrimination, we can say that this is really the "unexplained by covariates" gap. Maybe these results are due to the nature of our sample or to omitted variables biases.

In order to have a more complete picture we used a broader set of covariates controlling for both 1-digit and 2-digit occupations.[14] Again, as we can see in table 3 (column 5 and 7), experience is working towards widening

[13]Notice that the returns to potential experience is much lower for women than for men. This is certainly due to the fact that women, due to child rearing activities, tend to acquire less market experience than men.

[14]Table 3 reports the results for the 1-digit occupation model. The results for the 2-digit occupation wage regression will not be reported here and can be provided upon request.

the gap and in a smaller scale schooling contributes to a narrowing of the wage differential. Here we have to take into account the fact that potential experience could overstate labor market experience for women, since it does not account for leaves due childrearing related activities. This is probably why experience is working towards widening the gap, since from the data there is no difference in average experience between men and women that can explain the wage gap. A good proxy for experience for females could actually show a different result. Gender differences among the distribution of white and married workers also contributes towards widening the wage differential. It is interesting to notice the role of union status in both set of regressions. In the personal characteristic regression (table 2, columns 5 and 7) union status worked towards narrowing the wage gap. In the occupation controlled model (Table 3) it is the opposite. Moreover, all of the variables controlling for occupations contribute to narrowing the gender wage gap.

After controlling for 1-digit occupation, the estimated effect of discrimination, $ln(\hat{D}+1)$, turns out to be 94%, 4 percent points lower than the baseline model result (bottom of column 5 and 7 at table 3). When controlling for 2-digit occupations the effect of the individual variables on the wage gap does not change qualitatively. The overall effect of discrimination, after taking into account productivity differences with a broader set of occupations, is a lot lower than the two previous cases. Now, $ln(\hat{D}+1)$, accounts for 79% of the total gender wage gap. A result much more similar to the ones found by Oaxaca (1973). It is worth noting that, as mentioned by Oaxaca (1973), "By controlling for broadly defined occupation, we eliminate some of the effects of occupation barriers as sources of discrimination. As a result, we are likely to underestimate the effects of discrimination." (Oaxaca, 1973, p. 699).

These are interesting results in the sense that we would have speculated that more than 30 years after the equal pay, equal opportunity acts, the non-explained wage gap would be a lot lower. In fact, what the results are showing here is that the part of the gender wage gap not explained by differences in productivity is higher today than the one found almost 30 years ago by Oaxaca (1973).

The question remains: what is driving this unexplained gender wage gap? In order to shed some light on this question, we investigate a sample 20 years older than ours (also from the CPS - Outgoing Rotation Group Files) and try to investigate the main changes in female and male observed productive characteristics over this period.

In table 4 we notice that the female/male wage ratio is 66%, much lower than the one found in 1999 (81%). The sample in 1979 is younger and less educated. Men and women have about the same age, but women are slightly better educated than men. The percentage of college graduates is lower than in 1999, 1% versus 8%. Also, in 1999 the percentage of women with a college degree is higher than men. Workers have also less experience in 1979.

Notice that due to differences in occupation codes across time, the categories are not exactly the same (Table 4 (Cont.)). Again women have a higher participation in professional, sales, clerical and private households services compared to men. Male workers, however, are much more present in operative, transport, craft and labor occupations. Notice that in 1979 women were much more likely to be in operative and transport related activities than in the end of the 1990s. Also, female workers today are much better represented in professional activities than 20 years ago. As pointed by O'Neil and Polacheck (1993), both men and women reduced their representation in blue-collar jobs, but men remained in a higher percentage in this kind of occupations compared to women. So in the last twenty years female workers increased their share in predominantly male occupations and decreased their participation in female occupations.

In terms of personal characteristics, there was a general upgrading in terms of education and experience within the whole sample. Women 20 years ago still had more education on average, were more likely to have college degrees and were less likely to be married than men. Moreover, female workers were also more likely to be in public jobs than their male counterparts. From the difference in means in table 4, we notice that both samples are statistically different, except with respect to region of the country and manager/administrative occupations. Finally, today women and men are much more alike in terms of level of education, labor market experience and attachment to the labor force.

All in all, the wage gap decreased and from the remaining gap the part not explained by differences in skills seems to be a lot higher today. Notice that the Oaxaca coefficient of discrimination after controlling for 1-digit occupation represents in 1979, 79% of the wage gap,[15] while today as we saw it accounts for 94% of the gap (Table 3(Cont.), bottom of columns 5 and 7). From the summary statistics for 1999 and 1979 (tables 1 and 4), it seems that 20 years later the gender gap decreased and male and female workers are much more alike, with women increasing their participation in occupations before dominated by men. The direct consequence of that is that the small wage gap that remains does not seem to be explained by differences in workers characteristics. This is what Oaxaca (1973) would claim is the effect of discrimination. Now we turn to our quantile regression results and analyze how the results presented here change across the quantiles of the conditional wage distribution.

4.1 Quantile Results

After stressing how the mean coefficient of discrimination changed over time we turn to the analysis of the evolution of those coefficients for different quantiles of the conditional wage distribution. Fist we analyze the estimates of the quantile wage model conditional on workers productive characteristics. The specification of our quantile model follows the one used for the estimation of the OLS coefficients. Recall that the hourly wage, our dependent variable, is defined by the total weekly wage divided by the total numbers of hours worked during that week. The control variables we use are: education, potential experience defined as age - schooling - 6 (linear and quadratic terms), and categorical variables for marital status, race, union status, public job, 1-digit occupation, urban area and region of the country. In Figure 1 the solid line represents the 21 point estimates of the quantile coefficient for $\tau's$ ranging from 0.01 to 0.99, while the dotted lines reports their 95% confidence intervals. The dashed lines show the ordinary least squares estimates for each covariate and their 95% confidence interval. As posed by Koenker and Hallock (2001): "Covariates may influence the conditional distribution of the response in myriad other ways: expanding its dispersion as in traditional models of heteroskedasticity, stretching one tail of the distribution, compressing the other tail, and even inducing multimodality." . We would, therefore, like to verify how the covariates are influencing the conditional

[15]The Oaxaca coefficient of discrimination was also estimated for the 1979 sample but the results are not presented here. These results can also be obtained upon request.

distribution for both men and women.

First , in Figure 1, we may notice that the returns to schooling are positive and increase monotonically as we move from lower to higher quantiles. They range from 4% to slightly more than 7% for men, and from 5% to 8% for women. These results are consistent with the findings of Bushinsky (1994), Chamberlain (1994), Arias, Hallock and Sosa (2001) and Machado and Mata (2000). They imply that education increases the wage dispersion, i.e., samples with more educated individuals show a higher wage dispersion than samples with less educated workers.

Experience entered the regression both with linear and quadratic terms. The result on the plots represent its effect evaluated at the variable's mean. The plots for the returns to experience show the same pattern as the returns to education. With the return increasing from the lower to the upper quantiles. Again, the returns to experience for males show a much more dispersed pattern than the ones for females. Notice also that just as for the OLS results, the quantile estimates show a higher return to experience for males than for females, which is probably reflecting the fact that actual experience is in fact lower for females than what is showing by potential experience [16].

The coefficients on the white indicator are much more homogeneous across quantiles and are higher for males than for females. The same is true for returns to marriage, with married men receiving higher returns than married women. Notice also that for both male and female, workers at the lower quantile of the conditional wage distribution tend to receive higher returns for being married compared to other individuals. The omitted category for marital status is single. Also, divorced workers tend to earn more than the ones never married. This pattern again is relatively constant across quantiles, but people at the lower quantiles receive higher returns from being divorced compared to other individuals in the conditional wage distribution.

Returns to union tend to be higher for workers at the lower quantiles of the conditional wage distribution. This is true for both men and women.[17] Notice however that men on average tend to receive higher wage premia compared with women. But this is not true for different quantiles. Women receive

[16] See the appendix for the results using Buchinky's (1998) proxy for actual experience.
[17] These results are in accordance with Chamberlain (1994).

higher returns at lower quantiles and lower returns at the upper quantiles. This implies that women in lower paying jobs receive a higher premium for begin unionized than men, but this is not true for higher paying jobs. Men tend to earn less in governmental jobs, specially in higher paying jobs. Notice that women in the upper quantiles receive higher returns for being at public occupations.

The returns for different occupation show some interesting patterns. Notice that the omitted occupation is sales so that the coefficients can be interpreted relative to this category. The plots for the returns to executive occupations show that for both men and women, workers at lower quantiles receive higher returns for being in an executive position, relative to workers in the upper quantiles. For women however, not only is this stronger for the middle quantiles, but they also tend to receive higher returns relative to men, at all quantiles. It might be the case that very few women make it into executive occupations so they are, in general, high ability workers and receive higher returns than male executives. In order to analyze this more closely, however, we would have to control for omitted ability here.[18]

Returns for professional positions show a similar pattern for both men and women. Workers in technical occupations tend to do better in low paying jobs relative to high paying ones. This is also true for both men and women. But notice that the returns for women tend to be higher up to the 6^{th} percentile.

Female workers tend to receive higher returns from clerical occupations than men, however both male and female workers do better in clerical occupations at the lower quantiles of the conditional wage distribution. For the rest of the occupations the returns look a lot more dispersed for men than for women. The returns for the most blue-collar kind of occupations follow a similar pattern for men, since workers at the lower quantiles earn higher returns for holding such occupations relatively to those at the upper quantiles of the conditional wage distribution.

For both men and women the returns for being in urban areas are greater for the upper quantiles and they are higher for women at all quantiles of the conditional wage distribution. It might be the case that women tend to

[18]For further detail on gender wage differential in top corporate jobs see Bertrand and Hallock (2001).

take advantage of the amenities of the metropolitan areas such as more job opportunities and child care facilities. The indicators for regions do not show any interesting pattern. Notice that the omitted category was west region, so none of the other regions show any higher returns than the west for both male and females.

Following Koenker and Hallock (2001) we can interpret the intercept as the estimated conditional quantile function of the wage distribution, for both men and women, for a single non-white worker, with 13 years of schooling, 18 of labor market experience, working in sales occupation and living in the west region in rural area. Both men and women in this case are doing better at the upper quantiles of the conditional wage distribution.

The first plot in figure 2 shows the conditional male-female wage gap across different quantiles. Notice that the gap is higher for the lowest quantiles, it decreases and start increasing again up to the 90^{th} percentile, falling afterwards. So women at the upper quantiles, or on high paying jobs, suffer from higher wage gaps than the ones in the lower-middle portions of the conditional wage distribution. We may observe, however, that despite the changes the actual variation across quantile is not large. From the lower to the upper quantile we observe a variation in the gap of 20%.

The percentage of the male-female wage gap not explained by differences in productive characteristics is presented in the second plot of figure 2.[19] Not only is the unexplained portion of the gap is very high (close to 100%) but it increases as we go from the lower quantiles of the conditional wage distribution, to the upper quantiles. We can say that the high male-female wage gap in the upper quantiles are almost totally explained by factors other than differences in measured attributes. Once more, we call attention to the fact that labor market experience can be driving those results, since it is not reflecting the actual experience for females. Furthermore we may have an incomplete set of covariates.

Finally, the third plot in figure 2 reports the calculated Oaxaca coefficient of discrimination for different quantiles. Not surprisingly, they follow the

[19]The confidence intervals in this graph were obtained using the Oaxaca coefficients with both male and female weights. Estimates with male weights are at the upper confidence intervals.

same pattern as the estimated gender gap given that very small fraction of the gender gap is explained by differences in productive characteristics.

5 Concluding Remarks

We agree with Bushinsky (1994) who states that *"on the average has never been a satisfactory statement with which to conclude a study on heterogeneous populations. Characterization of the conditional mean constitutes only a limited aspect of possibly more extensive changes involving the entire distribution"*.

In this paper, we review the most-often used measure of labor market discrimination provided by Oaxaca (1973). Many of the empirical studies in the economics of discrimination in the last several decades have estimated Oaxaca's coefficient based on OLS regression. Nevertheless, for these studies based on heterogeneous individuals, the OLS approach only characterizes the mean, giving a very limited picture of the problem. Given this set up, we have proposed an extension of the Oaxaca coefficient based on quantile regression. Besides the robustness and the equivariance to monotone transformation properties, the quantile approach seems to be more informative in the sense that it provides a more complete picture of labor market discrimination and also deals with the heterogeneity problem in an informative and constructive way. Our empirical results suggests that discrimination increases as we move from low to high quantiles of the conditional wage distribution. This result was expected since the female wage distribution is less dispersed than the male distribution. Moreover, it confirms the results found in the OLS estimations for the Oaxaca coefficient, since they show that most of the conditional wage gap cannot be accounted by differences in productive characteristics between men and women.

Nearly 40 years after the Equal Pay Act of 1963 and the Civil Rights Act of 1964, men and women are much more alike in terms of level of education, labor market experience and and attachment to the labor force. From the summary statistics we notice that in the last twenty years female workers increased their share in predominantly male occupations and decreased their participation in female occupations. What our results show however is that, while the gender wage gap decreased, a high fraction of this remaining gap cannot be explained by differences in productive characteristics among men and women. Moreover this is not only true on average but is also observed across quantiles of the conditional wage distribution, being stronger at the upper quantiles. That is, after 30 years women and men are much more alike, and from the little wage gap remaining almost nothing can be explained by differences in observable skills, and women at higher paying jobs are the most affected by this unexplained wage differential.

6 References

Aigner, D and G. Cain (1977). Statistical Theories of Discrimination in Labor Markets. *Industrial and Labor Relation Review*, 30(2), pp. 175-87

Arias, O. K. F. Hallock and W. Sosa-Escudero (2001). Individual Heterogeneity in the Returns to Shooling: Instrumental Variables Quantile Regression Using Twins Data. *Empirical Economics.*

Arrow, K. (1973). The Theory of Discrimination. In: textitDiscrimination in Labor Markets, Orley Ashenfelter and Albert Rees editors, Princeton University Press, pp. 3-33.

Ashenfelter, O. and R. Oaxaca (1987). The Economics of Discrimination: Economists Enter the Courtroom. *The American Economic Review*, 77(2), pp. 321-325.

Bertrand, M. and K. F. Hallock (2000). "The Gender Gap in Top Corporate Jobs". *Industrial and Labor Relations Review.*, forthcoming.

Bushinsky, M. (1994). Changes in the Wage Structure 1963-1987: Application of Quantile Regression. *Econometrica*, 62(2), pp. 405-458.

Buchinsky, M. (1998). The Dynamics of Changes in the Female Wage Distribution in the USA: A Quantile Regression Approach. *Journal of Applied Econometrics.* vol.13, 1-30.

Blau, F. and M. Ferber (1987). Discrimination: Empirical Evidence from the United States. *The American Economic Review*, 77(2), pp. 316-320.

Blinder, A. (1973). Wage Discrimination : Reduced Form and Structural Estimates. *Journal of Human Resources.* (8), pp. 436-55.

Cain, G. (1986). The Economic Analysis of Labor Market Discrimination: A Survey. In: *The Handbook of Labor Economics*, O. Ashenfelter and R. Layard editors, vol.1, pp. 693-785.

Card, D. (1995) Earnings, Schooling and Ability Revisited, *Research in Labor Economics*, S. Polacheck ed., JAI Press, V. 14, pp. 23-48.

Chamberlain, G. (1994). Quantile Regression, Censoring, and the Structure of Wages. In: *Advanced in Econometrics, Sixth World Congress*, C. A. Sims editor, vol. I, pp. 171-210.

D'Amico, T. (1987). The Conceit of Labor Market Discrimination. *The American Economic Review*, 77(2), pp. 310-315.

Garcia, J. P. Hernandez and A. Lopez (1997): "How wide is the gap? An investigation of gender wage differences using quantile regression." *Empirical Economics,* 26:87-113

Greene, W. H. (1997). *Econometric Analysis.* Third Edition, Prentice Hall.

Koenker, R. and K. F. Hallock (2001). Quantile Regression and Introduction. Working Paper, University of Illinois.

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

Koenker, R. and G. Bassett (1982). Robust Test for Heteroskedasticity based on Quantile Regression. *Econometrica*, 50(1), pp. 43-61.

Koenker, R. and S. Portnoy (1996). Quantile Regression. *Office of Research Working Paper #97-0100* College of Commerce and Business Administration, University of Illinois at Urbana-Champaign.

Kuhn, P. (1987). Sex Discrimination in Labor Markets: The Role of Statistical Evidence. *American Economic Review*, 77(4), pp.567-583.

Machado, J. A. and Jose Mata (2000). Counterfactual decomposition of changes in wage distributions using quantile regression. *Empirical Economics*, forthcoming.

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

Rouse, C. (1997) Further Estimates of the Economic return to Schooling from a New Sample of Twins, *Princeton University Industrial Relations Section Working Paper, 388*, July.

7 TABLES AND FIGURES

Table 1: Summary Statistics-1999

	All sample	**Men**	**Women**	**T-Statistics for Diff. in Means**
Observations	146796	75558	71238	
Hourly Wage	15.00	16.49	13.39	(53.66)
	(11.21)	(11.73)	(10.39)	
(log) Hourly Wage	2.51	2.61	2.40	(64.35)
	(0.63)	(0.63)	(0.62)	
Age	37.78	37.60	37.96	(-5.78)
	(11.91)	(11.85)	(11.97)	
Schooling	13.27	13.21	13.33	(-9.20)
	(2.50)	(2.39)	(2.59)	
College Graduates	0.08	0.07	0.09	
Experience	18.50	18.39	18.62	(-3.70)
	(11.88)	(11.74)	(12.03)	
White	0.82	0.84	0.81	(15.50)
	(0.37)	(0.36)	(0.38)	
Married	0.56	0.59	0.53	(22.48)
	(0.49)	(0.49)	(0.53)	
Divorced	0.10	0.08	0.13	(-31.21)
	(0.3)	(0.28)	(0.33)	
Union	0.13	0.16	0.11	(28.12)
	(0.34)	(0.37)	(0.31)	
Government	0.15	0.13	0.18	(-26.51)
	(0.36)	(0.34)	(0.38)	
Urban	0.82	0.82	0.82	(0)
	(0.38)	(0.37)	(0.38)	
Northeast	0.18	0.18	0.19	(-4.97)
	(0.38)	(0.38)	(0.39)	
Midwest	0.24	0.24	0.24	(0)
	(0.42)	(0.42)	(0.42)	
South	0.35	0.34	0.35	(-7.97)
	(0.47)	(0.24)	(0.24)	
West	0.21	0.22	0.21	(4.72)
	(0.41)	(0.41)	(0.40)	

Notes: (1) Sample averages; sample standard errors in parenthesis (2) T-Statistics for difference in means.

Table 1: (Cont.) Summary Statistics for Occupations-1999

	All sample	**Men**	**Women**	**T-Statistics for Diff. in Means**
Observations	146796	75558	71238	
Executive	0.136	0.133	0.139	(-3.37)
	(0.34)	(0.34)	(0.34)	
Professional	0.15	0.13	0.17	(-21.20)
	(0.36)	(0.34)	(0.38)	
Technical	0.036	0.032	0.039	(-7.42)
	(0.18)	(0.17)	(0.19)	
Sales	0.11	0.10	0.12	(-12.33)
	(0.31)	(0.30)	(0.32)	
Clerical	0.15	0.06	0.24	(-98.22)
	(0.36)	(0.24)	(0.43)	
Private Household	0.006	0.0006	0.013	(-29.38)
	(0.08)	(0.025)	(0.11)	
Services	0.13	0.11	0.15	(-22.75)
	(0.34)	(0.31)	(0.36)	
Product. Repair	0.10	0.19	0.021	(111.72)
	(0.31)	(0.39)	(0.14)	
Operatives	0.063	0.078	0.048	(24.38)
	(0.24)	(0.26)	(0.21)	
Transport	0.04	0.07	0.008	(61.74)
	(0.20)	(0.26)	(0.09)	
Laborers	0.04	0.07	0.018	(57.01)
	(0.20)	(0.25)	(0.018)	

Notes: (1) Sample averages; sample standard errors in parenthesis (2) T-Statistics for difference in mean.

Table 2: Standard Oaxaca Estimates - Personal Characteristics Regression

				Female Weights		Male Weights	
Variable	$\hat{\beta_m}$	$\hat{\beta_f}$	$\Delta\beta$	$ln(\hat{D}+1) = -\bar{X_m}'\Delta\beta$		$ln(\hat{D}+1) = -\bar{X_f}'\Delta\beta$	
Schooling	0.094 (0.00079)	0.106 (0.0009)	0.01	-0.156	-74.06%	-0.157	-74.76%
Experience	0.034 (0.0006)	0.025 (0.0006)	-0.009	0.173	82.49%	0.176	83.55%
Experience2	-0.0005 (0.00001)	-0.0004 (0.00001)	0.0001	-0.071	-33.95%	-0.073	-35.06%
White	0.137 (0.005)	0.048 (0.005)	-0.008	0.075	35.80%	0.072	34.49%
Married	0.157 (0.005)	0.048 (0.005)	-0.108	0.064	30.60%	0.057	27.40%
Divorced	0.067 (0.008)	0.048 (0.007)	-0.023	0.001	0.89%	0.003	1.46%
Other	0.067 (0.013)	-0.01 (0.010)	-0.08	0.001	0.94%	0.004	2.19%
Union	0.12 (0.005)	0.13 (0.006)	0.008	-0.0014	-3.64%	-0.01	-14.36%
Government	-0.04 (0.006)	0.01 (0.005)	0.05	-0.007	-3.64%	-0.01	-4.97%
Urban	0.11 (0.005)	0.14 (0.005)	0.03	-0.03	-14.40%	-0.03	-14.36%
Northeast	-0.006 (0.006)	0.001 (0.006)	0.007	-0.001	-0.67%	-0.001	-0.69%
Midwest	-0.01 (0.005)	-0.039 (0.005)	-0.02	0.005	2.49%	0.005	2.48%
South	-0.05 (0.005)	-0.05 (0.005)	-0.001	0.0003	0.17%	0.0003	0.17%
Constant	0.699 (0.013)	0.54 (0.014)	-0.158	0.158	75.14%	0.158	75.14%
$ln(\hat{D}+1)$				0.21	101.11%	0.20	96.57%

Notes : (1) Standard errors in parenthesis; (2) Conditional Wage Gap : $ln(G+1)$ =0.21;
(3) $ln(\hat{D}+1)$: Wage differential not explained by difference in skills.

Table 3: Standard Oaxaca Estimates - 1 Digit Occupation Indicators also Controlled

				Female Weights		Male Weights	
Variable	$\hat{\beta}_m$	$\hat{\beta}_f$	$\Delta\beta$	$ln(\hat{D}+1) = -\bar{X}_m{}'\Delta\beta$		$ln(\hat{D}+1) = -\bar{X}_f{}'\Delta\beta$	
Schooling	0.068 (0.0009)	0.071 (0.001)	0.003	-0.044	-21.04%	-0.044	-21.24%
Experience	0.031 (0.0006)	0.022 (0.0006)	-0.008	0.16	76.37%	0.162	77.35%
Experience2	-0.0005 (0.00001)	-0.0003 (0.00001)	0.0001	-0.068	-32.37%	-0.07	-33.43%
White	0.097 (0.005)	0.021 (0.005)	-0.07	0.064	30.43%	0.061	29.31%
Married	0.12 (0.005)	0.029 (0.005)	-0.09	0.058	27.81%	0.052	24.90%
Divorced	0.057 (0.008)	0.035 (0.007)	-0.02	0.001	0.86%	0.002	1.40%
Other	0.05 (0.013)	0.03 (0.009)	-0.06	0.001	0.76%	0.003	1.78%
Union	0.16 (0.005)	0.15 (0.006)	-0.012	0.002	0.96%	0.001	0.65%
Government	-0.018 (0.006)	-0.015 (0.005)	0.002	-0.0003	-0.18%	-0.0005	-0.24%

Notes : (1) Standard errors in parenthesis;(2) Conditional Wage Gap : $ln(G+1)$ =0.21; (3) $ln(\hat{D}+1)$: Wage differential not explained by difference in skills.

Table 3: (Cont.)

Executive	0.24 (0.007)	0.35 (0.007)	0.11	-0.014	-6.97%	-0.015	-7.31%
Professional	0.17 (0.008)	0.28 (0.008)	0.10	-0.014	-6.94%	-0.019	-9.26%
Technical	0.15 (0.011)	0.24 (0.011)	0.08	-0.002	-1.36%	-0.003	-1.66%
Clerical	-0.07 (0.009)	0.07 (0.006)	0.14	-0.009	-4.43%	-0.03	-17.04%
Priv.Hous.	-0.42 (0.07)	-0.30 (0.017)	0.12	-8.05e-5	-0.03%	0.001	-0.8%
Service	-0.23 (0.008)	-0.12 (0.007)	0.11	-0.01	-5.98%	-0.017	-8.32%
Productive	0.07 (0.007)	0.14 (0.014)	0.07	-0.01	-6.5%	-0.001	-0.74%
Operative	-0.08 (0.009)	-0.009 (0.010)	0.07	-0.005	-2.60%	-0.003	-1.60%
Transport	-0.07 (0.009)	-0.01 (0.021)	0.06	-0.004	-2.24%	-0.0005	-0.26%
Laborers	-0.18 (0.009)	-0.10 (0.01)	0.08	-0.005	-2.80%	-0.001	-0.7%
Urban	0.09 (0.005)	0.12 (0.005)	0.03	-0.026	-12.67%	-0.02	-12.63%
Northeast	-0.002 (0.005)	0.007 (0.006)	0.009	-0.001	-0.86%	-0.001	-0.88%
Midwest	-0.014 (0.005)	-0.035 (0.005)	-0.02	0.005	2.44%	0.005	2.42%
South	-0.05 (0.005)	-0.05 (0.005)	-0.001	0.0006	0.29%	0.0006	0.30%
Constant	1.118 (0.016)	0.97 (0.016)	-0.14	0.14	66.93%	0.14	66.93%
$ln(\hat{D}+1)$				0.21	99.77%	0.18	88.88%

Notes : (1) Standard errors in parenthesis;(2) Conditional Wage Gap : $ln(G+1)$ =0.21; (3) $ln(\hat{D}+1)$: Wage differential not explained by difference in skills.

Table 4: Summary Statistics-1979

	All sample	**Men**	**Women**	**T-Statistics for Diff. in Means**
Observations	88620	48480	40140	
Hourly Wage	0.05	0.06	0.04	(7.60)
	(0.41)	(0.46)	(0.32)	
(log) Hourly Wage	-3.00	-2.84	-3.20	(112.71)
	(0.51)	(0.50)	(0.45)	
Age	34.62	34.33	34.98	(-7.28)
	(13.22)	(13.21)	(13.22)	
Schooling	11.82	11.65	12.04	(-24.02)
	(2.45)	(2.58)	(2.25)	
College Graduates	0.018	0.018	0.018	
Experience	16.85	16.74	16.99	(-2.64)
	(14.02)	(14.12)	(13.89)	
White	0.86	0.86	0.85	(4.35)
	(0.34)	(0.34)	(0.34)	
Married	0.59	0.62	0.56	(18.31)
	(0.49)	(0.48)	(0.49)	
Other	0.40	0.37	0.43	(-18.31)
	(0.49)	(0.48)	(0.49)	
Government	0.12	0.11	0.13	(-9.07)
	(0.32)	(0.31)	(0.34)	
Urban	0.65	0.65	0.66	(-3.15)
	(0.47)	(0.47)	(0.47)	
Northeast	0.21	0.21	0.20	(3.66)
	(0.40)	(0.41)	(0.40)	
Northcentral	0.29	0.29	0.29	(0)
	(0.45)	(0.45)	(0.45)	
South	0.31	0.31	0.31	(0)
	(0.46)	(0.46)	(0.46)	
West	0.17	0.17	0.17	(0)
	(0.38)	(0.38)	(0.38)	

Note: (1)sample averages; sample standard errors in parenthesis; (2) Wage rates in 1979 current values; (3) T-Statistics for differences in means.

Table 4: (Cont.)Summary Statistics for Occupations-1979

	All sample	**Men**	**Women**	**T-Statistics for Diff. in Means**
Observations	88620	48480	40140	
Manag./Adm.	0.028	0.028	0.028	(0)
	(0.16)	(0.16)	(0.16)	
Prof./Tech.	0.07	0.05	0.09	(-26.28)
	(0.26)	(0.22)	(0.23)	
Sales	0.041	0.021	0.06	(-27.84)
	(0.19)	(0.14)	(0.25)	
Clerical	0.19	0.06	0.34	(-107.43)
	(0.39)	(0.25)	(0.47)	
Private Household	0.008	0.0003	0.019	(-28.59)
	(0.09)	(0.018)	(0.13)	
Services	0.15	0.10	0.21	(-44.74)
	(0.35)	(0.30)	(0.41)	
Craftsmen	0.18	0.30	0.02	(126.6)
	(0.38)	(0.46)	(0.15)	
Operat./Transp	0.19	0.21	0.17	(15.23)
	(0.39)	(0.40)	(0.38)	
Laborers	0.07	0.11	0.01	(64.50)
	(0.25)	(0.31)	(0.13)	

Note: (1) sample averages; sample standard errors in parenthesis; (2) T-Statistics for Differences in Means.

Table 5: Standard Oaxaca Estimates - 1 Digit Occupation Indicators also Controlled- 1999 March CPS

				Female Weights		Male Weights	
Variable	$\hat{\beta_m}$	$\hat{\beta_f}$	$\Delta\beta$	$ln(\hat{D}+1) = -\bar{X_m}'\Delta\beta$		$ln(\hat{D}+1) = -\bar{X_f}'\Delta\beta$	
Schooling	0.0724 (0.0003)	0.0739 (0.004)	0.001	-0.0209	-8.64%	-0.0202	-8.75%
Experience	0.032 (0.002)	0.024 (0.002)	-0.008	0.168	69.47%	0.1678	69.33%
$Experience^2$	-0.0005 (0.00005)	-0.0004 (0.00005)	0.00009	-0.048	-20.16%	-0.048	-20.13%
Exper.Child	0.0018 (0.001)	-0.00133 (0.001)	-0.003	0.047	19.57%	0.039	16.16%
$Exp^2.Child$	-0.00004 (0.00003)	0.00005 (0.00005)	0.00009	-0.029	-12.36%	-0.022	-9.45%
White	0.124 (0.019)	-0.0099 (0.019)	-0.1339	0.116	48.04%	0.11	45.48%
Married	0.132 (0.02)	0.054 (0.02)	-0.078	0.051	21.22%	0.043	18.05%
Other	0.052 (0.026)	0.014 (0.024)	-0.038	0.0044	1.83%	0.008	3.32%
Government	-0.0461 (0.02)	0.0377 (0.019)	0.008	-0.0121	-5.02%	-0.0173	-7.14%
Executive	0.18 (0.018)	0.25 (0.029)	0.071	-0.010	-4.26%	-0.012	-5.14%
Professional	0.165 (0.02)	0.218 (0.03)	0.053	-0.007	-3.12%	-0.010	-4.38%

Notes : (1) Standard errors in parenthesis;(2) Conditional Wage Gap : $ln(G+1)$ =0.242; (3) $ln(\hat{D}+1)$: Wage differential not explained by difference in skills.

Table 5: (Cont.)

Technical	0.143 (0.042)	0.204 (0.043)	0.0619	-0.001	-0.80%	-0.0023	-0.96%
Clerical	-0.056 (0.03)	0.037 (0.02)	0.09	-0.005	-2.26%	-0.024	-9.95%
Priv.Hous.	0.254 (0.3)	-0.41 (0.08)	-0.66	-0.0002	0.09%	0.005	2.22%
Priv. services	0.02 (0.35)	0.028 (0.07)	0.0082	-0.0002	-0.11%	-0.000082	-0.03%
Service	-0.32 (0.034)	-0.23 (0.031)	0.09	-0.005	-2.38%	-0.01	-4.51%
Craft	0.087 (0.02)	0.0821 (0.05)	-0.0052	0.001	0.46%	0.00001	0.04%
Operative	-0.05 (0.03)	-0.06 (0.03)	-0.01	0.001	0.43%	0.0006	0.27%
Transport	-0.058 (0.03)	-0.12 (0.07)	-0.06	0.004	1.84%	0.0005	0.24%
Laborers	-0.19 (0.03)	-0.07 (0.06)	0.11	-0.006	-2.70%	-0.001	-0.70%
Urban	0.10 (0.016)	0.18 (0.017)	0.077	-0.06	-24.93%	-0.06	-24.87%
Northeast	0.0105 (0.019)	0.05 (0.021)	-0.047	0.01	-4.17%	0.009	4.13%
Nc./ Midwest	0.014 (0.019)	0.001 (0.02)	-0.01	0.003	1.36%	0.003	1.33%
South	-0.07 (0.018)	-0.05 (0.019)	-0.02	-0.007	-2.91%	-0.007	-3.04%
Constant	0.996 (0.06)	0.907 (0.06)	-0.088	0.08	36.72%	0.08	36.72%
$ln(\hat{D}+1)$				0.259	107.20%	0.216	89.97%

Notes : (1) Standard errors in parenthesis;(2) Conditional Wage Gap : $ln(G+1)$ =0.2420; (3) $ln(\hat{D}+1)$: Wage differential not explained by difference in skills.

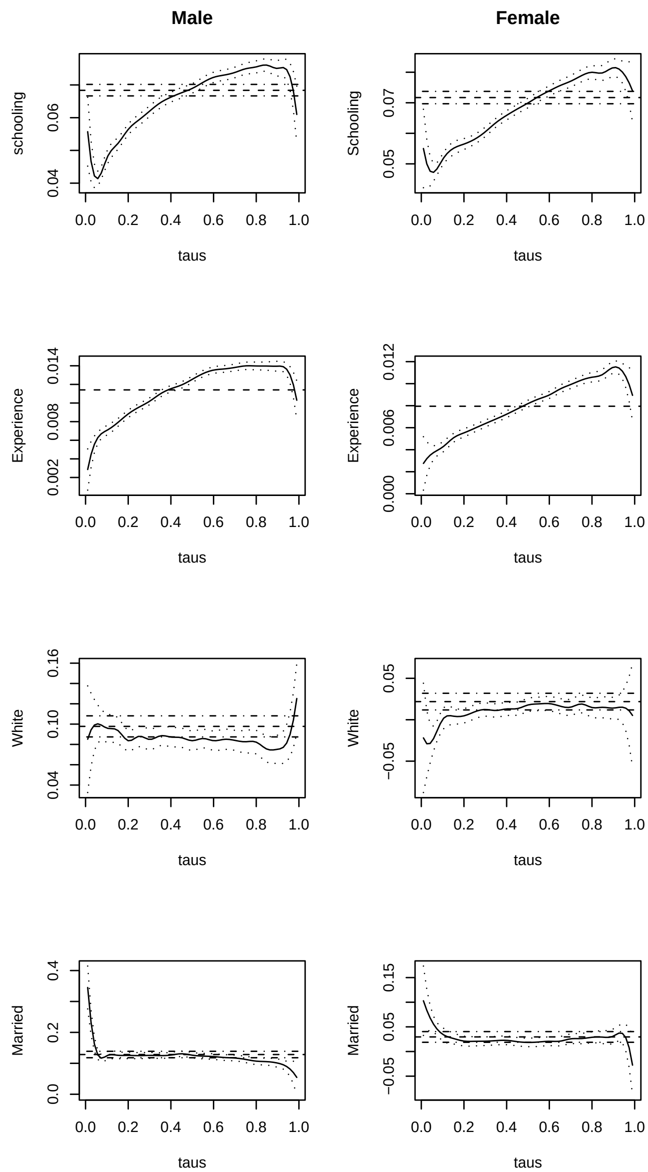

Figure 1: Quantile Regression Estimates

The above plots present the quantile regression estimates for each individual covariate indicated. The solid lines are the quantile estimates, the dotted lines the 95% confidence intervals. The dashed lines present the Ordinary Least Square estimates and the 95% confidence interval. The plots on the left report the estimates for males, the plots on the right for females.

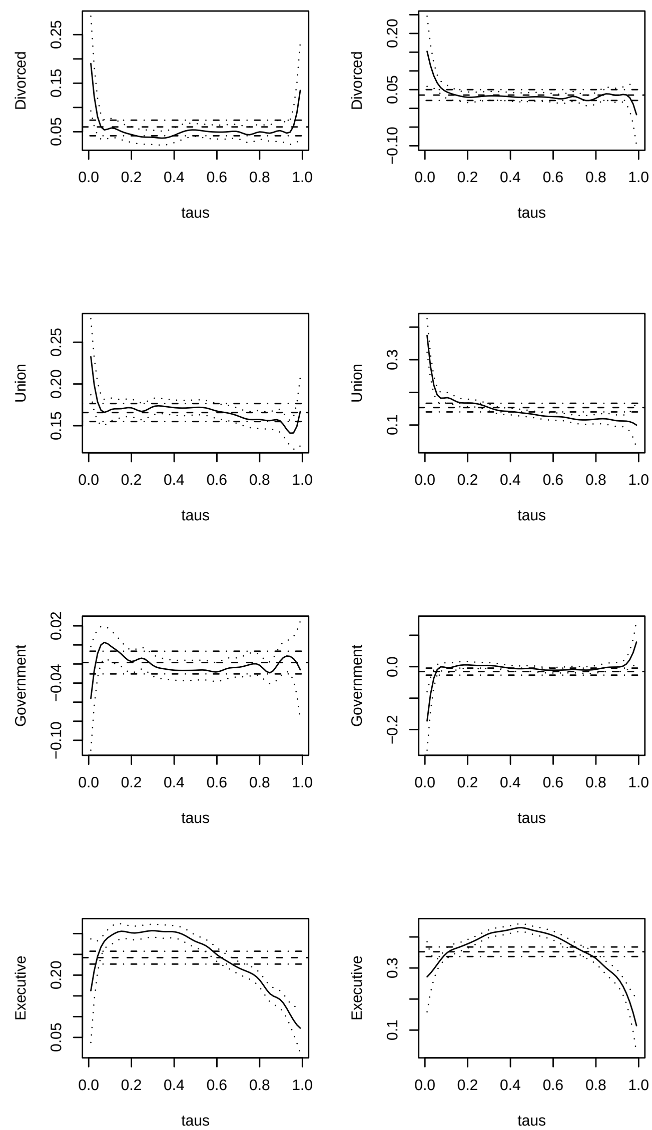

Figure 1 (Cont.): The above plots present the quantile regression estimates for each individual covariate indicated. The solid lines are the quantile estimates, the dotted lines the 95% confidence intervals. The dashed lines present the Ordinary Least Square estimates and the 95% confidence interval. The plots on the left reports the estimates for males, the plots on the right for females.

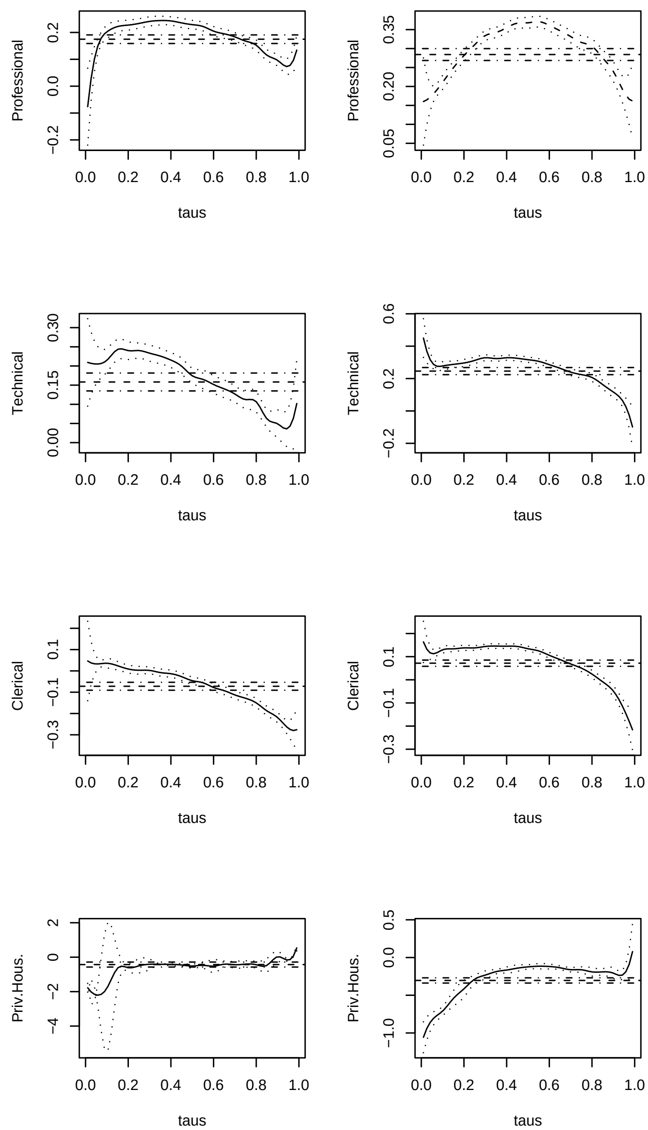

Figure 1 (Cont.): The above plots present the quantile regression estimates for each individual covariate indicated. The solid lines are the quantile estimates, the dotted lines the 95% confidence intervals. The dashed lines present the Ordinary Least Square estimates and the 95% confidence interval. The plots on the left reports the estimates for males, the plots on the right for females.

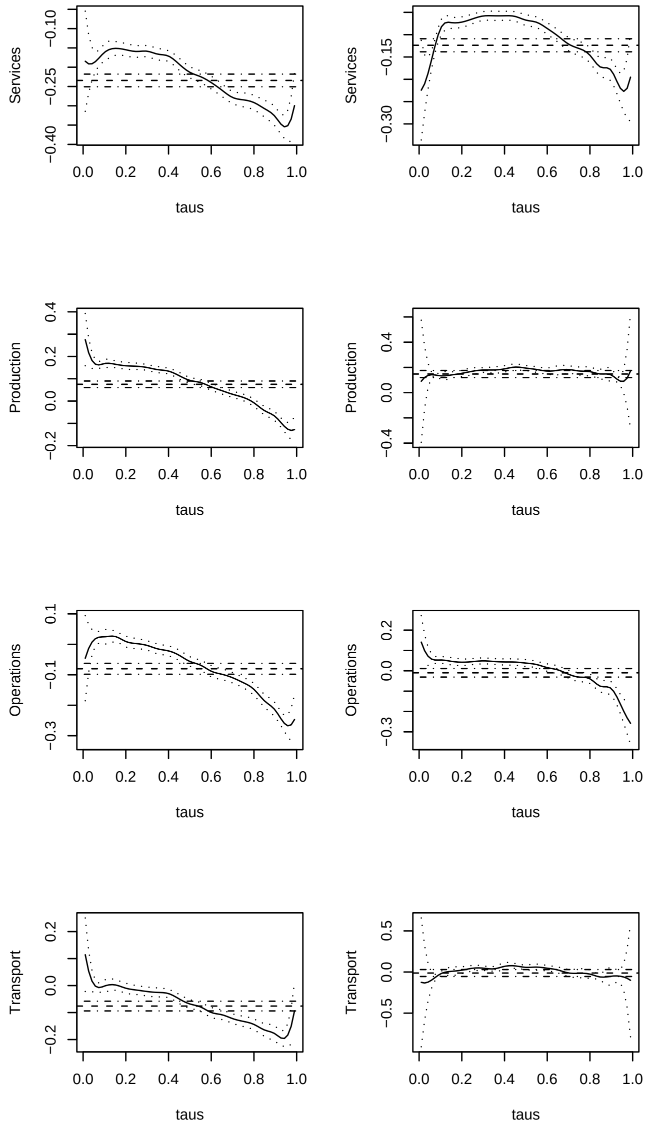

Figure 1 (Cont.): The above plots present the quantile regression estimates for each individual covariate indicated. The solid lines are the quantile estimates, the dotted lines the 95% confidence intervals. The dashed lines present the Ordinary Least Square estimates and the 95% confidence interval. The plots on the left reports the estimates for males, the plots on the right for females.

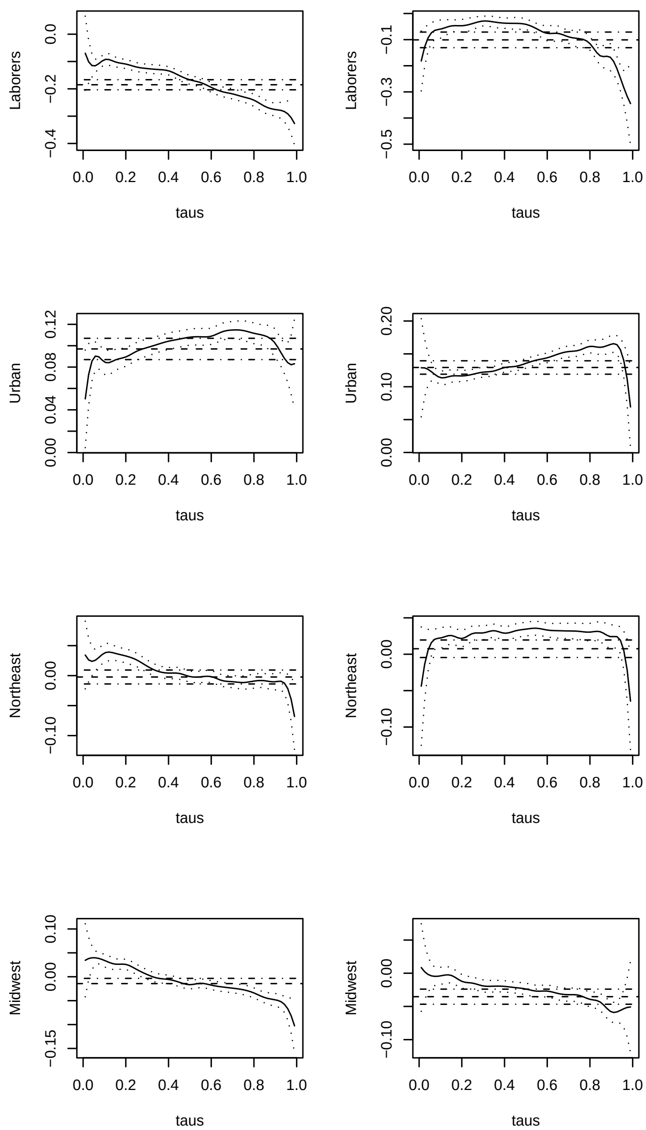

Figure 1 (Cont.): The above plots present the quantile regression estimates for each individual covariate indicated. The solid lines are the quantile estimates, the dotted lines the 95% confidence intervals. The dashed lines present the Ordinary Least Square estimates and the 95% confidence interval. The plots on the left reports the estimates for males, the plots on the right for females.

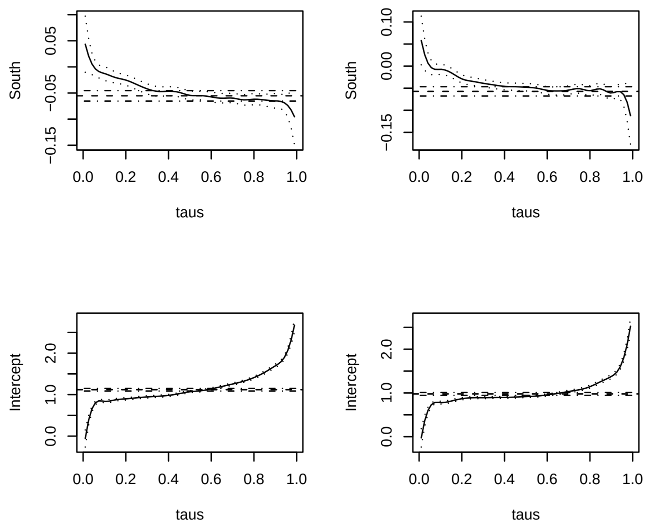

Figure 1 (Cont.): The above plots present the quantile regression estimates for each individual covariate indicated. The solid lines are the quantile estimates, the dotted lines the 95% confidence intervals. The dashed lines present the Ordinary Least Square estimates and the 95% confidence interval. The plots on the left reports the estimates for males, the plots on the right for females.

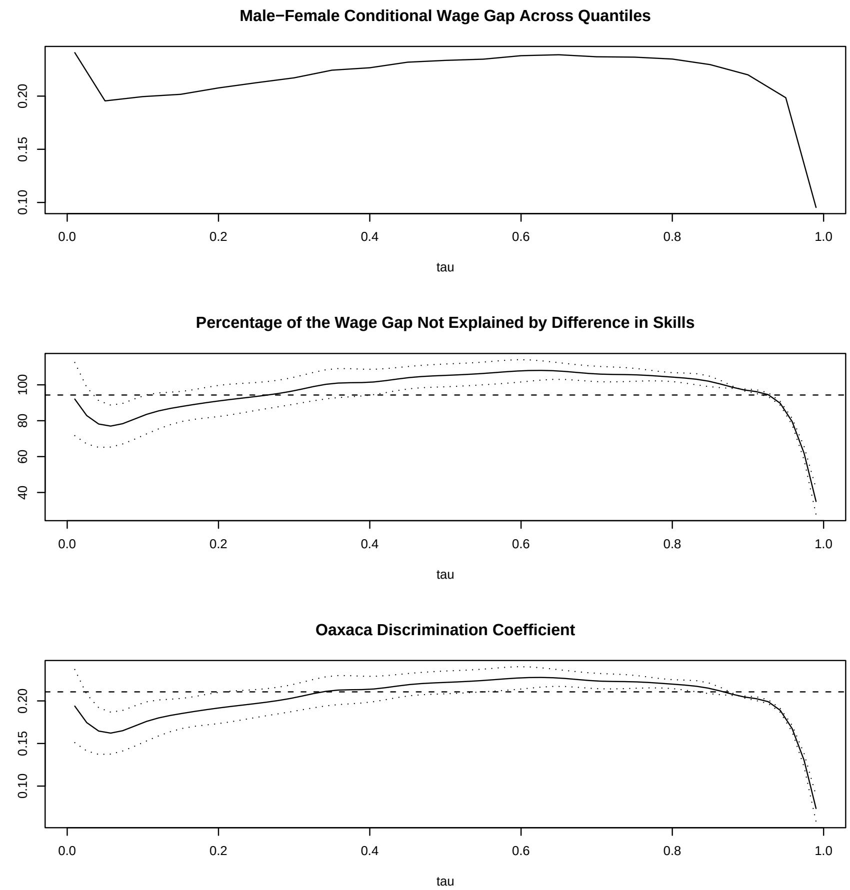

Figure 2: Oaxaca Coefficient Using Quantile - ORG files:

The first plot presents the wage gap for each quantile of the conditional wage distribution. The second plot shows the percentage of the gap not explained by differences in productive characteristics for each quantile. The third plot brings the estimations of the Oaxaca Coefficient of Discrimination for each quantile of the conditional wage distribution.Here we control for potential experience defined as age - sch -6.

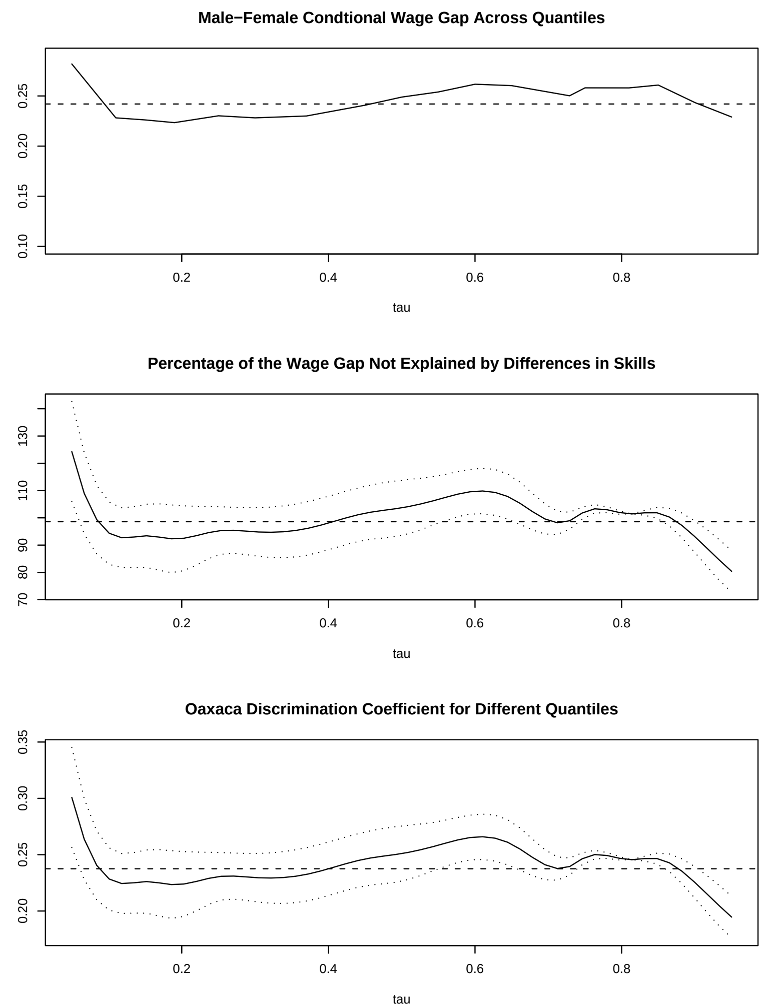

Figure 3: Oaxaca Coefficient Using Quantile - March CPS files:

The first plot presents the wage gap for each quantile of the conditional wage distribution. The second plot shows the percentage of the gap not explained by differences in productive characteristics for each quantile. The third plot brings the estimations of the Oaxaca Coefficient of Discrimination for each quantile of the conditional wage distribution. All of the estimations were performed using March CPS files controlling for Buchinsky's (1998) proxy for actual experience.

8 Appendix

In order to have an idea of the influence of experience on the gender wage gap we decided to look for a better proxy for actual labor market experience. As we mentioned before, women relative greater specialization in child rearing and other home production has been at the expense of a more continuous labor market participation. As a result, women not only have acquired fewer years of work experience but also have had less incentive than men to invest in length training with a distant payoff. As shown by O'Neil and Polacheck (1993), using the NLS data, the proportion of years worked by women in 1987 was 70% to 83% on average, while men continued to work more than 90% of their potential working years.

Oaxaca (1973) attempted to overcome the problem of measurement error on women's labor market experience by controlling for the number of children born to the female. The children variable entered his wage equation as a linear term and is supposed to reflect the cost of lost experience due to child care, including the costs from the depreciation of skills during the periods of absence from the labor force. As such, the estimated effect of the number of children on earnings is expected to have a negative sign.

In a similar way, Buchinsky (1998) tries to overcome this problem also by introducing the variable number of children under 18 in the family, and interacting it with the variable potential experience, defined as age - sch -6. He assumes that the main alternative use for women's time is child rearing and other home activities related to that. As such he defines a new proxy as:

$$X_p = \gamma_1 x_p + \gamma_2 x_p^2 + \delta_1 x_p.chld + \delta_2 x_p^2.chld =$$

$$= \gamma_1 x_p(1 - \tfrac{\delta_1}{\gamma_1}.chld) + \gamma_2 x_p^2(1 - \tfrac{\delta_2}{\gamma_2}.chld)$$

where xp denotes potential experience estimated as age -sch -6, and $chld$ the number of children under 18 in the family. According to him $(1-\frac{\delta_1}{\gamma_1}.chld)$ would be the adjustment one needs to make to potential experience in order to obtain actual experience. Here we use the same proxy for the March CPS files and re-estimated our results for the full- scale model that controls for 1-digit occupation variables. The statistics for both files, basically means and standard deviation, are very similar for both samples [20].

The OLS results are presented at Table 5 in the same fashion as the ones in Table 3. According to our results the mean logarithm of conditional hourly wages are 2.65 for men and 2.41 for women, so the wage difference in logarithm terms, $ln(D+1)$ is 0.24. Recall that the unexplained portion of the wage gap is obtained by subtracting the effect due to differences in workers productive characteristics from the conditional wage gap. The results did not change much ager the introduction of the new proxy.

Just as in the previous results, gender differences in the distribution of experience is contributing to widening the gap, even when we control for the total effect of experience. [21] Notice however, that the effect of experience on the wage gap is higher than on the previous specification using the ORG files with proxy for potential experience.

Again, gender differences in the distribution of race and marital status also contributes to a widening of the conditional gender gap. Schooling works towards narrowing the gap, however the effect is significantly smaller compared to the previous estimations. Also, the fact that the workers lives in a urban are contributes to the narrowing of the gender gap.

Most of the variables controlling for occupations contribute to narrowing the gender gap, except for private household, craft, operative and transport related occupations. Here we should call attention to the fact that in controlling for occupations we may be neglecting the fact that discrimination may take the form of occupational barriers , or even barriers in the access to promotion and training that take place within each occupation category.

After accounting for the contribution of productive characteristics, our estimations show that the unexplained wage gap $ln(D+1)$ accounts, on average, to 98% of the conditional wage gap. Notice that, the specification that controls solely for potential experience, shows an estimated unexplained by the covariates wage gap of 94% on average, 4 points percent lower than the new estimates using Buchinsky's (1998) proxy for actual experience.

[20] The statistics for the 1999 March CPS files will not be presented here but can be provided upon request.

[21] Notice that the total effect of experience is given by the derivative $\frac{\partial \omega}{\partial exp.} = \beta_2 + 2 * \beta_3 * exp + \beta_4 chld + 2 * \beta_5 exp.chld$ where the variables are evaluated at their average values.

This results call attention to the fact that indeed gender differences in the distribution of experience contributes to a widening of the gap, certainly because women acquire on average less experience than men. However, after using a broader proxy for labor market experience we could not find qualitatively different results regarding the unexplained portion of the gender gap. Still, most of the conditional wage gap cannot be explained by differences in skill among men and women.

Our quantile results show somehow higher male-female conditional wage gap for all quantiles of the conditional wage distribution, compared to the previous specification The quantile regression estimates for each covariates present the same pattern of change as the ORG estimates, implying that the samples are closely comparable and the inclusion of the new proxy on our specification does not change our results significantly [22]. The gap ranges from 0.28 at the lower quantiles to 0.22 at the top of the distribution. Notice, however, that just as before, the conditional wage gap is higher for the lowest quantiles, it increases up to the 80^{th} percentile, and start decreasing afterwards. This means that women at both the very bottom and the middle-upper quantiles of the conditional wage distribution suffer from higher gaps than the ones at the lower and very high quantiles. Again, despite the change in the conditional wage gap across quantiles, the actual variation is not large. From the lower to the upper quantiles, the conditional wage gap varies around 20%.

The following graph shows the unexplained by covariates portion of the wage gap by quantiles. Once more we observe the same pattern as the one for the ORG results. Most of the conditional wage gap shown on the previous Figure cannot be explained by differences in productive characteristics across gender. Moreover, the unexplained portion of the gap is very high at the bottom of the conditional wage distribution and is lower at the very top. The unexplained gap increases as we move from the 10^{th} up to the 60^{th} percentile of the wage distribution. Basically, for almost all the quantiles the male-to-female wage gap cannot be explained by differences in skills across gender. Notice that after controlling for a broader proxy for actual experience we do not have significant differences compared to our previous results using potential experience, except for the lower tail. The gap for the workers at the lower tail cannot be explained at all by differences in skill across gender,

[22] Again, the quantile estimates for the March CPS sample can be provided upon request.

not even after controlling for a broader proxy for experience. That is, factors other than productive characteristics, controlled in our model, are driving the wage gap both for low-paying workers and middle-class workers.

Not surprisingly the Oaxaca's (1973) discrimination coefficient follows the same pattern of change across quantiles as the unexplained wage gap, given the very small portion of the gap that can be explained by differences in skills. Therefore, after using a proxy for experience that controls for the effect of children on women's potential working years, we found basically the same results as before. The portion of the wage gap not explained by differences in skills is not only very high, but increases as we go from the 10^{th} to the 60^{th} percentile, being higher for low paying female workers and lower for women at the very top of the distribution. Therefore, women at the middle-upper quantiles, and the very bottom of the distribution suffer from the highest unexplained wage gap.

Notice however, that in order to have a broader picture of the problem of discrimination, or unexplained by skills wage gap, one should take into account differences in careers opportunities such as, access to promotion and training within each occupation. Moreover, we should also consider size of the firms, since as shown by Bertrand and Hallock (2000) women in high paying jobs tend to be employed in small-size firms. By solely controlling for occupation we are not considering these kinds of gender occupation segregation. Therefore, the high unexplained wage gap found here may account for the fact that those points are not being considered here. Future research should focus not only on the variation of the unexplained conditional wage gap across quantiles of the wage distribution but also, how more robust specifications that take into account occupational barriers, firm size and gender differences in career's decisions, present different results regarding the unexplained by skills gender wage gap, than the ones presented here.