# Decomposing the Gender Wage Differences using Quantile Regressions

Anja Heinze
Centre for European Economic Research (ZEW Mannheim)

Using linked employee-employer data, this paper measures and decomposes the differences in earnings distribution between male and female employees in Germany. I propose to extend the traditional decomposition to disentangle the effect of human capital characteristics and the effect of firm characteristics in explaining the gender wage gap. Three decomposition methods are used: Oaxaca-Blinder decomposition using OLS and quantile regressions as well as the decomposition proposed by Machado and Mata (2002). I find that the difference in the returns to the firm characteristics forms the largest part of the gender wage gap while the difference in returns to the human capital characteristics constitutes the smallest part. Furthermore, the observed gender wage gap and the decomposed parts of them vary across the wage distribution.

JEL Classification: J16 and J31

Keywords: gender wage gap, quantile regression, discrimination.

# 1 Motivation

It is a widely documented fact that male employees earn higher wages than females even after controlling for measurable characteristics related to their productivity (see, e.g., Blau and Kahn 1997). The usual methodological approach in most studies to investigate the gender wage gap is to decompose it into a part attributable to differences in the vector of worker characteristics and a part attributable to differences in the returns to these characteristics. For this purpose, they use the conditional wage distributions of male and female employees and conclude that a substantial percentage of the wage gap is due to differences in the returns to observable characteristics that favour men.

However, the analysis of the gender wage gap at the mean is limited because it could lead us to conclude that the size of the wage gap and the explaining factors are constant throughout the whole wage distribution. I propose to measure and decompose the gender wage gap at each percentile of the wage distribution which allows me to analyze how the part attributed to different characteristics and the part attributable to differential returns to these characteristics are distributed across the wage distribution.

Furthermore, the aim of the paper is not only to explain the gender wage gap on the basis of human capital characteristics. It is already accepted that firm characteristics also affect the wage level as well as the wage distribution (see e.g. Abowd, Kramarz and Margolis 1999). Moreover, some studies show that firm-specific characteristics have different effects on the wages of male and female employees.[1] Heinze and Wolf (2006) first provide a comprehensive study on the effects of various firm characteristics and the institutional framework on the gender wage gap in Germany by looking at within-firm gender wage differentials. Meng (2004) and Meng and Meurs (2004) extend the traditional decomposition of the observed wage gap in an endowment and a remuneration effect to an additional firm effect. In this setting, the firm effect represents the difference between the firm's premiums paid to male and female employees and can be interpreted as employer discrimination. In a second step, the impact of firm characteristics on this discrimination term is determined. This decomposition is implemented at the conditional mean, though.

The contribution of this paper is two-fold. First, I propose to extend the traditional Oaxaca-Blinder decomposition to disentangle the effect of human capital characteristics (afterwards denoted as

[1] Reilly and Wirjanto (1999) as well as Datta Gupta and Rothstein (2005) include both personal and establishment-level information to point out the effect of segregation on the earnings differences between men and women in Canada and Denmark. Drolet (2002) investigates how much of the Canadian pay gap can be attributed to specific workplace characteristics, such as high-performance workplace practices or training expenditures. Datta Gupta and Eriksson (2004) analyze the relationship between new workplace practices and the GWG. Simón and Russell (2005) analyze the GWG in a set of EU countries with a cross-national survey of matched employer-employee data. They show that workplace characteristics are very relevant in explaining wage differences between males and females in all countries.

individual characteristics) and the effect of firm characteristics in explaining the gender wage gap. This decomposition results in four terms: a part attributable to differences in the human capital characteristics, a part attributable to differences in the returns to human capital characteristics, a part that captures differences in firm-specific characteristics as well as a part that results difference in the returns to these characteristics. As a second contribution, this decomposition is implemented for the whole wage distribution. Quantile regressions are used to estimate the returns to the different characteristics at each percentile. Combined with a bootstrap method, these estimates allow for determination the four parts of the extend Oaxaca-Blinder decomposition for each percentile.

The remainder of the paper is organized as follows: Section 2 briefly discusses the literature of decomposing the gender wage gap throughout the wage distribution. The econometric methodology is expounded in Section 3. Section 4 describes the data source and in the following section, the results are presented. Section 6 concludes.

# 2 Background

While the mean gender wage gap has been extensively studied in the labour economics literature, attention has shifted only relatively recently has attention shifted to investigating the degree to which the gender gap might vary across the wages distribution. Blau and Kahn (1996, 1997) explained the international differences in female wage deficiency and their evolution in time using the methodology proposed by Juhn et al. (1993). This methodology allowed them to take into account the role played by the wage structure in the explanation of a wage inequality. Fortin and Lemieux (1998) decomposed at various wage percentiles changes in the US gender wage gap using rank regressions. Bonjour and Gerfin (2001) applied the methodology proposed by Donald et al. (2000) to decompose the wage gap in Switzerland. Most recently, other papers have used quantile regressions in order to decompose the gender wage gap at different points of the wage distribution. García et al. (2001) point out that considering the return to unobserved characteristics is important to measure the discrimination. They propose to use quantile regressions in order to compare quantiles of male and female wage distributions conditional on the same set of characteristics as an approximation to the returns to unobserved and observed characteristics. Their decomposition of the Spanish gender wage gap evaluates the vector of characteristics of men and women at the same point, the unconditional mean, regardless of which quantile is considered. Gardeazabel and Ugidos (2005) state that it might be considered more appropriate to weight the difference in returns to a certain characteristic (for example primary education) at a given quantile according to the proportion of individuals with this characteristic at that quantile. Based on this methodological approach, their findings at the Spanish wage gap contradict the results by García et al. (2001). While in the analysis of García et al. (2001) the part of the gender wage gap attributed to the different returns to characteristics increases across the wagedistribution, Gardeazabel and Ugidos (2005) find a the opposite.

By considering only the mean of the regressors, however some important factors explaining the difference between two distributions are neglected. Assume for example, that the sample means of the covariates are the same for males and females, but the variance is much higher for males. Then, ceteris paribus, the distribution of the dependent variable will also have a higher variance for males but this difference can not be analysed with the decomposition above. Machado and Mata (2005) (MM) propose an alternative decomposition procedure which combines a quantile regression and a bootstrap approach in order to estimate counterfactual density function. For the first time Albrecht et al. (2003) applied this method to decompose the gender wage gap in Sweden. They show that the gender wage gap in Sweden increases throughout the wage distribution and accelerates in the upper tail. The authors interpret this as a strong glass ceiling effect. The increasing pattern persists to a considerable extent after controlling for gender difference in characteristics. Using the same estimation strategy, de la Rica et al. (2005) show that the gender wage gap in Spain is much flatter than in Sweden. However, this pattern hides a composition effect when the sample is split by education. There is also a glass ceiling effect for the individuals with high educational attainment. By contrast, the gender wage gap decreases across the wage distribution for workers with low education. Albrecht et al. (2004) investigate the gender wage gap in the Netherlands with the MM decomposition method taking into account selection of women in a full time employment. Thus the authors purpose this to make statements for all employed women regardless of their employment status. Applying also the MM decomposition method Arulampalam et al. (2006) explore the wage differential for the eleven European countries. Their results show a u-shaped raw wage gap for Germany. However, in the private sector the gender wage gap is wider at the bottom end. They interpret this as sticky floor effect in contrast to the glass ceiling effect. Besides Beblo et al. (2003) this is the only analysis of the gender wage gap across the wage distribution in Germany. These studies primarily focus upon on the differences in individual characteristics, though.

# 3 Methodology

## *3.1 Wage Regression*

OLS and most estimation approaches focus on the mean effects. They restrict the effect of covariates to operate as a simple “location shift”. The quantile regression model introduced by Koenker and Bassett (1978) is more flexible than OLS and allows for studying effects of covariates on the whole distribution of the dependent variable. There is a rapidly expanding empirical quantile regression (QR) literature. Fitzenberger et al. (2001) and Koenker and Hallock (2001) have surveyed it.

Let $w_i$ denote the log wage of worker *i*, $X_i$ a vector of covariates representing the individual characteristics and $Z_i$ a vector of covariates representing the firm characteristics. The statistical model specifies the $\theta th$ quantile of the conditional distribution of $w_i$ given $X_i$ and $Z_i$ as a linear function of the covariates,

$$Q_\theta\left(w_i \middle| X_i, Z_i\right) = X_i\beta_\theta + Z_i\delta_\theta, \quad \theta \in (0,1). \tag{1}$$

As shown by Koenker and Bassett (1978), the quantile regression estimators of $\beta_\theta$ and $\delta_\theta$ solve the following minimization problem

$$\begin{bmatrix} \hat{\beta}_\theta \\ \hat{\delta}_\theta \end{bmatrix} = \arg\min_{\beta,\delta} \left[ \sum_{i:w_i \geq X_i\beta + Z_i\delta} \theta \left| w_i - X_i\beta - Z_i\delta \right| + \sum_{i:w_i < X_i\beta + Z_i\delta} (1-\theta) \left| w_i - X_i\beta - Z_i\delta \right| \right]. \tag{2}$$

This minimization problem can be transferred into a GMM framework which has been used to prove consistency and asymptotic normality of the estimators as well as to find its asymptotic covariance matrix (Buchinsky 1998).[2]

Since the wage data at use are censored from above at the social security taxation threshold $c_s$, one observes only $\tilde{w}_{s,i} = \min\left\{w_{s,i}, c_s\right\}$. Powell (1984, 1986) developed censored quantile regressions as a robust extension to the censored regression problem. In the case of censoring from above the minimization problem is extended to

$$\min_{\beta,\delta} \left[ \sum_{i:w_i \geq X_i\beta + Z_i\delta} \theta \left| \tilde{w}_i - \min\left\{X_i\beta - Z_i\delta, c_s\right\} \right| + \sum_{i:w_i < X_i\beta + Z_i\delta} (1-\theta) \left| \tilde{w}_i - \min\left\{X_i\beta - Z_i\delta, c_s\right\} \right| \right]. \tag{3}$$

There are different algorithms to solve the non-convex optimization problem in the literature, see Buchinsky (1994), Fitzenberger (1997a, 1997b), or Koenker and Park (1996). In order to get the best estimation it is necessary to test different starting values. Because of the limited access to the data[3] and the amount of data it is not possible to implement censored quantile regressions. Alternatively, I apply quantile regressions after imputation of uncensored wage data. As described in the next section, right-censored observations are replaced by wages randomly drawn from a truncated normal distribution whose moments are constructed by the predicted values from the Tobit regressions and whose (lower) truncation point is given by the contribution limit to the social security system. In the Tobit regression model the same exogenous variable are used as in quantile regression model.

Heteroscedasticity consistent standard errors can be obtained by means of the design matrix bootstraps. Again, because of the limited access to the data, I do not calculate consistent standard

---

[2] Although the estimator in (2) is consistent and asymptotically normal, it is not efficient. An efficient estimator requires the use of ant estimator for the unknown density function $f_{u\theta}(0|X,Z)$

[3] The limited access to the data means the data are only available at Research Data Centre (FDZ) of the Federal Employment Services (BA) at the Institute for Employment Research (IAB) in Nuremberg. It is only possible to work with data there and the places for visiting sojourn are limited.

errors. After all, the focus of this analysis is a decomposition gender wage gap interpreting the results of the single quantile regressions.

## 3.2 Decomposition

The above regression analysis provides detailed insights into remuneration of observed worker and firm characteristics for men and women and for different parts of the wage distribution. In general, decomposition analyses are well-suited to complement the regression evidence by answering the question whether differences in observed distributions result from differences in estimated coefficients or from difference in the composition of the workforce. In an Oaxaca (1973) and Blinder (1973)-type (OB) decomposition the gender wage gap are evaluated at the average characteristics of male (m) and female (f) employees:

$$\overline{w}_m - \overline{w}_f = \left(\overline{X}_m - \overline{X}_f\right)\hat{\beta}_m - \overline{X}_m\left(\hat{\beta}_m - \hat{\beta}_f\right), \tag{4}$$

where $\overline{w}_j$ is the mean of the log wage, $\overline{X}_j$ the vector of average characteristics o employees and $\hat{\beta}_j$ the estimated vector of returns to the characteristics. The first term on the right hand side of equation (4) shows the difference in characteristics and the second term states the difference in the estimated coefficients. In order to distinguish between individual (X) and firm characteristics (Z) I extend the Oaxaca-Blinder (OB) decomposition in the following way:

$$\overline{w}_m - \overline{w}_f = \underbrace{\left(\overline{X}_m - \overline{X}_f\right)\hat{\beta}_m}_{(i)} + \underbrace{\overline{X}_m\left(\hat{\beta}_m - \hat{\beta}_f\right)}_{(ii)} + \underbrace{\left(\overline{Z}_m - \overline{Z}_f\right)\hat{\delta}_m}_{(iii)} + \underbrace{\overline{Z}_m\left(\hat{\delta}_m - \hat{\delta}_f\right)}_{(iv)} \tag{5}$$

The first and the third term on the right hand side show the difference in the individual characteristics (i) and the in firm characteristics (iii). The second and the fourth term state the difference in the remuneration for the individual (ii) and the firm characteristics (iv). When decomposing the gender wage gap in this paper, I choose the counterfactual $X_f\beta_m$ and $Z_f\delta_m$ respectively to answer the question what the log wage would have been, had a population with the same distribution of characteristics as female employees faced returns to characteristics as male employees.[4] The approach assumes that the male returns are the relevant benchmark for the distribution in the absence of any "discrimination".

This approach considers only differences at the mean of the two earnings distributions. As mentioned in section 2, there is evidence that the decomposition for the average wage gap is not representative of the gap between different quantiles of the wage distribution. Garcia et al. (2001) suggest to combine the decomposition technique with quantile regressions to determine the rent component at various points of the wage distribution. The disadvantage of this approach is that they only consider the mean

[4] It is well known that the partition depends on the ordering of the effects and that the decomposition results my not be invariant with respect to the choice of the involved counterfactual. See the surveys of Oaxaca and Ransom (1994) and Silber and Weber (1999). Therefore, the choice of a counterfactual should be guided by the questions of economic interest.

of the covariates. Differences in higher moments of the distribution of the independent variables are not controlled for.

Machado and Mata (2005) introduce an alternative decomposition procedure which combines a quantile regression model with a bootstrap approach. In a first step, the conditional quantiles of $w$ are given by equation (1) and can be estimated by quantile regressions. The second idea underlying their technique is the probability integral transformation theorem from elementary statistics: If U is uniformly distributed on $\left[0,1\right]$, then $F^{-1}\left(U\right)$ hast distribution $F$. Thus, for given $\left[X_i : Z_i\right]$ and a random $\theta \sim U\left[0,1\right]$, $X_i\beta_\theta + Z_i\delta_\theta$ has the same distribution as $w_i \left| X_i, Z_i\right.$. If $\left[X : Z\right]$ are randomly drawn from the population, instead of keeping $\left[X_i : Z_i\right]$ fixed, $X\beta_\theta + Z\delta_\theta$ has the same distribution as $w$. In order to save computation time I apply a simplification of the MM techniques as suggested in Albrecht et al. (2003). Formally, the estimation procedure involves four steps:

1. Estimate for male and female employees quantile regression coefficients for each single percentile: $\begin{pmatrix}\hat{\beta}^m_\theta \\ \hat{\delta}^m_\theta\end{pmatrix}, \begin{pmatrix}\hat{\beta}^f_\theta \\ \hat{\delta}^f_\theta\end{pmatrix}; \quad \theta = 1,...,99.$
2. Generate the following samples of size M=10000 with replacement from the covariates of $\left[X : Z\right]$ for each estimated coefficient vector: $\left\{\tilde{X}^m_i : \tilde{Z}^m_i\right\}_{i=1}^M$; $\left\{\tilde{X}^f_i : \tilde{Z}^f_i\right\}_{i=1}^M$; $\left\{\tilde{X}^f_i : \tilde{Z}^m_i\right\}_{i=1}^M$
3. Calculate $\left\{\tilde{w}^m_i = \tilde{X}^m_i \hat{\beta}^m_\theta + \tilde{Z}^m_i \hat{\delta}^m_\theta\right\}_{i=1}^M$ and $\left\{\tilde{w}^f_i = \tilde{X}^f_i \hat{\beta}^f_\theta + \tilde{Z}^f_i \hat{\delta}^f_\theta\right\}_{i=1}^M$ for each estimated coefficient vector. These data sets are random samples of size $M \times 99$ from the marginal wage distributions of $w$ consistent with the linear model in (1).
4. Generate the following random sample of the counterfactual distributions with the estimated coefficients of each percentile:

   $\left\{\tilde{w}^1_i = \tilde{X}^f_i \hat{\beta}^m_{\theta_i} + \tilde{Z}^m_i \hat{\delta}^m_{\theta_i}\right\}_{i=1}^M$, $\left\{\tilde{w}^2_i = \tilde{X}^f_i \hat{\beta}^f_{\theta_i} + \tilde{Z}^m_i \hat{\delta}^m_{\theta_i}\right\}_{i=1}^M$ and $\left\{\tilde{w}^3_i = \tilde{X}^f_i \hat{\beta}^f_{\theta_i} + \tilde{Z}^f_i \hat{\delta}^m_{\theta_i}\right\}_{i=1}^M$

   $\tilde{w}^1$ states the hypothetical log wage for female employees if they had the firm characteristics of the male employees and they had been paid as male employees. $\tilde{w}^2$ is the hypothetical log wage for female employees if they had the firm characteristics of the male employees and only those characteristics had been paid like male employees. Finally, $\tilde{w}^3$ is a hypothetical log wage for female employees if their firm characteristics had been paid as for male employees.

   The first counterfactual wage distribution would have prevailed for female employees if their firm characteristics had been distributed similar to male employees.

Now I can decompose the gender wage gap into the contribution of the individual and firm characteristics as well as the contribution of the returns to individual and firm characteristics. Machado and Mata analyze the changes in the wage densities. In order to simplify the comparison to the QB-decomposition, I will decompose the quantiles of the wage distribution:

$$\begin{aligned} Q_\theta\left(w^m\right)-Q_\theta\left(w^f\right) &= \underbrace{\left[Q_\theta\left(\tilde{w}^m\right)-Q_\theta\left(\tilde{w}^1\right)\right]}_{(i)}+\underbrace{\left[Q_\theta\left(\tilde{w}^1\right)-Q_\theta\left(\tilde{w}^2\right)\right]}_{(ii)} \\ &+\underbrace{\left[Q_\theta\left(\tilde{w}^2\right)-Q_\theta\left(\tilde{w}^3\right)\right]}_{(iii)}+\underbrace{\left[Q_\theta\left(\tilde{w}^3\right)-Q_\theta\left(\tilde{w}^f\right)\right]}_{(iv)}+R \end{aligned} \qquad (6)$$

Analogue to (5), the first term is the contribution of the individual characteristics and the third term is the contribution of the corresponding coefficients to the difference between the $\theta th$ quantile of the male wage distribution and the $\theta th$ quantile of the female wage distribution. The second term refers to the contribution of the firm characteristics and the fourth term is the contribution of the corresponding coefficients. The last term is a residual term. It includes sampling errors which disappear with more observations, simulation errors which disappear with more simulations and specification errors by estimating a linear quantile regression. Assuming the correct specification, the residual term asymptotically tends to zero and (6) is a true decomposition of the gender wage gap in quantiles.

# 4 Data

The present paper uses a representative German employer – employee linked data set which is a combination of two separate data sets. The first data set, the *IAB Establishment Panel*, is an annual survey of West-German establishments administered since 1993.[5] The database is a representative sample of German establishments employing at least one employee who pays social security contributions. During the time of analysis about 84% of all employed persons in Germany are covered by the social security system. The survey was administered through personal interviews and provides general information on the establishment, such as, for example, investment, revenues, the size and composition of their work forces, salaries and wages.

The second data set, the so-called *Employment Statistics Register*, is an administrative register data set of all employees in Germany paying social security contributions.[6] The data set is based on the notifying procedure for the health insurance, statutory pension scheme and unemployment insurance, which was introduced in 1973. In order to comply with legal requirements, employers have to provide information to the social security agencies for all employees required to pay social security contributions. These notifications are required for the beginning and ending of any employment relationship. In addition, employers are obliged to provide an annual report for each employee covered by social insurance who is employed on the 31st December of each year. Due to its administrative

[5] Detailed information on the *IAB Establishment Panel* is given by Bellmann et al. (1994), Bellmann (1997) and Kölling (2000).
[6] Information on the Employment Statistics Register is given by Bender et al. (1996, 2000).

nature, this database has the advantage of providing reliable information on the daily earnings that are subject to social security contributions.

The construction of the Linked Employer-Employee data set occurs in two steps: First, I select establishments from the establishment panel data set. From the available waves 1993 to 2003, I use the year 2002, since the estimation procedure does not allow for more observations and the information from the matched individuals were not completed for the year 2003. I exclude firms from East Germany and non-profit firms. Furthermore, I only consider firms with a least 10 employees.

In the second step, the establishment data are merged with notifications for all employees who are employed by the selected establishments on 30th June of each year. From the worker data I drop observations for apprentices, part-time workers and homeworkers. In order to avoid modelling human capital formation and retirements decisions, I exclude individuals younger than 20 and older than 60. Since I consider only full-time workers, I also eliminate those whose wage is less than twice the lower social security contribution limit and employees with more than one employment. The final sample comprises 477160 male and 124488 female employees in 4021 establishments.[7]

The individual data include information on the daily wage, age, gender, nationality, employment status, education[8] and the date of entry into the establishment. The latter is used to approximate tenure by subtracting the entry date from the ending date of the employer's notification which is also available in the worker data. Table 1 presents summary statistics for the individual variables used in the subsequent analysis. The choice of the individual characteristics is limited to the typical variables of a Mincer wage equation. The set contains formal skill dummies, age, age squared, job tenure as well a dummy for foreigners. The summary statistic shows that, on average, women have lower educational attainments and lower job tenures than male employees. Comparing male and female employees in the bottom of the wage distributions shows, however, that more women have higher educational attainments than men (see table A1 in the appendix).

[Table 1 here]

When choosing the establishment variables I confine myself to variables which have been shown to affect the wage level as well as the wage distribution (see e.g. Davis and Haltiwanger 1991; Bronars and Famulari 1997; Abowd et al. 1999). First the vector of firm characteristics includes variables indicating the workforce of the establishment. These are total employment and squared of total employment, the employment share of females as well as the share of highly qualified employees. Furthermore, I take into account variables describing the revenue and production situation. This encompasses the wage bill and the sales per employee, the share of exports on total sales, two dummy variables indicating whether the revenues of the establishment increased or decreased during the last year, a discrete choice variable indicating the state-of-the-art of the production technology used in the establishment, the number of the average agreed working hours at as well as a dummy variables

[7] Note that the exclusion of certain individual groups entails a loss of 229 establishments.
[8] The categories are: No degree, vocational training degree, high school degree (Abitur), high school degree and vocational training, technical college degree and university degree.

indicating whether the establishment has been found after 1989 and 10 industry dummies. Finally, I consider also the institutional environment by including a dummy variable indicating whether the firm is covered by an industry-wide or firm-specific wage agreement. In addition, I include a  and dummy variable for the existence of works council. The descriptive statistics of these variables are given in Table 2.

[Table 2 here]

The dependent variable in the subsequent analysis will be the real gross daily wage. Since there is an upper contribution limit to the social security system, gross daily wages are top-coded. In my sample, top-coding affects 10.9 per cent of all observations. While in the subsample of the male employees the wage is censored above the $86^{th}$ quantile of the male wage distribution, the censoring appears above the $95^{th}$ quantile of the female wage distribution. To address this problem, a tobit regression is estimated for each gender with log daily wages as the dependent variable and individual and establishment covariates as explanatory variables (see Table 3). As described in Gartner (2005), right-censored observations are replaced by wages randomly drawn from a truncated normal distribution whose moments are constructed by the predicted values from the Tobit regressions and whose (lower) truncation point is given by the contribution limit to the social security system.

[Table 3 here]

# 5 The empirical results

## 5.1 The gender wage gap

The usual procedure for measuring the male-female wage gap is to consider the difference between the average mal wage and its female counterpart. In my sample, the average log male wage after imputation is 4.6302 whereas the female log wage is 4.3894. Therefore, the male-female average wage differential is 0.2408.

Figure 1 shows nonparametric estimates of the density functions of male and female (log) wages. The male wage density is placed rightward with respect to the female wage distribution, indicating a non negligible gender wage gap. The gender wage gap is better viewed in Figure 2, which shows the empirical cumulative density function of male and female (log) wages. The horizontal distance between the two functions is the gender wage gap at that quantile. Figure 3 plots the raw gender wage gap as a function of the quantile index. The gender gap is sharply decreasing within the first three deciles, then the decrease decelerates until the 70th percentile, then increases up to 80th percentile, and from then on the gap is decreasing. The gender wage gap is far from being constant within the wage distribution.

[Figure 1 here]

[Figure 2 here]

[Figure 3 here]

## 5.2 Regression Results

The estimated log wage equations include a set of individual characteristics and a set of characteristics of the firm in which the individuals are employed. Separate earnings equations for male and female employees have been estimated using standard OLS and quantile regressions. The vector of regressors is described in section 3. Table 4 and Table 5 show the OLS coefficients with their standard errors and the coefficient estimated for quantile regressions at subset deciles of the distributions[9].

[Table 4 here]

[Table 5 here]

All estimated effects in the OLS regressions are significantly different from zero. The individual variables have the expected effect on the wage for both male and female employees. That is, the wage increases with the education level, age indicating potential experience and job tenure indicating job specific human capital. The comparison of the male and female OLS coefficients shows that the effects of the individual characteristics are slightly smaller for female employees. Moreover, the estimated QR coefficients for the individual characteristics generally vary across the distribution and differ from the OLS estimates regarding the sizes but not regarding the signs. The effects estimated by QRs are also smaller for female employees than for male workers.

Turning to the establishment variables show that the wage rate increases with the number of employees (but with a decreasing rate) and with the share of highly qualified employees for both men and women. The impact of the wage bill and sales per employee is also positive. Furthermore, good results in the last year lead also to an increasing wage rate. The firm characteristics indicating the institutional environment (wage agreement and works council) have a strong positive effect which is stronger for female employees than for male workers. The share of female employees affects the wage rate negatively whereas this effect is also stronger for female employees. Note, that in high quantiles the effect of this variable is positively for male employees. Moreover, the estimated QR coefficients for the establishment characteristics generally vary across the distribution. Thus the impact of the institutional variables decreases across the wage distribution for both male and female employees. It seems that unions and works council rather support employees at the lower tail of the wage distribution.

## 5.3 Results of the decomposition

[9] The results for the other percentiles are available upon request from the author.

Table 6 present the results of the OB decomposition and a preliminary version of decomposition in quantiles. The implemented OB decomposition is equal to (5). So far the decomposition is not the described MM decomposition in (6). I decompose the differences in the $\theta th$ quantile of the log wage distribution between the men and women as follow:

$$Q_\theta(w_i^m) - Q_\theta(w_i^f) = \underbrace{\left(\bar{X}_i^m - \bar{X}_i^f\right)\hat{\beta}_\theta^m}_{(i)} + \underbrace{\bar{X}_i^f\left(\hat{\beta}_\theta^m - \hat{\beta}_\theta^f\right)}_{(ii)} + \underbrace{\left(\bar{Z}^m - \bar{Z}^f\right)\hat{\delta}_\theta^m}_{(iii)} + \underbrace{\bar{Z}^f\left(\hat{\delta}_\theta^m - \hat{\delta}_\theta^f\right)}_{(iv)} + residual \quad (7)$$

where $Q_\theta\left(w_j\right)$ is the $\theta th$ empirical quantile of the wage distribution, $\bar{X}_j$ the vector of average individual characteristics of employees, $\bar{Z}_j$ the vector of average firm characteristics of employees, $\hat{\beta}_j$ the estimated vector of returns to the individual characteristics and $\hat{\delta}_j$ the estimated vector of returns to the firm characteristics. The first term on the right-hand side is the component of the log wage differential due to difference in the human capital endowments between male and female employees. The second term shows the component due to difference in the returns to these human capital endowments**.** The third term presents the part attributed to the difference in the firm characteristics. Finally, the fourth term is the part attributable to the difference in the returns to the firm characteristics. Note, that if the properties of the OLS estimators ensure that the predicted log wage evaluated at the sample average vector of characteristics is equal to the sample average log wage, the estimators for the quantile regression model do not have any comparable properties. That is, the conditional quantile evaluated at the mean of the covariates is not equal to the unconditional quantile. Thus a residual term occurs in (7).

The OB decomposition on the basis of the estimated OLS regressions shows that the largest part of the observed mean wage gap is explained by the difference in the returns to the firm characteristics. By contrast, the differences in returns to individual characteristics are the smallest part of the gap.

The OLS regression does not consider the entire wage distribution. The quantile regression is a more informative approach. The decomposition using the estimated quantile regression coefficients shows that the parts due to difference in the returns to the characteristics vary strongly with $\theta$. There is a male wage premium for the firm characteristics across the whole distribution while the part attributed to the difference in the returns to the individual characteristics shows that female employees are paid better for their individual characteristics between the 23th and 72th percentile.

The characteristics components show that male employees should earn more than the female employees at all points of the wage distributions. Compared with female employees, men are better educated and have more years of job tenure and are older.

The decomposition method described above evaluates the conditional quantiles at the covariates sample mean. Because of this, it makes the decomposition very similar to the OB decomposition. As proved in Koenker and Bassett (1982), $\theta_1 < \theta_2 \rightarrow \bar{X}\hat{\beta}_{\theta_1} < \bar{X}\hat{\beta}_{\theta_2}$, while the monotonicity of the conditional quantiles evaluated at another point is not guaranteed. However, I neglect some important factors explaining the difference between the distributions if I consider only the mean of the covariates. If, for example, the means of the regressors are the same for male and female employees, but the variance is higher in the group of the men, then ceteris paribus the wage distribution will also have a higher variance for males. However, this decomposition type is not able to analyze this pattern. An alternative is the MM decomposition. RESULTS come later!

# 6 Conclusion

The differences in the wage distribution between men and women have been decomposed into an explained and unexplained component with respect to human capital characteristics as well into an explained and unexplained part with respect to firm characteristics using different methodologies. The first is the simple OB decomposition using OLS estimates. The limit of this technique is that only the mean differences are considered. The second is a simple combination of the OB decomposition and quantile regression. The problem with this technique is that only one point in the covariates distribution is taken into account: the mean. Differences in higher moments of the distributions of the independent variables are not controlled for. The third decomposition method, proposed by Machado and Mata (2005), combine a quantile regression and a bootstrap approach to stochastically simulate counterfactual wage distributions.

The wage structure of male and female employees in the private sector in West Germany has been examined using data from the LIAB, a representative German employer – employee linked data set, for the year 2002.

The unconditional gender gap is sharply decreasing within the first three deciles of the wage distribution, then the decrease decelerates until the 70th percentile, then increases up to 80th percentile, and from then on the gap is decreasing. The gender wage gap is far from being constant within the wage distribution.

The decomposition using the estimated quantile regression coefficients shows that the parts due to difference in the returns to the characteristics vary strongly with $\theta$. There is a male wage premium for the firm characteristics across the whole distribution while the part attributed to the difference in the returns to the individual characteristics shows that female employees are paid better for their individual characteristics between the 23th and 72th percentile. The characteristics components show that male employees should earn more than the female employees at all points of the wage distributions. Compared with female employees, men are better educated and have more years of job tenure and are older.

## References

Abowd, J. M., Kramarz, D. and D. N. Margolis (1999): High wage workers and high wage firms, *Econometrica*, Vol. 67(2): 251-333.

Albrecht, J., A. Björklund and S. Vroman (2003): Is there a Glass Ceiling in Sweden? J*ournal of Labor Economics*, Vol. 21(1), 145-177.

Albrecht, J., A. van Vuuren and S. Vroman (2004): Decomposing the Gender Wage Gap in the Netherlands with Sample Selection Adjustments, IZA-DP 1400.

Arulampalam, W., A.L. Booth and M.L. Bryan (2006): Is There a Glass Ceiling over Europe? Exploring the Gender Wage Gap across the Wages Distribution, mimeo.

Beblo, M. D. Beninger, A. Heinze and F. Laisney (2003): Measuring Selectivity-Corrected Gender Wage Gaps in the UE, ZEW-DP No. 03 – 74, ZEW, Mannheim.

Bellmann, L. (1997): The IAB Establishment Panel with an Exemplary Analysis of Employment Expectations, IAB-Topics No. 20, IAB, Nürnberg.

Bellmann, L., S. Kohaut and J. Kühl (1994): Enterprise Panels and the Labour Market: Using Enterprise Panels to Meet the Needs of the White Paper, in: E. Ojo (ed.): *Enterprise Panels and the European Commission's White Paper*. Luxenbourg: Eurostat, 57-74.

Bender, S. A. Haas and C. Klose (2000): IAB Employment Subsample 1975-1995. Opportunities for Analysis Provided by the Anonymised Subsample, IZA DP No. 117, IZA, Bonn.

Bender, S., J. Hilzendegen, G. Rohwer and H. Rudolf (1996): Die IAB-Beschäftigtenstichprobe 1975-1990. Eine praktische Einführung, *Beiträge zur Arbeitsmarkt und Berufsforschung*, Vol. 197, IAB, Nürnberg.

Blau, F. and L.M. Kahn (1996): Wage Structure and Gender Earnings Differentials: An International Comparison, *Economica*, Vol. 63: 29-62.

Blau, F. and L.M. Kahn (1997): Swimming Upstream: Trends in the Gender Wage Differential in the 1980s, *Journal of Labor Economics*, Vol. 15(1): 1:42.

Blinder, A. (1973): Wage Discrimination: Reduced Form and Structural Estimates. J*ournal of Human Resources*, Vol. 8: 436-455.

Bonjour, D. and M. Gerfin (2001): The Unequal Distribution of Unequal Pay – An Empirical Analysis of the Gender Wage Gap in Switzerland, *Empirical Economics*, Vol. 26: 407-427.

Bronars, S. G. und M. Famulari (1997): Wage, tenure, and wage growth variation within and across establishments, *Journal of Labor Economics*, Vol. 15(2), 285 – 317.

Buchinsky, M. (1994): Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regressions, *Econometrica*, Vol. 62(2): 405-458.

Buchinsky, M. (1998): Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research, *Journal of Human Resources*, Vol. 33: 88-126.

Datta Gupta, N. and D.S. Rothstein (2005): The impact of worker and establishment-level characteristics on male-female wage differentials: Evidence from Danish matched employee-employer data, *Labour: Review of Labour Economics and Industrial Relations*, Vol. 19(1): 1- 34.

Datta Gupta, N. and T. Eriksson (2004): New workplace practices and the gender wage gap, Working Paper, 04-18, Aarhus.

Davis, S. J. and J. Haltiwanger (1991): Wage dispersion between and within US manufacturing plants, 1963 – 1986, Brookings papers on economic activity: 115 – 180.

De la Rica, S., J.J. Dolado and V.Llorens (2005): Ceiling and Floors: Gender Wage Gaps by Education in Spain, IZA-DP 1483.

Donald, S.G., D.A. Green and H.J. Paarsch (2000): Differences in Wage Distribution Between Canada and the United States: An Application of a Flexible Estimator of Distributions Functions in the Presence of Covariates, *Review of Economic Studies*, Vol. 67: 609-633.

Drolet, M. (2002): Can the workplace explain Canadian gender pay differentials? *Canadian Public Policy*, Supplement May 2002, Vol. 28: 41-63.

Fitzenberger, B. (1997a): A Guide to Censored Quantile Regressions, in Handbook of Statistics, ed. by G. S. Maddala and C. R. Rao, Vol. 15: Robust Inferences 405-437. Elsevier Science.

Fitzenberger, B. (1997b): Computational Aspects of Censored Quantile Regression, in $L_1$ - Statistical Procedure and Related Topics, ed. by Y. Dodge, Vol. 31 of IMS Lecture Notes – Monograph Series, 171-186. Institute of Mathematical Statistics, Hayward, CA.

Fitzenberger, B., R. Koenker and J.A.F. Machado (2001): Special Issue on Economic Applications of Quantile Regression, *Empirical Economics*, Vol. 26: 1-324.

Fortin, N.M. and T. Lemieux (1998): Rank Regressions, Wage Distributions and the Gender Wage Gap, *Journal of Human Resources*, Vol. 33: 610-643.

García, J., P.J. Hernández and A. López-Nicolás (2001): How Wide is the Gap? An Investigation of Gender Differences Using Quantile Regression, *Empirical Economics*, Vol. 26: 46-67.

Gardezable, J. and A. Ugidos (2005): Gender Wage Discrimination at Quantiles, *Journal of Population Economics*, Vol. 18: 165-179.

Gartner, H. (2005): The Imputation of Wages Above the Contribution Limit With the German IAB Employment Sample, FDZ Methodenreport 2/2005.

Heinze, A. and E. Wolf (2006): Gender Earnings Gap in German Firms: The Impact of Firm Characteristics and Institutions, ZEW-DP 06-020, ZEW, Mannheim.

Juhn, C. Murphy, K. and B. Pierce (1993): Wage Inequality and the Rise in Returns to Skill, *Journal of Political Economy*, Vol. 1001(3): 410-442.

Koenker, R. and B. J. Park (1996): An Interior Point Algorithm for Nonlinear Quantile Regression, *Journal of Econometrics*, Vol. 71: 265-283.

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

Koenker, R. and G. Bassett (1982): Robust Tests for Heteroskedasticity Based on Regression Quantilies, *Econometrica*, Vol. 50, 43-61.

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

Kölling, A (2000): The IAB Establishment Panel, *Applied Social Science Studies*, Vol. 120, No. 2, 291-300.

Machado, J.A.F and J. Mata (2005): Counterfactual Decomposition of Changes in Wage Distribution Using Quantile Regression, *Journal of Applied Econometrics*, Vol. 20: 445-465.

Meng, X. (2004): Gender earnings gap: the role of firm specific effects, *Labour Economics*, Vol. 11: 555-573.

Meng, X. and D. Meurs (2004): The gender earnings gap: effects of institutions and firms – a comparative study of French and Australian private firms, *Oxford Economic Papers*, Vol. 56: 189-208.

Oaxaca, R. (1973): Male-Female Wage Differentials in Urban Labor Markets, I*nternational Economic Review*, Vol. 14: 693-709.

Oaxaca, R. and M. R. Ransom (1994): On Discrimination and the Decomposition of Wage Differentials, *Journal of Econometrics*, Vol. 61: 5-21.

Powell, J.L. (1984): Least Absolute Deviations for the Uncensored Regression Model, *Journal of Econometrics*, Vol. 25: 303-325.

Powell, J.L. (1986): Censored Regression Quantiles, *Journal of Econometrics*, Vol. 32: 143-155.

Reilly, K.T. and T.S. Wirjanto (1999): Does more less? The male/female wage gap and proportion of females at the establishment level, *Canadian Journal of Economics*, Vol. 32, No.4: 906-929.

Silber, J. and M. Weber (1999): Labour Market Discrimination: Are There Significant Differences Between the Various Decomposition Procedure?, *Applied Economics*, Vol. 31: 359-365.

Simón, H. and H. Russell (2005): Firms and the gender pay gap: A cross-national comparison, Pay Inequalities and Economic Performance Working Paper, 15. February 2005, 1-43.

Table 1: Descriptive Statistic of Individual Characteristics

| Variables | Men | | Women | |
|---|---|---|---|---|
| | Mean | Std. Dev. | Mean | Std. Dev. |
| log daily wage (obs.) | 4.6008 | 0.2724 | 4.3795 | 0.3477 |
| log daily wage (imp.) | 4.6302 | 0.3226 | 4.3894 | 0.3680 |
| age | 40.8961 | 9.4491 | 39.1671 | 10.1378 |
| foreigner | 0.0983 | 0.2977 | 0.0885 | 0.2841 |
| low education without vocational training | 0.1518 | 0.3589 | 0.2378 | 0.4257 |
| vocational training | 0.6847 | 0.4646 | 0.6002 | 0.4899 |
| secondary school without vocational training | 0.0067 | 0.0817 | 0.0137 | 0.1161 |
| secondary school with vocational training | 0.0292 | 0.1685 | 0.0682 | 0.2521 |
| college of higher education | 0.0648 | 0.2462 | 0.0296 | 0.1695 |
| university | 0.0627 | 0.2424 | 0.0506 | 0.2192 |
| job tenure (in month)/100 | 1.3825 | 1.0130 | 1.1658 | 0.9575 |
| Observations | 477,160 | | 124,488 | |

Table 2: Descriptive Statistic of Firm Characteristics

| | Men | | Women | |
|---|---|---|---|---|
| Variables | Mean | Std. Dev. | Mean | Std. Dev. |
| number of employees/1000 | 2.4305 | 4.0604 | 1.7236 | 3.0120 |
| female quota (all employees) | 0.2071 | 0.1598 | 0.3980 | 0.2353 |
| quota of highly qualified employees (all employees) | 0.6770 | 0.2562 | 0.6419 | 0.2637 |
| business start-up after 1989 | 0.1463 | 0.3534 | 0.1471 | 0.3542 |
| export quota (sales) | 0.3096 | 0.2955 | 0.2515 | 0.2806 |
| wage bill per employee/1000 | 5.7900 | 2.0801 | 5.2848 | 2.3213 |
| sales per employee/100000 | 4.9998 | 13.4875 | 5.1867 | 19.3045 |
| good results last year | 0.3547 | 0.4784 | 0.3528 | 0.4778 |
| bad results last year | 0.2821 | 0.4500 | 0.2878 | 0.4528 |
| average results last year | 0.3632 | 0.4809 | 0.3594 | 0.4798 |
| technical state | 2.9735 | 0.7133 | 2.9948 | 0.7104 |
| industry-wide wage agreement | 0.7805 | 0.4139 | 0.7332 | 0.4423 |
| firm-specific wage agreement | 0.1110 | 0.3141 | 0.1018 | 0.3024 |
| no wage agreement | 0.1085 | 0.3110 | 0.1650 | 0.3712 |
| works council | 0.9152 | 0.2786 | 0.8713 | 0.3349 |
| agreed working hours per week | 36.7906 | 1.8830 | 37.2031 | 1.7786 |
| agriculture and forestry; electricity, gas and water supply, mining | 0.0358 | 0.1858 | 0.0234 | 0.1511 |
| manufacturing I | 0.2257 | 0.4180 | 0.1766 | 0.3813 |
| manufacturing II (reference) | 0.4967 | 0.5000 | 0.4189 | 0.4934 |
| construction | 0.0345 | 0.1825 | 0.0147 | 0.1204 |
| wholesale and retail trade | 0.0527 | 0.2234 | 0.1329 | 0.3395 |
| transport and communication | 0.0684 | 0.2524 | 0.0451 | 0.2075 |
| financial intermediation | 0.0012 | 0.0347 | 0.0009 | 0.0308 |
| real state, renting and business activities | 0.0518 | 0.2216 | 0.0686 | 0.2528 |
| education | 0.0029 | 0.0538 | 0.0062 | 0.0783 |
| other service activities | 0.0303 | 0.1714 | 0.1127 | 0.3163 |
| Berlin-West | 0.0426 | 0.2020 | 0.0579 | 0.2336 |
| Schleswig Holstein | 0.0492 | 0.2163 | 0.0584 | 0.2346 |
| Hamburg | 0.0570 | 0.2319 | 0.0500 | 0.2179 |
| Niedersachsen | 0.0796 | 0.2707 | 0.0718 | 0.2582 |
| Bremen | 0.0285 | 0.1664 | 0.0341 | 0.1815 |
| North Rhine-Westphalia (reference) | 0.2001 | 0.4001 | 0.1640 | 0.3703 |
| Hesse | 0.1341 | 0.3408 | 0.1350 | 0.3417 |
| Rhineland-Palatinate | 0.0463 | 0.2102 | 0.0540 | 0.2260 |
| Baden-Wurttemberg | 0.1354 | 0.3422 | 0.1658 | 0.3719 |
| Bavaria | 0.1660 | 0.3721 | 0.1713 | 0.3768 |
| Saarland | 0.0611 | 0.2395 | 0.0377 | 0.1904 |
| Observations | 477,160 | | 124,488 | |

Table3: Tobit regression

| Variables | Men | | Women | |
|---|---|---|---|---|
| | Coefficient | Standard Errors | Coefficient | Standard Errors |
| age | 0.0328** | 0.0003 | 0.0285** | 0.0006 |
| $(age)^2$ | -0.0326** | 0.0003 | -0.0313** | 0.0007 |
| foreigner | -0.0399** | 0.0011 | -0.0377** | 0.0028 |
| low education without vocational training | -0.1471** | 0.0010 | -0.1648** | 0.0020 |
| vocational training (reference) | - | - | - | - |
| secondary school without vocational training | 0.0879** | 0.0040 | 0.0441** | 0.0066 |
| secondary school with vocational training | 0.1935** | 0.0020 | 0.1232** | 0.0031 |
| college of higher education | 0.3707** | 0.0015 | 0.2933** | 0.0047 |
| university | 0.4435** | 0.0017 | 0.4014** | 0.0039 |
| job tenure (in month)/100 | 0.0362** | 0.0004 | 0.0502** | 0.0010 |
| number of employees/1000 | 0.0201** | 0.0003 | 0.0276** | 0.0008 |
| $(\text{number of employees}/1000)^2$ | -0.0008** | 0.0000 | -0.001** | 0.0001 |
| female quota (all employees) | -0.0074** | 0.0024 | -0.1051** | 0.0043 |
| quota of highly qualified employees (all employees) | 0.1151** | 0.0015 | 0.1932** | 0.0033 |
| business start-up after 1989 | 0.0371** | 0.0011 | 0.0393** | 0.0023 |
| export quota (sales) | 0.0082** | 0.0015 | 0.0375** | 0.0038 |
| wage bill per employee/1000 | 0.0245** | 0.0002 | 0.0294** | 0.0004 |
| sales per employee/100000 | 0.0006** | 0.0000 | 0.0002** | 0.0000 |
| good results last year | 0.0113** | 0.0008 | 0.0212** | 0.0019 |
| bad results last year | -0.0092** | 0.0009 | 0.0042** | 0.0019 |
| average results last year (reference) | - | - | - | - |
| technical state | 0.0145** | 0.0005 | 0.0096** | 0.0012 |
| industry-wide wage agreement | 0.0217** | 0.0013 | 0.0421** | 0.0025 |
| firm-specific wage agreement | 0.0061** | 0.0016 | 0.0168** | 0.003 |
| no wage agreement (reference) | - | - | - | - |
| works council | 0.0817** | 0.0014 | 0.1471** | 0.0027 |
| agreed working hours per week | -0.0120** | 0.0002 | -0.0195** | 0.0006 |
| observations | 477,160 | | 124,488 | |
| uncensored | 408,746 | | 118,211 | |
| right-censored | 68,414 | | 6,277 | |

Note: The dummy variables for regions and industries are also included in the estimation. The results are available on inquiry. ** significant on 5%-level, * significant on 10%-level.

Table 4: Results of the OLS and Quantile Regressions for Male Employees

| | OLS Regression | | Quantile Regression | | | |
|---|---|---|---|---|---|---|
| Variables | Coefficient | Std. Errors | θ = 0.1 | θ = 0.25 | θ = 0.5 | θ = 0.75 |
| age | 0.0330** | 0.0003 | 0.0269 | 0.0273 | 0.0285 | 0.0327 |
| $(age)^2$ | -0.0327** | 0.0003 | -0.0294 | -0.0289 | -0.0286 | -0.0303 |
| foreigner | -0.0406** | 0.0011 | -0.0161 | -0.0216 | -0.0305 | -0.0440 |
| low education without vocational training | -0.1488** | 0.0010 | -0.1030 | -0.1077 | -0.1262 | -0.1617 |
| vocational training (reference) | - | - | - | - | - | - |
| secondary school without vocational training | 0.0923** | 0.0040 | -0.0813 | 0.0222 | 0.1459 | 0.1880 |
| secondary school with vocational training | 0.1983** | 0.0019 | 0.0918 | 0.1572 | 0.2342 | 0.2391 |
| college of higher education | 0.3785** | 0.0014 | 0.4004 | 0.3958 | 0.3919 | 0.3688 |
| university | 0.4517** | 0.0014 | 0.4657 | 0.4685 | 0.4669 | 0.4400 |
| job tenure (in month)/100 | 0.0363** | 0.0004 | 0.0477 | 0.0422 | 0.0385 | 0.0294 |
| number of employees/1000 | 0.0200** | 0.0003 | 0.0254 | 0.0241 | 0.0230 | 0.0149 |
| $(\text{number of employees}/1000)^2$ | -0.0008** | 0.0000 | -0.0010 | -0.0010 | -0.0009 | -0.0006 |
| female quota (all employees) | -0.0047** | 0.0024 | -0.1118 | -0.0746 | -0.0155 | 0.0560 |
| quota of highly qualified employees (all employees) | 0.1159** | 0.0015 | 0.0903 | 0.0886 | 0.0987 | 0.0957 |
| business start-up after 1989 | 0.0371** | 0.0010 | 0.0220 | 0.0376 | 0.0511 | 0.0353 |
| export quota (sales) | 0.0089** | 0.0015 | 0.0106 | 0.0054 | -0.0010 | -0.0015 |
| wage bill per employee/1000 | 0.0248** | 0.0002 | 0.0223 | 0.0291 | 0.0329 | 0.0348 |
| sales per employee/100000 | 0.0006** | 0.0000 | 0.0007 | 0.0008 | 0.0009 | 0.0009 |
| good results last year | 0.0114** | 0.0008 | 0.0130 | 0.0110 | 0.0092 | 0.0136 |
| bad results last year | -0.0091** | 0.0009 | -0.0128 | -0.0088 | -0.0096 | -0.0087 |
| average results last year (reference) | - | - | - | - | - | - |
| technical state | 0.0146** | 0.0005 | 0.0125 | 0.0135 | 0.0124 | 0.0138 |
| industry-wide wage agreement | 0.0218** | 0.0012 | 0.0531 | 0.0371 | 0.0213 | 0.0061 |
| firm-specific wage agreement | 0.0062** | 0.0016 | 0.0136 | 0.0142 | 0.0089 | -0.0019 |
| no wage agreement (reference) | - | - | - | - | - | - |
| works council | 0 .0818** | 0.0014 | 0.1072 | 0.0875 | 0.0702 | 0.0614 |
| agreed working hours per week | -0.0121** | 0.0002 | -0.0137 | -0.0124 | -0.0110 | -0.0112 |
| Observations | 477,160 | | | | | |

Note: The dummy variables for regions and industries are also included in the estimation. The results are available on inquiry. ** significant on 5%-level, * significant on 10%-level.

Table 5: Results of the OLS and Quantile Regressions for Female Employees

| | OLS Regression | | Quantile Regression | | | |
|---|---|---|---|---|---|---|
| Variables | Coefficient | Std. Errors | θ = 0.1 | θ = 0.25 | θ = 0.5 | θ = 0.75 |
| age | 0.0287** | 0.0006 | 0.0120 | 0.0229 | 0.0321 | 0.0386 |
| $(age)^2$ | -0.0314** | 0.0007 | -0.0142 | -0.0270 | -0.0365 | -0.0420 |
| foreigner | -0.0378** | 0.0028 | -0.0154 | -0.0289 | -0.0347 | -0.0413 |
| low education without vocational training | -0.1656** | 0.0020 | -0.0867 | -0.1144 | -0.1581 | -0.2056 |
| vocational training (reference) | - | - | - | - | - | - |
| secondary school without vocational training | 0.0468** | 0.0066 | -0.0937 | -0.0169 | 0.0601 | 0.1122 |
| secondary school with vocational training | 0.1248** | 0.0031 | 0.0775 | 0.0878 | 0.1083 | 0.1280 |
| college of higher education | 0.2985** | 0.0046 | 0.2587 | 0.2833 | 0.3010 | 0.3022 |
| university | 0.4105** | 0.0037 | 0.3715 | 0.3996 | 0.4063 | 0.4239 |
| job tenure (in month)/100 | 0.0504** | 0.0010 | 0.0575 | 0.0535 | 0.0488 | 0.0409 |
| number of employees/1000 | 0.0276** | 0.0008 | 0.0394 | 0.0312 | 0.0251 | 0.0182 |
| $(\text{number of employees}/1000)^2$ | -0.001** | 0.0000 | -0.0016 | -0.0013 | -0.0010 | -0.0005 |
| female quota (all employees) | -0.1044** | 0.0043 | -0.0985 | -0.1080 | -0.1131 | -0.1050 |
| quota of highly qualified employees (all employees) | 0.1941** | 0.0034 | 0.2182 | 0.1624 | 0.1477 | 0.1425 |
| business start-up after 1989 | 0.0403** | 0.0023 | 0.0091 | 0.0182 | 0.0371 | 0.0546 |
| export quota (sales) | 0.0372** | 0.0038 | 0.0303 | 0.0312 | 0.0356 | 0.0318 |
| wage bill per employee/1000 | 0.0296** | 0.0004 | 0.0222 | 0.0357 | 0.0406 | 0.0448 |
| sales per employee/100000 | 0.0002** | 0.0000 | 0.0002 | 0.0000 | -0.0001 | 0.0002 |
| good results last year | 0.0212** | 0.0019 | 0.0160 | 0.0274 | 0.0259 | 0.0240 |
| bad results last year | 0.0043** | 0.0020 | -0.0123 | 0.0054 | 0.0084 | 0.0116 |
| average results last year (reference) | - | - | - | - | - | - |
| technical state | 0.0096** | 0.0012 | 0.0087 | 0.0083 | 0.0052 | 0.0041 |
| industry-wide wage agreement | 0.0423** | 0.0025 | 0.0831 | 0.0548 | 0.0407 | 0.0250 |
| firm-specific wage agreement | 0.0166** | 0.0034 | 0.0610 | 0.0248 | 0.0174 | 0.0042 |
| no wage agreement (reference) | - | - | - | - | - | - |
| works council | 0.1472** | 0.0028 | 0.2397 | 0.1712 | 0.1330 | 0.1072 |
| agreed working hours per week | -0.0195** | 0.0006 | -0.0256 | -0.0217 | -0.0193 | -0.0174 |
| Observations | 124,488 | | | | | |

Note: The dummy variables for regions and industries are also included in the estimation. The results are available on inquiry. ** significant on 5%-level, * significant on 10%-level.

Table 6: Decomposition

| Quantile | Obs. Gender Wage Gap | Diff. in individual characteristics (% of the obs. gap) | Diff. in returns to individual characteristics (% of the obs. gap) | Diff. in firm characteristics (% of the obs. gap) | Diff. in returns to firm characteristics (% of the obs. gap) |
|---|---|---|---|---|---|
| 10 | 0,3203 | 0,0456 | -0,0047 | 0,0638 | 0,1856 |
| | | (14,22%) | (-1,47%) | (19,90%) | (57,95%) |
| 20 | 0,2502 | 0,0433 | 0,0253 | 0,0537 | 0,1241 |
| | | (17,29%) | (10,13%) | (21,48%) | (49,59%) |
| 30 | 0,2280 | 0,0423 | -0,0190 | 0,0472 | 0,1600 |
| | | (18,56%) | (-8,35%) | (20,70%) | (70,17%) |
| 40 | 0,2191 | 0,0417 | -0,0317 | 0,0411 | 0,1706 |
| | | (19,04%) | (-14,45%) | (18,77%) | (77,84%) |
| 50 | 0,2168 | 0,0418 | -0,0361 | 0,0348 | 0,1771 |
| | | (19,26%) | (-16,66%) | (16,05%) | (81,68%) |
| 60 | 0,2150 | 0,0429 | -0,0344 | 0,0283 | 0,1785 |
| | | (19,96%) | (-15,99%) | (13,18%) | (83,03%) |
| 70 | 0,2234 | 0,0451 | -0,0143 | 0,0209 | 0,1655 |
| | | (20,19%) | (-6,38%) | (9,37%) | (74,09%) |
| 80 | 0,2376 | 0,0476 | 0,0670 | 0,0135 | 0,0955 |
| | | (20,01%) | (28,19%) | (5,67%) | (40,19%) |
| 90 | 0,1927 | 0,0498 | 0,1020 | 0,0066 | 0,0648 |
| | | (25,83%) | (52,91%) | (3,44%) | (33,64%) |
| 25 | 0,2327 | 0,0427 | -0,0109 | 0,0504 | 0,1549 |
| | | (18,33%) | (-4,70%) | (21,64%) | (66,56%) |
| 75 | 0,2303 | 0,0463 | 0,0246 | 0,0171 | 0,1322 |
| | | (20,09%) | (10,70%) | (7,41%) | (57,41%) |
| OB | 0.2408 | 0.0469 | 0.0291 | 0.0330 | 0.1318 |
| | | (19,48%) | (12.08%) | (13,70%) | (54,73%) |

Figure 1: Male and female wage densities

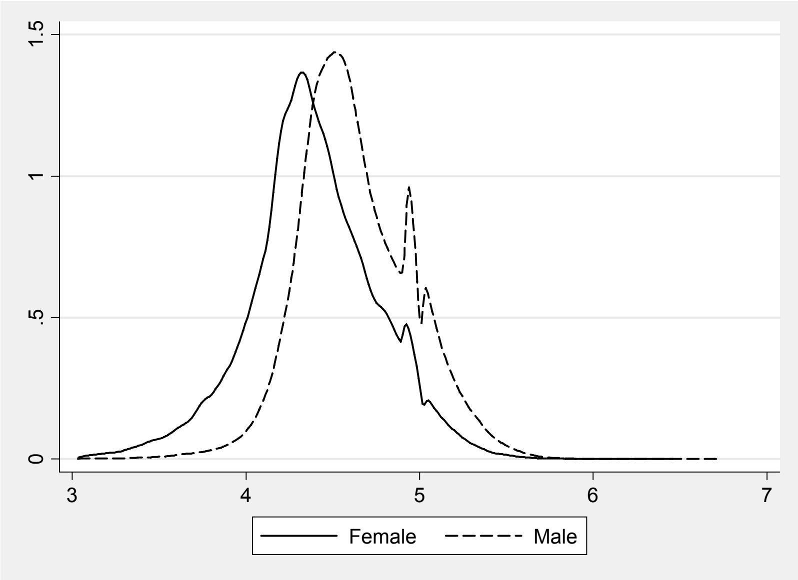



Figure 2: Male and female wage distribution functions

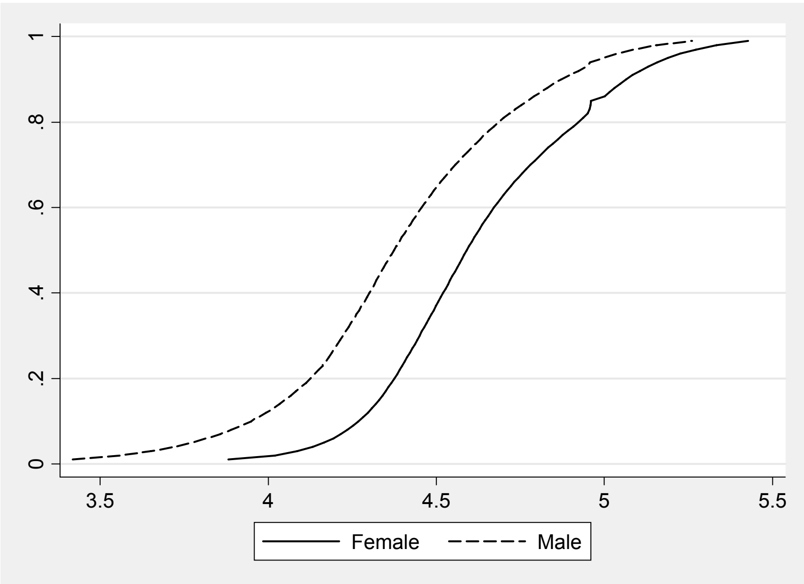
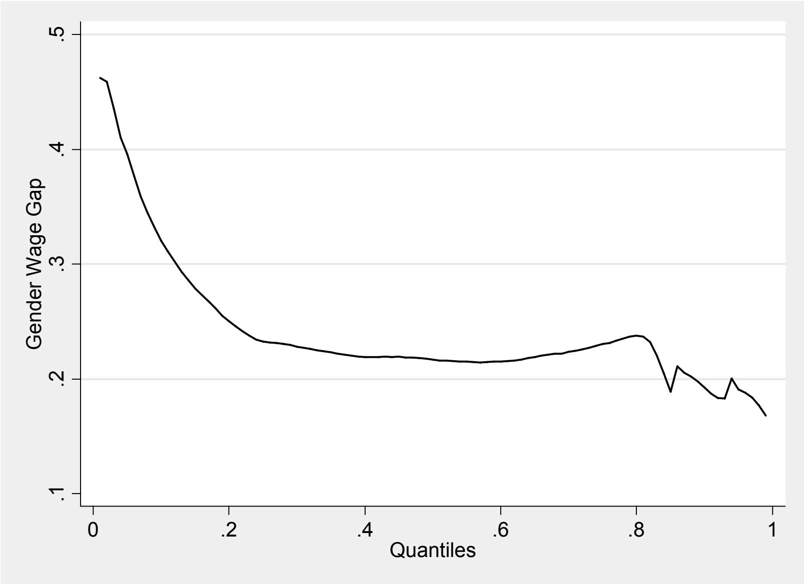


Figure 3: Gender wage gap at quantiles



# Appendix

Table A1: Descriptive statistics of individual characteristics

| | all | | $\ln w \leq \ln w_{0,25}$ | | $\ln w_{0,25} < \ln w \leq \ln w_{0,5}$ | | $\ln w_{0,5} < \ln w \leq \ln w_{0,75}$ | | $\ln w > \ln w_{0,75}$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variables | males | females | males | females | males | females | males | females | males | females |
| log daily wage (obs.) | 4.6008 | 4.3795 | 4.2524 | 3.9355 | 4.5025 | 4.2797 | 4.7037 | 4.4847 | 4.9446 | 4.8179 |
| log daily wage (imp.) | 4.6302 | 4.3894 | 4.2524 | 3.9355 | 4.5025 | 4.2797 | 4.7037 | 4.4847 | 5.0622 | 4.8575 |
| age | 40.8961 | 39.1671 | 37.8161 | 38.1482 | 40.2561 | 38.5245 | 41.4736 | 39.3392 | 44.0389 | 40.6563 |
| foreigner | 0.0983 | 0.0885 | 0.1426 | 0.1110 | 0.1301 | 0.1152 | 0.0840 | 0.0854 | 0.0364 | 0.0424 |
| low education without vocational trainig | 0.1518 | 0.2378 | 0.2896 | 0.3639 | 0.2000 | 0.3315 | 0.0998 | 0.2007 | 0.0178 | 0.0551 |
| vocational training | 0.6847 | 0.6002 | 0.6798 | 0.5743 | 0.7661 | 0.5953 | 0.7958 | 0.6697 | 0.4970 | 0.5613 |
| secondary school without vocational training | 0.0067 | 0.0137 | 0.0068 | 0.0115 | 0.0037 | 0.0098 | 0.0055 | 0.0121 | 0.0108 | 0.0213 |
| secondary school with vocational trainig | 0.0292 | 0.0682 | 0.0163 | 0.0357 | 0.0163 | 0.0489 | 0.0290 | 0.0760 | 0.0553 | 0.1122 |
| college of higher education | 0.0648 | 0.0296 | 0.0042 | 0.0063 | 0.0091 | 0.0076 | 0.0419 | 0.0215 | 0.2042 | 0.0830 |
| university | 0.0627 | 0.0506 | 0.0032 | 0.0084 | 0.0049 | 0.0069 | 0.0279 | 0.0200 | 0.2148 | 0.1671 |
| job tenure (in month)/100 | 1.3825 | 1.1658 | 1.0243 | 0.9104 | 1.4702 | 1.1845 | 1.5832 | 1.3062 | 1.4524 | 1.2622 |
| Observations | 477,160 | 124,488 | 119,296 | 31,122 | 119,286 | 31,122 | 119,285 | 31,119 | 119,293 | 31,125 |

Table A2: Descriptive statistics of firm characteristics

| | all | | $lnw \le lnw_{0,25}$ | | $lnw_{0,25} < lnw \le lnw_{0,5}$ | | $lnw_{0,5} < lnw \le lnw_{0,75}$ | | $lnw > lnw_{0,75}$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variables | males | females | males | females | males | females | males | females | males | females |
| number of employees/1000 | 2.4305 | 1.7236 | 1.2077 | 0.6560 | 2.3404 | 1.4256 | 3.1534 | 2.0984 | 3.0207 | 0.5613 |
| female quota (all employees) | 0.2071 | 0.3980 | 0.2359 | 0.4810 | 0.1805 | 0.4273 | 0.1854 | 0.3574 | 0.2265 | 0.3265 |
| quota of highly qualified employees (all employees) | 0.6770 | 0.6419 | 0.5967 | 0.5367 | 0.6477 | 0.6057 | 0.7019 | 0.6676 | 0.7616 | 0.7577 |
| business start-up after 1989 | 0.1463 | 0.1471 | 0.1464 | 0.1412 | 0.1000 | 0.1101 | 0.1473 | 0.1235 | 0.1914 | 0.2136 |
| export quota (sales) | 0.3096 | 0.2515 | 0.2332 | 0.1655 | 0.3199 | 0.2558 | 0.3218 | 0.2780 | 0.3635 | 0.3067 |
| wage bill per employee/1000 | 5.7900 | 5.2848 | 0.2332 | 4.0868 | 5.6295 | 4.8846 | 6.0580 | 5.5657 | 6.6520 | 6.6019 |
| sales per employee/100000 | 4.9998 | 5.1867 | 4.8205 | 3.1643 | 4.3380 | 4.3127 | 4.9808 | 5.2282 | 7.0872 | 8.0413 |
| good results last year | 0.3547 | 0.3528 | 0.2956 | 0.2893 | 0.3608 | 0.3208 | 0.3752 | 0.3754 | 0.3871 | 0.4254 |
| bad results last year | 0.2821 | 0.2878 | 0.3282 | 0.3159 | 0.2894 | 0.3017 | 0.2637 | 0.2850 | 0.2471 | 0.2487 |
| average results last year | 0.3632 | 0.3594 | 0.3761 | 0.3947 | 0.3498 | 0.3775 | 0.3611 | 0.3396 | 0.3657 | 0.3259 |
| technical state | 2.9735 | 2.9948 | 2.8619 | 2.9094 | 2.9244 | 2.9432 | 2.9963 | 3.0134 | 3.1113 | 3.1131 |
| industry-wide wage agreement | 0.7805 | 0.7332 | 0.7194 | 0.6049 | 0.8087 | 0.7717 | 0.8032 | 0.7819 | 0.7908 | 0.7743 |
| firm-specific wage agreement | 0.1110 | 0.1018 | 0.1107 | 0.1029 | 0.1093 | 0.0871 | 0.1169 | 0.1048 | 0.1072 | 0.1124 |
| no wage agreement | 0.1085 | 0.1650 | 0.1699 | 0.2921 | 0.0820 | 0.1412 | 0.0799 | 0.1133 | 0.1021 | 0.1133 |
| works council | 0.9152 | 0.8713 | 0.8198 | 0.6921 | 0.9354 | 0.9095 | 0.9492 | 0.9328 | 0.9562 | 0.9507 |
| agreed working hours per week | 36.7906 | 37.2031 | 37.4088 | 37.8909 | 36.6897 | 36.9797 | 36.5753 | 36.9901 | 36.4886 | 36.9515 |
| agriculture and forestry; electricity, gas and water supply, mining | 0.0358 | 0.0234 | 0.0310 | 0.0145 | 0.0277 | 0.0120 | 0.0398 | 0.0256 | 0.0448 | 0.0413 |
| manufacturing I | 0.2257 | 0.1766 | 0.2408 | 0.1210 | 0.2647 | 0.1610 | 0.2043 | 0.1783 | 0.1929 | 0.2461 |
| manufacturing II (reference) | 0.4967 | 0.4189 | 0.4122 | 0.4033 | 0.4822 | 0.4580 | 0.5370 | 0.4371 | 0.5554 | 0.3772 |
| construction | 0.0345 | 0.0147 | 0.0540 | 0.0170 | 0.0390 | 0.0138 | 0.0251 | 0.0154 | 0.0199 | 0.0126 |
| wholesale and retail trade | 0.0527 | 0.1329 | 0.0901 | 0.1687 | 0.0314 | 0.1691 | 0.0343 | 0.0839 | 0.0551 | 0.1100 |
| transport and communication | 0.0684 | 0.0451 | 0.0603 | 0.0300 | 0.0970 | 0.0389 | 0.0816 | 0.0658 | 0.0347 | 0.0457 |
| financial intermediation | 0.0012 | 0.0009 | 0.0001 | 0.0002 | 0.0001 | 0.0002 | 0.0007 | 0.0005 | 0.0039 | 0.0028 |
| real state, renting and business activities | 0.0518 | 0.0686 | 0.0687 | 0.0914 | 0.0286 | 0.0423 | 0.0470 | 0.0588 | 0.0629 | 0.0818 |
| education | 0.0029 | 0.0062 | 0.0033 | 0.0076 | 0.0019 | 0.0062 | 0.0027 | 0.0053 | 0.0036 | 0.0056 |
| other service activities | 0.0303 | 0.1127 | 0.0395 | 0.1463 | 0.0275 | 0.0985 | 0.0273 | 0.1293 | 0.0268 | 0.0769 |
| Berlin-West | 0.0426 | 0.0579 | 0.0397 | 0.0508 | 0.0357 | 0.0527 | 0.0598 | 0.0626 | 0.0353 | 0.0657 |
| Schleswig Holstein | 0.0492 | 0.0584 | 0.0615 | 0.0709 | 0.0512 | 0.0637 | 0.0430 | 0.0589 | 0.0410 | 0.0402 |
| Hamburg | 0.0570 | 0.0500 | 0.0407 | 0.0277 | 0.0424 | 0.0324 | 0.0793 | 0.0532 | 0.0657 | 0.0866 |
| Niedersachsen | 0.0796 | 0.0718 | 0.1179 | 0.1108 | 0.0880 | 0.0693 | 0.0672 | 0.0638 | 0.0452 | 0.0433 |
| Bremen | 0.0285 | 0.0341 | 0.0259 | 0.0392 | 0.0258 | 0.0275 | 0.0277 | 0.0338 | 0.0346 | 0.0360 |
| North Rhine-Westphalia (reference) | 0.2001 | 0.1640 | 0.1665 | 0.1118 | 0.2132 | 0.1707 | 0.1953 | 0.1645 | 0.2254 | 0.2088 |
| Hesse | 0.1341 | 0.1350 | 0.1314 | 0.1261 | 0.1279 | 0.1229 | 0.1385 | 0.1305 | 0.1387 | 0.1605 |
| Rhineland-Palatinate | 0.0463 | 0.0540 | 0.0561 | 0.0781 | 0.0508 | 0.0549 | 0.0413 | 0.0482 | 0.0371 | 0.0348 |
| Baden-Wurttemberg | 0.1354 | 0.1658 | 0.0987 | 0.1423 | 0.1277 | 0.1555 | 0.1389 | 0.2032 | 0.1765 | 0.1621 |
| Bavaria | 0.1660 | 0.1713 | 0.1943 | 0.1918 | 0.1605 | 0.2080 | 0.1408 | 0.1475 | 0.1684 | 0.1378 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Saarland | 0.0611 | 0.0377 | 0.0673 | 0.0505 | 0.0768 | 0.0423 | 0.0681 | 0.0337 | 0.0321 | 0.0243 |
| Observations | 477,160 | 124,488 | 119,296 | 31,122 | 119,286 | 31,122 | 119,285 | 31,119 | 119,293 | 31,125 |

Orginalmethode:

1. Draw M numbers at random from a uniform distribution $U[0,1]$: $\theta_1,\ldots,\theta_M$
2. Estimate for male and female employees M different quantile regression coefficients: $\begin{pmatrix}\hat{\beta}^m_{\theta_i}\\ \hat{\delta}^m_{\theta_i}\end{pmatrix}, \begin{pmatrix}\hat{\beta}^f_{\theta_i}\\ \hat{\delta}^f_{\theta_i}\end{pmatrix}; \; i=1,\ldots,M.$
3. Generate the following samples of size with replacement from the covariates of $[X:Z]$: $\left\{\tilde{X}^m_i:\tilde{Z}^m_i\right\}_{i=1}^M; \; \left\{\tilde{X}^f_i:\tilde{Z}^f_i\right\}_{i=1}^M; \; \left\{\tilde{X}^f_i:\tilde{Z}^m_i\right\}_{i=1}^M$
4. $\left\{\tilde{w}^m_i=\tilde{X}^m_i\hat{\beta}^m_{\theta_i}+\tilde{Z}^m_i\hat{\delta}^m_{\theta_i}\right\}_{i=1}^M$ and $\left\{\tilde{w}^f_i=\tilde{X}^f_i\hat{\beta}^f_{\theta_i}+\tilde{Z}^f_i\hat{\delta}^f_{\theta_i}\right\}_{i=1}^M$ are random sample of size M from the marginal wage distributions of $w$ consistent with the linear model in (1).
5. Generate the following random sample of the counterfactual distributions: $\left\{\tilde{w}^1_i=\tilde{X}^f_i\hat{\beta}^m_{\theta_i}+\tilde{Z}^m_i\hat{\delta}^m_{\theta_i}\right\}_{i=1}^M, \; \left\{\tilde{w}^2_i=\tilde{X}^f_i\hat{\beta}^f_{\theta_i}+\tilde{Z}^m_i\hat{\delta}^m_{\theta_i}\right\}_{i=1}^M$ and $\left\{\tilde{w}^3_i=\tilde{X}^f_i\hat{\beta}^f_{\theta_i}+\tilde{Z}^f_i\hat{\delta}^m_{\theta_i}\right\}_{i=1}^M$