



TI 2004-123/3

Tinbergen Institute Discussion Paper

Decomposing the Gender Wage Gap in the Netherlands with Sample Selection Adjustments

James Albrecht¹

Aico van Vuuren²

Susan Vroman¹

¹ *Georgetown University, Washington, USA,*

² *Vrije Universiteit Amsterdam, the Netherlands.*

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam

Roetersstraat 31

1018 WB Amsterdam

The Netherlands

Tel.: +31(0)20 551 3500

Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam

Burg. Oudlaan 50

3062 PA Amsterdam

The Netherlands

Tel.: +31(0)10 408 8900

Fax: +31(0)10 408 9031

Please send questions and/or remarks of non-scientific nature to driessen@tinbergen.nl.

Most TI discussion papers can be downloaded at <http://www.tinbergen.nl>.

Decomposing the Gender Wage Gap in the Netherlands with Sample Selection Adjustments

James Albrecht¹ Aico van Vuuren² Susan Vroman³

November 2004

¹Georgetown University, Department of Economics, Washington, DC 20057, USA, e-mail:albrecht@georgetown.edu.

²Free University Amsterdam, Department of Economics, De Boelelaan 1105, 1081HV Amsterdam, The Netherlands, email:vuuren@tinbergen.nl

³Georgetown University, Department of Economics, Washington, DC 20057, USA, e-mail: vromans@georgetown.edu.

Keywords: Gender, quantile regression, selection. *JEL-codes:* C24, J22, J31, J71. We thank participants at the ESPE meeting in New York and the Econometric Society meetings in San Diego and at seminars at the University of British Columbia and in Amsterdam and The Hague, in particular Rob Euwals, for useful comments.

Abstract

In this paper, we use quantile regression decomposition methods to analyze the gender gap between men and women who work full time in the Netherlands. Because the fraction of women working full time in the Netherlands is quite low, sample selection is a serious issue. In addition to shedding light on the sources of the gender gap in the Netherlands, we make two methodological contributions. First, we prove that the Machado-Mata quantile regression decomposition procedure yields consistent and asymptotically normal estimates of the quantiles of the counterfactual distribution that it is designed to simulate. Second, we show how the technique can be extended to account for selection.

We find that there is a positive selection of women into full-time work in the Netherlands, i.e., women who get the greatest return to working full time do work full time. We find that about two-thirds of this selection is due to observables such as education and experience with the remainder due to unobservables. Our decompositions show that the majority of the gender log wage gap is due to differences between men and women in returns to labor market characteristics rather than to differences in the characteristics. This is true across the wage distribution, particularly in the top half of the distribution.

1 Introduction

In this paper, we use quantile regression decomposition techniques based on Machado and Mata (2004) to analyze the gender gap across the log wage distributions for men and women who work full time in the Netherlands. In addition to shedding light on the sources of the gender gap in the Netherlands, we make two methodological contributions. First, we prove that the Machado-Mata procedure yields consistent and asymptotically normal estimates of the quantiles of the counterfactual distribution that it is designed to simulate. Second, we show how Machado and Mata (2004) can be extended to account for selection. Since an important application of decomposition techniques is to differences between male and female wage distributions and since selection is often important for women, this is an important practical contribution. This is particularly true in our application since many Dutch women work part time.

Studies of other European countries have found significant differences in the gender gap at different quantiles of the log wage distribution.¹ As in Sweden and Denmark, but unlike, for example, Switzerland, we find a strong glass ceiling effect in the Netherlands. That is, comparing the distributions of log wages of men and women who work full time, the gender gap is greatest at the highest quantiles, although this effect is not as pronounced as in the Scandinavian countries. As a first step to understand this pattern, we use the Machado-Mata method as applied in Albrecht, Björklund, and Vroman (2003) to decompose the difference between the male and female full-time log wage distributions into a component due to differences in the distributions of observable characteristics between genders and a component due to differences in the distribution of rewards to these characteristics between genders. As in the Swedish case, differences in the distributions of characteristics explain relatively little of the difference between the two log wage distributions.

This, however, ignores an important part of the story. As noted above, part-time work is common among women in the Netherlands, so sample

¹See Albrecht, Björklund and Vroman (2003) for Sweden, Blundell, Gosling, Ichimura and Meghir (2002) and Fitzenberger and Wunderlich (2001) for the UK, Bonjour and Gerfin (2001) for Switzerland, Datta Gupta and Smith (2002) for Denmark, Dolado and Llorens (2003) for Spain, and Fitzenberger and Wunderlich (2002) for Germany.

selection is a serious issue.² Accordingly, we extend the Machado-Mata technique to construct a counterfactual distribution of full-time log wages for women, namely, the distribution that would have prevailed had all women worked full-time. The overall selection effect is strongly positive; that is, the women who actually work full time are those with the highest potential wages. Had all women worked full time in the Netherlands, the gender gap would have been considerably larger. Our technique also allows us to decompose the selection effect into a portion due to observables (about two thirds in our application) and a portion due to unobservables. Finally, we use the Machado-Mata technique to construct a counterfactual to the counterfactual, namely, the distribution of log wages that women would earn if all women worked full time and had the male distribution of labor market characteristics.

The rest of the paper is organized as follows. In the next section, we present the data used in our paper. In Section 3, we estimate a series of quantile regressions to find the marginal contributions of observable characteristics to log wages for both genders. We apply Buchinsky (1998a) to correct for the effect of selection into full-time work in the women's quantile regressions. In Section 4, we give a general description of the Machado-Mata method of quantile regression decomposition, we show that the procedure leads to consistent and asymptotically normal estimates of the quantiles of the counterfactual distribution, and we extend the procedure to allow for sample selection. Section 5 presents the results of applying the Machado-Mata decomposition technique without adjusting for selection to the Dutch data. We find that the gender log wage gap between men and women who

²In terms of gender and the labor market, the Netherlands has changed dramatically over the past 20 years. In 1980, the gender gap in employment in the Netherlands was about 40%, similar to that of Spain, Greece, Italy, and Ireland. By 2000, this gap had fallen to a level (about 18%) more like that of most other Western European countries, albeit still above the levels observed in the U.S., the U.K., and the Scandinavian countries. Most of this change was due to an expansion of part-time work for women. In 2000, the Netherlands had the highest rate of part-time work (defined as fewer than 30 hours per week) among women in all the OECD countries. About 56% of employment among women aged 25-54 in the Netherlands was part time. This is considerably higher than the corresponding rates for Germany and Belgium (35%), France (23%), the U.K. (38%) and the U.S. (14%). Only Switzerland (47%) is remotely comparable. This difference relative to other countries was mainly caused by high part-time employment rates for mothers. For example, 83% of working mothers aged 25-54 with 2 or more children worked part time in the Netherlands. The comparable figure for the U.S. is 24%. The part-time employment rate of men in the same age group in the Netherlands (6%) is not substantially different from the corresponding rates in other countries, especially when we look at fathers. These figures are taken from the July 2002 OECD Economic Outlook.

work full time is primarily due to differences in the returns to observed characteristics. In Section 6, we discuss the results after adjusting for selection. We find that there is a positive selection into full-time work for women in our data set. That is, the women who would get the greatest return from working full time are the ones who in fact did work full time. We find that after adjusting for selection and for the difference in observed characteristics between men and women working full time, there is a significant positive gender log wage gap across the entire distribution and that it is larger at the top of the distribution, indicating a glass ceiling effect. Section 7 concludes.

2 Data

We use data from the OSA (Dutch Institute for Labor Studies) Labor Supply Panel. This panel dates from 1985 and is based on biannual (1986, 1988, etc.) interviews of a representative sample of about 2,000 households. All members of these households between the ages of 16 and 65 who were not in (daytime) school and who could potentially work were interviewed. The survey focuses on respondents' labor market experiences. Van den Berg and Ridder (1998) give a detailed description.³ We use data from the 1992 survey for all of our analysis. This year is particularly rich in terms of variables that can be used to explain participation in full-time work.⁴ The sample size for 1992 is 4536. In order to focus on those who are most likely to be working full time, i.e., on those who are least likely to still be in school or already retired, we restrict our sample to individuals between 25 and 55 years of age inclusive. We deleted 1238 individuals who fell outside this range. In addition, there are relatively many individuals with missing data for years of work experience among those who were not working at the survey date. We set work experience equal to 0 for those who did not work in the previous 2 years. All others lacking experience data were dropped from our dataset. This resulted in an additional 113 deletions. This leaves a sample of 1617 females and 1568 males. Of course, not all of these individuals worked full time. In terms of full-time wage data, we deleted observations lacking a reported wage. These include 1463 individuals who did not work full-time plus 94 nonresponses. We also deleted two individuals reporting wages below

³See also <http://www.uvt.nl/osa>.

⁴We have repeated our analysis using data from the 1998 survey and have found qualitatively similar results. We prefer to use the 1992 data because they are better suited for dealing with the selection issue.

2 euros or exceeding 20 euros per hour.⁵ Finally, we deleted two individuals who reported contractual hours per week above 60 hours, as this is likely to be due to measurement error. This leaves 391 women and 1233 men who reported usable full-time wages. We use the data on all 1617 women to carry out the selection analysis.

In Table 1, we give some descriptive statistics for the key variables for all women, women working part time, women working full time and women working full time with reported wages. Similar descriptive statistics for men are presented in Table 2 (likewise for all men, men working part time, men working full time and men working full time with reported wages). Most men between the ages of 25 and 55 work, and among those who are working, almost all work full time. Among women, the situation is quite different. In terms of the variables that we can use to explain variation in wages, there are also some important differences between the genders. Among those working full time and reporting wages, men are on average almost 3 years older than women, and years of work experience are much higher for men than for women. We measure education using four categories – (i) up to elementary education, (ii) lower secondary education, (iii) upper secondary education, and (iv) bachelors/masters degree.⁶ Overall, men have achieved higher levels of education than women have, but if we compare men and women who work full time, this pattern is reversed. In general, women are slightly more likely to be married than men are. This is simply because women tend to marry men who are older than they are. However, women who work part time are much more likely to be married than are women in general, and women who work full time are much less likely to be married. Finally, men and women are approximately equally likely to live in cities, but women who work full time are much more likely to do so than are men who work full time.

The data include variables concerning attitudes about working and about the relationship between family and work.⁷ We use reactions to the statement, “Parents should be willing to reduce working hours for childcare.” Respondents could reply that they completely disagreed, disagreed, were

⁵The minimum wage in the Netherlands was the equivalent of 983 Euros per month for a full-time worker over 22 years of age. This implies a gross wage of about 5.5 Euros per hour. The reported wages are net wages, but this cannot explain wages below 2 Euros per hour.

⁶Our education variable is based on the first digit of the ISCED codes. Lower secondary education is level 3, upper secondary education is level 4 and our last category includes all education levels exceeding level 4.

⁷These attitude variables are the reason that we use the 1992 data for our analysis.

indifferent, agreed or completely agreed. We count those individuals who either agreed or completely agreed as agreeing with the statement. Women who work full time are less likely to agree with this statement than are other women. We also have data on whether there are children in the household. We report whether there are children living at home and whether the youngest child is (i) below 5 years of age, (ii) age 5-11, or (iii) age 12-17. Relatively few women who are working full time have children living at home (32%); relatively many women working part time do. This phenomenon is even more apparent when we look at the youngest age group. Only 8 percent of the full-time working women who report wages have children below the age of 5 living at home, while among all women in the sample, this figure is 18 percent. The same holds true for children 5 - 11 and 12 - 17, but the percentage differences are smaller. Finally, we have a variable that summarizes religious attitudes. Overall, women seem to be a bit more “religious” than men, but this difference disappears when we compare full-time workers among men and women.

In sum, full-time working women are more educated, less likely to be married, less likely to have (young) children at home, and have different attitudes towards working than women who do not work full time.

Table 3 gives wages for women who work part time, for women who work full time, and for men. These are net hourly wages excluding extra payments for overtime, shift work, bonuses, tips etc. As expected, men’s wages are on average higher than women’s wages. The average wage among women working full time is slightly higher than the corresponding average among women who work part time.

Figure 1 plots the estimated kernel densities of men’s and women’s wages. In this figure, we use all wages, including those of part-time workers. The gender gap, i.e., the difference in log wages between males and females at each quantile of their respective distributions, is plotted in Figure 2. This figure shows the gap for all working men versus all working women, i.e., both the male and female distributions include part-time workers. We show the 95% confidence bands in Figure 2. The variance used in the calculation of the confidence interval for the q^{th} quantile is estimated using $\frac{q(1-q)}{\hat{f}(\hat{\theta}(q))^2}$, where $\hat{\theta}(q)$ is our estimate of the q^{th} quantile and the density $f(\cdot)$ is estimated using a kernel density method. Figure 3 shows the corresponding gap for men versus full-time women. Figures 2 and 3 show different patterns for the gender gap across the distribution. In Figure 2, the gender gap is relatively flat; but in Figure 3, the gender gap is larger at higher quantiles. That is,

the gender gap between men and women who work full time exhibits a glass ceiling effect; full-time working women do relatively well at the bottom of the wage distribution but not as well at the top. This pattern is presented in a different form in Figure 4, which plots the gap between the log wages of women who work full time and those who work part time. We focus on the pattern in Figure 3. That is, we focus on men versus women who work full time.

3 Quantile Regressions

In this section, we present log wage quantile regressions for women and for men who work full time. We regress log wage on the basic human capital variables, years of work experience and education (with less than secondary education as the left out category), as well as marital status and whether the individual lived in a city.

In comparing the quantile regressions for men and women, we need to account for the selection of women into full-time work. Considerable attention has been devoted to sample selection in the literature.⁸ Much of this work extends Heckman's (1979) classic model to allow for non-normality. For our purpose, since we want to correct for selection at various quantiles, we cannot restrict ourselves to a symmetric (e.g., normal) distribution to explain the gap over the wage distribution.

We use the method introduced by Buchinsky (1998a) to correct for selection in quantile regression. We make this adjustment only for women. There are other techniques we could have used to correct for sample selection. In particular, Blundell, Gosling, Ichimura and Meghir (2002) use the approach suggested in Manski (1994) to bound the possible impact of selection. The idea is simple: even if we know nothing about the productivity of non-workers, bounds can be obtained by assuming either that all non-workers are more productive than workers (resulting in the upper bound) or that all non-workers are less productive than workers (resulting in the lower bound). Although the method of Blundell et. al. requires fewer assumptions than the ones we adopt in our paper, it has the disadvantage of being less precise. Our method is based on the argument that we have good instruments for predicting the incidence of full-time work among women – indeed, this is the reason that we use data from 1992.

We begin with our estimates of the determinants of full-time work among

⁸See Gallant and Nychka 1987, Newey 1988, Buchinsky 1998a, Blundell, Gosling, Ichimura and Meghir 2002, Das, Newey and Vella 2003. Vella 1998 gives a survey.

women.⁹ These are given in Table 4. The first column gives probit results, while the second column contains the results for an estimation using a single-index technique.¹⁰ Table 4 indicates that older and married women are less likely to work full time. Highly educated women are more likely to work full time, as are women with a relatively high level of work experience. Having a young child at home has a strong negative impact on the propensity to work full time and is more important the lower the age of the youngest child. Finally, women with children who respond that it is better for women to stay at home when they have children are less likely to work full time. Whether or not a woman lives in a city seems to have little impact on the incidence of full-time work nor does the religion variable.

Table 5 presents the uncorrected quantile regressions results for full-time working women, while Table 6 presents the corresponding results with the Buchinsky correction. The quantile regression results for men are in Table 7. In the uncorrected estimates for women, the basic human capital variables are the most important and have the anticipated effects. That is, work experience has a positive effect, which rises across the quantiles, and education has a strong positive effect. Marital status and the dummy for living in a city have insignificant effects at almost all the quantiles. Correcting for selection has several effects. First, the estimated constant terms decrease once we adjust for selection, especially towards the top of the distribution. Second, the rewards to education increase in the upper part of the distribution after correcting for selection. Finally, Table 7 indicates that men also receive a positive return to education. As is the case for women, the coefficients on years of experience are strongly positive and increase across the distribution. Finally, the data show a strong marriage premium for men in the Netherlands, which is relatively constant across the distribution.

4 Machado-Mata Decompositions: Theory

In this section, we explain the Machado and Mata (2004) decomposition method. We give a proof of the consistency and asymptotic normality of

⁹We model full-time work as a matter of choice for women. We base this on our comparison of histograms for desired versus contractual working hours for both men and women (not shown). In fact, the match between desired and contractual hours is much closer for women than for men.

¹⁰Details of the estimation technique are given in Appendix A. Note that the constant and the coefficient of one of the continuous variables are not identified in a single-index model. Hence, we normalize by setting the constant and the coefficient on age equal to their values in the probit model so that the results are comparable.

the estimated quantiles of the counterfactual distribution generated by this method. We conclude this section with an extension of the method to account for selection.

We start with a general description of the Machado-Mata (M-M) method. Their procedure can be viewed as a generalization of the Oaxaca-Blinder decomposition (Oaxaca 1973).¹¹ Consider two groups, A and B , with characteristics given by the stochastic vectors X_A for group A and X_B for group B . Realizations of these stochastic vectors will be denoted by x_A and x_B . Assume that X_A and X_B both have dimension k and have distribution functions G_{X_A} and G_{X_B} , respectively. The endogenous variable is Y_A for group A and Y_B for group B with unconditional distribution functions F_{Y_A} and F_{Y_B} , respectively. Let the sizes of the two samples be N_A and N_B , and suppose that the outcomes as well as the characteristics are observed for both groups.

The M-M assumption is that the regression quantiles are $\beta^A(u)$ for group A and $\beta^B(u)$ for group B for each $u \in [0, 1]$; that is, $\text{Quant}_u(Y_A|X_A = x_A) = x_A\beta^A(u)$ for each $u \in [0, 1]$ and likewise for Y_B . The conditional distribution of $Y_A|X_A = x_A$ is completely characterized by the collection of regression quantiles $\{\beta^A(u); u \in [0, 1]\}$, and likewise for the conditional distribution of $Y_B|X_B = x_B$.

Consider a counterfactual random variable Y_{AB} with the property that its quantiles conditional on x_A are given by

$$\text{Quant}_u(Y_{AB}|X_A = x_A) = x_A\beta^B(u) \quad u \in [0, 1].$$

The M-M method generates a sample from the unconditional distribution of Y_{AB} as follows:

1. Sample u from a standard uniform distribution.
2. Compute $\hat{\beta}^B(u)$, i.e., estimate the u^{th} regression quantile of y^B on x_B .
3. Sample x_A from the empirical distribution \hat{G}_{X_A} .
4. Compute $\hat{y}_{AB} = x_A\hat{\beta}^B(u)$.
5. Repeat steps 1 to 4 M times.

Of course, the sample generated in this way is not a *real* sample from the stochastic variable Y_{AB} since it is based on estimates rather than on

¹¹There are other techniques that have the same objective, e.g., Dinardo, Fortin and Lemieux (1996).

the *real* parameters of the distribution. This implies that estimators (like sample means, medians, etc.) based on the sample generated by the method cannot be interpreted as estimates based on the population Y_{AB} . However, as N_A and N_B become large, this problem should become unimportant. We now make this more precise by comparing the sample quantiles computed from the sample generated using the M-M method to the corresponding population quantiles of Y_{AB} .

Let $\theta_0(q)$ be the q^{th} quantile of the unconditional distribution of Y_{AB} , i.e.,

$$\theta_0(q) = F_{Y_{AB}}^{-1}(q)$$

where $F_{Y_{AB}}(y) = \int F_{Y_{AB}}(y|X_A = x_A)dG_{X_A}(x_A)$. Let $\hat{\theta}(q)$ be the corresponding estimate obtained using the M-M technique. We make the following assumptions about the distributions of Y_B , X_A , and X_B :

Assumption A. Let

A1. F_{Y_B} have a compact support $[\inf \{\text{supp}(Y_B)\}, \sup \{\text{supp}(Y_B)\}]$ on \mathbb{R} and be continuously differentiable on its support, with positive density f_{Y_B}

A2. G_{X_A} and G_{X_B} have a compact support on \mathbb{R}^K

A3. $N_B^{-1} X_B^T X_B$ converge in probability to a positive definite matrix

A4. $\forall u \in [0, 1]$ and $x_A \in \text{supp}(X_A) : \frac{dx_A \beta^B(u)}{du} > 0$

A5. $X_A \perp X_B$

These assumptions ensure that the coefficient estimates that result from quantile regressions of Y_B on X_B are consistent and asymptotically normal (see, for example, Van der Vaart 1998, page 307 and Koenker and Bassett 1978). In addition, the joint compactness assumption (together with condition 4) guarantees that the support of $F_{Y_{AB}}$ is also a convex and compact subset of \mathbb{R} (not proven here).¹² This is necessary to prove consistency and asymptotic normality of $\hat{\theta}(q)$. Condition 4 states that the quantile regression lines cannot cross on the support of X_A . Condition 5 is made for convenience and is only necessary for the computation of the covariance. We make this assumption in order to satisfy the condition that the moments of the different subsamples of populations A and B are uncorrelated (for more details, see section 6.2 of Newey and McFadden, 1994).

We prove the following in Appendix B:

¹²The conditions are somewhat stronger than strictly necessary. Identification may still be satisfied even if the support of Y_{AB} is not a convex compact subset of \mathbb{R} . Details about this can be found in for example Van der Vaart, Lemma 21.4.

Theorem 1 *Let Assumptions A1-A4 be satisfied, and let $M, N_A, N_B \rightarrow \infty$ with $M/N_A \rightarrow I_A < \infty$, $M/N_B \rightarrow I_B < \infty$. Then $\hat{\theta}(q) \xrightarrow{P} \theta_0(q)$.*

The most important step in the proof is relatively simple and is based on the inverse transformation method (see, for example, Law and Kelton 1991). However, this method assumes that the underlying population distributions are known. In the M-M approach, these distributions are estimated. Hence, most of our proof is devoted to showing that when the sample sizes of both datasets are large, the impact of the estimation method is negligible. We note that if groups A and B are the same, then the M-M approach simply replicates the distribution of A .

In addition to consistency, it is important to prove asymptotic normality. This was not an issue in Albrecht, Björklund and Vroman (2003) since their sample sizes were extremely large. Our sample sizes are not so large. Therefore, in Appendix B, we prove the following:

Theorem 2 *Let Assumptions A1-A5 be satisfied, and let $M, N_A, N_B \rightarrow \infty$ with $M/N_A \rightarrow I_A < \infty$, $M/N_B \rightarrow I_B < \infty$. Then*

$$\sqrt{M}(\hat{\theta}(q) - \theta_0(q)) \rightsquigarrow N(0, \Omega)$$

$$\text{with } \Omega = \frac{q(1-q)I_A + I_B \mathbf{E}_{X_A, U, V} \left\{ f_{Y_{AB}}^2(\theta(q)) X_A \Lambda(\beta^B(U), \beta^B(V)) X_A^T \right\}}{f_{Y_{AB}}^2(\theta(q))}$$

and $\Lambda(\beta^B(u), \beta^B(v))$, $u, v \in [0, 1]$ given by

$$\Lambda(\beta^B(u), \beta^B(v)) = \begin{cases} \mathbf{E}_{X_B} \left(f_{Y_B}(X_B \beta^B(u)) X_B X_B^T \right)^{-1} u(1-v) \mathbf{E}_{X_B} (X_B X_B^T) \\ \quad \times \mathbf{E}_{X_B} \left(f_{Y_B}(X_B \beta^B(v)) X_B X_B^T \right)^{-1} & \text{if } u \leq v \\ \mathbf{E}_{X_B} \left(f_{Y_B}(X_B \beta^B(v)) X_B X_B^T \right)^{-1} v(1-u) \mathbf{E}_{X_B} (X_B X_B^T) \\ \quad \times \mathbf{E}_{X_B} \left(f_{Y_B}(X_B \beta^B(u)) X_B X_B^T \right)^{-1} & \text{otherwise} \end{cases}$$

where U and V are independent standard uniform random variables.

We have two comments about the results derived in Theorem 2. First, the expression for $\Lambda(\beta^B(u), \beta^B(v))$, which is derived in a separate lemma in Appendix B (see also Koenker and Bassett 1978, Theorem 4.2), is an extension of the usual expression for the covariance matrix for regression quantiles. This covariance matrix can be derived when $u = v$ is substituted into the expression above. When u and v are different, the expression gives

the covariance between regression quantiles at different points in the distribution. It can be easily checked that the expression is largest when u and v are close to each other and its maximum occurs when $u = v$. This makes sense since if u and v are close to each other, we are essentially comparing nearby quantiles, so the regression quantiles are likely to be close to each other as well. When u and v are far away from each other, the regression quantile $\beta^B(u)$ gives little information about $\beta^B(v)$. In that case the covariance between the two regression quantiles is low. The second comment about Theorem 2 is related to the two terms in the numerator of Ω . The first term is the standard deviation of the estimated quantile at q based on a sample of size N_A . Indeed if $I_B \rightarrow 0$, the variance converges to this term since in that case the only randomness comes from the X_A 's. The second term, which is quite complicated, takes into account the estimation of the $\beta^B(u)$'s across the distribution. The complexity of this term is mainly due to the fact that even though we sample independent draws from a uniform distribution, the resulting quantile regression estimates are not independent of each other.

We use a kernel density method to estimate the covariance matrix $\Lambda(\cdot)$ (see Buchinsky, 1998b). In addition, we use the following estimates to complete the calculation of Ω :

$$\widehat{I}_A = \frac{M}{N_A}$$

$$\widehat{I}_B = \frac{M}{N_B}$$

$$\begin{aligned} & \widehat{E}_{X_A, U, V} \{ f_{Y_{AB}}^2(\theta(q)) X_A \Lambda(\beta^B(U), \beta^B(V)) X_A^T \} \\ &= \frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M \widehat{f}_{Y_{AB}}(\widehat{\theta}(q)) x_{A_i} \widehat{\Lambda}(\beta^B(u_i), \beta^B(u_j)) x_{A_j}^T \end{aligned}$$

$$\begin{aligned} \widehat{\Lambda}(\beta^B(u_i), \beta^B(u_j)) &= \left[\frac{1}{M h_M} \sum_{l=1}^M K \left(\frac{y_{B_l} - x_{B_l} \widehat{\beta}^B(u_i)}{h_M} \right) x_{B_l} x_{B_l}^T \right]^{-1} \times \\ & \left[u_i(1 - u_j) \frac{1}{M} \sum_{l=1}^M x_{B_l} x_{B_l}^T \right] \left[\frac{1}{M h_M} \sum_{l=1}^M K \left(\frac{y_{B_l} - x_{B_l} \widehat{\beta}^B(u_j)}{h_M} \right) x_{B_l} x_{B_l}^T \right]^{-1} \\ \widehat{f}_{Y_{AB}}(y) &= \frac{1}{M h_M} \sum_{l=1}^M K \left(\frac{x_{A_l} \widehat{\beta}^B(u_l) - y}{h_M} \right) \end{aligned}$$

where h_M is the bandwidth and K is the kernel function. We use the standard normal density as a kernel function and regression cross validation techniques to find the right levels of h_M (see, for example, Silverman 1988).

In our analysis, we are interested in log wage gaps, i.e., differences between quantiles of two log wage distributions. The quantiles of the distributions of Y_A and Y_{AB} are correlated. The covariance between a particular quantile of these two distributions can be shown to be equal to $q(1 - q) / (f_{Y_A}(\theta(q))f_{Y_{AB}}(\theta(q)))$. This results in a reduction in the variance of the wage gap. The reduction does not arise when we look at the difference in quantiles of the distributions Y_B and Y_{AB} .¹³

Next we extend the M-M procedure to allow for selection. In our application, we consider the issue of selection of women into full-time work. In the above notation, groups A and B could stand for any arbitrary groups so, for example, when we do the decomposition without selection correction, we consider A to be women and B to be men. In adjusting for selection, we let A denote all women and B denote women who are actually working full time.

Let Y_A be the counterfactual random variable representing the wage that a randomly selected woman would earn were she to work full time. The quantiles of Y_A conditional on x_A are given by

$$\text{Quant}_u(Y_A|X_A = x_A) = x_A\beta(u) \quad u \in [0, 1],$$

where $\beta(u)$ is the true value of the coefficient correcting for selection. We estimate $\beta(u)$ using the following model based on Buchinsky (1998a):

$$\text{Quant}_u(Y_B|Z_B = z_B) = x_B\beta(u) + h(z_B\gamma) \quad u \in [0, 1].$$

Here h plays the role that the Mill's ratio plays in the usual Heckman procedure, but it is more general so as not to assume normality. The vector z_B is the realization of the random vector Z_B that corresponds to x_B . Here Z is the set of characteristics that influence the probability that a woman will work full time (but not necessarily the outcome variable Y_A directly). For

¹³The intuition is as follows. The conditional quantiles of Y_A given x_A are given by $Y_A = x_A\beta^A(u)$ for $u \in [0, 1]$, while those of Y_{AB} given x_A are given by $Y_{AB} = x_A\beta^B(u)$ for $u \in [0, 1]$. We can use the M-M method to recover the unconditional distribution of both Y_A and Y_{AB} . The sample from X_A is used in constructing both distributions. This means that the correlation between the two generated stochastic variables is essentially that between $X_A\beta^A(u)$ and $X_A\beta^B(u)$. The correlation between Y_B and Y_{AB} does not have this feature because in that case, we generate Y_B and Y_{AB} by sampling from X_B and X_A , respectively. Note that a similar correlation arises in the Oaxaca decomposition method. Details are available from the authors on request.

identification, Z should contain at least one element that is not included in X . Hence the set of variables in X must be a strict subset of those in Z . This model then gives the wage that a woman with characteristics x_A who is working full time would earn after adjusting for selection.

The M-M algorithm is changed as follows:

1. Estimate γ using a single-index method
2. Sample u from a standard uniform distribution.
3. Compute $\hat{\beta}(u)$ and $\hat{h}(\cdot)$ using the Buchinsky technique.
4. Sample x_A from the empirical distribution \hat{G}_{X_A} .
5. Compute $\hat{y}_A = x_A \hat{\beta}(u)$.
6. Repeat steps 2 - 5 M times.

Since population B in our case is a subset of population A , Assumption A5 is not likely to be satisfied. Therefore we make an alternative assumption for this problem.

Assumption A5': Let $\text{Quant}_u(Y_B - x_B \beta(u) - h(z_B \gamma) | Z_B = z_B) = 0$ $u \in [0, 1]$.

Assumption A5' is necessary to estimate $\beta(u)$ consistently. Thus, for consistency, as in Theorem 1, this assumption is necessary when dealing with selection (unlike Assumption A5 in the original problem). Assumption A5' was also made by Buchinsky (1998a). Finally, we can extend Theorem 2 by replacing $\Lambda(\cdot)$ by the covariance matrix that is computed using the technique of Buchinsky (1998a).

Following the above procedure generates the distribution of wages for all women if all women worked full time. The difference between this distribution and the actual wage distribution for full-time women gives the effect of selection on the wage distribution. We can decompose this effect into a part due to observables and a part due to unobservables. To do this, we construct another hypothetical distribution by modifying step 4 and sampling from the data on women who work full time. We then obtain the distribution that would result if women who do not work full time had the same distribution of observed characteristics as those who do work full time. The difference between these two distributions is the portion of the selection effect due to observed characteristics. The remainder of the sample selection effect is the part due to unobserved characteristics. This latter portion can be obtained by comparing the distribution obtained by sampling from full-time working women with the original distribution of observed women's wages.

5 Machado-Mata Decompositions: Results without Selection Correction

In this section, we present the results of the decomposition of the gender gap done without correcting for selection. In order to see the results over the whole distribution, it is best to view them graphically. For reference, recall that Figure 3 presented the gender log wage gap for men versus women working full time based on the raw data. This was derived by simply subtracting the log wage of the full-time women at each quantile of their log wage distribution from the log wage of the men at the corresponding quantiles of the male log wage distribution. As noted above, without conditioning on covariates or adjusting for selection, there is a significant glass ceiling effect in the data, i.e., the gender gap is significantly higher at the higher quantiles of the distribution. We use the M-M procedure to analyze what proportion of the gap is due to differences in labor market characteristics between the genders and what proportion is due to differences in the returns to these characteristics between the genders.

Figure 5 plots the wage gap taking account of the differences in distributions of observed characteristics between men and women. That is, we construct a counterfactual distribution using the M-M method that gives the log wage distribution that women would earn if they had the same characteristics as men, but were still paid for those characteristics like women. The characteristics that we include are those given in the quantile regression tables, namely, experience, education, marital status, and whether the individual lived in a city. As can be seen in Figure 5, there remains a significant positive gender gap across the whole distribution after accounting for differences in observed characteristics. Comparing Figure 5 to Figure 3, we can see that at most one third of the gender gap for full-time workers in the Netherlands is due to differences in characteristics between men and women who work full time. Accounting for the differences in observed characteristics reduces the gap most at the higher quantiles of the distribution.

To confirm that the largest portion of the gender gap is due to differences in returns to these characteristics, we also construct a counterfactual distribution of log wages that represents the wages that full-time women would earn if they retained their observed characteristics but were paid for them like men. The gap between the log wages of men and the log wages given in this counterfactual distribution is presented in Figure 6. The gap is smaller over the whole distribution. It appears from Figures 5 and 6 that on average about three quarters of the gender log wage gap between men

and full-time women is explained by differences in the returns to observed characteristics across the genders and about one quarter is due to differences in these characteristics.

This analysis, of course, does not account for the selection into full-time work by women. The results accounting for this selection are given in the next section.

6 Machado-Mata Decompositions: Results with Selection Correction

Next, we investigate the effect of sample selection on the women's log wage distribution and on the counterfactual distribution. To see the direct effect of selection, we first look at the gap between the log wage distribution for women working full time and the log wage distribution that we would have observed had all women worked full time. Figure 7 shows that the overall selection effect is positive. That is, women who actually work full time have higher earnings potential in full-time work than do women in general. This difference is significantly positive everywhere but at the extremes of the distribution.

In Figure 8, we show the gender log wage gap after adjusting for selection, i.e., we plot the difference between the male log wage distribution and the log wage distribution for women if all women worked full time. This can be compared with the gender log wage gap in the raw data shown in Figure 3. This gap after adjustment for selection is significantly greater at almost all quantiles. That is, after accounting for the fact that women with the highest rewards to full-time work are the ones engaging in full-time work, the gender gap is larger.

The results taking the selection effects into account are based on sampling characteristics from the data set for all women. In Step 4 of our modification of the M-M procedure to deal with selection, if we had only sampled women working full time, we would have ignored the impact of the difference between the observed characteristics of women who do and do not work full time. Although this would have resulted in the wrong distribution and hence the wrong wage gaps, it is a useful exercise in order to explain the sample selection effect. This distribution can be interpreted as the distribution of wages that would have resulted if the women who do not work full time had the same characteristics as those who do work full time. Hence, comparing the distribution with the proper sample selection correction with this distribution allows us to view the portion of the selec-

tion effect due to observed characteristics. The remainder of the sample selection effect is the part due to unobserved characteristics. The portion due to the unobserved characteristics can be obtained by comparing the distribution obtained from sampling from full-time working women with the original distribution of observed women’s wages. Figures 9 and 10 present the observed and unobserved parts of the selection effect. The observed part is positive and significantly different from zero over the bottom two thirds of the distribution. This positive effect presumably results from the fact that women who work full time are in general better educated and have more work experience than women who are not working full time. The unobserved part of the sample selection effect is also positive and significant from approximately the 30th to the 90th percentiles of the distribution. On average, observables account for about two thirds of the sample selection bias.

Finally, we can address the issue of what proportion of the wage gap shown in Figure 8, the gap between men and women if all women worked full time, is due to differences in characteristics between men and women and what proportion is due to differences in returns to these characteristics between the genders. Figure 11 shows the difference between the male log wage distribution and the distribution of log wages that women would earn if all women worked full time and had the characteristics of men, but received the returns of women (adjusted for sample selection). As can be seen from the figure, accounting for the difference in characteristics reduces the gender gap by about one third on average with the greatest effect in the middle of the distribution.

Concluding, we find a positive and significant selection effect for full-time work in the Netherlands. A large part of the selection effect is due to observables. Compared with the results of the previous section, accounting for these observables has a larger impact on the gender log wage gap. This result reflects the fact that, in contrast to what we observe in the population of women working full time, education and years of work experience in the population of all women are lower than the corresponding levels for males (see Table 1 for details).

7 Conclusions

In this paper, we have made two contributions. First, we have contributed to the econometrics underlying the Machado-Mata quantile regression decomposition technique. This method is an intuitively appealing generalization of

the Oaxaca-Blinder approach, which decomposes differences between groups in average outcomes into differences between average characteristics and differences in rewards to those characteristics. The M-M method is designed to simulate counterfactual distributions; for example, what would the distribution of full-time log wages for women have been if working women had the same distribution of labor market characteristics as men do? We have shown that the M-M method leads to consistent estimators for the quantiles of the counterfactual distribution that it is designed to simulate, and we have developed asymptotic standard errors for these estimators. We have also extended the M-M technique to account for selection. The idea is to use the technique to simulate another counterfactual distribution; for example, what would the distribution of full-time log wages for women have been if all women worked full time? Our method for accounting for selection also allows us to decompose the selection effect into a component due to observables and one due to unobservables.

Our second contribution has been to apply our extension of the M-M technique to help understand the gender gap in the Netherlands. Specifically, we examined the difference between the male and female distributions of log wages among full-time workers. Taking the population of women working full time as given, we found an average log wage gap on the order of 20%, and we documented that this gap increases as we move up the distribution. That is, a glass ceiling effect is present in the Netherlands. Approximately one quarter of this gender gap can be attributed to differences between men and women in labor market characteristics; approximately three quarters can be ascribed to gender differences in rewards to those characteristics. However, relatively many women work part time in the Netherlands, so the sample of women working full time is a selected one. We addressed several questions in connection with this selection process. First, what would the difference between the male and female distributions of log wages among full-time workers have looked like if all women had worked full time? There is an important practical reason to be interested in the answer to this question. If we want to compare the gender gap among full-time workers in the Netherlands to, say, the corresponding gap in Sweden, we need to adjust for the fact that relatively fewer women work full time in the Netherlands. Correcting for selection makes it possible to compare “apples with apples.” In fact, correcting for selection turns out to be very important. Were all Dutch women to work full time, the average log wage gap between the genders would be much higher; that is, there is a strong positive selection into full-time work among women in the Netherlands. Second, how can we explain the selection we observe? Our technique allows us to ascribe about two thirds of

the selection effect to observables – women who are working full time have higher education and more work experience than other women do – and about one third to unobservables. Finally, if we compare the distribution of log wages that we would observe if women worked full time to the same extent as men do with the corresponding distribution across men, to what extent would the difference between the counterfactual female distribution and the actual male distribution be ascribed to differences in characteristics? To what extent would it be ascribed to differences in rewards? We found that approximately one third of this counterfactual difference would be ascribed to differences in the distributions of characteristics between men and women. That is, once we correct for selection, differences in labor market characteristics between men and women play a larger role in explaining the gender gap in the Netherlands. Nonetheless, most of the gender gap across the distribution continues to be accounted for by differences in how men and women are rewarded.

References

- Albrecht, J., A. Björklund and S. Vroman (2003), Is there a glass ceiling in Sweden?, *Journal of Labor Economics*, 21, 145–177.
- Andrews, D.W. and M.A. Schafgans (1998), Semiparametric estimation of the intercept model, *Review of Economic Studies*, 65, 497–517.
- Blundell, R., A. Gosling, H. Ichimura and C. Meghir (2002), Changes in the distribution of male and female wages accounting for employment composition, working paper, London.
- Bonjour, D. and M. Gerfin (2001), The unequal distribution of unequal pay - an empirical analysis of the gender wage gap in Switzerland, *Empirical Economics*, 26, 407–427.
- Buchinsky, M. (1998a), The dynamics of changes in the female wage distribution in the USA: a quantile regression approach, *Journal of Applied Econometrics*, 13, 1–30.
- Buchinsky, M. (1998b), Recent advances in quantile regression models: a practical guideline for empirical research, *Journal of Human Resources*, 33, 88–126.
- Das, M., W.K. Newey and F. Vella (2003), Nonparametric estimation of sample selection models, *Review of Economic Studies*, 70, 33–58.
- Datta Gupta, N. and N. Smith (2002), Swimming upstream, floating downstream: trends in the U.S. and Danish gender wage gaps, working paper, Aarhus.
- Dinardo, J., N.M. Fortin and T. Lemieux (1996), Labor market institutions and the distribution of wages 1973–1992: a semiparametric approach, *Econometrica*, 64, 1001–1024.
- Dolado, J.J. and V. Llorens (2003), Gender wage gaps by education in Spain: glass floors versus glass ceilings, Universidad Carlos III, Madrid.
- Fitzenberger, B. and G. Wunderlich (2001), The changing gender gap across the wage distribution in the UK, working paper, Mannheim.
- Fitzenberger, B. and G. Wunderlich (2002), Gender wage differences in West Germany: a cohort analysis, *German Economic Review*, 3, 379–414.

- Gallant, A.R. and D.W. Nychka (1987), Semi-Parametric Maximum Likelihood Estimation, *Econometrica*, 55, 363-390.
- Heckman, J. (1979), Sample Selection Bias as a Specification Error, *Econometrica*, 47, 153-161.
- Horowitz, J. (2001), The Bootstrap, in: J.J. Heckman and E.E. Leamer, *Handbook of Econometrics*, 5, Elsevier B.V.
- Ichimura, H. (1993), Semiparametric least squares (SLS) and weighted SLS estimation of single index models, *Journal of Econometrics*, 58, 71-120.
- Koenker, R. and G. Bassett (1978), Regression Quantiles, *Econometrica*, 46, 33-50.
- Law, A.M. and W.D. Kelton (1991), *Simulation modelling and analysis*, McGraw-Hill, New York.
- Machado, J.A.F. and J. Mata (2004), Counterfactual decomposition of changes in wage distributions using quantile regression, *Journal of Applied Econometrics*, forthcoming.
- Manski, C. (1994), The selection problem, in C. Sims, *Advances in Econometrics, Sixth World Congress*, 1, Cambridge University Press.
- Newey, W.K. (1988), Two step series estimation of sample selection models, working paper, Cambridge.
- Newey, W.K. and D. McFadden (1994), Large Sample Estimation and Hypothesis Testing, in R.H. Engle and D. McFadden (eds.), *Handbook of Econometrics, volume 4*, North-Holland, Amsterdam.
- Oaxaca, R. (1973), Male-Female Differentials in Urban Labor Markets, *International Economic Review*, 14, 693-709.
- OECD (2002), *Employment Outlook 2002*, OECD, Paris.
- Silverman, B.W. (1986), *Density estimation*, Chapman and Hall, London.
- Van den Berg, G.J. and G. Ridder (1998), An empirical equilibrium search model of the labor market, *Econometrica*, 66, 153-161.
- Van der Vaart, A. (1998), *Asymptotic statistics*, Cambridge University Press, Cambridge.

Vella, F. (1998), Estimating models of sample selection bias, *Journal of Human Recourses*, 61, 191–211.

Appendix A: The Selection Adjustment

As in the main text, lower case letters are realizations of stochastic vectors in upper case letters. Let $Y^R(X_1, U)$ be a woman's reservation wage for full-time work as a function of her observed characteristics X_1 and unobserved characteristics U . Let $Y^*(X_2, V)$ be her wage for full-time work as a function of her observed characteristics X_2 and unobserved characteristics V .

Assume that a woman works full time if $Y^* \geq Y^R$. Define $D^* = Y^* - Y^R$ and let $D = 1$ if and only if $D^* \geq 0$. The variable D is equal to zero if $D^* < 0$. Let Y be the observed wage. This variable is observed and equal to Y^* if and only if $D = 1$. If $D = 0$, we do not observe Y . Assume the following specifications for Y^* and Y^R :

$$Y^* = X_2\beta + V$$

and

$$Y^R = X_1\alpha + U$$

We rewrite Y^* in the kernel regression form $q \in [0, 1]$

$$Y^* = X_2\beta(q) + V_q$$

with $\text{Quant}_q(Y^*|X_2 = x_2) = x_2\beta(q)$. It follows directly that

$$\text{Quant}_q(Y|X_2 = x_2, D = 1) = x_2\beta(q) + \text{Quant}_q(V_q|X_2 = x_2, D = 1) \neq x_2\beta(q).$$

Hence, a quantile regression of Y on X_2 is biased. Now define $h_q(x_1) = \text{Quant}(v_q|X_1 = x_1, D = 1)$ and let

$$Y = X_2\beta(q) + h_q(X_1) + V_q^*$$

By definition $V_q^* = V_q - h(X_1)$, so $\text{Quant}_q(V_q^*|X_1 = x_1, D = 1) = 0$. Note that

$$h_q^*(x_1) = \text{Quant}(V_q|X_1 = x_1, D = 1) = \text{Quant}(V_q|U - V \leq x_1\gamma, X_1 = x_1, D = 1) \equiv h_q^*(z),$$

where $\gamma_q = \beta' - \alpha$ and $Z = X_1\gamma$. β' is equal to β for all elements included in both X_1 and X_2 and equal to zero if excluded from X_2 . From this we see that h_q^* is only a function of z and that $h_q^{*'} > 0$. The form of h_q^* depends on the joint distribution of V and U . With specifying this joint distribution, we cannot know the form of h_q^* . Buchinsky (1998a) suggests using the following series estimator (see also Newey 1988):

$$\widehat{h}_q^*(z) = \delta_0(q) + \delta_1(q)\lambda(z) + \delta_2(q)\lambda(z)^2 + \dots$$

where λ is the inverse Mill's ratio. The function $\widehat{h}_q^*(z)$ is a power series of $\widehat{h}_q^*(z)$. Thus for appropriate values of the δ 's, $\widehat{h}_q^*(z) \rightarrow h_q^*(z)$ as the number of terms goes to infinity. That is, we can make the difference between $\widehat{h}_q^*(z)$ and $h_q^*(z)$ arbitrarily small when we increase the number of terms in the equation above, using appropriate values of the δ 's. Of course, the use of the inverse Mill's ratio is not necessary. Any function of z could be used (including z itself).

The only remaining problem is to obtain estimates of z given x_1 . That is, we need to estimate γ . From the equation above, it is clear that if we could observe Y^R then regressing Y^R on X_1 would give consistent estimates of γ . However, we only observe whether the difference between the wage and the reservation wage is positive (i.e. whether $D = 1$). Hence, minimization of the squared distance between D and the probability $P(D = 1 | X_1 = x_1) = G(x_1\gamma)$ - where G is the distribution of Y^R - gives a consistent estimate of γ . As we do not know the distribution of Y^R , we estimate G using kernel regression. This is the procedure described in Ichimura (1993).

When estimating a semiparametric sample selection model as described above, the intercept in the wage equation is not identified. This can be seen most easily by taking into account that both the intercept as well as $\delta_0(q)$ appear in the linear regression equation. As in Buchinsky (1998a) and Andrews and Schafgans (1998), we estimate this intercept through an identification at infinity approach. This implies that we restrict ourselves to a sample of women with a relatively high probability (and in the limit with probability equal to one) of working full time to estimate the intercept. For more details, see Andrews and Schafgans (1998).

Appendix B: Proofs of theorems

Proof of theorem 1

We first consider the estimator $\tilde{\theta}(q)$, which is the sample quantile obtained from the sample $\tilde{y}_i; i = 1 \dots M$. This sample is obtained as follows

1. Sample u_i from a standard uniform distribution
2. Sample \tilde{x}_i from the distribution G_{X_A}
3. Compute $\tilde{y}_i = \tilde{x}_i \beta^B(u_i)$
4. Repeat steps 1 to 3 M times.

The realizations of this method can itself be seen as realizations of a stochastic variable. Denote this variable by \tilde{Y}_{AB} . The distribution function of this variable is

$$\begin{aligned}
 F_{\tilde{Y}_{AB}}(y) &= \int_0^1 \int_{\text{supp}(X_A)} \mathbb{P}(\tilde{Y}_{AB} \leq y | U = u; X_A = x_A) dG_{X_A}(x_A) du \\
 &= \int_0^1 \int_{\text{supp}(X_A)} \mathbb{1}_{\{x_A \beta^B(u) \leq y\}} dG_{X_A}(x_A) du \\
 &= \int_0^1 \int_{\text{supp}(X_A)} \mathbb{1}_{\{F_{Y_{AB}}^{-1}(u|X_A=x_A) \leq y\}} dG_{X_A}(x_A) du \quad (1) \\
 &= \int_{\text{supp}(X_A)} \int_0^{F_{Y_{AB}}(y|X_A=x_A)} du dG_{X_A}(x_A) \\
 &= \int_{\text{supp}(X_A)} F_{Y_{AB}}(y|X_A=x_A) dG_{X_A}(x) = F_{Y_{AB}}(y)
 \end{aligned}$$

Hence, $\tilde{Y}_{AB} \stackrel{d}{=} Y_{AB}$. This implies that the observations from the sampling method are sampled from the population distribution Y_{AB} . For the remainder of this proof we use v_i as a k -dimensional vector that is used to sample from the distribution G_{X_A} and \hat{G}_{X_A} . The elements of v_i are sampled from the standard uniform distribution.¹⁴ Hence $\tilde{X}_{A,i} = \tilde{X}_A(v_i) = G_{X_A}^{-1}(v_i)$ and likewise for $\hat{X}_{A,i}$. Let $\tilde{\Psi}_M(\theta(q)) = \frac{1}{M} \sum_i m_q(\tilde{y}_{AB,i}, \theta(q))$ and $\hat{\Psi}_M(\theta(q)) = \frac{1}{M} \sum_i m_q(\hat{y}_{AB,i}, \theta(q))$, where

¹⁴It is always possible to sample from a k -dimensional distribution with a known distribution function based on repetitive conditioning and a draw from a k -dimensional vector sampled from univariate uniform distributions. Although it is also possible to do this using the empirical distribution function of a k -dimensional stochastic vector, bootstrapping from the data would be a much easier way to obtain such a sample. For the remainder of the proof it does not matter.

$$m(y_{AB}, \theta(q)) = \begin{cases} q(y_{AB} - \theta(q)) & \text{if } y_{AB} > \theta(q) \\ (1 - q)(\theta(q) - y_{AB}) & \text{if } y_{AB} \leq \theta(q) \end{cases}$$

In addition, define $\Psi(\theta(q))$ as

$$\Psi(\theta(q)) \equiv \int_{\text{supp}(Y_{AB})} m(y_{AB}, \theta(q)) dF_{y_{AB}}(y) = (1 - q) \int_{-\infty}^{\theta(q)} F_{y_{AB}}(y) dy + q \int_{\theta(q)}^{\infty} \bar{F}_{y_{AB}}(y) dy$$

Taking derivatives, we obtain

$$\Psi_{\theta(q)}(\theta(q)) = (1 - q)F_{y_{AB}}(\theta(q)) + q\bar{F}_{y_{AB}}(\theta(q))$$

It can be easily checked that under assumption A-2 and A-4, this derivative has a single root. This is a sufficient condition for the identification for quantiles. It remains to show that

$$\sup_{\theta(q)} \left| \widehat{\Psi}_M(\theta(q)) - \Psi(\theta(q)) \right| = o_P(1)$$

By the triangle inequality

$$\sup_{\theta(q)} \left| \widehat{\Psi}_M(\theta(q)) - \Psi(\theta(q)) \right| \leq \sup_{\theta(q)} \left| \widehat{\Psi}_M(\theta(q)) - \widetilde{\Psi}_M(\theta(q)) \right| + \sup_{\theta(q)} \left| \widetilde{\Psi}_M(\theta(q)) - \Psi(\theta(q)) \right| \quad (2)$$

The last term on the last line is $o_P(1)$ by the law of large numbers and equation (1). It remains to show that the first term is $o_P(1)$ as well. We have that (dropping the subscripts AB for \widehat{y} and \widetilde{y} for the remainder of the proof)

$$\begin{aligned}
\sup_{\theta(q)} \left| \widehat{\Psi}_M(\theta(q)) - \widetilde{\Psi}_M(\theta(q)) \right| &= \sup_{\theta(q)} \left| \frac{1}{M} \sum_{\widehat{y}_i < \theta} q(\widehat{y}_i - \theta(q)) - \frac{1}{M} \sum_{\widetilde{y}_i < \theta} q(\widetilde{y}_i - \theta(q)) + \right. \\
&\quad \left. \frac{1}{M} \sum_{\widehat{y}_i \geq \theta} (1-q)(\theta(q) - \widehat{y}_i) - \frac{1}{M} \sum_{\widetilde{y}_i \geq \theta} (1-q)(\theta(q) - \widetilde{y}_i) \right| \\
&\leq q \sup_{\theta(q)} \left| \frac{1}{M} \sum_{\widehat{y}_i < \theta(q)} (\widehat{y}_i - \widetilde{y}_i) \right| + (1-q) \sup_{\theta(q)} \left| \frac{1}{M} \sum_{\widetilde{y}_i \geq \theta(q)} (\widetilde{y}_i - \widehat{y}_i) \right| \\
&+ q \sup_{\theta(q)} \left| \frac{1}{M} \sum_{\widehat{y}_i < \theta(q) \leq \widetilde{y}_i} (\widehat{y}_i - \theta(q)) \right| + q \sup_{\theta(q)} \left| \frac{1}{M} \sum_{\widetilde{y}_i < \theta(q) \leq \widehat{y}_i} (\widehat{y}_i - \theta(q)) \right| \\
&+ (1-q) \sup_{\theta(q)} \left| \frac{1}{M} \sum_{\widehat{y}_i < \theta(q) \leq \widetilde{y}_i} (\theta(q) - \widehat{y}_i) \right| \\
&+ (1-q) \sup_{\theta(q)} \left| \frac{1}{M} \sum_{\widetilde{y}_i < \theta(q) \leq \widehat{y}_i} (\theta(q) - \widehat{y}_i) \right| \tag{3}
\end{aligned}$$

Making use of the definition of \widehat{y}_i , \widetilde{y}_i , \widehat{x}_i and x_i and using the triangle inequality again, we obtain (dropping superscript B for β and A for x)

$$\begin{aligned}
q \sup_{\theta(q)} \left| \frac{1}{M} \sum_{\widehat{y}_i < \theta(q)} (\widehat{y}_i - \widetilde{y}_i) \right| &\leq \sup_{\theta(q)} \left| \frac{1}{M} \sum_{\widehat{y}_i < \theta(q)} \widehat{x}_i(v_i) \left(\widehat{\beta}(u_i) - \beta(u_i) \right) \right| \\
&+ \left| \frac{1}{M} \sum_{\widehat{y}_i < \theta(q)} \left(\widehat{G}_{X_A}^{-1}(v_i) - G_{X_A}^{-1}(v_i) \right) \beta(u_i) \right| \tag{4}
\end{aligned}$$

We have that $\widehat{\beta}(u_i) \xrightarrow{P} \beta(u_i)$; $i = 1, \dots, M$ for $N_B \rightarrow \infty$ (consistency of quantile regressions). Since $M/N_B \rightarrow I_B < \infty$ when $M, N_B \rightarrow \infty$ the first term is $o_P(1)$. For the second term, we have that $\widehat{G}_{X_A}^{-1}(v_i) \xrightarrow{P} G_{X_A}^{-1}(v_i)$ when $n \rightarrow \infty$ due to a combination of the Glivenko-Cantelli theorem (satisfied by A-2) and the continuous mapping theorem.¹⁵ Hence, the second term in equation (4) is $o_P(1)$ as well (using $M/N_A \rightarrow I_A < \infty$ as $M, N_A \rightarrow \infty$). This implies that the right-hand side of equation (4) and

¹⁵For the use of the continuous mapping theorem we need continuity of $G_{X_A}(\cdot)$. This is not an assumption in A1-A4. However, in case X_A is strictly discrete with mass x_{A_j} ; $j = 1, \dots, l$, then the second part of this equation changes into $\sum_i \sum_j x_j (\widehat{p}_j - p_j) \beta(u_i)$, where the \widehat{p}_j 's and p_j 's are the sample and population frequencies of individuals with characteristics equal to x_j . This term obviously

hence the first term on the right-hand side of equation (3) converges in probability to zero. The proof that the other terms on the right-hand side of (3) are $o_p(1)$ as well is along the same lines. Hence $\sup_{\theta}(q) \left| \widehat{\Psi}_M(\theta(q)) - \widetilde{\Psi}_M(\theta(q)) \right|$ converges in probability to zero when $N_A, N_B \rightarrow \infty$. This implies that $\sup_{\theta}(q) \left| \widehat{\Psi}_M(\theta(q)) - \Psi(\theta(q)) \right|$ (see equation (2)) converges to zero in probability since $M/N_A \rightarrow I_A < \infty$. Using theorem 5.7 of Van der Vaart (1998) completes the proof.

■

Proof of theorem 2

Before we are able to prove theorem 2 we need to prove the following lemma.

Lemma 3 *Let $q_1, q_2 \dots q_m \in (0, 1)$, X be a k -dimensional random vector with compact support on \mathbb{R}^k . In addition let $\beta(q)$ be such that*

$$\text{Quant}_q(Y|X=x) = x\beta(q)$$

where Y is a random variable with compact support on \mathbb{R} . Then the regression quantiles $\widehat{\beta}(q_1), \dots, \widehat{\beta}(q_m)$ satisfy

$$\sqrt{n} \begin{pmatrix} \widehat{\beta}(q_1) - \beta(q_1) \\ \vdots \\ \widehat{\beta}(q_m) - \beta(q_m) \end{pmatrix} \rightarrow N(0, \Lambda(\beta))$$

with $\Lambda(\beta)$ an $m \times k$ by $m \times k$ dimensional matrix, where the elements of $\Lambda(\beta)$ are equal to

$$\Lambda(\beta(q_i), \beta(q_j)) = \begin{cases} \mathbb{E}_X (f_Y(X\beta(q_i))XX^T)^{-1} q_i(1 - q_j)\mathbb{E}_X(XX^T) \\ \quad \times \mathbb{E}_X (f_Y(X\beta(q_i))XX^T)^{-1} & \text{if } i \leq j \\ \mathbb{E}_X (f_Y(X\beta(q_i))XX^T)^{-1} q_j(1 - q_i)\mathbb{E}_X(XX^T) \\ \quad \times \mathbb{E}_X (f_Y(X\beta(q_i))XX^T)^{-1} & \text{otherwise} \end{cases}$$

Proof

When we define $\varepsilon = Y - X\beta$ then if $f_{\varepsilon}(\cdot|X=x) = f_{\varepsilon}(\cdot)$ the lemma is the same as theorem 4.2. of Koenker and Bassett (1978). The regression quantiles are estimated as the solution to

converges to zero in probability. Of course the most general case is when X_A contains both discrete as well as continuous elements. We do not prove this case in this paper.

$$\max_{\beta(q_1), \beta(q_2), \dots, \beta(q_m)} z(y_i, \beta(q_1), \dots, \beta(q_m), q_1, \dots, q_m)$$

where

$$z(x, y, \beta(q_1), \dots, \beta(q_m), q_1, \dots, q_m) = \sum_{i=1}^m q_i (y - x\beta(q_i)) \mathbf{1}_{\{y > x\beta(q_i)\}} + (1 - q_i) (x\beta(q_i) - y) \mathbf{1}_{\{y \leq x\beta(q_i)\}}$$

Define

$$\Psi(\beta(q_1), \dots, \beta(q_m), q_1, \dots, q_m) = \mathbf{E}_{Y, X} (z(X, Y, \beta(q_1), \dots, \beta(q_m), q_1, \dots, q_m))$$

It is possible to show that

$$\frac{\partial \Psi}{\partial \beta(q_i)} = \mathbf{E}_{Y, X} [X F_Y(X\beta(q_i)) - x q_i]$$

We denote the hessian of Ψ by H . Its elements are equal to (*i.e.* $k \times k$ matrices)

$$H_{ij} = \frac{\partial^2 \Psi}{\partial \beta(q_i) \partial \beta(q_j)} = \begin{cases} \mathbf{E}_{Y, X} [X X^T f_Y(X\beta(q_i))] & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

The first order derivative of $z(x, y, \beta(q_1), \dots, \beta(q_m), q_1, \dots, q_m)$ with respect to $\beta(q_i)$ is equal to

$$z_{\beta(q_i)}(y, x, \beta(q_1), \dots, \beta(q_m), q_1, \dots, q_m) = \begin{cases} -q_i x & y > x\beta(q_i) \\ (1 - q_i) x & y \leq x\beta(q_i) \end{cases}$$

Hence the (expected) cross derivative matrix Z has the following elements $i, j = 1, \dots, m; j \geq i$ (*i.e.* $k \times k$ matrices)

$$Z_{ij} = \mathbf{E}_{Y, X} (z_{\beta(q_i)} z_{\beta(q_j)}^T) = q_i (1 - q_j) \mathbf{E}_X (X X^T)$$

The covariance matrix of $\sqrt{n}(\hat{\beta}(q_1) - \beta(q_1), \dots, \hat{\beta}(q_m) - \beta(q_m))$ is equal to $\Lambda(\beta) = H^{-1} Z H^{-1}$. Taking the results above, the lemma follows immediately.

■

Proof of theorem 2: The criterion function to obtain $\hat{\theta}(q)$ is equal to

$$\min_{\theta(q)} \frac{1}{M} \sum_{\hat{y}_{AB_i} > \theta(q)} q \|\hat{y}_{AB_i} - \theta\| + (1 - q) \sum_{\hat{y}_{AB_i} \leq \theta(q)} \|\hat{y}_{AB_i} - \theta(q)\| = \min_{\theta} \sum_{i=1}^M m(\hat{y}_{AB_i}, \theta(q))$$

where¹⁶

$$m(y_{AB}, \theta(q)) = \begin{cases} q(y_{AB} - \theta(q)) & \text{if } y_{AB} > \theta(q) \\ (1 - q)(\theta(q) - y_{AB}) & \text{if } y_{AB} \leq \theta(q) \end{cases}$$

and \hat{y}_{AB_i} is the estimated level of y_{AB_i} (*i.e.* the i -th observation from the sample using the Machado-Mata technique). The first order condition is equal to

$$\sum_{i=1}^M \frac{\partial m(\hat{y}_{AB_i}, \hat{\theta}(q))}{\partial \theta(q)} = 0$$

In general we should take account of the fact that we are taking bootstrap samples from X_A and hence the convergence of the expression above is dependent on this. As is proven in Van der Vaart (1998, page 333-334), this expression has the same limit as the statistic that results when we sample from the population instead of the data set. Hence we will not take this into account in the remainder of our analysis. Taking a first order Taylor series expansion of this results in

$$0 = \frac{1}{\sqrt{M}} \sum_{i=1}^M \frac{\partial m(\hat{y}_{AB_i}, \theta(q))}{\partial \theta(q)} + \frac{1}{M} \sum_{i=1}^n \frac{\partial^2 m(\hat{y}_{AB_i}, \theta(q))}{\partial \theta(q)^2} \sqrt{M}(\hat{\theta}(q) - \theta(q)) + o_P(1) \quad (5)$$

We have that (see for example Van der Vaart (1998))

$$\frac{1}{M} \sum_{i=1}^n \frac{\partial^2 m(\hat{y}_{AB_i}, \theta(q))}{\partial \theta(q)^2} \xrightarrow{P} f_{y_{AB_i}}(\theta(q))$$

The first part on the right-hand side of equation (5) can be further expanded using a Taylor series expansion around the true value of the $\beta(u)'s$

$$\begin{aligned} \frac{1}{\sqrt{M}} \sum_{i=1}^M \frac{\partial m(\hat{y}_{AB_i}, \theta(q))}{\partial \hat{\theta}(q)} &= \frac{1}{\sqrt{M}} \sum_{i=1}^M \frac{\partial m(y_{AB_i}, \theta(q))}{\partial \theta(q)} + \\ &\quad \sqrt{\frac{M}{N_B}} \frac{1}{M} \sum_{i=1}^M \frac{\partial^2 m(y_{AB_i}, \theta(q))}{\partial \theta(q) \partial \beta^B(u_i)} \sqrt{N_B}(\hat{\beta}^B(u_i) - \beta^B(u_i)) + o_P(1) \end{aligned} \quad (6)$$

¹⁶Note that the function presented is not differentiable everywhere. This is not sufficient for the proofs presented below. In general it suffices to show that a Lipschitz condition holds. It is not difficult to show that this condition is satisfied. See Van der Vaart (1998) for more details.

Without loss of generality, we assume that the u_i 's are in ascending order. The second expression of equation (6) is non-standard because of the dependence between the regression quantiles. By the definition of Y_{AB} , we have

$$\begin{aligned} \mathbb{E}_{Y_{AB}}(m(Y_{AB}, \theta(q))) &= q \int_{\theta(q)}^{\infty} \bar{F}(y) dy + (1-q) \int_{-\infty}^{\theta(q)} F(y) dy \\ &= q \int_0^1 \int_{\text{supp}X_A} \int_{\theta(q)}^{x_A \beta^B(u)} dy \mathbf{1}_{\{x_A \beta^B(u) > \theta(q)\}} dG_{X_A}(x_A) du \\ &\quad + (1-q) \int_0^1 \int_{\text{supp}X_A} \int_{x_A \beta^B(u)}^{\theta(q)} dy \mathbf{1}_{\{x_A \beta^B(u) \leq \theta(q)\}} dG_{X_A}(x_A) du \end{aligned}$$

Hence, the partial derivative with respect to $\beta^B(u)$ is

$$\begin{aligned} \mathbb{E}_{X_A, U} \left(\frac{\partial m(Y_{AB}, \theta(q))}{\partial \beta^B(U)} \right) &= q \int_0^1 \int_{\text{supp}X_A} x_A \mathbf{1}_{\{x_A \beta^B(u) > \theta(q)\}} dG_{X_A}(x_A) du \\ &\quad - (1-q) \int_0^1 \int_{\text{supp}X_A} x_A \mathbf{1}_{\{x_A \beta^B(u) \leq \theta(q)\}} dG_{X_A}(x_A) du \end{aligned}$$

The cross derivative with respect to $\theta(q)$ is

$$\begin{aligned} \mathbb{E}_{X_A, U} \left(\frac{\partial^2 m(Y_{AB}, \theta(q))}{\partial \beta^B(U) \partial \theta(q)} \right) &= -q \int_0^1 \int_{\text{supp}X_A} x_A \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \mathbf{1}_{\{\theta(q) < x_A \beta^B(u) \leq \theta(q) + \varepsilon\}} dG_{X_A}(x_A) du \\ &\quad - (1-q) \int_0^1 \int_{\text{supp}X_A} x_A \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \mathbf{1}_{\{\theta(q) < x_A \beta^B(u) \leq \theta(q) + \varepsilon\}} dG_{X_A}(x_A) du \\ &= - \int_0^1 \int_{\text{supp}X_A} x_A \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \mathbf{1}_{\{\theta(q) < x_A \beta^B(u) \leq \theta(q) + \varepsilon\}} dG_{X_A}(x_A) du \\ &= - \int_{\text{supp}X_A} x_A f_{Y_{AB}}(\theta(q)) dG_{X_A}(x_A) \\ &= -\mathbb{E}_{X_A, U}(X_A f_{Y_{AB}}(\theta(q))) \end{aligned}$$

Combining this with lemma 3 we find that the second part converges to a 1 dimensional normal distribution with variance

$$\mathbb{E}_{X_A, U, V} \{ f_{Y_{AB}}^2(\theta(q)) X_A \Lambda(\beta^B(U), \beta^B(V)) X_A^T \}$$

The first part of equation (6) is standard and can be shown to converge to a zero mean normal distribution with variance $I_A q(1-q)$. Making assumption A-5, it can be shown that the criterion functions of the first step (i.e. z as in lemma 3) and

the criterion function of the second step, m are independent.¹⁷ Hence the right hand side of equation (6) converges to a normal distribution with zero mean and covariance matrix

$$I_A q(1 - q) + I_B E_{X_A, U, V} \{ f_{Y_{AB}}^2(\theta(q)) X_A \Lambda(\beta^B(U), \beta^B(V)) X_A^T \}$$

The result of the theorem follows directly.

■

¹⁷Although the intuition behind this is easy, the notation is somewhat tedious. Define $z(y_B, x_B, \beta^B(u), u)$ in the same way as in lemma 3. This implies that $E(z(y_B, x_B, \beta^B(u), u) m(y_{AB}, \theta(q))) = \int_{X_A, Y_B, X_B, U} z(y_B, x_B, \beta^B(u), u) m(y_{AB}, \theta(q)) dG_{X_A, Y_B, X_B, U}(x_A, y_B, x_B, u) = \int_{X_A, U} m(y_{AB}, \theta(q)) \int_{Y_B, X_B} z(y_B, x_B, \beta^B(u), u) dG_{Y_B, X_B}(y_B, x_B) dG_{X_A, U}(x_A, u)$. The inner integral has derivative with respect to $\beta(u)$ being equal to zero. This completes the proof of independence.

Appendix C: Comparison of Asymptotic Results with Bootstrap Standard Errors

In order to obtain an idea about the performance of our asymptotic results we compare them with those derived from a bootstrap sampling procedure. Although we do not prove it in this paper, the method of bootstraps is likely to be able to consistently estimate the distribution of the statistics that we consider using the Machado-Mata technique. The motivation behind this is that the M-M method mainly consists of the estimation of regression quantiles and it has been shown that for these regression quantiles the conditions for consistency are met (see for example Van der Vaart, 1998 and Horowitz, 2002). The rate of convergence is in general equal to \sqrt{n} , where n is the number of observations from an arbitrary sample. This implies that we compare two \sqrt{n} -approximations of the unknown finite sample distribution with each other. In general, if the number of observations is not too small, this implies that the difference in the approximations should not be large. In order to proceed we draw a replacement sample of size N_A from the data of population A (X_A) and a replacement sample of size N_B from the data of population B (X_B and Y_B). For every sample we proceed through steps 1 to 5 of the Machado-Mata method and compute the quantiles of the sample obtained from this procedure at different locations of the distribution. For reasons of convenience we take $M = N_A$, making our results and those obtained from the bootstrap sampling procedure directly comparable. Both methods converge at rate $\sqrt{N_A}$ in this case. We repeat this procedure 1000 times.

Table 8 presents results from our bootstrap exercise. These are the results when we sample from the data of the characteristics of men and use the regression quantiles of women who are working full time. This is the exercise that is described in section 5 as well. The first column reports the different locations of the distribution. The second column reports the results of the M-M method, and the third column reports the average over the bootstrap samples. Comparing the M-M quantile estimates with the bootstrapped estimates, we find one interesting feature: although the average from the bootstraps are roughly similar to those obtained directly from the M-M method in general, this is not the case for the lowest quantiles. It seems that the M-M method has an upward bias for these low quantiles. The fourth column of Table 8 lists the levels of the standard errors using the asymptotic results as presented in section 4. The last column of the table lists the results of the standard errors from the bootstraps. We find the standard errors to be quite close to each other. This implies that the asymptotics derived in theorem 2 of section 4 work quite well with the sample sizes we use in our analysis.

Table 1: Descriptive Statistics for Women in OSA Data

	All	Part-time	Full-time	Full-time wage obs.
Age	39.3 (8.36)	39.94 (7.48)	36.07 (8.56)	35.97 (8.61)
Married	0.87	0.95	0.76	0.76
Number of years of work experience	10.9 (7.32)	12.1 (7.89)	12.8 (7.84)	12.2 (7.91)
Agree, parents should reduce hours	0.65	0.64	0.55	0.55
Living in city	0.32	0.32	0.39	0.40
Religious	0.60	0.60	0.53	0.52
Education levels				
Up to elementary school	0.12	0.10	0.06	0.06
Lower secondary education	0.44	0.49	0.30	0.27
Higher secondary education	0.30	0.25	0.40	0.40
Bachelors/masters	0.14	0.16	0.23	0.26
<i>Age of youngest child living at home</i>				
Below 5 years	0.18	0.19	0.08	0.08
Between 5 and 11 years	0.20	0.23	0.13	0.11
Between 12 and 17 years of age	0.19	0.26	0.11	0.12
Number of observations	1617	336	410	391

Table 2: Descriptive Statistics for Men in OSA Data

	All	Part-time	Full-time	Full-time wage obs.
Age	39.26 (8.39)	43.20 (8.74)	38.90 (8.28)	38.74 (8.22)
Married	0.86	0.78	0.87	0.87
Number of years of work experience	19.53 (9.76)	23.74 (10.56)	19.61 (9.67)	19.45 (9.64)
Agree, parents should reduce hours	0.61	0.70	0.61	0.62
Living in city	0.34	0.39	0.32	0.33
Religious	0.55	0.74	0.54	0.54
<i>Education levels</i>				
Up to elementary school	0.10	0.05	0.10	0.10
Lower secondary education	0.36	0.39	0.36	0.36
Higher secondary education	0.33	0.39	0.33	0.33
Bachelors/masters	0.21	0.17	0.21	0.21
<i>Age of youngest child living at home</i>				
Below 5 years	0.19	0.13	0.19	0.20
Between 5 and 11 years	0.19	0.22	0.19	0.19
Between 12 and 17 years of age	0.17	0.17	0.17	0.17
Number of observations	1568	23	1312	1233

Table 3: Net Hourly Wages in the OSA Labor Supply Survey in 1992 as Measured in Euros

	<u>Women</u>		<u>Men</u>
	Part time	Full time	
Wages	6.33 (2.15)	6.36 (2.12)	7.72 (2.38)

Table 4: Estimates of the Incidence of Full-Time Work

	Probit	Single-index
Constant	2.226 (0.282)	2.226 (·) [‡]
Age	-0.085 (0.006)	-0.085 (·) [‡]
Married	-0.259 (0.115)	-0.176 (0.123)
Religious	-0.046 (0.082)	0.013 (0.098)
Number of years of work experience	0.065 (0.006)	0.063 (0.005)
Living in city	0.076 (0.084)	-0.020 (0.102)
Agree, parents should reduce hours interacted with children present	-0.237 (0.113)	-0.329 (0.168)
<i>Education</i>		
Lower secondary education	0.085 (0.140)	0.296 (0.200)
Upper secondary education	0.511 (0.144)	0.684 (0.212)
Bachelors/masters	0.735 (0.157)	0.947 (0.225)
<i>Age of youngest child</i>		
Younger than 5 years of age	-1.300 (0.145)	-1.360 (0.221)
Between 5 and 12 years of age	-0.654 (0.127)	-0.658 (0.160)
Between 12 and 18 years of age	-0.290 (0.134)	-0.404 (0.180)

[‡] We normalize the constant and age coefficient in the single index model.

Table 5: Quantile Regressions for Women Without Corrections for Selectivity

	10%	20%	30%	40%	50%	60%	70%	80%	90%
Constant	1.30 (0.12)	1.40 (0.08)	1.41 (0.07)	1.45 (0.06)	1.52 (0.08)	1.54 (0.07)	1.59 (0.07)	1.67 (0.08)	1.81 (0.07)
Married	-0.00 (0.05)	0.03 (0.04)	0.02 (0.04)	0.02 (0.03)	0.02 (0.04)	0.02 (0.03)	-0.00 (0.03)	-0.01 (0.04)	-0.09 (0.03)
Yrs of experience/100	0.52 (0.31)	0.72 (0.21)	0.76 (0.19)	0.93 (0.15)	1.00 (0.21)	1.13 (0.17)	1.11 (0.17)	1.04 (0.19)	1.22 (0.14)
City	-0.02 (0.05)	-0.01 (0.03)	-0.00 (0.03)	0.02 (0.02)	0.03 (0.03)	0.02 (0.03)	0.01 (0.03)	0.00 (0.03)	0.00 (0.03)
<i>Education</i>									
Lower secondary	0.15 (0.09)	0.05 (0.07)	0.11 (0.06)	0.08 (0.05)	0.06 (0.07)	0.08 (0.06)	0.11 (0.06)	0.11 (0.07)	0.10 (0.06)
Upper secondary	0.21 (0.09)	0.14 (0.07)	0.20 (0.06)	0.20 (0.05)	0.16 (0.07)	0.18 (0.06)	0.19 (0.06)	0.18 (0.07)	0.14 (0.06)
Bachelors/masters	0.33 (0.10)	0.32 (0.07)	0.36 (0.06)	0.33 (0.05)	0.31 (0.07)	0.35 (0.06)	0.40 (0.07)	0.42 (0.07)	0.41 (0.06)

Table 6: Quantile Regressions for Women with Corrections for Selectivity

	10%	20%	30%	40%	50%	60%	70%	80%	90%
Constant	1.27 (0.02)	1.28 (0.03)	1.34 (0.02)	1.35 (0.02)	1.38 (0.02)	1.40 (0.02)	1.44 (0.02)	1.52 (0.03)	1.56 (0.03)
Married	0.00 (0.04)	0.00 (0.03)	-0.00 (0.03)	0.01 (0.04)	-0.01 (0.04)	-0.02 (0.04)	-0.03 (0.04)	-0.05 (0.04)	-0.05 (0.04)
Yrs of experience/100	0.39 (0.40)	0.60 (0.24)	0.72 (0.19)	0.80 (0.18)	0.89 (0.18)	1.03 (0.21)	0.94 (0.21)	0.88 (0.21)	1.19 (0.30)
City	-0.01 (0.04)	-0.02 (0.03)	-0.00 (0.03)	-0.02 (0.03)	0.01 (0.03)	0.03 (0.03)	0.02 (0.03)	0.01 (0.03)	0.01 (0.04)
<i>Education</i>									
Lower secondary	0.13 (0.12)	0.07 (0.06)	0.08 (0.06)	0.11 (0.07)	0.10 (0.06)	0.11 (0.06)	0.10 (0.06)	0.11 (0.06)	0.18 (0.07)
Upper secondary	0.19 (0.13)	0.16 (0.07)	0.20 (0.06)	0.23 (0.07)	0.23 (0.06)	0.23 (0.06)	0.21 (0.06)	0.19 (0.06)	0.21 (0.07)
Bachelors/masters	0.31 (0.13)	0.33 (0.07)	0.34 (0.07)	0.40 (0.07)	0.38 (0.07)	0.39 (0.06)	0.41 (0.07)	0.43 (0.07)	0.49 (0.08)

Table 7: Estimates of the Quantile Regressions for Men Without Correction

	10%	20%	30%	40%	50%	60%	70%	80%	90%
Constant	1.36 (0.05)	1.42 (0.04)	1.48 (0.04)	1.51 (0.04)	1.55 (0.02)	1.59 (0.04)	1.61 (0.04)	1.67 (0.05)	1.75 (0.07)
Married	0.11 (0.03)	0.10 (0.02)	0.11 (0.02)	0.12 (0.02)	0.13 (0.02)	0.11 (0.02)	0.13 (0.03)	0.13 (0.03)	0.10 (0.05)
Yrs of experience/100	0.49 (0.13)	0.65 (0.09)	0.65 (0.09)	0.68 (0.09)	0.69 (0.06)	0.80 (0.08)	0.86 (0.10)	1.03 (0.12)	1.18 (0.16)
City	-0.00 (0.02)	-0.01 (0.02)	-0.01 (0.02)	0.00 (0.02)	0.01 (0.01)	0.01 (0.02)	0.00 (0.02)	0.02 (0.02)	-0.00 (0.03)
<i>Education</i>									
Lower secondary	0.14 (0.04)	0.12 (0.03)	0.11 (0.03)	0.11 (0.03)	0.10 (0.02)	0.10 (0.03)	0.10 (0.03)	0.07 (0.04)	0.12 (0.06)
Upper secondary	0.19 (0.04)	0.18 (0.03)	0.18 (0.03)	0.20 (0.03)	0.20 (0.02)	0.21 (0.03)	0.24 (0.03)	0.21 (0.04)	0.27 (0.06)
Bachelors/masters	0.33 (0.04)	0.37 (0.03)	0.39 (0.03)	0.42 (0.03)	0.42 (0.02)	0.43 (0.03)	0.45 (0.04)	0.47 (0.04)	0.54 (0.06)

Table 8: Results of the Bootstrap Exercise for the MM-method for the Counterfactual Distribution of Women with the Characteristics of Men but paid for these Characteristics like Women Using 1000 Bootstrap Samples

	<u>Quantile estimates</u>		<u>Standard errors</u>	
	M-M	M-M bootstrap	computed	boot- straped
10 %	1.553	1.554	0.021	0.030
20 %	1.648	1.659	0.019	0.024
30 %	1.732	1.734	0.020	0.021
40 %	1.798	1.798	0.019	0.020
50 %	1.858	1.856	0.019	0.019
60 %	1.922	1.921	0.019	0.020
70 %	1.986	1.986	0.020	0.021
80 %	2.073	2.064	0.020	0.024
90 %	2.205	2.181	0.022	0.031

Figure 1: Kernel Density Estimates for Wages in OSA Labor Supply Panel Survey

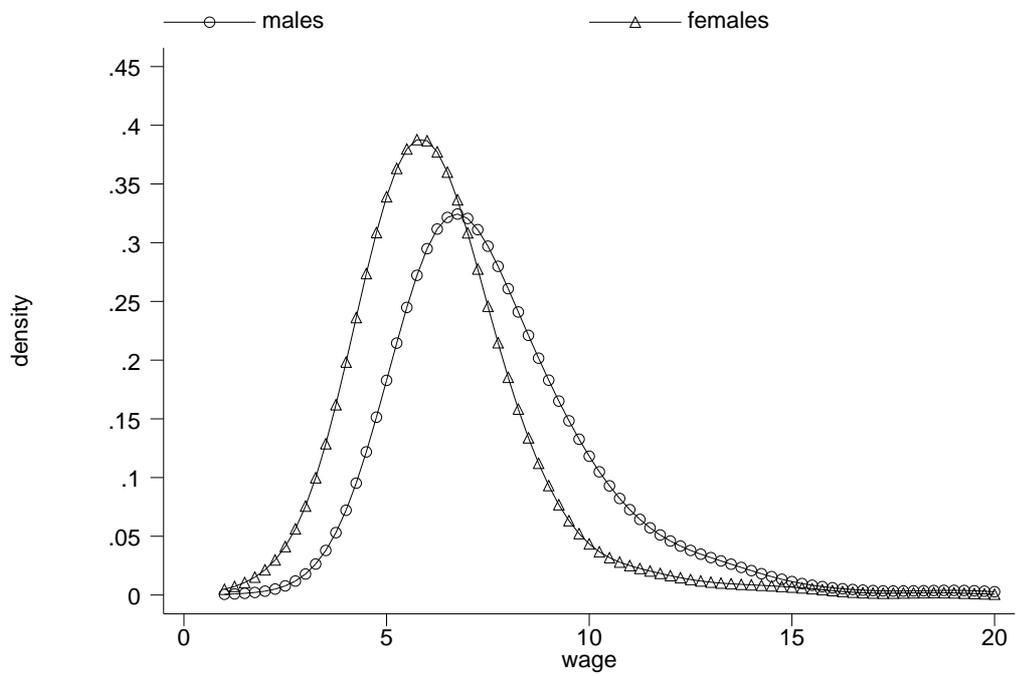


Figure 2: Gender Log Wage Gap from Raw Data

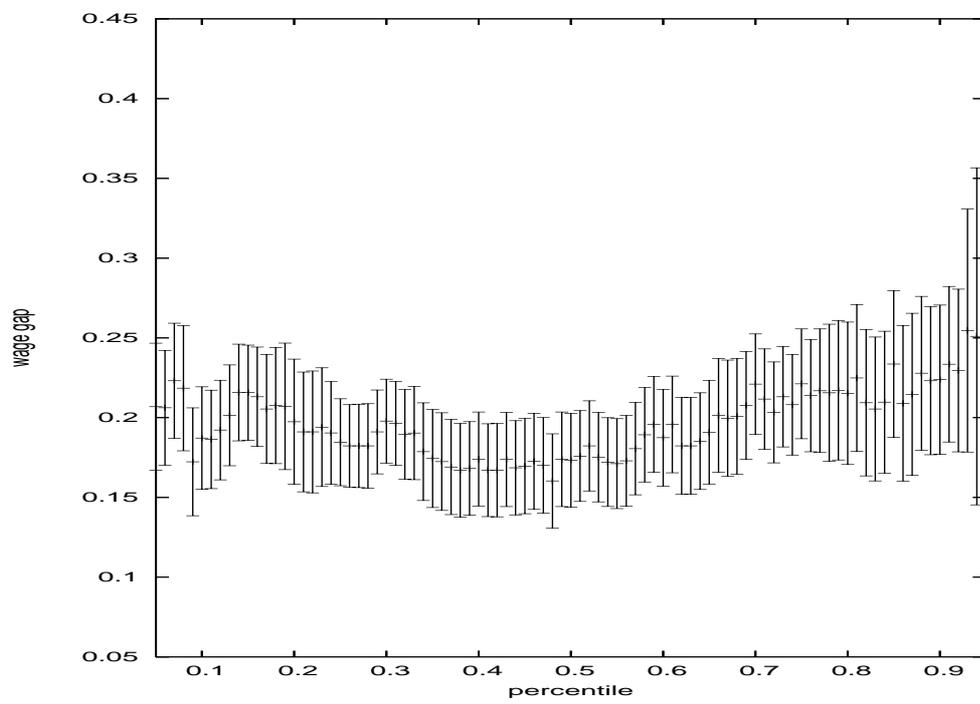


Figure 3: Gender Log Wage Gap from Raw Data for Women Working Full Time

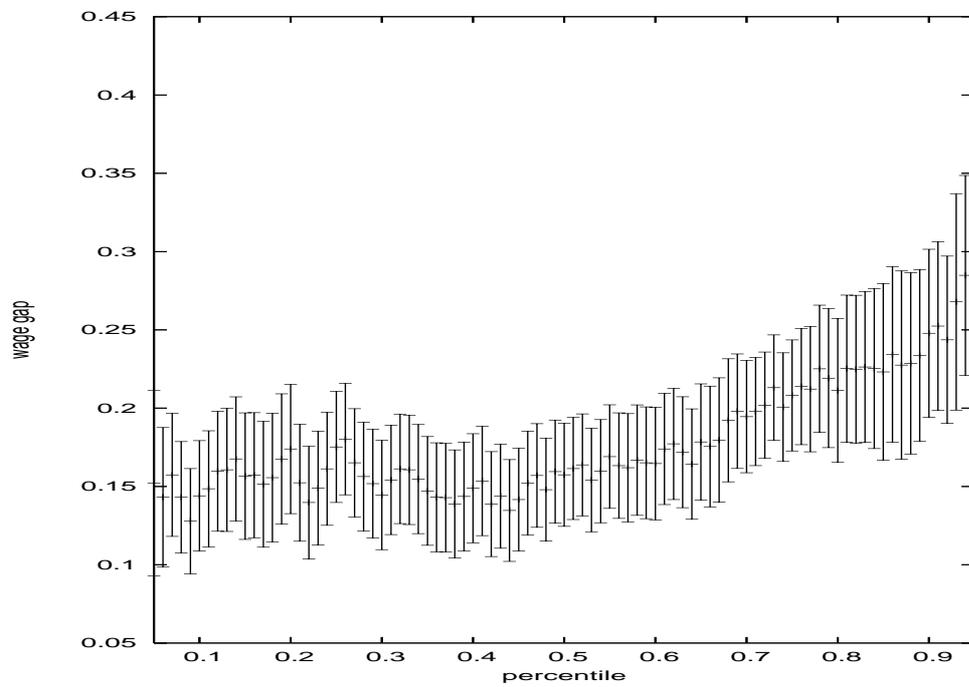


Figure 4: Log Wage Gap Between Women Working Full Time and Part Time

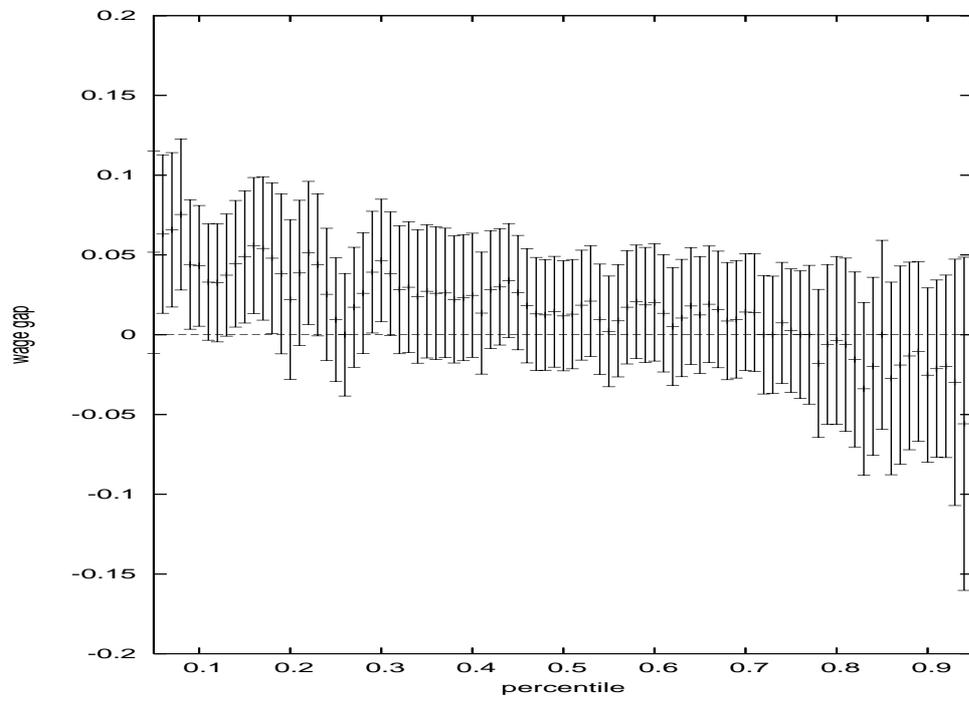


Figure 5: Log Wage Gap Between Men's Wages and Wages that Women Would Earn if They had Men's Characteristics and Women's Returns to Those Characteristics

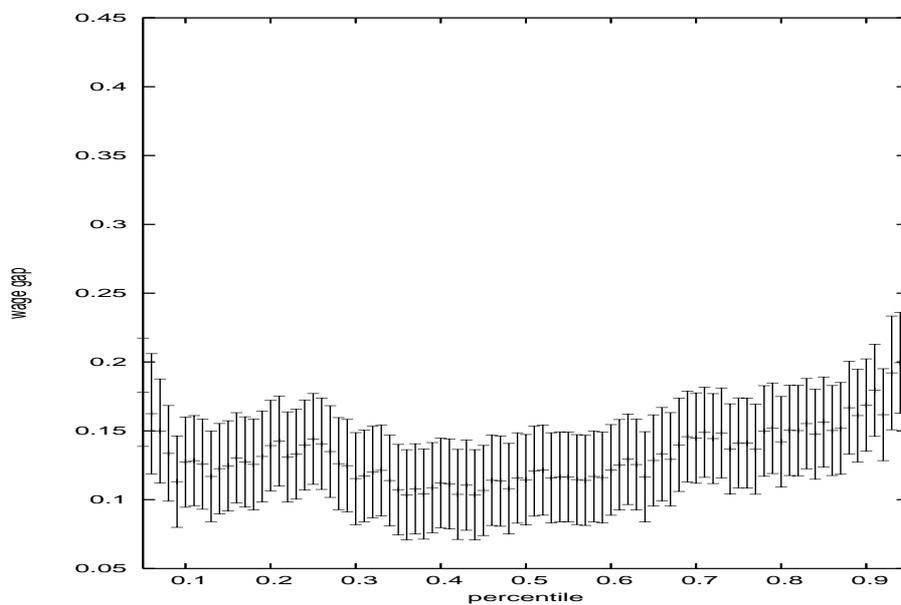


Figure 6: Log Wage Gap Between Men and Full-Time Women Paid Like Men

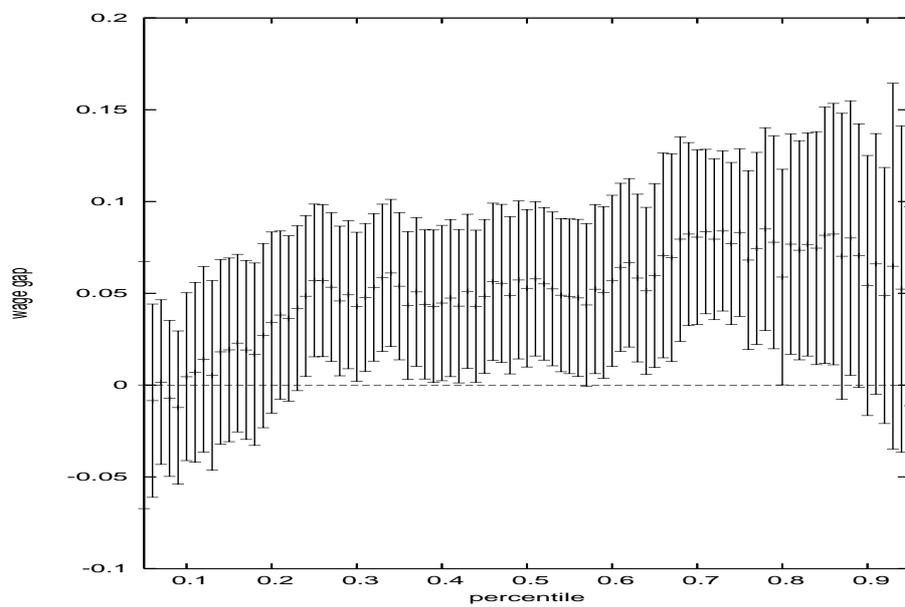


Figure 7: Log Wage Gap Between Full-Time Women's Wages Before and After Selection Correction

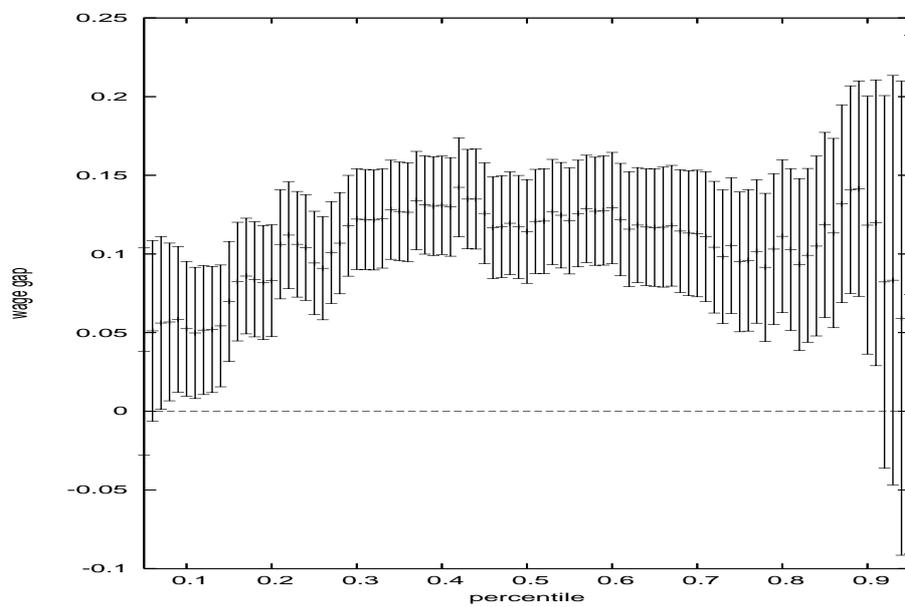


Figure 8: Log Wage Gap Between Men and Full-Time Women Corrected for Selection

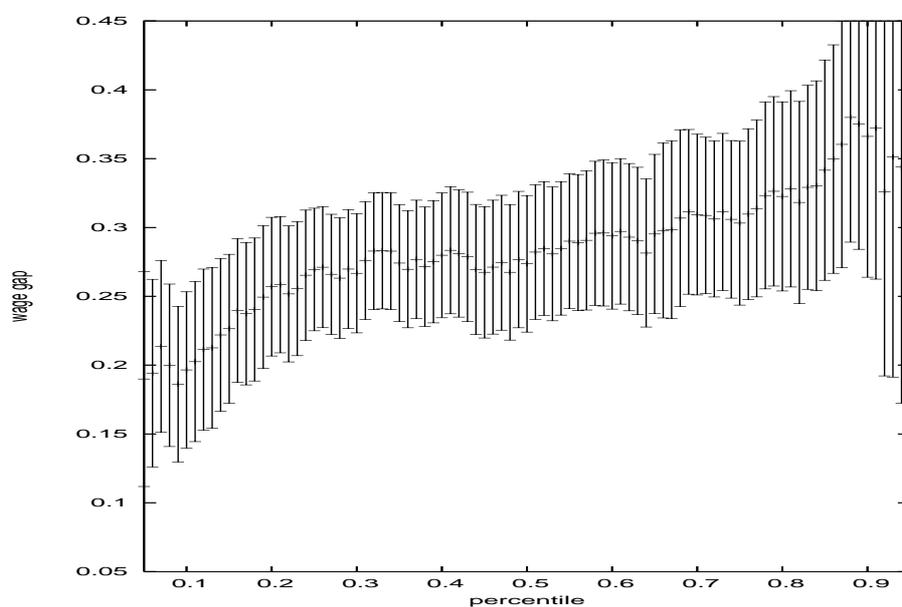


Figure 9: Sample Selection Based on Observed Characteristics

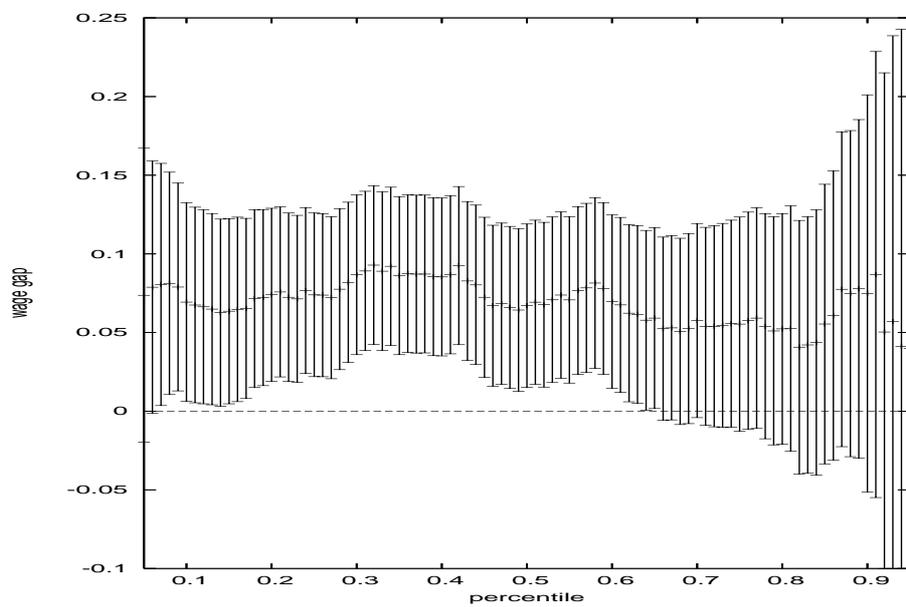


Figure 10: Sample Selection Based on Unobserved Characteristics

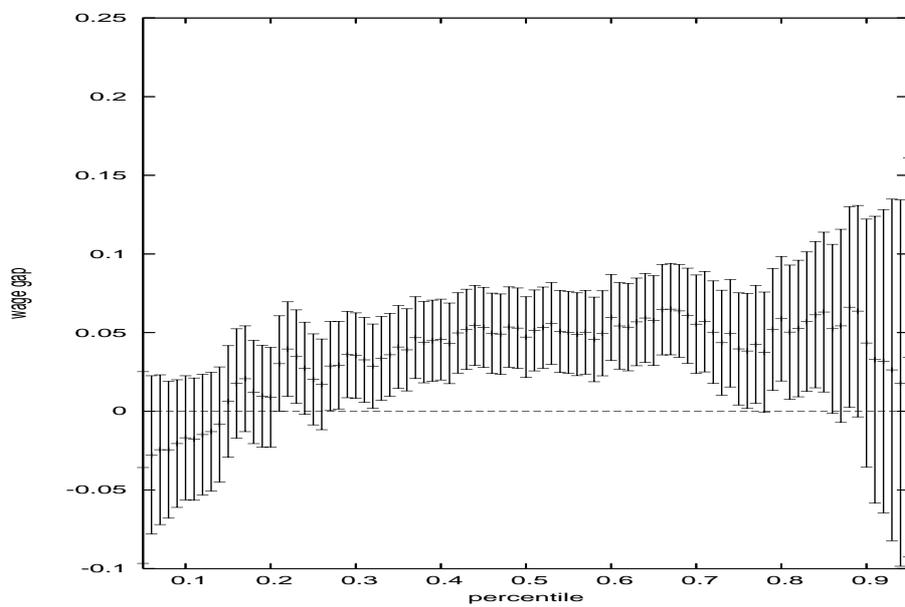


Figure 11: Difference Between Men's Log Wages and the Distribution of Wages That Women Would Earn If All Women Worked Full Time and Had Male Characteristics but Women's (Selection Adjusted) Returns

