The gender pay gap

ECON4150

Victoria Jensen
Seminargruppe 1

[Semesteroppgave ved Økonomisk institutt/SV-fakultet]

UNIVERSITETET I OSLO

[13.03.2019]

Table of Contents

Introduction

Gender pay gap in the labor market is one of the key focus points of labor economics. The need for mobility and heterogeneity of labor forces make gender policy gap important for the development of labor policy. To which extent does gender play a role in determining wages is important to know, to target specific labor policies in the right direction. In addition, it is important to observe to which extent such demographic characteristics as race, occupation, marital status, differently affect wages between men and women. The results show that such differences exist and should be taken into consideration when forming future labor policies. The main purpose of this term paper is, therefore, to explore the extent to which impact of gender and other demographic characteristics as well as occupation choice has influence for determination of the hourly wage in US. By doing so, the following research question will be explored - to what extent the information in this available dataset can identify gender discrimination in the US labor market. The focus of this term paper is, therefore, around the main hypothesis set, which is as follows: there is gender discrimination in the US labor market based on the available dataset. This hypothesis will be explored using Blinder-Oaxaca decomposition based on quantile regression model to test if it is truly correct.

In the first part of the paper the brief analysis of available literature about the gap will be presented. Furthermore, the empirical approach and data with the model for estimation outlining its strengths and weaknesses will be analyzed. Finally, it will be given interpretation to the finding and the answer to the research question to what extent the information in this dataset can identify the gender discrimination in the US labor market. The analysis argues that the causes and consequences of the gender pay gap need to be addressed, as it is an obstacle to the economic independence of women in old age, when they face a higher risk of poverty than men.

Empirical approach

The gender pay gap or gender wage gap is the average difference between the remuneration for employed men and women. There are two distinct numbers regarding the pay gap: unadjusted versus adjusted pay gap. It is unadjusted pay gap that is pure gap without any differences in working hours, occupation or education chosen as well as job experience. The reasons for lower pay include both individual choice and other innate and external factors. People can voluntary prefer part-time vs full time employment. On the other side, people can involuntary prefer low-skilled jobs due to inability to access high education. Discrimination is one of the external factors for lower pay. An example of a voluntary choice is choosing the part-time work when the full-time employment is available. An example of an involuntary choice is having a low-skill job because of an inability to access higher education. An example of an external factor is discrimination. Even when the reason for the gap is entirely voluntary the gender pay gap can be a problem from a public policy perspective. It reduces economic output and indicates that women, who earn less than men, are more likely to be dependent upon welfare payment especially in old age. The definition of the unadjusted gender pay gap, expressed as a percentage, is as follows: $\frac{Mean\ (gross) hourly\ earning\ of\ men - Mean\ (gross) hourly\ earning\ of\ women}{Mean\ (gross) hourly\ earning\ of\ men}$. As an unadjusted indicator, the gender pay gap gives an overall picture of the differences between men and women in pay. A part of the difference in earnings of men and women can be explained by differences in the average characteristics of male and female employees. The differences in the average characteristics can result from many factors, including the concentration of one sex in certain economic activities or the concentration of one sex in certain occupations. The first phenomenon is called sectoral gender segregation and the second one is called occupational gender segregation. Sectoral gender segregation may explain only partially the difference in earnings for men and women, when one sex tends to be concentrated in low-paying economic sectors and the other sex tends to be concentrated in high-paying sectors. Similarly, occupational gender segregation may explain the difference in earnings of men and women, when one sex tends to be concentrated in low-payed and the other sex is in high-payed occupations. Gender segregation based on occupation may also be partially caused by men being more often promoted to supervisory and management positions than women due to discrimination. The unadjusted gender pay gap is, therefore, rather complex indicator and measures both possible discrimination between men and women through unequal pay for equal work and the differences in the average characteristics of male and female employees. To separate out the different factors at work in the gender pay gap, this analysis will use a methodology to decompose the unadjusted gender pay gap based on the Oaxaca (1973) decomposition, also called the Blinder-Oaxaca decomposition. This method is carried out in two stages: a regression analysis and decomposition analysis of the structure of earnings. In the first stage, a regression analysis is conducted to estimate the earnings equations separately for men M and women W as detailed in the following equations:

$$\ln y_i^M = \beta_0^M + \sum_{k=1}^{K} x_{ki}^M \beta_k^M + u_i^M$$
$$\ln y_i^W = \beta_0^W + \sum_{k=1}^{K} x_{ki}^W \beta_k^W + u_i^W$$

where i) ln y_i represents the natural log of hourly earnings for observation i; ii) x_{ki}, 1<= k<=K, are explanatory variables of the observed characteristics. It may impact on the log hourly earnings of individual i; iii) β_0 is a constant responsible for observed characteristics; β_k, 1<= k<=K, are the parameters of the observed characteristics; iv) u_i is a disturbance term, independent and normally distributed.

The regression analysis includes several explanatory variables covering the observed personal, and demographic characteristics (Table 1, appendix). The regression equations are a result of adjusting and expanding the standard log wage regression, like Mincer (1974) earnings equation, which relates, in a linear way, the log hourly earnings to years of education but not include a traditional linear and quadratic function of experience as far as our dataset does not have such information. Several variables from dataset have duplet information. In order to describe occupation, it was chosen variable describing major occupation codes divided into 10 categories. To exclude the problem of the perfect multicollinearity, the variable “employed” with value 1, is excluded from the regression equation. All other variables were recoded into dummies with value from 0 to 1 as an assumption from Oaxaca decomposition to find whether there is gender discrimination in the US labor market based on the available dataset. The identification strategy is based on the following steps: i) estimation of the quantile regression with quantiles 0,25, 0,5 and 0,75; ii) the Blinder–Oaxaca decomposition to explain the difference in the means a dependent variables between men and women by decomposing the gap into that part that is due to differences in the mean values of the independent variable within the groups, on the one hand, and group differences in the effects of the independent variable, on the other hand.

OLS regression estimates the conditional mean of the responsible variable given the certain values of the predictor variables, $Y_i = E(Y_i|X_i) + \varepsilon_i$. If regular assumptions about uncorrelated ε_i with the zero mean and constant variance σ^2 are satisfied, then OLS estimator $\hat{\beta}$ for β is the BLUE estimator. OLS regression is built upon the following assumptions: i) linearity; ii) normality; iii) homogeneity of variance (homoscedasticity); iv) independence; v) errors in variables; vi) model specification. Among problems there are possible omitted variable bias, measurement error and multicollinearity. Different tests after OLS regression help to solve it. Quantile regression, the extension of linear regression, is a type of regression analysis aimed to estimate either the conditional median or other quantiles of the response variable and used when conditions of linear regression are not applicable. One of advantage of quantile regression, relative to OLS regression, is that the quantile regression estimates are more robust against outliers in the response measurements. In our case due to the fact that there are possible outliers that can be not normally distributed, different measures of central tendency and statistical dispersion can help to obtain a more comprehensive analysis of the relationship between variables. In our analysis different covariates influence response variable in different quantiles by the regression: $Q(\tau|X_i) = \alpha + \beta(\tau)X_i$,

where $\widehat{\beta(\tau)}$ and τ = 0,5, or 0,75 or 0,25 is estimated by solving the following equation :

$$\widehat{\beta(\tau)} = \sum_{i;Y_i \le \alpha_\tau + \beta_\tau X_i} \tau(Y_i - \alpha_\tau - \beta_\tau X_i) + \sum_{i:Y_i > \alpha_\tau + \beta_\tau X_i} (1+\tau)(Y_i - \alpha_\tau - \beta_\tau X_i)$$

After estimation of quantile regression with quantiles of interest 0,5, 0,25 and 0,75 the Blinder-Oaxaca decomposition will be done. Blinder-Oaxaca decomposition helps to decompose the gap into the part that is due to the differences in the mean values of the independent variable within groups and group differences on the effects of the independent variable. Our groups of interest are female. The unexplained wage difference for the same value of explanatory variables cannot be explained only as wage difference due to discrimination due to existence of other unobserved explanatory variables may also account for wage difference. After fitting separate regression models for men and women, a decomposition analysis of the difference between the means of log hourly earnings of men and women will be preceded:

$$\Delta = \overline{\ln y}^M - \overline{\ln y}^W$$

The Oaxaca decomposition uses the following regression property for the means of log hourly wages for men and women:

$$\overline{\ln y}^M = \hat{\beta}_0^M + \sum_{k=1}^{K} \bar{x}_k^M \hat{\beta}_k^M$$

$$\overline{\ln y}^W = \hat{\beta}_0^W + \sum_{k=1}^{K} \bar{x}_k^W \hat{\beta}_k^W$$

These equations provide insights into the male and female earnings structures by showing the relationship between the mean of log hourly earnings and the observed average characteristics for men and women ($\bar{x}_k^M$) and ($\bar{x}_k^W$), respectively. The estimated constant $\hat{\beta}_0^M$ and $\hat{\beta}_0^K$ and coefficients $\hat{\beta}_k^M$ and $\hat{\beta}_k^K$ measure the financial returns to the characteristics of male and female employees respectively. Within the decomposition approach, it must be decided which earnings structure constitutes the non-discriminatory benchmark against which to decompose the difference Δ between the means of log hourly earnings of men and women (Bazen 2011). It is assumed, in accordance with the definition of the unadjusted GPG, that the male earnings structure constitutes this benchmark. The following other options are possible: i) the female earnings structure constituting a non-discriminatory benchmark (Oaxaca 1973); ii) both the female and male earnings structures constituting non-discriminatory benchmarks with some weighted average applied (Cotton 1988 and Reimers 1983); iii) the whole population earnings structure constituting a non-discriminatory benchmark (Neumark 1988). The estimated constant and coefficients in the men's equation are treated as the non-discriminatory benchmarks for the financial returns to characteristics of employees. Because of this, a counterfactual equation is constructed where the constant and coefficients in the women's equation are replaced by those of the men's equation:

$$\overline{\ln y}^{W*} = \hat{\beta}_0^M + \sum_{k=1}^{K} \bar{x}_k^W \hat{\beta}_k^M$$

This equation can be interpreted as what the average female would have earned if she had been paid on the same basis as an equivalent male worker. The difference between the means of log hourly earnings between men and women can then be decomposed as follows:

$$\Delta = E + U$$

$$\text{where } E = \overline{\ln y}^M - \overline{\ln y}^{W*} \text{ and } U = \overline{\ln y}^{W*} - \overline{\ln y}^W$$

Where *E* is the difference between the actual mean of the log hourly earnings of men and the counterfactual mean of the log hourly earnings of women and can be expresses as

$$E = \sum_{k=1}^{K} \hat{\beta}_k^M (\bar{x}_k^M - \bar{x}_k^W)$$

It measures the part of Δ that is due to differences in the average characteristics of men and women weighted by the male coefficients and show the explained part, *E*, of the difference in earnings between men and women, Δ. The part of the equation *U* is the difference between the counterfactual and actual means of log hourly earnings of women and can be expressed as:

$$U = (\hat{\beta}_0^M - \hat{\beta}_0^W) + \sum_{k=1}^{K} \bar{x}_k^W (\hat{\beta}_k^M - \hat{\beta}_k^W)$$

It measures the part of Δ that is due to the difference in the estimated constants for men and women, plus the difference in the estimated coefficients for men and women weighted by the average characteristics of women. This second component corresponds to the different financial returns paid to men versus women for each variable. The part *U* shows what a female worker with average characteristics would have earned if she had been treated in the same way as a typical male worker and compares that with what she earns (Bazen 2011). In the decomposition, the part *U*, unexplained part of the difference in gender earnings (Δ). U cannot be explained only as a discrimination part because we do not have some other possible explanatory variables in the dataset (e.g. age of children in a family, personal abilities or negotiating skills, job experience) that would most likely change the unexplained part. This limitation should be taken into consideration when interpreting the unexplained part, when there is a low coefficient of determination R-squared. Therefore, it seems to be more appropriate to view the part *U* as a residual in that it is the part of the difference in earnings between men and women (Δ) that is not explained by the difference in average observed gender characteristics. The final decomposition for the difference between the means of log hourly earnings of men and women is:

$$\overline{\ln y}^M - \overline{\ln y}^W = \underbrace{(\hat{\beta}_0^M - \hat{\beta}_0^W) + \sum_{k=1}^{K} \bar{x}_k^W (\hat{\beta}_k^M - \hat{\beta}_k^W)}_{\text{Unexplained}} + \underbrace{\sum_{k=1}^{K} \hat{\beta}_k^M (\bar{x}_k^M - \bar{x}_k^W)}_{\text{Explained}}$$

Data

To decompose the unadjusted gender pay gap the micro dataset covers two broad areas: the earnings of individual employees and the observed characteristics of individual employees. These observed characteristics include: (i) the personal characteristics of individual employees, e.g. age, education, race, children in the household (ii) the types of occupation, and (iii) place of residence (states). The scope of the microdata in the decomposition analysis is the same as the scope and coverage of the unadjusted gender pay gap. Data from table 2 (appendix) shows data only for 2017 for those who are employed and missing information for service sector. The dataset represents information about 51 % of men and 49 % female of total amount of respondents. To estimate first OLS and then quantile regressions, several dummies for different covariates were prepared. To avoid omitted variable bias, several reference categories were chosen for comparison: white race, variable explaining those who have 2 children, New York state, high school education and professional and other occupation. Coefficient for 10^{th} children in the household is omitted because of missing data. Data from the table 2 shows that the wage gap is higher for hourly wages for covariates responsible for occupation Farming, Fishery and forestry and number of children. Farming and fishery are occupations dependent on a season, that make week pay very volatile. The same concern the number of children in the household. This variable is very volatile and changing from week to week. Families with several children in the household are very dependent on the time when they can work. That is why they often choose to have not fixed week hours for work to combine care for the children with work. It indicates that the effect of wage gap is stronger for these covariates. For all other covariates the wage gap is more is less similar or even weaker for these covariates. To find out whether gender discrimination exists we need to calculate the effect of different factors affecting gender pay gap both pooled and for each gender separately. Instead of looking into the averages it needs to know what these effects of gender pay gap are if we use different quantiles of dependent variable log hourly wage.

There are several ways to test the evidence against the null hypothesis that t=0 in var(e)=σ^2 exp(t) or heteroscedasticity. In the normal version, performed by default, the null hypothesis also includes the assumption that the regression disturbances are independent-normal draws with var(e)=σ^2. Test statistic Breusch-Pagan is one is of them. Another way is to use White’s general test for heteroscedasticity as a special case of Breusch-Pagan test. White's test is usually very similar to the first term of the Cameron-Trivedi decomposition normally reported by imtest and computes the general test for heteroscedasticity in the error distribution by regressing the squared residuals on all distinct regressors, cross-products, and squares of regressors. The test statistic, a Lagrange multiplier measure, is distributed Chi-squared (p) under the null hypothesis of homoscedasticity. During regression estimation it was used Stata routines for estimating robust standard errors. Heteroscedasticity causes biased standard errors. OLS assumes that errors are both independent and identically distributed; robust standard errors relax either or both of those assumptions. Hence, when heteroscedasticity exists, robust standard errors tend to be more trustworthy. According to the IM test the null hypothesis of homoscedasticity is not rejected as according to the hettest. That is why only IM test was performed (table 3, appendix).

One of the main assumptions for OLS regression is the homogeneity of variance of residuals. If the model is well-fitted, there should be no pattern to the residuals plotted against the fitted values. As far as the p-value is very small, we would have to reject the hypothesis and accept the alternative hypothesis that the variance is not homogenous. So in our case, the evidence is against the null hypothesis that the variance is homogeneous. RVF plot was performed to detect non-linearity, unequal error variances, and outliers. The plot suggests that there is a weak decreasing linear relationship between residuals and fitted values. It also suggests that there are no unusual data points in the data set and illustrates that the variation around the estimated regression line is constant suggesting that the assumption of equal error variances is reasonable. (figure 1, appendix). Linktest was performed for the model specification. The t-test for hatsq is now insignificant, indicating that we have fixed the model so that it passes the link test. (table 3, appendix). The variance inflation factor is a useful way to look for multicollinearity amongst the independent variables. Vif 1,46 looks fine here.

Log hourly wage was use as a dependent variable to measure "percent" changes in wage (which is what log (wage) is), rather than absolute changes in wage. The wage distribution is truncated at 0 and is highly right-skewed. After estimating of the effects of various things on wages, we can probably receive estimates with negative wages for people that make no sense. As far as the goal is to measure a change in marginal utility by Mincer wage equation, log wage provides with a good approximation for this. To interpret results from OLS log regression three assumptions need to be hold: i) Gauss-Markov assumptions hold; ii) estimated coefficients need to be statistically significant or practically significant; iii) all other independent variables need to be hold as constant. Results from table 4 (appendix) show that the effect is stronger for regression with higher quantiles and coefficient are mostly statistically significant from zero. Keeping variables that are not statistically significant can reduce the model's precision. On the other hand, a p-value that is greater than the significance level indicates that there is insufficient evidence in your sample to conclude that a non-zero correlation exists. That is why it is difficult to explain not statistically significant coefficients but as far as we need them in a regression for better model prediction, they are going to be the part of the model. Regressions results indicate that the effect monotonically increases when the quantiles of regression increases, e.g. the effect is higher for individuals with higher log hourly wage gap. However, starting from quantile 0,5, estimates becomes larger than those of after OLS regression. It shows that starting from quantile 0,5 the effect is not significant from OLS coefficients. Higher quantiles can be not significantly different from OLS regression (or larger than OLS coefficient).

Results

Results from both OLS and quantile regressions are mostly significant. By OLS model, switching from female to male workers, its expected ca. 19 % decrease in the geometric mean of wages. By estimating the $0{,}25^{th}$ q. regression the decrease will be lower, i.e. only 16 % less for wages for females, ceteris paribus, by $0{,}25^{th}$ q the decrease will be higher than OLS estimate, i.e. ca 21 % less for wages for female and by 0,75 % the decrease will be higher than OLS estimate, i.e. ca 23 % less for wages for females. The quantile regression results indicate that the effect of switching from female to male has a larger negative impact on the lower quantiles. The linear regression model underestimates this effect at the $0{,}5^{th}$ q. and higher. Standard errors and confidence limits for the quantile regression coefficient estimates can be obtained with asymptotic and bootstrapping methods. Both methods provide robust results (Koenecker and Hallock 2001). For standard errors and confidence limits it will be used the bootstrap method as a more practical (Hao and Naiman, 2007).

Estimates from the pooled OLS regression show that by switching from black to white workers the wage will be ca. 12 % lower for black workers. Again, the linear regression model underestimates this effect at the $0{,}5^{th}$ q., i.e. by estimating of $0{,}5^{th}$ q. and higher quantiles the wage of the black workers will be little bit over than 12 % lower than for the white ones. Results for workers with Spanish origin are similar as for the black ones. By separating male from female, OLS estimates show that black males receive 16 % less salary than the white ones. The OLS underestimates this effect at the $0{,}25^{th}$ q. and higher. Both results from OLS and quantile regression are statistically significant, but all quantiles are not significant as far as concern OLS. By separating only results for female black from OLS, women black receive only 8 % lower salary than males and males black receive ca 18 % lower wage than women.

Married workers, both men and women, will increase their salary only by 5 % comparing with those who are not married. The OLS estimate gives the whole effect. This effect is overestimated by $0{,}75^{th}$ q. However, by separating men from women, only married men receive 10 % higher salary than those who are not married. Married women receive only ca. 3 % salary increase than those of non-married.

The OLS estimates the total effect for households with no children comparing to households with 2 children. Results from OLS and quantile regressions for households from 0 to 3 children (households with 2 children is the reference category) are significant. Results for households from 4 to 10 children comparing to households with 2 children are not statistically significant that indicate that these variables are not significant to explain regression.

Males working in public sector, earns 0,4 % more than those who are not in the public sector. Women wages in public sector are 5 % lower than those who are not in the public sector. The unobservable difference can occur due to the discrimination. Oaxaca decomposition will help to distinguish this unobservable difference due to the discrimination and will help to separate the discrimination for men and women. Men, who are the members of trade unions, receive 16 % higher wage than those who are not in any trade union. By comparison, women, the members of trade unions, receive only 7 % higher salary. To estimate how place of residence influences wage gap, OLS and quantile regression were performed with reference state New-York. OLS regression shows that those people who live in Massachusetts, New Jersey, Minnesota, Maryland, District of Columbia, Virginia, Washington, California and Alaska have higher wages than those who live in New-York. However, men in Rhode Island earns a little bit less than in men New York, but women, in the opposite, receive a little higher wage then in New York. Men from Illinois earn more than men in New York, but women in Illinois receive lower salary than women in New-York. Such gender gap can be explained by the different occupation structure in the states in comparison with New York.

Regression results indicate that men working in management, business and financial sector, receive ca 8 % higher salary than men working in other occupation. Women working in management, business and finance, receive 7 % more wages than those from other industrial sectors. Results from table 2 indicate existence of the gender wage gap and this is not only women who earns less than men. However, it is obvious that women being in work the last week (hours worked last week), earn 0,5 % more in salary, but men earn less, only 0,4 % respectively. By looking into the agricultural sector, regression results show that men earn 10 % less than men not working in agriculture. Simultaneously, women receive only ca. 4 % less than those who are not from the same sector of the economy.

Table 5 shows results from the decomposition of the model. The decomposition shows how much of the wage gap is due to differing endowments between male and females groups, and how much is due to discrimination (regarded as the portion of the wage gap due to the combined effect of coefficients and slope intercepts for the two groups). Results from OLS regression shows that -19,8 % is the share gap due to discrimination and that cannot be explained by the individual’s characteristics. Results from 0,25 quantile regression shows a little less result, -18,6 % gender wage gape due to discrimination.

By comparing the output from the two regression equations female vs male (table 5), it is obvious that male workers have higher constants, and this is reflected in the 20.9% disadvantage of unexplained part for female (the shift coefficient). Nevertheless, net advantage in only coefficient of 1,1 % for female workers after offsetting of other factors. There is little difference in endowments between the two groups, female and males, something evident from a comparison of the himod and lomod output, which shows that there is little difference between the average group characteristics of female and male workers. This lack of group differences is reflected in the small figure for endowments, only -3,5%. Consequently, there is little difference between the raw differential (-23.3%) and the adjusted differential (-19.8%) because the difference in endowments between female and male workers is so small. In other words, almost all the difference (85%) is partly due to discrimination between female and male, and this is made up of the difference in the shift coefficient of unexplained part and differences in how the endowments are rewarded.

By comparing the output from the two regression equations it’s clear that male workers have higher constants, and this is reflected in the 20,9% advantage in unobservable or the shift coefficient. After offsetting, male workers leave with a net disadvantage in the observables of -3.3 %. There is little difference in endowments between the two groups, something evident from a comparison of the himod and lomod output, which shows that there is little difference between the average group characteristics of male and female workers. This lack of group differences is reflected in the small figure for gaps in endowments, just 5.8%. Consequently, there is little difference between the raw differential (23.3%) and the adjusted differential (17.6%) because the difference in endowments between male and female workers is so small. In other words, two third of difference, almost 75 % is due to discrimination between male and female, the part that also substitute the effect of group differences in unobserved predictors. This is made up of the difference in the shift coefficient and differences in how the endowments are rewarded or observables.

After decomposition and comparison of himod with lomod the Oaxaca decomposition computes the Blinder-Oaxaca decomposition to analyze gender wage gaps. It was used different legends under estimation. Swap legend reverses the order of the group female. Eform legend specifies that the results will be displayed in exponential form. Afterward it was estimated the pooled model that computes the twofold decomposition using the coefficients from a pooled model over both groups as the reference coefficients. The Blinder-Oaxaca decomposition breaks down the wage gap between high-wage (males) and low-wage (females) workers into several components. The unexplained component is the difference in the shift coefficients (or constants) between the two wage equations. Being inexplicable, this component can be attributed to discrimination. However, Blinder also argued that the explained component of the wage gap also contains a portion that is due to discrimination. To examine this Blinder decomposed the explained component into: i) the differences in endowments between the two groups, "as evaluated by the high-wage group's wage equation; ii) the difference between how the high-wage equation would value the characteristics of the low-wage group, and how the low-wage equation actually values them. Blinder called the first part the amount "attributable to the endowments" and the second part the amount "attributable to the coefficients" and argued that the second part should also be viewed as reflecting discrimination, i.e. it only exists because the market evaluates differently the identical bundle of traits if possessed by members of different demographic groups. It reflects discrimination as much as the shift coefficient is. Blinder-Oaxaca decomposition closely follows Blinder's exposition and uses both his method and his terminology and takes the average endowment differences between the two groups and weights them by the high-wage workers' estimated coefficients. The differences in the estimated coefficients are weighted the average characteristics of the low-wage workers.

Conventionally, the high-wage group's wage structure is regarded as the "non-discriminatory norm", that is, the reference group.

Counterfactual decomposition of differences in distribution (Table 6) shows the decomposition of the median difference in log hourly wage between men and women. The positive explained part means that the differences in average characteristics between male and female workers are in favor of men, whereas the negative explained part means that the differences in average characteristics between male and female workers are in favor of women. The observed median gender gap is about 35 %, 19,9 % is explained by gender differences and 80,1 % is due to discrimination and differing median coefficients between men and women if we take average between men and women. It needs to pay attention that 80,1 % includes not only discrimination. 19,9 % of the difference between log hourly earnings of men and women can be attributed to the differences in average characteristics between male and female workers that is in favor of men at the US. The US explained part is mostly driven by such explanatory factors as occupation, education and hours of work last week. The proportions of the overall explained part (positive or negative) and the unexplained part sum up to 100 %. The differences in average characteristics for categorical variables occupation and education can be interpreted as gender segregation. The decomposition of the unadjusted gender pay gap does not capture all segregation effects between men and women in the labor market. Women often work, on average, fewer hours per month than men and the lower proportion of women than men participate in the labor market. All these factors are not captured by the unadjusted gender pay gap calculated on an hourly basis. Gender segregation effect includes lower employment rate of women, lower number of hours worked and sectoral and occupational segregation. Unadjusted gender pay gap include unequal pay for equal work and again sectoral and occupational segregation. So, the sectoral and occupational segregation is a part both gender segregation effect and unadjusted gender pay gap.

Unequal pay for male and female for equal work is only one of the possible causes of the unadjusted gender pay gap and that is why understanding all its causes is very important. The results show that there are clear policy and statistical reasons to decompose the unadjusted gender pay gap into the explained and unexplained parts. The unadjusted gender pay gap indicator, together with the explained gap allow for a better identification and interpretation of the causes of the gender pay gap for better targeting of the policy actions. The explained gender pay gap shows the gap between male and female hourly earnings due to differences in the observed average characteristics of male and female employees. The unexplained gender pay gap can be viewed as a gap in residuals i.e. the part of the unadjusted gender pay gap that is not explained by those differences. Being a residual, it is not plausible to interpret the unexplained gender pay gap as a measurement of a possible discrimination by unequal pay for equal work. Including such additional variables as f. ex. the total work experience in the regression analysis may substantially change the results.

The decomposition also makes it possible to identify the main factors behind the explained gender pay gap. This analysis has shown that the explained gender pay gap is strongly driven by occupation, education and hours worked by last week as the main explanatory variables. However, these factors have different explanatory effects on the decomposition. On the one hand, the explained gender pay gap for education is negative that means that employed women have, on average, a higher level of education than men in the US labor market in 2017. On the other hand, the explained gender pay gap for occupation is mostly positive indicating that men tend to be employed in better paid professions than women (sectoral segregation). Negative gap for occupation could indicate that women tend to work in better paid occupations than men, due to "self-selection" effects as well. That is why occupational gender segregation have not so smooth effects as the sectoral gender segregation. The decomposition of the unadjusted gender pay gap does not capture all segregation effects between men and women. Women work, on average, fewer hours per month than men that is not captured by the unadjusted gender pay gap calculated on an hourly basis. Moreover, a lower proportion of women than men participate in the labor market. "Self-selection" of women into paid employment can be an issue that affects the measurement of the unadjusted gender pay gap and its decomposition. If both a low unadjusted gender pay gap and a low employment rate for women is the case, then the negative explained gender pay gap can be explained due to significant low-skilled women who are out of the job market when there are plenty different job opportunities.

Appendix

Figure 1. RVF plot (residuals versus fits plot)

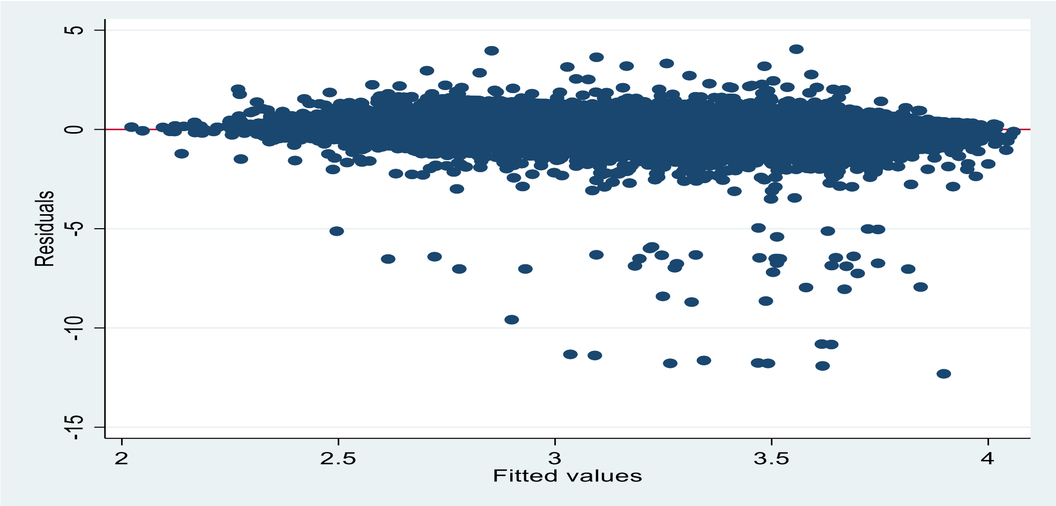

Figure 2. Histogram for distribution of weekly pay and hourly wages

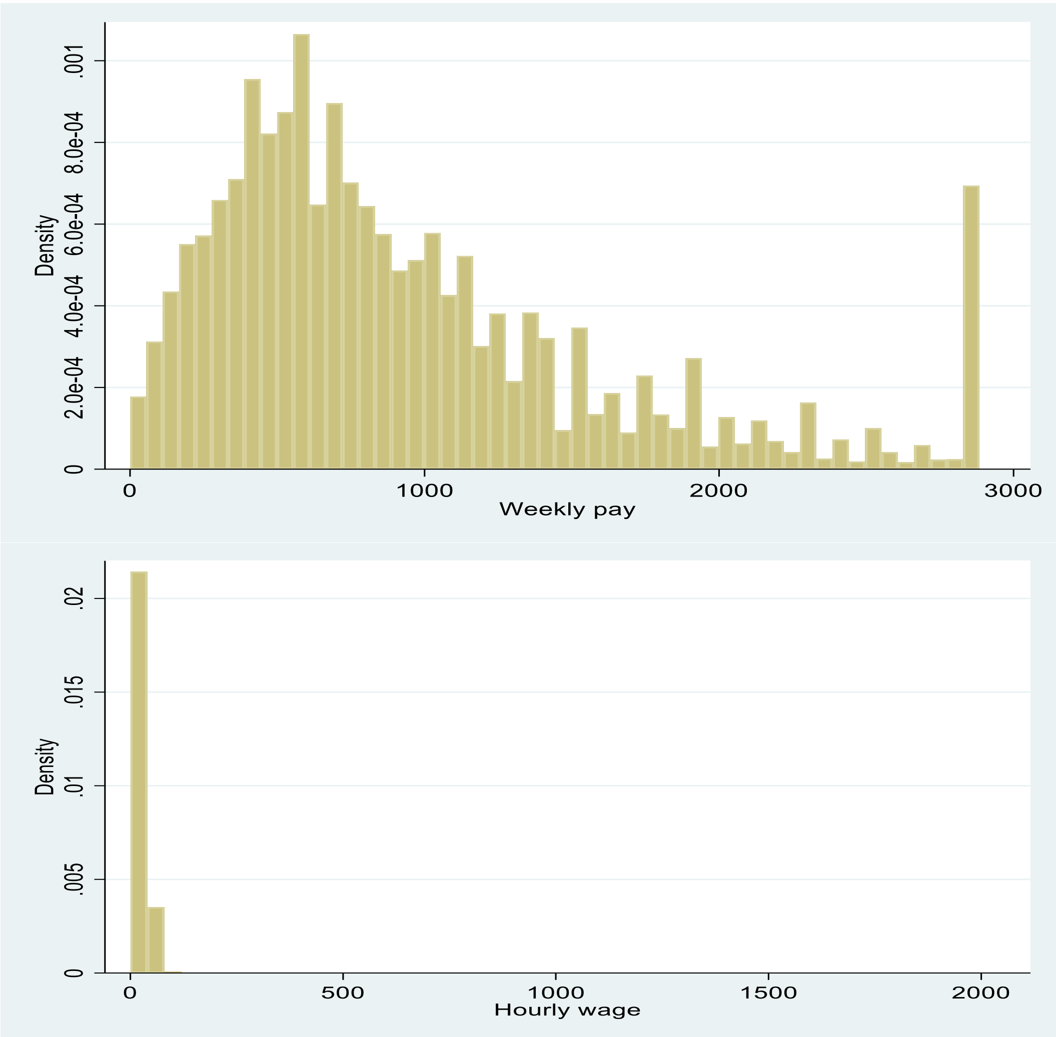

Table1. Observed characteristics used in the regression analysis

Personal characteristics	Value
Age	People of different ages
Female	if 1=female, if 0=men
Race	Divided into 4 races: white, black, Spanish and other
Married	If 1=married, if 0=do not married
Number of children	Divided into groups of 11 children
Employed	If 1=employed.
Public sector	If 1=employed in public sector, if 0= employed not in public sector
Union	If 1= participate in a union, if 0= not participate in union
Has more than one job	If 1= participate in a union, if 0= not participate in union, if 0= not participate in union
State	Divided into 51 states
Education	Divided into 5 levels: Lower then High School, High School, Some College, College, Advanced
Agriculture	If 1= employed in agriculture, if 0= not employed in agriculture
Manufacturing	If 1= employed in manufacturing, if 0= not employed in manufacturing
Major occupation	Major occupation recode. Divided into 10 categories
Hours last week, all jobs	Hours last week, all jobs
Weekly pay	Wage per week
Hourly pay	Wage per hour

Table 2: Table of statistics for gender wage gap

Major Occupation	**Weekly wage female**	**Weekly wage male**	**Gap**	**% Gap**	**Hourly wage female**	**Hourly wage male**	**Gap**	**% Gap**
Management, business, and financial occupations	1226,81	1 611,59	-384,78	15 %	30,39	37,48	-7,09	19 %
Professional and related occupations	1042,13	1 430,19	-388,07	16 %	28,54	35,69	-7,15	19 %
Service occupations	450,29	642,98	-192,69	8 %	14,29	17,47	-3,18	9 %
Sales and related occupations	599,71	1 002,24	-402,53	16 %	17,00	24,04	-7,04	19 %
Office and administrative support occupations	681,27	782,20	-100,93	4 %	18,97	20,41	-1,44	4 %

Farming, fishing, and forestry occupations	449,70	662,72	-213,02	9 %	12,46	15,66	-3,20	1 %
Construction and extraction occupations	826,60	945,21	-118,61	5 %	21,75	23,85	-2,10	6 %
Installation, maintenance, and repair	808,90	984,66	-175,75	7 %	20,60	24,65	-4,05	2 %
Production occupations	596,85	876,40	-279,56	11 %	15,85	21,66	-5,81	16 %
Transportation and material moving occupations	538,14	784,33	-246,19	10 %	16,11	19,61	-3,50	9 %
Total			**-2 502,11**				**-37,32**	
Race								
White	848,37	1 144,37	-295,99	34 %	23,28	28,04	-4,76	32 %
Black	710,19	824,50	-114,31	13 %	19,11	20,97	-1,87	13 %
Hispanic	638,14	811,60	-173,45	20 %	17,94	20,56	-2,62	18 %
Other	885,52	1 171,74	-286,22	33 %	23,72	29,21	-5,49	37 %
Total			-869,97		84,05	98,79	-14,74	
Education								
LTHS	345,11	557,04	-211,93	14 %	11,78	15,52	-3,74	14 %
HS	569,06	833,05	-263,98	18 %	16,22	20,85	-4,63	18 %
Some college	651,26	916,67	-265,41	18 %	18,52	23,04	-4,52	17 %
College	1032,90	1 405,89	-372,99	25 %	27,74	33,84	-6,10	24 %
Advanced	1331,36	1 700,80	-369,44	25 %	34,36	41,24	-6,88	27 %
Total			**-1 483,76**				**-25,87**	
State								
Maine	770,78	987,10	-216,32	2 %	21,23	24,63	-3,41	2 %
New Hampshire	825,38	1 123,82	-298,44	2 %	22,51	27,91	-5,40	2 %
Vermont	813,72	1 004,57	-190,85	1 %	22,67	24,90	-2,23	1 %
Massachusetts	964,21	1 287,66	-323,45	2 %	26,50	31,42	-4,92	2 %
Rhode Island	830,46	1 054,94	-224,48	2 %	23,54	26,02	-2,48	1 %
Connecticut	899,37	1 208,60	-309,23	2 %	26,48	30,90	-4,42	2 %
New York	906,77	1 135,90	-229,13	2 %	24,33	27,90	-3,58	2 %
New Jersey	988,25	1 243,30	-255,05	2 %	26,19	31,17	-4,98	2 %
Pennsylvania	787,19	1 054,83	-267,64	2 %	21,80	26,09	-4,29	2 %
Ohio	746,61	975,95	-229,34	2 %	20,54	24,20	-3,66	2 %
Indiana	748,14	997,83	-249,69	2 %	21,86	24,46	-2,60	1 %
Illinois	832,42	1 123,11	-290,69	2 %	22,76	27,30	-4,54	2 %
Michigan	765,94	1 047,24	-281,29	2 %	21,12	25,50	-4,38	2 %
Wisconsin	744,93	1 015,44	-270,51	2 %	20,43	24,98	-4,56	2 %
Minnesota	858,76	1 134,42	-275,66	2 %	23,57	27,77	-4,21	2 %

Iowa	717,17	994,77	-277,61	2 %	19,88	24,04	-4,16	2 %
Missouri	774,49	997,85	-223,36	2 %	21,00	24,79	-3,79	2 %
North Dakota	743,35	1 019,89	-276,54	2 %	20,98	25,61	-4,63	2 %
South Dakota	682,19	941,69	-259,50	2 %	19,04	23,28	-4,24	2 %
Nebraska	765,74	971,10	-205,36	2 %	21,07	25,96	-4,89	2 %
Kansas	758,90	1 003,94	-245,05	2 %	20,69	25,15	-4,46	2 %
Delaware	794,21	1 037,33	-243,12	2 %	21,89	25,93	-4,04	2 %
Maryland	969,13	1 229,97	-260,84	2 %	26,16	30,47	-4,31	2 %
Columbia	1296,53	1 526,32	-229,78	2 %	33,89	36,58	-2,69	1 %
Virginia	871,50	1 200,22	-328,72	2 %	23,36	29,61	-6,26	3 %
West Virginia	704,71	953,40	-248,70	2 %	19,38	23,47	-4,09	2 %
North Carolina	777,85	1 028,36	-250,51	2 %	21,08	24,83	-3,75	2 %
South Carolina	745,35	1 046,41	-301,06	2 %	20,46	25,42	-4,96	2 %
Georgia	768,91	1 022,29	-253,38	2 %	20,50	24,89	-4,39	2 %
Florida	775,02	998,99	-223,97	2 %	20,67	24,51	-3,85	2 %
Kentucky	733,89	931,25	-197,36	1 %	19,82	24,20	-4,38	2 %
Tennessee	757,61	978,28	-220,68	2 %	20,08	23,63	-3,55	2 %
Alabama	704,35	995,44	-291,08	2 %	19,21	24,43	-5,22	2 %
Mississippi	696,28	932,15	-235,87	2 %	18,88	23,11	-4,23	2 %
Arkansas	727,81	920,78	-192,97	1 %	19,30	22,60	-3,30	2 %
Louisiana	723,09	1 029,22	-306,13	2 %	19,48	24,88	-5,40	2 %
Oklahoma	714,48	1 043,54	-329,06	2 %	19,66	25,00	-5,34	2 %
Texas	788,38	1 051,56	-263,18	2 %	20,96	25,51	-4,55	2 %
Montana	696,64	966,89	-270,25	2 %	20,56	25,10	-4,53	2 %
Idaho	701,92	986,27	-284,34	2 %	19,59	24,02	-4,43	2 %
Wyoming	693,97	1 047,96	-353,99	3 %	19,41	25,19	-5,78	3 %
Colorado	869,33	1 160,45	-291,13	2 %	23,70	28,24	-4,54	2 %
New Mexico	732,84	912,65	-179,81	1 %	20,27	23,16	-2,88	1 %
Arizona	767,98	1 018,55	-250,58	2 %	22,32	25,12	-2,80	1 %
Utah	670,06	1 042,69	-372,64	3 %	19,41	25,75	-6,34	3 %
Nevada	719,49	960,24	-240,75	2 %	20,24	24,06	-3,83	2 %
Washington	887,66	1 213,45	-325,80	2 %	24,93	29,88	-4,95	2 %
Oregon	805,84	1 043,91	-238,07	2 %	22,57	26,22	-3,65	2 %
California	902,85	1 172,60	-269,74	2 %	25,01	29,30	-4,29	2 %
Alaska	897,01	1 165,09	-268,09	2 %	24,45	29,60	-5,15	2 %
Hawaii	801,93	1 022,00	-220,07	2 %	21,78	25,87	-4,09	2 %
Total			**-13 340,85**				**-217,36**	
Children								
0	864,20	1 162,85	-298,65	6 %	23,40	28,99	-5,58	12 %
1	855,28	1 226,55	-371,28	8 %	22,89	29,24	-6,36	14 %
2	895,43	1 304,12	-408,70	8 %	24,35	30,71	-6,37	14 %
3	793,91	1 254,23	-460,32	10 %	22,86	29,32	-6,46	14 %
4	697,84	1 213,12	-515,28	11 %	20,81	28,62	-7,82	17 %
5	595,11	1 147,30	-552,20	11 %	17,58	26,68	-9,09	19 %
6	550,78	1 118,50	-567,72	12 %	50,64	25,85	24,79	-53 %
7	636,39	1 253,83	-617,44	13 %	20,40	28,56	-8,15	17 %
8	696,73	1 006,90	-310,17	6 %	18,54	22,73	-4,18	9 %

9	860,00	1 408,20	-548,20	11 %	10,00	33,04	-23,04	49 %
10	145,00	317,60	-172,60	4 %	-	19,85	-	-
Total			**-4 822,56**				**-46,68**	

Tabell 3. Cameron & Trivedi's decomposition of IM-test. Linktest

Source	chi2	df	p
Heteroskedasticity	3238.99	1965	0.0000
Skewness	156.74	84	0.0000
Kurtosis	20.99	1	0.0000
Total	3416.72	2050	0.0000

Linktest

```
      Source |       SS           df       MS      Number of obs   =    48,663
-------------+----------------------------------   F(2, 48660)     =   9294.83
       Model |  5303.01279         2  2651.50639   Prob > F        =    0.0000
    Residual |   13881.076    48,660  .285266667   R-squared       =    0.2764
-------------+----------------------------------   Adj R-squared   =    0.2764
       Total |  19184.0888    48,662  .394231408   Root MSE        =     .5341

------------------------------------------------------------------------------
      lwage4 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        _hat |   1.357907    .126547    10.73   0.000     1.109874    1.605941
      _hatsq |  -.0560117    .019771    -2.83   0.005    -.0947632   -.0172603
       _cons |  -.5656361    .201054    -2.81   0.005    -.9597045   -.1715678
------------------------------------------------------------------------------
```

Table 4. Results from OLS and quantile regressions estimated with 0.25, 0,5 and 0,75 quantiles.

Variable	Reg pool	qr2 pool	qr3 pool	qr4 pool	Reg male	qr2 male	qr3 male	qr4 male	Reg female	qr2female	qr3 female	qr4 female
	All genders	All genders , o,25 q.	All genders, 05 q.	All genders, 0,75 q.	Male	Male, 025 q	Male, 0,5 q.	Male, 0,75 q.	Female	Female, 025 q.	Female, 0,5 q.	Female, 0,75 q.
Age	0.0049***	0.0037***	0.0055***	0.0072***	0.0046***	0.0038***	0.0056***	0.0072***	0.0051***	0.0032***	0.0051***	0.0069***
Female	-0.1898***	-0.1601***	-0.2069***	-0.2252***	-	-	-	-	-	-	-	-
Black	-0.1195***	-0.1145***	-0.1202***	-0.1335***	-0.1634***	-0.1790***	-0.1800***	-0.1786***	-0.0844***	-0.0708***	-0.0797***	-0.0967***
Hispanic	-0.1059***	-0.0963***	-0.1068***	-0.1125***	-0.1239***	-0.1153***	-0.1309***	-0.1400***	-0.0820***	-0.0696***	-0.0817***	-0.0835***
Others	-0.0349***	-0.0311***	-0.0205**	-0.0183*	-0.0481***	-0.0469***	-0.0253*	-0.0247*	-0.0156	-0.0212*	-0.0098	-0.0020
Married	0.0550***	0.0504***	0.0539***	0.0642***	0.1040***	0.0935***	0.0921***	0.1037***	0.0293***	0.0334***	0.0383***	0.0439***
Child 0	-0.0747***	-0.0705***	-0.0783***	-0.0887***	-0.0791***	-0.0754***	-0.0858***	-0.0951***	-0.0638***	-0.0502***	-0.0617***	-0.0720***
Child 1	-0.0213***	-0.0204***	-0.0278***	-0.0309***	-0.0176*	-0.0181*	-0.0300***	-0.0311***	-0.0231***	-0.0159*	-0.0255***	-0.0246**
Child 3	-0.0105	-0.0192*	-0.0129	-0.0066	-0.0041	-0.0135	-0.0032	0.0027	-0.0158	-0.0256*	-0.0178	-0.0118
Child 4	0.0019	-0.0082	-0.0271*	-0.0114	0.0166	-0.0015	-0.0065	-0.0023	-0.0145	-0.0210	-0.0439*	-0.0128
Child 5	-0.0539*	-0.0329	-0.0979***	-0.0399	-0.0385	-0.0377	-0.0668	0.0031	-0.0806*	-0.0269	-0.1044*	-0.1171*

Child 6	-0.0434	-0.0719	-0.0924	-0.1053*	-0.0414	-0.0492	-0.0221	-0.0528	-0.0430	-0.1019	-0.1552*	-0.1466
Child 7	-0.0543	-0.1884	-0.0017	0.0737	-0.1111	-0.2645*	-0.2293*	-0.0680	0.0755	0.0791	0.1635	0.1205
Child 8	-0.2362	-0.2161	-0.3850*	-0.4912*	-0.2597	-0.2780	-0.3468	-0.4667	-0.1311	-0.1307	-0.3180	-0.0935
Child 9	-0.1088	-0.1705	-0.3176	-0.5549	0.0573	-0.0772	-0.3144	0.3837	-0.6013***	-0.4477	-0.6594	-0.8439
Child 10	0.5714***	0.8448	0.6134	0.3574	0.6046***	0.9712	0.7112	0.3773	(omitted)	(omitted)	(omitted)	(omitted)
Public sector	-0.0284***	-0.0124*	-0.0322***	-0.0329***	0.0045	0.0181*	0.0039	-0.0084	-0.0521***	-0.0220**	-0.0468***	-0.0498***
Union	0.1198***	0.1131***	0.1111***	0.1136***	0.1578***	0.1401***	0.1454***	0.1641***	0.0711***	0.0834***	0.0682***	0.0464***
More than 1 job	-0.0840***	-0.1206***	-0.0994***	-0.0678***	-0.0771***	-0.1222***	-0.1035***	-0.0642***	-0.0905***	-0.1194***	-0.1020***	-0.0723***
Maine	-0.0863***	-0.0625**	-0.1229***	-0.1569***	-0.0782**	-0.0289	-0.1345***	-0.1194***	-0.0925***	-0.1121***	-0.1074***	-0.1654***
New Hampshire	-0.0411*	-0.0219	-0.0432**	-0.0822***	-0.0103	0.0026	-0.0325	-0.0511	-0.0696**	-0.0398	-0.0555*	-0.0981***
Vermont	-0.0861***	-0.0558**	-0.0849***	-0.1450***	-0.0918***	-0.0628*	-0.0824**	-0.1559***	-0.0777***	-0.0542*	-0.0850***	-0.1109***
Massa-chusetts	0.0596***	0.0769***	0.0685***	0.0402*	0.0523	0.1102***	0.0917***	0.0436	0.0642**	0.0533**	0.0462*	0.0534*
Rhode Island	-0.0063	0.0151	-0.0151	-0.0355	-0.0297	-0.0073	-0.0282	-0.0109	0.0192	0.0269	0.0057	-0.0275
Connec-ticut	0.0492**	0.0731***	0.0434*	0.0073	0.0625*	0.1145***	0.0484	0.0233	0.0392	0.0362	0.0466	-0.0030
New Jersey	0.0726***	0.0719***	0.0841***	0.0513**	0.0822***	0.1099***	0.0879***	0.0591*	0.0637**	0.0181	0.0770***	0.0538*
Pennsyl-vania	-0.0346**	-0.0397**	-0.0469***	-0.0606***	-0.0268	-0.0270	-0.0415*	-0.0422	-0.0437*	-0.0563**	-0.0532**	-0.0676**
Ohio	-0.0840***	-0.0592***	-0.0878***	-0.1125***	-0.0883***	-0.0566*	-0.0933***	-0.0850***	-0.0785***	-0.0670***	-0.0924***	-0.1134***
Indiana	-0.0688***	-0.0814***	-0.1067***	-0.0970***	-0.0514**	-0.0464	-0.0912***	-0.0654*	-0.0854***	-0.1101***	-0.1144***	-0.1116***
Illinois	-0.0111	-0.0372**	-0.0208	-0.0079	0.0048	-0.0014	0.0144	0.0208	-0.0265	-0.0659***	-0.0521**	-0.0131
Michigan	-0.0752***	-0.0510**	-0.1058***	-0.1096***	-0.0696***	-0.0418	-0.1053***	-0.0779***	-0.0774***	-0.0603**	-0.0943***	-0.1260***
Wisconsin	-0.0396**	0.0051	-0.0530**	-0.0990***	-0.0286	0.0283	-0.0321	-0.0732**	-0.0504**	-0.0224	-0.0599**	-0.1123***
Minnesota	0.0099	0.0284	-0.0068	-0.0320	0.0221	0.0449	0.0034	-0.0015	0.0026	0.0048	-0.0185	-0.0392
Iowa	-0.1052***	-0.0699***	-0.0931***	-0.1169***	-0.1124***	-0.0793**	-0.0822**	-0.0951**	-0.0957***	-0.0739**	-0.0995***	-0.1068***
Missouri	-0.0789***	-0.0679***	-0.0855***	-0.1045***	-0.0703**	-0.0360	-0.0505*	-0.0690*	-0.0865***	-0.0869***	-0.1092***	-0.1295***
North Dakota	-0.0610***	-0.0387*	-0.0659***	-0.1264***	-0.0475*	-0.0472	-0.0608*	-0.0861**	-0.0706***	-0.0371	-0.0722**	-0.1447***
South Dakota	-0.1109***	-0.0958***	-0.1179***	-0.1687***	-0.1051***	-0.0946**	-0.1327***	-0.1143***	-0.1189***	-0.1029***	-0.1060***	-0.1885***
Nebraska	-0.0577***	-0.0471*	-0.0680***	-0.1128***	-0.0391	-0.0294	-0.0453	-0.1110***	-0.0724***	-0.0529*	-0.0917***	-0.0908***
Kansas	-0.0763***	-0.0742***	-0.1063***	-0.1266***	-0.0726***	-0.0563*	-0.1029***	-0.0973***	-0.0750***	-0.0921***	-0.1133***	-0.1239***
Delaware	0.0030	0.0112	-0.0158	-0.0262	-0.0001	0.0365	0.0075	-0.0108	0.0095	-0.0188	-0.0168	-0.0160
Maryland	0.1158***	0.0977***	0.1062***	0.1035***	0.1322***	0.1347***	0.1295***	0.1264***	0.0960***	0.0510*	0.0726**	0.1022***
District of Columbia	0.2057***	0.2252***	0.2266***	0.1636***	0.1776***	0.2651***	0.2126***	0.1430***	0.2169***	0.1826***	0.2191***	0.2078***
Virginia	0.0321*	0.0045	0.0157	0.0164	0.0704***	0.0416	0.0668**	0.0636*	-0.0068	-0.0545*	-0.0385	-0.0149
West Virginia	-0.1138***	-0.1239***	-0.1234***	-0.1575***	-0.1016***	-0.1071***	-0.1111***	-0.1214***	-0.1227***	-0.1256***	-0.1352***	-0.1681***
North Carolina	-0.0659***	-0.0539***	-0.0815***	-0.0974***	-0.0629**	-0.0484*	-0.0653**	-0.0839***	-0.0705***	-0.0696***	-0.0999***	-0.0898***
South Carolina	-0.0723***	-0.0651***	-0.0990***	-0.0989***	-0.0577*	-0.0108	-0.0661**	-0.0605*	-0.0885***	-0.1068***	-0.1320***	-0.1110***
Georgia	-0.0818***	-0.0963***	-0.1009***	-0.0895***	-0.0500*	-0.0679**	-0.0734***	-0.0236	-0.1146***	-0.1188***	-0.1211***	-0.1323***

Florida	-0.0595***	-0.0707***	-0.0766***	-0.0717***	-0.0635***	-0.0853***	-0.0862***	-0.0527**	-0.0545**	-0.0728***	-0.0689***	-0.0706***
Kentucky	-0.1198***	-0.1173***	-0.1277***	-0.1523***	-0.0982***	-0.0726*	-0.1130***	-0.1302***	-0.1395***	-0.1609***	-0.1559***	-0.1532***
Tennes-see	-0.1271***	-0.1213***	-0.1313***	-0.1505***	-0.1084***	-0.0948***	-0.1070***	-0.1073***	-0.1453***	-0.1374***	-0.1691***	-0.1723***
Alabama	-0.0812***	-0.0947***	-0.1211***	-0.1247***	-0.0355	-0.0380	-0.0663**	-0.0717**	-0.1267***	-0.1591***	-0.1787***	-0.1664***
Mississippi	-0.1058***	-0.1398***	-0.1324***	-0.1223***	-0.0703***	-0.1000***	-0.1067***	-0.0903***	-0.1397***	-0.1799***	-0.1495***	-0.1442***
Arkansas	-0.1156***	-0.1287***	-0.1325***	-0.1436***	-0.1030***	-0.1022***	-0.1250***	-0.0832***	-0.1295***	-0.1553***	-0.1457***	-0.1654***
Louisiana	-0.0617***	-0.0957***	-0.0882***	-0.0626***	-0.0130	-0.0424	-0.0290	-0.0155	-0.1106***	-0.1399***	-0.1350***	-0.1132***
Okla-homa	-0.0958***	-0.1040***	-0.1162***	-0.0983***	-0.0605**	-0.0599*	-0.0790***	-0.0438	-0.1317***	-0.1617***	-0.1338***	-0.1385***
Texas	-0.0266*	-0.0422***	-0.0513***	-0.0391**	0.0004	-0.0060	-0.0107	0.0137	-0.0545***	-0.0859***	-0.0893***	-0.0730***
Montana	-0.1127***	-0.0893***	-0.0951***	-0.1574***	-0.1107***	-0.0647*	-0.0757***	-0.1283***	-0.1133***	-0.1123***	-0.1153***	-0.1758***
Idaho	-0.0947***	-0.0759***	-0.1074***	-0.1201***	-0.0804***	-0.0404	-0.0729**	-0.0947***	-0.1089***	-0.1075***	-0.1455***	-0.1262***
Wyoming	-0.0584***	-0.0203	-0.0603***	-0.1022***	0.0028	0.0364	-0.0039	-0.0176	-0.1183***	-0.0716**	-0.0961***	-0.1652***
Colorado	0.0270	0.0379*	0.0134	0.0003	0.0333	0.0702*	0.0195	0.0082	0.0230	-0.0154	0.0071	0.0038
New Mexico	-0.0621***	-0.0621***	-0.0808***	-0.0615***	-0.0667*	-0.0600*	-0.0652**	-0.0178	-0.0591*	-0.0790***	-0.0881***	-0.0685**
Arizona	-0.0189	-0.0226	-0.0234	-0.0460*	-0.0155	-0.0265	-0.0222	-0.0308	-0.0205	-0.0250	-0.0324	-0.0611*
Utah	-0.0570***	-0.0391*	-0.0766***	-0.1203***	-0.0268	0.0168	-0.0523*	-0.0770**	-0.0882***	-0.0777***	-0.1021***	-0.1332***
Nevada	-0.0104	-0.0121	-0.0115	-0.0483*	-0.0051	0.0102	-0.0027	-0.0306	-0.0060	-0.0227	-0.0106	-0.0352
Washing-ton	0.0817***	0.0895***	0.0717***	0.0465**	0.0959***	0.1344***	0.0918***	0.0931***	0.0666***	0.0463*	0.0559*	0.0238
Oregon	-0.0043	-0.0157	-0.0131	-0.0438*	0.0055	0.0147	-0.0142	-0.0033	-0.0099	-0.0378	-0.0057	-0.0398
California	0.0980***	0.0840***	0.0747***	0.0767***	0.1034***	0.1007***	0.0819***	0.0885***	0.0967***	0.0656***	0.0671***	0.0798***
Alaska	0.0866***	0.0997***	0.0957***	0.0488*	0.0797**	0.1285***	0.0993***	0.0527	0.0922***	0.0867***	0.0957***	0.0687*
Hawaii	-0.0434*	-0.0269	-0.0823***	-0.0909***	-0.0343	0.0219	-0.0538*	-0.0532	-0.0550*	-0.0486*	-0.0983***	-0.1124***
Lowe	-0.1757***	-0.1352***	-0.1759***	-0.2229***	-0.1858***	-0.1624***	-0.1923***	-0.2204***	-0.1519***	-0.0922***	-0.1467***	-0.2102***
Some college	0.0750***	0.0638***	0.0831***	0.0938***	0.0635***	0.0628***	0.0795***	0.0815***	0.0821***	0.0654***	0.0790***	0.0947***
College	0.2971***	0.2468***	0.3167***	0.3703***	0.2725***	0.2273***	0.3027***	0.3611***	0.3162***	0.2531***	0.3183***	0.3702***
Others	0.3968***	0.3770***	0.4287***	0.4644***	0.3401***	0.3236***	0.3904***	0.4216***	0.4496***	0.4035***	0.4595***	0.5025***
Manage-ment, business, and financial occupations	0.0831***	0.1170***	0.0777***	0.0608***	0.0815***	0.0997***	0.0725***	0.0594***	0.0763***	0.1127***	0.0738***	0.0535***
Service occu-pations	-0.3840***	-0.3399***	-0.4078***	-0.4194***	-0.4054***	-0.4150***	-0.4619***	-0.4296***	-0.3877***	-0.3193***	-0.4008***	-0.4383***
Sales and related occupations	-0.2412***	-0.2707***	-0.2937***	-0.2324***	-0.1946***	-0.2539***	-0.2464***	-0.1648***	-0.2956***	-0.2933***	-0.3497***	-0.3175***
Office and administrative support occupapations	-0.2127***	-0.1609***	-0.2330***	-0.2669***	-0.3040***	-0.3132***	-0.3662***	-0.3138***	-0.1900***	-0.1181***	-0.2081***	-0.2671***
Farming, fishing, and forestry occupations	-0.4449***	-0.3738***	-0.4281***	-0.4879***	-0.4407***	-0.3837***	-0.4482***	-0.4622***	-0.4741***	-0.3407***	-0.4271***	-0.5210***
Construction and extraction occupations	-0.0845***	-0.0731***	-0.1176***	-0.0882***	-0.1050***	-0.1187***	-0.1423***	-0.1041***	-0.0397	-0.0439	-0.0861	0.0025

Installation, maintenance, and repair	-0.1063***	-0.0580***	-0.1018***	-0.1266***	-0.1238***	-0.1039***	-0.1274***	-0.1406***	-0.1742***	-0.1520**	-0.2522***	-0.0811
Production occupations	-0.2782***	-0.2522***	-0.3139***	-0.3194***	-0.2632***	-0.2715***	-0.3020***	-0.2893***	-0.3628***	-0.2897***	-0.3763***	-0.4399***
Transportation and material moving occupations	-0.3265***	-0.3134***	-0.3591***	-0.3586***	-0.3390***	-0.3542***	-0.3735***	-0.3488***	-0.3297***	-0.2995***	-0.3865***	-0.3990***
Agriculture	-0.0810**	-0.0598*	-0.0783***	-0.0514*	-0.1015*	-0.1182***	-0.0989***	-0.0754*	-0.0365	-0.0070	-0.0198	-0.0273
Manufacture	0.1004***	0.1185***	0.1056***	0.0784***	0.0961***	0.1219***	0.0945***	0.0598***	0.1178***	0.1051***	0.1234***	0.1323***
Hours worked last week	0.0046***	0.0068***	0.0062***	0.0041***	0.0038***	0.0064***	0.0056***	0.0028***	0.0050***	0.0067***	0.0066***	0.0050***
_cons	2.8095***	2.4626***	2.7497***	3.0642***	2.8240***	2.4668***	2.7498***	3.0611***	2.6151***	2.3336***	2.5605***	2.8386***
N	98223	98223	98223	98223	48663	48663	48663	48663	49560	49560	49560	49560
ll	-73608				-38533				-34529			
aic	147386	.	.	.	77233	.	.	.	69224	.	.	.
bic	148193	.	.	.	77972	.	.	.	69955	.	.	.
		legend: * p<0.05; ** p<0.01; *** p<0.001										

Table 5. Results from decomposition. Female vs Male.

Female vs Male					Male vs Female			
	OLS	Q.reg, q.0,25	Q.reg. q. 0,5	Q.reg., q. 0,75	OLS	Q.reg, q. 0,25	Q.reg., q. 0,5	Q.reg., q. 0,75
Summary of decomposition results	%)	%)	%)	%)	%)	%)	%)	%)
Amount attributable (CE):	-2,4	-8,9	-7,1	-5,1	2,4	8,9	7,1	-12,7
- due to endowments (E):	-3,5	-3,6	-3,5	-5,3	5,8	7,4	7	2,5
- due to coefficients (C):	1,1	-5,3	-3,6	0,2	-3,3	1,5	0,1	-15,2
Shift coefficient (U):	-20,9	-13,3	-18,9	-22,2	20,9	13,3	18,9	25,7
Raw differential (R) {E+C+U}:	-23,3	-22,2	-26	-27,3	23,3	22,2	26	13,1
Adjusted differential (D) {C+U}:	-19,8	-18,6	-22,5	-22	17,6	14,8	19	10,5
Endowments as % total (E/R):	15	16,2	13,6	19,4	24,8	33,2	26,9	19,4
Discrimination as % total (D/R):	85	83,8	86,4	8,6	75,2	66,8	73,1	80,6

Table 6. Results from the Oaxaca decomposition.

```
Summary of decomposition results:
High: female==      0.0000
Low:  female==      1.0000
-------------------------------------------------------------------
  Mean prediction high (H):   3.207
   Mean prediction low (L):   2.973
Raw differential (R) {H-L}:   0.233
    - due to endowments (E):  0.035
  - due to coefficients (C):  0.176
  - due to interaction (CE):  0.023
-------------------------------------------------------------------
                        D:    0       1       0.5     0.495   *
                              -------------------------------------
Unexplained (U){C+(1-D)CE}:   0.198   0.176   0.187   0.187   0.144
    Explained (V) {E+D*CE}:   0.035   0.058   0.046   0.046   0.089
       % unexplained {U/R}:    85.0    75.2    80.1    80.1    61.9
         % explained (V/R):    15.0    24.8    19.9    19.9    38.1
-------------------------------------------------------------------
Note: D in 4th column = relative frequency of high group
      * reference: pooled model over both categories
```

				explain	ned: D =		
Variables	E(D=0)	C	CE	1	0.5	0.495	*
Age	0.007	-0.022	-0.001	0.006	0.006	0.006	0.007
White	0.000	0.101	0.003	0.004	0.002	0.002	0.001
Black	0.003	0.008	-0.003	0.000	0.001	0.001	0.003
Hispanic	-0.001	0.014	0.002	0.001	-0.000	-0.000	-0.001
Other	0.000	0.009	-0.000	-0.000	-0.000	-0.000	0.000
Married	0.004	0.058	0.010	0.014	0.009	0.009	0.012
0 own child	-0.000	-0.040	0.000	0.000	-0.000	-0.000	-0.000
1 own child	-0.003	-0.013	0.002	-0.001	-0.002	-0.002	-0.006
2 own child	0.001	-0.012	-0.001	0.001	0.001	0.001	0.002
3 own child	0.001	-0.003	-0.001	0.001	0.001	0.001	0.003
4 own child	0.001	-0.001	-0.000	0.001	0.001	0.001	0.002
5 own child	0.000	-0.000	-0.000	0.000	0.000	0.000	0.000
6 own child	0.000	-0.000	-0.000	0.000	0.000	0.000	0.000
7 own child	0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
8 own child	0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000
9 own child	-0.000	0.000	0.000	0.000	-0.000	-0.000	0.000
10 own child	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Public sector	0.003	0.012	-0.003	-0.000	0.001	0.001	0.002
Union	0.001	0.010	0.002	0.003	0.002	0.002	0.002
Multiple jobs	0.000	0.001	-0.000	0.000	0.000	0.000	0.000
Maine	0.000	0.001	-0.000	0.000	0.000	0.000	0.000
New Hampshire	-0.000	0.002	0.000	0.000	-0.000	-0.000	0.000
Vermont	0.000	0.001	-0.000	0.000	0.000	0.000	-0.000
Massachusetts	-0.000	0.002	-0.000	-0.000	-0.000	-0.000	-0.000

Rhode Island	0.000	0.000	-0.000	-0.000	-0.000	-0.000	-0.000
Connecticut	-0.000	0.001	-0.000	-0.000	-0.000	-0.000	-0.000
New York	0.000	0.004	-0.000	-0.000	-0.000	-0.000	-0.000
New Jersey	0.000	0.002	0.000	0.000	0.000	0.000	0.000
Pennsylvania	0.000	0.003	-0.000	-0.000	0.000	0.000	-0.000
Ohio	0.000	0.002	-0.000	0.000	0.000	0.000	-0.000
Indiana	0.000	0.002	-0.000	-0.000	0.000	0.000	-0.000
Illinois	0.000	0.004	-0.000	-0.000	-0.000	-0.000	-0.000
Michigan	-0.000	0.002	0.000	0.000	-0.000	-0.000	0.000
Wisconsin	0.000	0.002	-0.000	-0.000	0.000	0.000	-0.000
Minnesota	0.000	0.002	-0.000	-0.000	-0.000	-0.000	-0.000
Iowa	0.000	0.001	-0.000	0.000	0.000	0.000	0.000
Missouri	0.000	0.002	-0.000	-0.000	0.000	0.000	-0.000
North Dakota	0.000	0.002	-0.000	-0.000	0.000	0.000	-0.000
South Dakota	0.000	0.001	-0.000	0.000	0.000	0.000	0.000
Nebraska	0.000	0.002	-0.000	-0.000	0.000	0.000	-0.000
Kansas	-0.000	0.001	0.000	0.000	-0.000	-0.000	0.000
Delaware	0.000	0.001	-0.000	-0.000	-0.000	-0.000	-0.000
Maryland	-0.000	0.002	-0.000	-0.000	-0.000	-0.000	-0.000
District of Columbia	-0.001	0.001	-0.000	-0.001	-0.001	-0.001	-0.001
Virginia	0.000	0.003	-0.000	-0.000	-0.000	-0.000	-0.000
West Virginia	-0.000	0.002	0.000	-0.000	-0.000	-0.000	-0.000
North Carolina	0.000	0.002	-0.000	-0.000	0.000	0.000	-0.000
South Carolina	0.000	0.002	-0.000	-0.000	0.000	0.000	-0.000
Georgia	0.000	0.004	-0.000	-0.000	0.000	0.000	-0.000
Florida	0.000	0.004	-0.000	-0.000	0.000	0.000	-0.000
Kentucky	0.000	0.002	-0.000	0.000	0.000	0.000	0.000
Tennessee	-0.000	0.002	0.000	-0.000	-0.000	-0.000	-0.000
Alabama	-0.000	0.004	0.000	0.000	-0.000	-0.000	0.000
Mississippi	0.000	0.003	-0.000	-0.000	0.000	0.000	0.000
Arkansas	0.000	0.002	-0.000	0.000	0.000	0.000	0.000
Louisiana	0.000	0.004	-0.000	-0.000	0.000	0.000	-0.000
Oklahoma	-0.000	0.002	0.000	0.000	-0.000	-0.000	-0.000
Texas	-0.000	0.008	0.001	0.000	0.000	0.000	0.000
Montana	0.000	0.002	-0.000	0.000	0.000	0.000	0.000
Idaho	-0.000	0.002	0.000	-0.000	-0.000	-0.000	-0.000
Wyoming	-0.000	0.003	0.000	0.000	-0.000	-0.000	0.000
Colorado	0.000	0.001	0.000	0.000	0.000	0.000	0.000
New Mexico	0.000	0.001	-0.000	-0.000	0.000	0.000	-0.000
Arizona	-0.000	0.001	0.000	0.000	0.000	0.000	0.000
Utah	-0.000	0.002	0.001	0.000	-0.000	-0.000	0.000
Nevada	-0.000	0.001	0.000	0.000	0.000	0.000	0.000
Washington	0.000	0.002	0.000	0.000	0.000	0.000	0.000
Oregon	-0.000	0.002	0.000	0.000	0.000	0.000	0.000
California	0.001	0.008	0.001	0.002	0.001	0.001	0.002
Alaska	-0.000	0.001	-0.000	-0.000	-0.000	-0.000	-0.000
Hawaii	0.000	0.002	-0.000	-0.000	0.000	0.000	-0.000
Low then High School	0.000	0.000	0.000	0.000	0.000	0.000	-0.012
High School	0.006	0.008	0.001	0.007	0.007	0.007	-0.016

Some college	-0.009	0.005	-0.001	-0.010	-0.010	-0.010	0.013
College	-0.006	-0.002	0.000	-0.006	-0.006	-0.006	0.001
Advanced	-0.004	-0.012	0.000	-0.003	-0.003	-0.003	0.000
Management, business and finance	0.010	-0.027	-0.003	0.007	0.008	0.008	0.002
Professional and related	-0.047	-0.052	0.017	-0.030	-0.038	-0.038	0.000
Service	-0.006	-0.034	0.012	0.007	0.000	0.000	0.024
Sales and related	0.000	-0.006	-0.000	0.000	0.000	0.000	-0.000
Office and administrative support	-0.041	-0.056	0.041	0.000	-0.020	-0.021	0.034
Farming, fishing, and forestry	0.000	-0.001	-0.001	-0.001	-0.000	-0.000	-0.003
Construction and extraction	0.039	-0.001	-0.021	0.018	0.029	0.029	0.002
Installation, maintenance, and repair	0.020	-0.000	-0.008	0.012	0.016	0.016	0.000
Production	0.005	-0.003	-0.003	0.002	0.004	0.004	-0.011
Transportation and material moving	0.011	-0.004	-0.013	-0.003	0.004	0.004	-0.018
Agriculture	-0.000	-0.000	-0.001	-0.001	-0.001	-0.001	-0.000
Manufacture	0.013	-0.002	-0.002	0.010	0.012	0.012	0.014
Hours worked last week	0.026	-0.044	-0.006	0.020	0.023	0.023	0.032
_cons	0.000	0.164	0.000	0.000	0.000	0.000	0.000
Total	0.035	0.176	0.023	0.058	0.046	0.046	0.089

References

1. Kitagawa, Evelyn M. (1955). "Components of a Difference between Two Rates". Journal of the American Statistical Association. 50 (272): 1168–1194.

2. Oaxaca, R. (1973). "Male-Female Wage Differentials in Urban Labor Markets". International Economic Review. 14 (3): 693–709.

3. Blinder, A. S. (1973). "Wage Discrimination: Reduced Form and Structural Estimates". Journal of Human Resources. 8 (4): 436–455.

4. O' Donnell, Owen A.; et al. (2008). "Explaining Differences between Groups: Oaxaca Decomposition". Analyzing Health Equity Using Household Survey Data: A Guide to Techniques and Their Implementation. World Bank Publications. pp. 147–158.

5. Koenker, Roger (2005). Quantile Regression. Cambridge University Press.

6. Wei, Y.; Pere, A.; Koenker, R.; He, X. (2006). "Quantile Regression Methods for Reference Growth Charts". Statistics in Medicine. 25 (8): 1369–1382.

7. Stock, James H. and Mark W. Watson (2007), Introduction to Econometrics, 2nd Edition, London: Pearson Education.

8. Walker, Ian and Zhu, Yu (2003), "Education, earnings and productivity: recent UK evidence", Labour Market Trends, Vol.111, Issue 3, p.145.

9. Huang, Tung-Chun (1999), "The impact of education and seniority on the male-female wage gap: is more education the answer?", International Journal of Manpower, Vol.20, No. 6, pp. 361-374.

10. Bazen S. (2011), "Econometric methods for labour economics", Oxford: Oxford University Press.

11. Cotton, J. (1988), "On the Decomposition of Wage Differentials", Review of Economics and Statistics, 70(2), p. 236-43.

12. Neumark, D (1988), "Employers' Discriminatory Behaviour and the Estimation of Wage Discrimination", Journal of Human Resources, 23(3), p. 279-95.

13. White (1980). "A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity". Econometrica, 1980, vol. 48, issue 4, 817-38.

14. . Koenker, R. and Hallock, K. (2001), Quantile Regression: An Introduction, Journal of Economic Perspectives, 15, 143–156.

15. Hao L. and Naiman D. Q.(2007), Quantile Regression, Sage Publications, Thousand Oaks. 3