Going behind the gender wage gap:
Are women less educated or are they in worse firms?

Anja Heinze
Centre for European Economic Research (ZEW Mannheim)

Using linked employee-employer data, this paper measures and decomposes the differences in the earnings distribution between male and female employees in Germany. I extend the traditional decomposition to disentangle the effect of human capital characteristics and the effect of firm characteristics in explaining the gender wage gap. Furthermore, I implement the decomposition across the whole wage distribution with the method proposed by Machado and Mata (2005). Thereby, I take into account the dependence between the human capital endowment of individuals and workplace characteristics. My decomposition detects that female employees are better educated then men in the lower tail of the wage distribution but that they work in inferior firms. In the upper tail of the distribution men and women work in similar firms but the female employees have less human capital.

JEL Classification: J16, J31

Keywords: gender wage gap, quantile regression, discrimination.

1 Motivation

The well-known fact that men generally earn higher wages than their female colleagues is one of the most studied issues in labour economics. Even though the pay differential tends to shrink over time, a sizeable gender wage gap persists that cannot be fully explained by differences in individual human capital characteristics which are considered as important determinates within the wage setting process (see e.g. Datta Gupta et al. 2006). During the last decade, the role of workplace characteristics and institutions in determining wage rates has moved more and more into the focus of scholarly interest (see e.g. Blau and Kahn 1996, 1997, Abowd, Kramarz and Margolis 1999). Due to the use of linked employer – employee data, the hitherto "unexplained" wage differential can then be assigned to individual as well as to workplace characteristics and the institutional environment. As Groshen has argued, this research helps us to understand whether women's disadvantage depends on 'who you are, what you do or where you work' (subtitle of Groshen, 1991, page 457).

The empirical evidence in this field consistently concludes that workplace characteristics are highly relevant in explaining wage differences between males and females (see e.g. Reilly amd Wirjanto 1999, Datta Gupta and Rothstein 2005, Drolet 2002, Data Gupta and Eriksson 2004). Most studies apply the traditional Oaxaca-Blinder decomposition which ascribes the observed gender wage gap to differences in the set of individual and firm-specific characteristics (endowment part) and to differences in the returns to these characteristics (remuneration part). The inclusion of workplace characteristics tends to increase the endowment component so that the explained fraction of the observed gap increases. Alternatively Heinze and Wolf (2006, 2007) investigate the effect of various firm characteristics, human resource practices and the organizational and institutional framework on the gender wage gap in Germany by looking at within-firm gender wage differentials. Meng (2004) and Meng and Meurs (2004) also differ in their research design and extend the traditional decomposition of the observed wage gap in France and Australia to an additional firm effect. In this setting, the firm effect represents the difference between the firm's premium paid to male and female employees and can be interpreted as employer discrimination.

One drawback common to all the studies that look at the role of workplace characteristics is that they focus on the average gender wage gap. Gender gap studies based solely on individual characteristics show, however, that earnings differences are very complex and vary over the wage distribution. Albrecht et al. (2003), for instance, detect that while the average gender wage gap is indeed relatively small in Sweden, the gap increases throughout the wage distribution and rises even more in the upper tail. They conclude that so-called glass ceiling effects limit the earnings potential of women in the upper part of the wage distribution. Hence, analyses of mean differences between male and female earnings are limited because they could lead us to conclude that the gender wage gap is of minor importance and that the size of the wage gap is constant throughout the whole wage distribution.

Furthermore, the traditional approach is based on the assumption that the importance of explanatory factors does not vary with the wage rate. This assumption is not very realistic. Among others, Albrecht et al. (2003) show an increasing impact of education on the wage differential across the wage distribution.

In fact, there are many good reasons to believe that male and female wages are also not equally affected by innovative human resource practices and institutional settings along the whole wage distributions. In particular, firm characteristics describing the institutional framework are supposed to be more important in the wage determination process of employees with low earnings because these workers belong to the main target group of unions. Furthermore, it is conceivable that firm's profits have a stronger impact on the wage rate of highly-paid employees because they are more likely to get corresponding bonus payments.

Therefore, I propose to use quantile regressions to adequately assess the effect of both individual and firm characteristics on the gender wage gap at each percentile of the wage distribution. To decompose the observed wage gap, I apply an extension of the traditional Oaxaca-Blinder (OB) decomposition to disentangle the effect of human capital characteristics (also denoted as individual characteristics) and the effect of firm characteristics in explaining the gender wage gap. This decomposition results in four terms: one fraction attributable to differences in the human capital characteristics, one part referring to differences in the returns to human capital characteristics, one component that captures differences in firm-specific characteristics as well as one fraction resulting from differences in the returns to these characteristics. On the basis of this decomposition I want to shed light on the causes of gender wage gap. Are women only less educated or do they work in "worse" firms in comparison to the men? The answer of this question has important implications for future strategies to reduce the gender wage gap. Should women are motivated to invest more in their education or they are motivated to select in better paying firms?

To accommodate differences along the wage distribution, I apply the Machado and Mata (2005) method. In a first step, quantile regressions are used to estimate the returns to the different characteristics at each percentile. Second, together with the estimated coefficients, I determine counterfactual marginal wage distributions by resampling the characteristics of male and female employees in random samples. These counterfactual distributions allow me to determine the four parts of the decomposition for each wage percentile.

The German LIAB data provide a very comprehensive data base to disentangle the effect of human capital characteristics and the effect of firm characteristics in explaining the gender wage gap. The LIAB data is a representative linked employer-employee panel including information on all employees of firms covered by the IAB establishment survey. The data set merges annual survey data, the IAB-establishment panel and process generated individual data, the Employment Statistical Register of the IAB which is based on administrative social security records).

The comparison of the wage information of 430269 male and 113466 female employees in 4010 establishments shows that the raw gender wage gap is sharply decreasing within the first quartile, then the decrease decelerates until the 70h percentile, and from then on the gap is increasing. In sum, there is a slightly u-shaped gender wage gap which explains why the average gap is rather small.

My decomposition detects that female employees are better educated then men in the lower tail of the wage distribution but they work in the inferior firms. In the upper tail of the distribution men and women work in similar firms but the female employees have less human capital. These results remain hidden when applying a decomposition approach at the mean. Based on my results, I provide new insights into the nature and the sources of gender wage inequality in Germany.

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 presented in Section 3. Section 4 describes the data source and section 5 presents and discusses the results are presented. Section 6 concludes.

2 Background for the decomposition

While the mean gender wage gap has been extensively studied in the labour economics literature, the attention has shifted towards investigating the degree to which the gender gap might vary across the wage distribution. Blau and Kahn (1996, 1997) explain the international differences in female wage deficiency and their evolution in time using the methodology proposed by Juhn et al. (1993). This methodology allows them to take into account the wage structure to explain wage inequality. Fortin and Lemieux (1998) decompose changes in the US gender wage gap at various wage percentiles using rank regressions. Bonjour and Gerfin (2001) apply the methodology proposed by Donald et al. (2000) to decompose the gender wage gap in Switzerland. Most recently, other papers use quantile regressions in order to decompose the gender wage gap at different points of the wage distribution. García et al. (2001) propose to use quantile regressions in order to compare quantiles of the male and female wage distribution conditional on the same set of characteristics as an approximation of the returns to unobserved and observed characteristics. However, their decomposition of the Spanish gender wage gap evaluates the vector of characteristics of men and women at the one 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 for the Spanish wage gap contradict the results of García et al. (2001). While in the analysis of García et al. (2001) the part of the gender wage gap that can be attributed to the different returns to characteristics increases across the wage distribution, Gardeazabel and Ugidos (2005) find the opposite.

By considering only the mean of the regressors like García et al. (2001), 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. In this setting, the distribution of the dependent variable will also have a higher variance for males. This feature can not be analysed with the method suggested by García et al. (2001) or the one used by Gardeazabel and Ugidos (2005). Machado and Mata (2005) (MM) hence propose an alternative decomposition procedure which combines a quantile regression and a bootstrap approach in order to estimate counterfactual density functions. 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 rises 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 differences in individual 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 using the MM decomposition method and taking into account a selection of women into full time employment. Thus, the authors' purpose is to make statements for all employed women regardless of their employment status. Also applying the MM decomposition method, Arulampalam et al. (2006) explore the wage differential for 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 left hand side. They interpret this as a sticky floor effect. By contrast, Hübler (2005) finds a decreasing raw wage gap at increasing quantiles. In this study the gender wage gap are considered over a time period from 1984 to 2002. Based on a combination of linear local matching and quantile regressions he shows that the unexplained wage differences between males and females are larger in the higher percentiles of the wage distribution with a decreasing importance over time. Beblo et al. (2003) also take into account the whole distribution in their analysis of the gender wage gap in Germany, but they primarily focus upon on the differences in individual characteristics. Fitzenberger and Kunze (2005) find like Hübler (2005) that the German gender wage gap is highest in the lower part and lowest in the upper part of the distribution. Their study highlights that occupational segregation and lower occupational mobility among females may explain the gender wage gap, a result that differs across the wage distribution.

My study differs from existing paper in three respects. First, apart from limiting the explanatory variables to individual characteristics, I include a set of detailed firm characteristics. Second, I extend the traditional OB decomposition to disentangle the effect of human capital characteristics and the effect of firm characteristics in explaining the gender wage gap. Finally, I implement the decomposition across the entire wage distribution with the MM method. Based on this most flexible parametric decomposition, I provide new insights into the nature and the sources of gender wage inequality in Germany.

3 Methodology

3.1 Wage Regression

OLS and most other estimation approaches focus on the mean effects. That is, 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 this literature.

Let w_i denote the log wage of worker *i*, X_i a vector of covariates representing his individual characteristics and Z_i a vector of covariates representing the characteristics of his workplace. The statistical model specifies the θth 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 β_θ and δ_θ 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).[1]

Since observed wage data are censored from above at the social security taxation threshold c_s one observes only $\tilde{w}_i = \min\{w_i, c\}$. Powell (1984, 1986) developed censored quantile regressions as a robust extension to the censored regression problem. 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 and to achieve convergence it is necessary to test different starting values. Unfortunately because access to the data[2] is limited and the large sample size it is not possible to implement censored quantile regressions. Alternatively, I apply quantile regressions after imputating estimated 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 of the social security system. In the Tobit regression model the same exogenous variable are used as in the 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 errors. Given that the main purpose of this paper is to propose and apply a new decomposition analysis and not the interpretation of the results of a single quantile regression, this limitation is justifiable.

3.2 Decomposition

The above regression analysis provides detailed insights into remuneration of observed worker and firm characteristics for men and women across the whole 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 is evaluated at the average characteristics of male (m) and female (f) employees:

$$\overline{w}_m - \overline{w}_f = \left(\overline{X}_m - \overline{X}_f\right)\hat{\beta}_m + \overline{X}_f\left(\hat{\beta}_m - \hat{\beta}_f\right), \tag{3}$$

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 (3) shows the difference in characteristics and the second term refers to the difference in the estimated coefficients. In order to distinguish between human capital endowment (X) and firm characteristics (Z) I extend the OB decomposition in the following way:

$$\overline{w}_m - \overline{w}_f = \underbrace{\left(\overline{X}_m - \overline{X}_f\right)\hat{\beta}_m}_{(i)} + \underbrace{\overline{X}_f\left(\hat{\beta}_m - \hat{\beta}_f\right)}_{(ii)} + \underbrace{\left(\overline{Z}_m - \overline{Z}_f\right)\hat{\delta}_m}_{(iii)} + \underbrace{\overline{Z}_f\left(\hat{\delta}_m - \hat{\delta}_f\right)}_{(iv)} \tag{4}$$

---

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

[2] The data means the data are only available at the Research Data Centre (FDZ) of the Federal Employment Agency (BA) in Nuremberg. It is only possible to work with data set there and the computation time for visiting scholars is limited.

The first and the third term on the right hand side of equation (4) show the difference in the human capital endowment (i) and the differences in firm characteristics (iii). The second and the fourth term refer to the difference in the remuneration for the human capital (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, in order to answer the question what the log wage would have been had female sample faced the same returns to characteristics as male employees.[3] The approach assumes that the male returns are the relevant benchmark for the distribution in the absence of any "discrimination".

The approach in equation (4) considers only differences of the average earnings. As stated above, 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 their approach is that they only consider the mean 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 $[0,1]$, then $F^{-1}(U)$ has distribution F. Thus, for given $[X_i : Z_i]$ and a random $\theta \sim U[0,1]$, $X_i\beta_\theta + Z_i\delta_\theta$ has the same distribution as $w_i | X_i, Z_i$. If $[X : Z]$ are randomly drawn from the population, instead of keeping $[X_i : Z_i]$ 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: $\begin{pmatrix} \hat{\beta}_\theta^m \\ \hat{\delta}_\theta^m \end{pmatrix}, \begin{pmatrix} \hat{\beta}_\theta^f \\ \hat{\delta}_\theta^f \end{pmatrix}$; $\theta = 1,\ldots,99$. Thus producing 99 coefficient vectors for males and 99 coefficient vectors for females.
2. Generate the following samples of size M=10000 with replacement from the set of covariates $[X : Z]$ for each estimated coefficient vector : $\left\{\tilde{X}_i^m : \tilde{Z}_i^m\right\}_{i=1}^M$; $\left\{\tilde{X}_i^f : \tilde{Z}_i^f\right\}_{i=1}^M$; $\left\{\tilde{X}_i^f : \tilde{Z}_i^m\right\}_{i=1}^M$

[3] 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.

3. Calculate $\left\{\tilde{w}_i^m = \tilde{X}_i^m \hat{\beta}_\theta^m + \tilde{Z}_i^m \hat{\delta}_\theta^m\right\}_{i=1}^M$ and $\left\{\tilde{w}_i^f = \tilde{X}_i^f \hat{\beta}_\theta^f + \tilde{Z}_i^f \hat{\delta}_\theta^f\right\}_{i=1}^M$ for each estimated coefficient vector. These data sets are random samples of $M \times 99(=\theta)$ observations from the marginal wage distributions of w which is consistent with the linear model in equation (1).
4. Generate the following random samples of the counterfactual distributions with the estimated coefficients of each percentile:
$\left\{\tilde{w}_i^1 = \tilde{X}_i^f \hat{\beta}_\theta^m + \tilde{Z}_i^m \hat{\delta}_\theta^m\right\}_{i=1}^M$, $\left\{\tilde{w}_i^2 = \tilde{X}_i^f \hat{\beta}_\theta^f + \tilde{Z}_i^m \hat{\delta}_\theta^m\right\}_{i=1}^M$ and $\left\{\tilde{w}_i^3 = \tilde{X}_i^f \hat{\beta}_\theta^f + \tilde{Z}_i^f \hat{\delta}_\theta^m\right\}_{i=1}^M$
$\tilde{w}^1$ states the hypothetical log wage for female employees if they had the firm characteristics of 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 male employees and only those characteristics had the same returns as for male employees. Finally, $\tilde{w}^3$ denotes the hypothetical log wage for female employees if their firm characteristics had been paid if they were man..

The empirical implementation of this procedure is, however, not straightforward. In the second step of the estimation procedure above, I have to constitute a random sample that contains random draw of human capital characteristics of female employees and firm characteristics drawn from the male employees. If the covariates were independent it would be possible to assign the randomly drawn female individual covariates to any drawn male firm covariate. However, it is not very realistic to assume independency between individual and workplace related characteristics. In contrast, it is much more likely that individuals select themselves into firms In order to assess the correlation between individual and firm covariates, I decide for the following assignment strategy guided by the economic meaning behind the counterfactual wage distributions in step 4: First I constitute a random sample of M female employees. After this I implement a Mahalanobis matching in order to assign each of these women to a similar male worker with respect to human capital characteristics. From the matched pairs I consider the human capital covariates from the female employees and the firm covariates from the male employees.

Based on the estimation results generated by the procedure described above, I can decompose the gender wage gap into the contribution of the human capital and firm characteristics as well as the contribution of the returns to human capital and firm characteristics. In order to simplify the comparison to the OB-decomposition, I will decompose the quantiles of the wage distribution as follows:

$$\begin{aligned} Q_\theta\left(w^m\right) - Q_\theta\left(w^f\right) = & \underbrace{\left[Q_\theta\left(\tilde{w}^m\right) - Q_\theta\left(\tilde{w}^1\right)\right]}_{(i)} + \underbrace{\left[Q_\theta\left(\tilde{w}^1\right) - Q_\theta\left(\tilde{w}^2\right)\right]}_{(ii)} \\ & + \underbrace{\left[Q_\theta\left(\tilde{w}^2\right) - Q_\theta\left(\tilde{w}^3\right)\right]}_{(iii)} + \underbrace{\left[Q_\theta\left(\tilde{w}^3\right) - Q_\theta\left(\tilde{w}^f\right)\right]}_{(iv)} + R \end{aligned} \tag{5}$$

Analogue to equation (4) there are four terms. The first term represents the contribution of the human capital characteristics and the third term denotes the contribution of the corresponding coefficients to the difference between the θth quantile of the male and 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.

Note that, these four terms have not exactly the same meaning as in the OB decomposition because the terms are based on counterfactual wage distributions as defined in step 4 of the estimation procedure. As a consequence the decomposition in equation (5) is well-defined. Accurately described, the first term, for example, is the difference between the log wages of the male employees and the counterfactual log wages of female employees if they had the firm characteristics of male employees and they had been paid as male employees. In the counterfactual wage distribution I don't take into account the whole distribution of male firm characteristics but rather a subsample (only men which are similar to female employees with respect to their human capital) as described above. As a result the difference between the wage distribution of males and the counterfactual wage distribution $\tilde{w}^1$ does not show the pure difference in the human capital between male and female employees. The first term additionally includes an effect resulting from the fact that not the whole sample of male employees is considered in the simulation.

The last term is a residual term in equation (5). 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 that my specification is correct, the residual term asymptotically tends to zero and equation (5) describes the true decomposition of the gender wage gap in quantiles.

4 Data

The analysis is based on a representative German linked employer – employee 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.[4] 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, investments, revenues, the size and composition of their workforces, salaries and wages.

[4] Detailed information on the *IAB Establishment Panel* is given by Bellmann et al. (1994), Bellmann (1997) and Kölling (2000).

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.[5] 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. Due to its administrative nature, this database has the advantage of providing reliable information on the daily earnings that are subject to social security contributions.

The sample for the subsequent analysis of the linked employer-employee data is constructed 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 because both the wage level as well as the wage setting process is still different in those firms which would require a separate analysis. Furthermore, I only consider firms with a least 10 employees.

In the second step, the establishment data are merged with the notifications for all employees who are employed by the selected establishments on 30th June of each year. From the worker data I drop foreigners, apprentices, part-time workers and homeworkers in order to ensure that the dependent and the independent variables are comparable for my sample. 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 430269 male and 113466 female employees in 4010 establishments.

The individual data include information on the daily wage, age, gender, nationality, employment status, education[6] and the date of entry into the establishment. The latter is used to approximate tenure by substracting the entry date from the ending date of the employer's notification which is also available in the individual data. Table 1 presents summary statistics for the human capital variables used in the subsequent analysis. The summary statistic shows that, on average, women have lower educational attainments and lower job tenures than male employees. It is interesting to note, however, that in the bottom tail of the wage distributions shows, that more women have a college or university degree than men (see table A1 in the appendix).

[Table 1 here]

The descriptive statistics of the firm characteristics are given in Table 2.

[5] Information on the Employment Statistics Register is given by Bender et al. (1996, 2000).

[6] The categories are: No degree, vocational training degree, high school degree (Abitur), high school degree and vocational training, technical college degree and university degree.

[Table 2 here]

The dependent variable in the subsequent analysis is the real gross log daily wage. Since there is an upper contribution limit to the social security system, gross daily wages are top-coded. In the sample, top-coding affects 24.5 per cent of all observations. While in the subsample of male employees the wage is censored above the 81th quantile of the male wage distribution, the censoring appears above the 93th quantile of the female wage distribution. To address this problem, a tobit regression is estimated by gender with log daily wages as the dependent variable and human capital 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 distribution of the gender wage gap

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 usual way to measure the male-female wage gap is to consider the difference between the average male wage and its female counterpart. In my sample, the average log male wage after imputation is 4.66 whereas the female log wage is 4.40. Therefore, the male-female average wage differential is 0.26. The gender wage gap is better assessed 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 a given quantile. Figure 3 plots the raw gender wage gap as a function at the quantiles. The gap is sharply decreasing within the first quartile, then remains rather stable until the 70th percentile, and then increases agian (apart from the 99th percentile). Hence, 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 an individual's workplace. The choice of the individual characteristics is limited to variables

indicating human capital. The set of individual characteristics contains formal skill dummies, age, age squared describing general human capital as well as job tenure indicating firm specific human capital.
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 describing the workforce of the establishment. These are the number of employees, the employment share of females as well as the share of highly qualified employees. Second, 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 as well as a dummy variable 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 dummy variable for the existence of a works council.
Separate earnings equations for male and female employees have been estimated using standard OLS and quantile regressions. Table 4 and Table 5 show the OLS coefficients with their standard errors and the coefficient estimated by QRs for subset deciles of the distributions[7].
[Table 4 here]
[Table 5 here]
All estimated effects in the OLS regressions are significantly different from zero. The variables describing the human capital have the expected effects 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 human capital characteristics generally vary across the distribution and differ from the OLS estimates in size but not in signs. The returns to educational attainment increases across the wage distribution. The impact of job tenure on the wage rate decreases across the wage distribution for both sexes. Most effects estimated by QRs are also smaller for female employees than for male workers.
Turning to the establishment variables, I find that wage rates increase with the number of employees and with the share of highly qualified employees for both men and women. The OLS regressions indicate that the share of female employees affects the wage rate of the women negatively and the wage rate of the men positively. The QR reveals that the impact of this variable is also negative for male employees in the lower quantiles. An explanation could be that the share of the female

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

employees indicates the downgrading of jobs in firms where a lot of females are employed regardless of whether a men or women do the job, as for example in retail. This is not true for male employees in higher positions of the wage distribution who are often the superiors.

Firms with higher sales per employee, good results in the last year and a state-of-the-art production technology tend to have higher wage rates while there are no clear trends across the wage distribution. The export quota has a positive impact on the wage rate in the OLS regression. The QR shows that this impact increases across the quantiles for female employees while it decreases for male employees.

The firm characteristics describing the institutional environment (wage agreement and works council) have a strong positive effect which is even stronger for female employees than for male workers. Note that the impact of the institutional variables decreases across the wage distribution for both male and female employees. It seems that unions and works councils rather support employees at the lower tail of the wage distribution. Which is not surprising, unionized employees tend to have lower and middle wage rates. Furthermore, employees at the higher tail of the wage distribution are often paid outside collective wage agreements.

5.3 Results of the decomposition

Table 6 and Figure 4 include the results of the OB decomposition and the MM decomposition.

[Table 6]

[Figure 4]

In the presented MM decomposition I control for the dependence between the human capital endowment and firm characteristics to determine the counterfactual wage distributions $\tilde{w}^1$ and $\tilde{w}^2$. I match to every randomly drawn female employee the most similar male employee with respect to the human capital endowment. Then I consider the human capital covariates of the female and the firm characteristics of the matched male.

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 the firm 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 MM decomposition using the estimated quantile regression coefficients shows that the part due to the difference of human capital characteristics and the part due to the returns to these covariates vary strongly with θ while the other two parts are more stable. The wage gap attributed to the different individual characteristics increases across the distribution. The other parts decrease. There is a male wage premium for the firm characteristics across the whole distribution while the gap due to difference in human capital characteristics shows that female employees are endowed with the better paid human capital upon the 60^{th} percentile. This suggests that women in the lower tail of the wage distribution should get a higher wage rate concerning their better human capital endowment but it seems they do not market their human capital enough. This interesting result is not identifiable in the OB decomposition. The difference in firm characteristics almost converges to zero in the upper tail of the wage distribution. It seems male and female employees at the upper tail of the wage distribution work in similar firms. Unfortunately, I cannot say anything about the statistical significance because the calculation of significance bonds with a bootstrap method is not possible given the computation time. Anyway,

My decomposition detects that female employees are better educated then men in the lower tail of the wage distribution but they work in “worse” firms. In the upper tail of the distribution men and women work in similar firms but female employees have less human capital endowment.

6 Conclusion

This study differs from existing paper examining the decomposition of the gender wage gap in three respects. First, apart from limiting the explanatory variables to individual characteristics, I include a set of detailed firm characteristics. Second, I extend the traditional Oaxaca Blinder decomposition to disentangle the effect of human capital characteristics and the effect of firm characteristics in explaining the gender wage gap. More specifically, I decompose the differences in the wage distribution between men and women into an explained and unexplained component with respect to human capital characteristics as well as into an explained and unexplained part with respect to firm characteristics. This approach yields new insights into what causes the gender wage gap Are women less educated or do they work in worse firms in comparison to men? Moreover, I implement the decomposition across the entire wage distribution with the Machado Mata method. Based on this most flexible parametric decomposition method, I provide new insights into the nature and the sources of gender wage inequality in Germany using.

I use data from the LIAB, a representative German employer – employee linked data set, for the year 2002 to exam the wage structure of male and female employees in the private sector in West Germany. The unconditional gender gap is sharply decreasing within the first quartile of the wage distribution, then the decrease decelerates until the 70th percentile, and from then on the gap is increasing. The gender wage gap is far from being constant within the wage distribution.

My decomposition throughout the whole wage distribution detects that female employees are better educated then men in the lower tail of the wage distribution but they work in inferior firms. In the upper tail of the distribution men and women work in similar firms but female employees have less human capital.

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., R. Oaxaca and N. Smith (2006): Swimming Upstream, Floating Downstream: Comparing Women's Relative Wage Progress in the United States and Denmark, *Industrial and Labor Relations Review*, Vol. 59 (2), 243-266.

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.

Fitzenberger, B. and A. Kunze (2005): Vocational Training and Gender: Wages and Occupational Mobility Among Youth Workers, *Oxford Review of Economic Policy*, Vol. 21 (3): 392 – 415.

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.

Heinze, A. and E. Wolf (2007): Personnel Policy ant the Gender Wage Gap within German Firms, mimeo.

Hübler, O. (2005): Is There a Varying Unexplained Gender Wage Gap in Germany?, *Applied Economics Quarterly*, Vol. 51: 29 – 48.

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 human capital characteristics

Variables	Men		Women	
	Mean	Std. Dev.	Mean	Std. Dev.
log daily wage (obs.)	4.6110	0.2686	4.3876	0.3487
log daily wage (imp.)	4.6600	0.3525	4.4039	0.3815
age	41.0615	9.3321	39.0688	10.0759
low education without vocational training	0.1182	0.3229	0.2012	0.4009
vocational training	0.7090	0.4542	0.6321	0.4822
secondary school without vocational training	0.0066	0.0811	0.0131	0.1139
secondary school with vocational training	0.0307	0.1726	0.0717	0.2579
college of higher education	0.0697	0.2547	0.0311	0.1736
university	0.0657	0.2478	0.0508	0.2195
job tenure (in month)/100	1.3899	1.0145	1.1580	0.9472
Observations	430,269		113,466	

Table 2: Descriptive statistic of firm characteristics

Variables	Men		Women	
	Mean	Std. Dev.	Mean	Std. Dev.
number of employees/1000	2.3724	3.9798	1.7078	3.0082
female quota (all employees)	0.2093	0.1609	0.3997	0.2378
quota of highly qualified employees (all employees)	0.6852	0.2523	0.6501	0.2623
business start-up after 1989	0.1483	0.3554	0.1471	0.3542
export quota (sales)	0.3064	0.2954	0.2467	0.2815
wage bill per employee/1000	5.7880	2.0689	5.2790	2.3054
sales per employee/100000	5.0858	14.0341	5.2793	19.9938
good results last year	0.3589	0.4797	0.3522	0.4777
bad results last year	0.2829	0.4504	0.2877	0.4527
average results last year	0.3582	0.4795	0.3600	0.4800
technical state	2.9779	0.7135	2.9941	0.7142
industry-wide wage agreement	0.7771	0.4162	0.7264	0.4458
firm-specific wage agreement	0.1134	0.3170	0.1050	0.3065
no wage agreement				
works council	0.9142	0.2801	0.8686	0.3379
agreed working hours per week	36.7963	1.8823	37.2314	1.7702
agriculture and forestry; electricity, gas and water supply, mining	0.0373	0.1896	0.0249	0.1559
manufacturing I	0.2174	0.4124	0.1745	0.3795
manufacturing II (reference)	0.4993	0.5000	0.4078	0.4914
construction	0.0354	0.1847	0.0157	0.1241
wholesale and retail trade	0.0548	0.2275	0.1368	0.3436
transport and communication	0.0682	0.2520	0.0464	0.2104
financial intermediation	0.0013	0.0355	0.0010	0.0314
real state, renting and business activities	0.0527	0.2235	0.0704	0.2559
education	0.0030	0.0548	0.0058	0.0759
other service activities	0.0307	0.1724	0.1167	0.3211
Berlin-West	0.0439	0.2048	0.0586	0.2349
Schleswig Holstein	0.0511	0.2203	0.0614	0.2400
Hamburg	0.0593	0.2362	0.0517	0.2214
Niedersachsen	0.0842	0.2776	0.0752	0.2636
Bremen	0.0301	0.1708	0.0353	0.1846
North Rhine-Westphalia	0.1964	0.3973	0.1630	0.3694
Hesse	0.1318	0.3383	0.1332	0.3398
Rhineland-Palatinate	0.0472	0.2121	0.0552	0.2284
Baden-Wurttemberg	0.1301	0.3364	0.1569	0.3637
Bavaria	0.1671	0.3731	0.1714	0.3769
Saarland	0.0588	0.2352	0.0381	0.1914
Observations	430,269		113,466	

Table3: Tobit regression

Variables	Men		Women	
	Coefficient	Standard Errors	Coefficient	Standard Errors
age	0.0369**	0.0003	0.0308**	0.0007
$(age)^2$	-0.0363**	0.0004	-0.0336**	0.0008
low education without vocational training	-0.1677**	0.0012	-0.1767**	0.0023
vocational training (reference)	-	-	-	-
secondary school without vocational training	0.1105**	0.0050	0.0560**	0.0077
secondary school with vocational training	0.2253**	0.0024	0.1330**	0.0035
college of higher education	0.4676**	0.0021	0.3320**	0.0054
university	0.5562**	0.0025	0.4656**	0.0047
job tenure (in month)/100	0.0391**	0.0005	0.0534**	0.0011
number of employees/1000	0.0214**	0.0004	0.0292**	0.0009
$(\text{number of employees}/1000)^2$	-0.0008**	0.0000	-0.0010**	0.0001
female quota (all employees)	0.0195**	0.0029	-0.0933**	0.0048
quota of highly qualified employees (all employees)	0.1228**	0.0018	0.1969**	0.0038
business start-up after 1989	0.0511**	0.0013	0.0497**	0.0027
export quota (sales)	0.0079**	0.0018	0.0322**	0.0043
wage bill per employee/1000	0.0302**	0.0003	0.0350**	0.0005
sales per employee/100000	0.0009**	0.0000	0.0002**	0.0000
good results last year	0.0140**	0.0010	0.0241**	0.0022
bad results last year	-0.0103**	0.0011	0.0042*	0.0022
average results last year (reference)	-	-	-	-
technical state	0.0162**	0.0006	0.0099**	0.0013
industry-wide wage agreement	0.0279**	0.0015	0.0455**	0.0027
firm-specific wage agreement	0.0084**	0.0019	0.0186**	0.0038
no wage agreement (reference)	-	-	-	-
works council	0.0800**	0.0017	0.1468**	0.0031
agreed working hours per week	-0.0123**	0.0003	-0.0188**	0.0006
constant	3.7050**	0.0129	3.8408**	0.0283
observations	430,269		113,466	
uncensored	311,086		99,454	
right-censored	119,183		14,012	

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	0.0354**	0.0003	0.0290	0.0284	0.0292	0.0350
$(age)^2$	-0.0351**	0.0004	-0.0314	-0.0298	-0.0290	-0.0326
low education without vocational training	-0.1577**	0.0011	-0.1126	-0.1157	-0.1324	-0.1711
vocational training (reference)	-	-	-	-	-	-
secondary school without vocational training	0.0902**	0.0044	-0.0678	0.0407	0.1451	0.1828
secondary school with vocational training	0.2085**	0.0021	0.1015	0.1665	0.2381	0.2516
college of higher education	0.4549**	0.0015	0.4069	0.4333	0.4767	0.4829
university	0.5483**	0.0015	0.4885	0.5352	0.5740	0.5760
job tenure (in month)/100	0.0388**	0.0004	0.0476	0.0424	0.0399	0.0323
number of employees/1000	0.0211**	0.0003	0.0251	0.0247	0.0233	0.0160
$(\text{number of employees}/1000)^2$	-0.0008**	0.0000	-0.0010	-0.0010	-0.0010	-0.0006
female quota (all employees)	0.0083**	0.0026	-0.0848	-0.0525	0.0006	0.0662
quota of highly qualified employees (all employees)	0.1165**	0.0016	0.0896	0.0910	0.1013	0.1011
business start-up after 1989	0.0514**	0.0011	0.0308	0.0514	0.0615	0.0495
export quota (sales)	0.0058**	0.0016	0.0039	-0.0019	-0.0050	-0.0003
wage bill per employee/1000	0.0293**	0.0002	0.0276	0.0335	0.0367	0.0372
sales per employee/100000	0.0010**	0.0000	0.0011	0.0012	0.0013	0.0011
good results last year	0.0140**	0.0009	0.0138	0.0134	0.0130	0.0177
bad results last year	-0.0094**	0.0009	-0.0107	-0.0073	-0.0094	-0.0084
average results last year (reference)	-	-	-	-	-	-
technical state	0.0152**	0.0006	0.0135	0.0134	0.0132	0.0158
industry-wide wage agreement	0.0288**	0.0014	0.0567	0.0444	0.0255	0.0100
firm-specific wage agreement	0.0091**	0.0017	0.0164	0.0170	0.0075	-0.0019
no wage agreement (reference)	-	-	-	-	-	-
works council	0.0782**	0.0015	0.0985	0.0797	0.0672	0.0601
agreed working hours per week	-0.0120**	0.0003	-0.0133	-0.0118	-0.0103	-0.0109
constant	3.7354**	0.0117	3.7059	3.7432	3.7800	3.8097
Observations	430,269					

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	0.0296**	0.0006	0.0126	0.0236	0.0325	0.0399
$(age)^2$	-0.0325**	0.0008	-0.0150	-0.0280	-0.0369	-0.0433
low education without vocational training	-0.1723**	0.0022	-0.0927	-0.1184	-0.1619	-0.2105
vocational training (reference)	-	-	-	-	-	-
secondary school without vocational training	0.0468**	0.0073	-0.0936	-0.0089	0.0653	0.1147
secondary school with vocational training	0.1249**	0.0033	0.0761	0.0856	0.1085	0.1330
college of higher education	0.3148**	0.0049	0.2592	0.2833	0.3124	0.3386
university	0.4454**	0.0039	0.3755	0.4027	0.4510	0.4835
job tenure (in month)/100	0.0527**	0.0011	0.0579	0.0558	0.0511	0.0441
number of employees/1000	0.0292**	0.0009	0.0395	0.0318	0.0251	0.0192
$(\text{number of employees}/1000)^2$	-0.0010**	0.0001	-0.0016	-0.0013	-0.0010	-0.0005
female quota (all employees)	-0.0936**	0.0046	-0.0799	-0.0963	-0.0997	-0.0992
quota of highly qualified employees (all employees)	0.1907**	0.0036	0.2112	0.1615	0.1482	0.1420
business start-up after 1989	0.0460**	0.0025	0.0064	0.0197	0.0430	0.0598
export quota (sales)	0.0358**	0.0041	0.0164	0.0220	0.0341	0.0370
wage bill per employee/1000	0.0338**	0.0004	0.0289	0.0387	0.0439	0.0472
sales per employee/100000	0.0002**	0.0000	0.0001	0.0000	0.0000	0.0004
good results last year	0.0227**	0.0021	0.0172	0.0281	0.0266	0.0259
bad results last year	0.0032	0.0021	-0.0118	0.0049	0.0081	0.0100
average results last year (reference)	-	-	-	-	-	-
technical state	0.0095**	0.0012	0.0078	0.0078	0.0048	0.0046
industry-wide wage agreement	0.0451**	0.0026	0.0820	0.0552	0.0418	0.0275
firm-specific wage agreement	0.0197**	0.0036	0.0602	0.0245	0.0194	0.0068
no wage agreement (reference)	-	-	-	-	-	-
works council	0.1466**	0.0030	0.2442	0.1720	0.1308	0.1051
agreed working hours per week	-0.0184**	0.0006	-0.0235	-0.0202	-0.0176	-0.0161
constant	3.8594**	0.0272	4.0133	3.8940	3.7998	3.7525
Observations	113,466					

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: MM decomposition at selected quantiles and OB decomposition

Quantile	Obs. Gender Wage Gap	Diff. in human capital characteristics	Diff. in returns to human capital characteristics	Diff. in firm characteristics	Diff. in returns to firm characteristics
0.1	0.3246	-0.0936	0.1902	0.0515	0.1545
0.2	0.2521	-0.0713	0.1500	0.0405	0.1421
0.3	0.2333	-0.0531	0.1260	0.0323	0.1357
0.4	0.2247	-0.0345	0.1081	0.0270	0.1282
0.5	0.2219	-0.0135	0.0936	0.0224	0.1198
0.6	0.2214	0.0122	0.0813	0.0187	0.1099
0.7	0.2311	0.0431	0.0736	0.0154	0.0955
0.8	0.2396	0.0766	0.0762	0.0119	0.0747
0.9	0.2527	0.0939	0.1024	0.0072	0.0540
OB	0.2560	0.0582	0.0595	0.0349	0.1035

Figure 1: Male and female wage densities

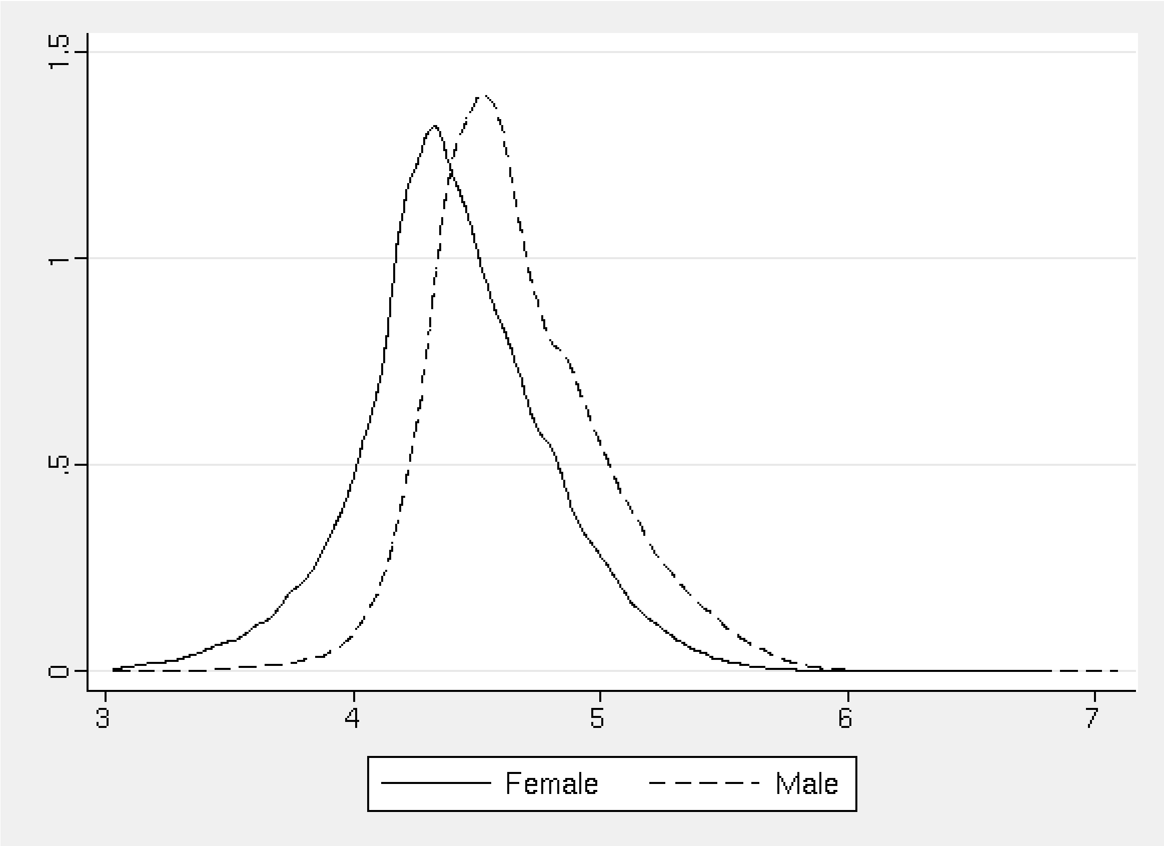


Figure 2: Male and female wage distribution functions

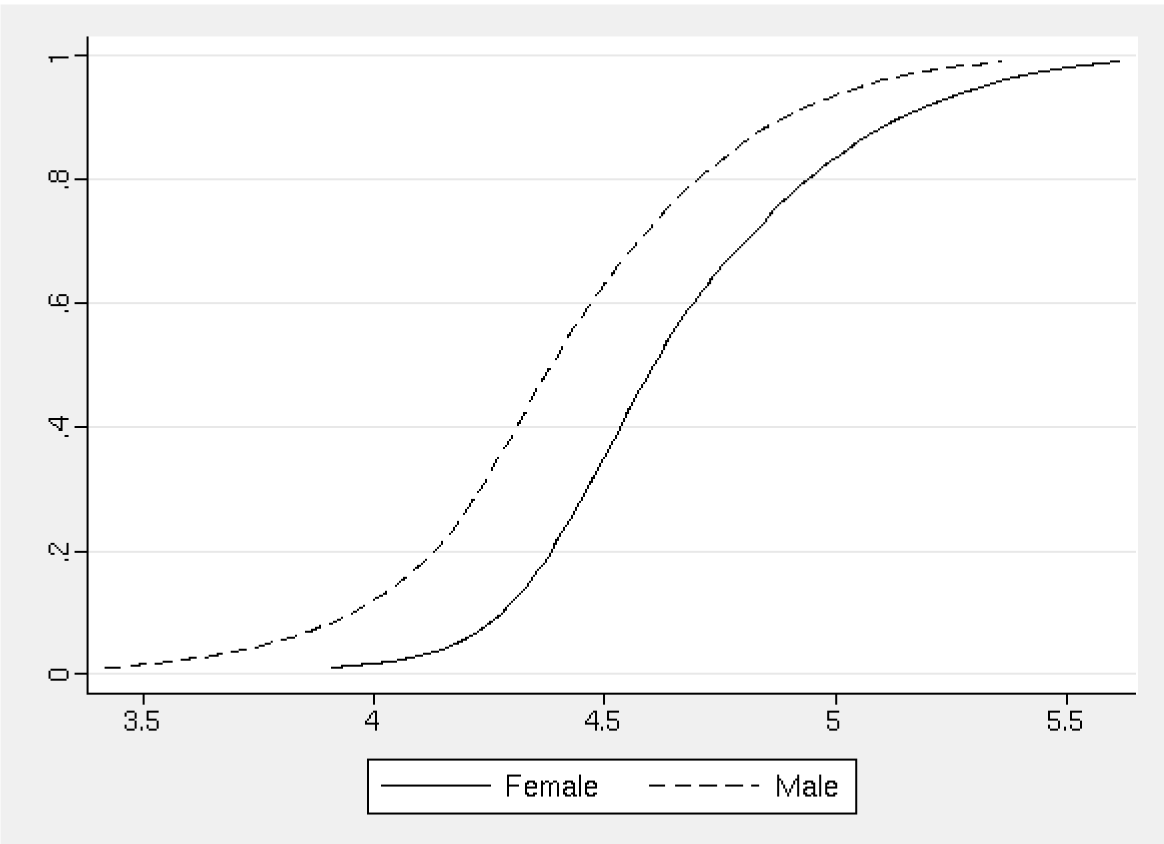


Figure 3: Gender wage gap at quantiles

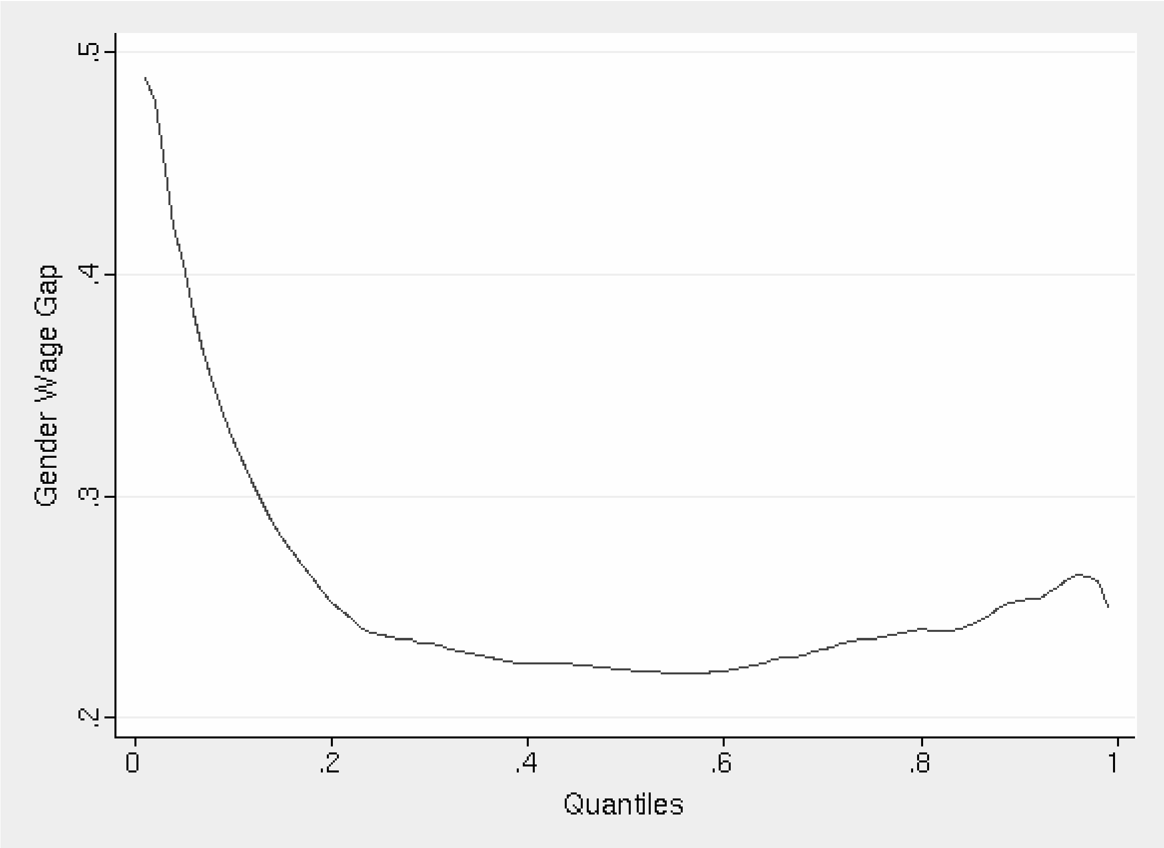
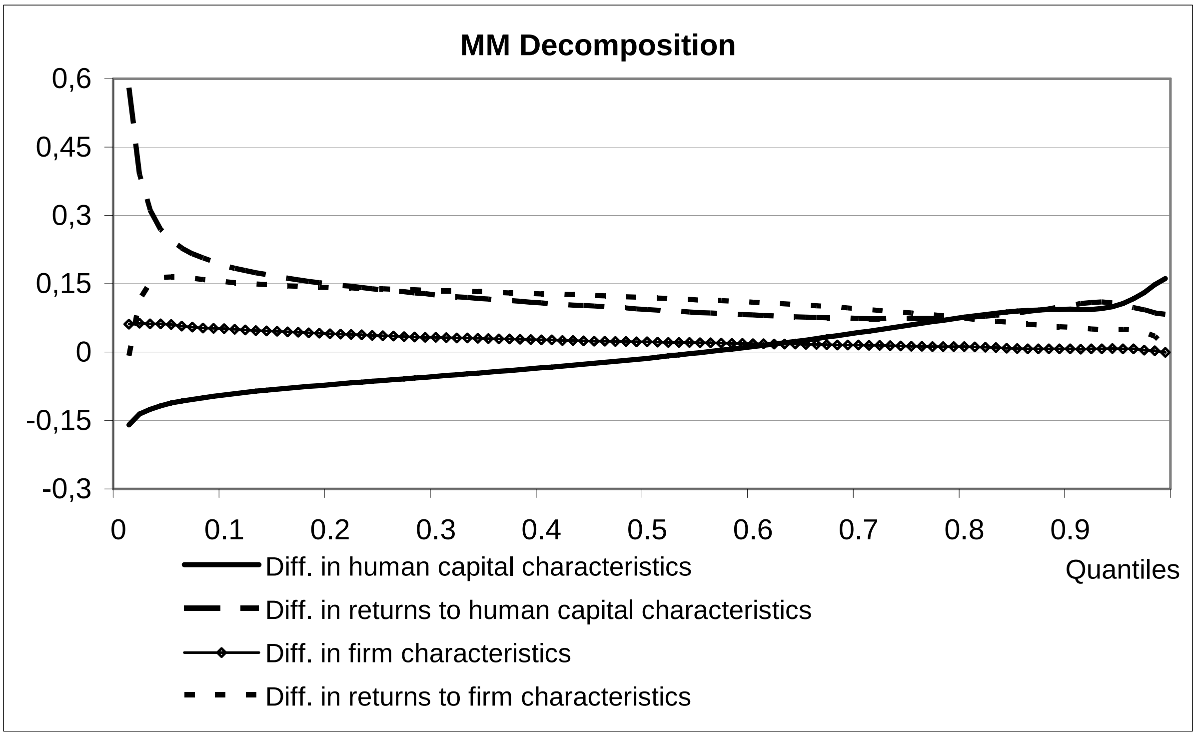


Figure 4: MM Decomposition



Appendix

Table A1: Descriptive statistics of human capital 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.)	4.6110	4.3876	4.2627	3.9396	4.5153	4.2874	4.7325	4.4990	4.9335	4.8244
log daily wage (imp.)	4.6600	4.4039	4.2627	3.9396	4.5153	4.2874	4.7245	4.4990	5.1374	4.8897
age	41.0615	39.0688	37.9303	38.1112	40.4757	38.2104	41.8201	39.1815	44.0197	40.7716
low education without vocational trainig	0.1182	0.2012	0.2366	0.3222	0.1493	0.2765	0.0722	0.1554	0.0149	0.0508
vocational training	0.7090	0.6321	0.7321	0.6137	0.8136	0.6458	0.8123	0.7058	0.4778	0.5631
secondary school without vocational training	0.0066	0.0131	0.0065	0.0107	0.0036	0.0091	0.0063	0.0122	0.0101	0.0205
secondary school with vocational trainig	0.0307	0.0717	0.0172	0.0384	0.0173	0.0533	0.0326	0.0813	0.0559	0.1136
college of higher education	0.0697	0.0311	0.0044	0.0068	0.0107	0.0084	0.0467	0.0239	0.2172	0.0853
university	0.0657	0.0508	0.0032	0.0082	0.0055	0.0070	0.0299	0.0212	0.2242	0.1666
job tenure (in month)/100	1.3899	1.1580	1.0455	0.9108	1.4881	1.1586	1.5721	1.2923	1.4539	1.2703
Observations	430,269	113,466	107,569	28,368	107,555	28,363	107,578	28,364	107,567	28,371

Table A2: Descriptive statistics of firm characteristics

	all		$lnw \leq lnw_{0,25}$		$lnw_{0,25} < lnw \leq lnw_{0,5}$		$lnw_{0,5} < lnw \leq lnw_{0,75}$		$lnw > lnw_{0,75}$	
Variables	males	females	males	females	males	females	males	females	males	females
number of employees/1000	2.3724	1.7078	1.1924	0.6546	2.3141	1.4059	2.9350	2.0874	3.0482	2.6831
female quota (all employees)	0.2093	0.3997	0.2341	0.4821	0.1837	0.4303	0.1924	0.3584	0.2268	0.3280
quota of highly qualified employees (all employees)	0.6852	0.6501	0.6107	0.5490	0.6592	0.6139	0.7074	0.6786	0.7636	0.7589
business start-up after 1989	0.1483	0.1471	0.1411	0.1398	0.1024	0.1088	0.1553	0.1244	0.1945	0.2155
export quota (sales)	0.3064	0.2467	0.2341	0.1629	0.3120	0.2503	0.3170	0.2703	0.3627	0.3033
wage bill per employee/1000	5.7880	5.2790	4.8343	4.0702	5.6013	4.8607	6.0358	5.5766	6.6805	6.6082
sales per employee/100000	5.0858	5.2793	3.6641	3.1983	4.3564	4.5293	5.0819	5.2681	7.2409	8.1211
good results last year	0.3589	0.3522	0.2993	0.2836	0.3724	0.3265	0.3747	0.3735	0.3891	0.4254
bad results last year	0.2829	0.2877	0.3304	0.3181	0.2900	0.3005	0.2648	0.2827	0.2465	0.2495
average results last year	0.3582	0.3600	0.3703	0.3983	0.3376	0.3730	0.3605	0.3438	0.3644	0.3251
technical state	2.9779	2.9941	2.8688	2.9082	2.9322	2.9446	2.9968	3.0121	3.1139	3.1113
industry-wide wage agreement	0.7771	0.7264	0.7175	0.5981	0.8067	0.7620	0.7938	0.7723	0.7906	0.7731
firm-specific wage agreement	0.1134	0.1050	0.1144	0.1063	0.1098	0.0906	0.1195	0.1099	0.1097	0.1132
no wage agreement										
works council	0.9142	0.8686	0.8199	0.6891	0.9331	0.9043	0.9444	0.9303	0.9593	0.9505
agreed working hours per week	36.7963	37.2314	37.3951	37.8905	36.7083	37.0351	36.6073	37.0494	36.4746	36.9505
agriculture and forestry; electricity, gas and water supply, mining	0.0373	0.0249	0.0303	0.0147	0.0295	0.0134	0.0424	0.0286	0.0472	0.0430
manufacturing I	0.2174	0.1745	0.2329	0.1155	0.2466	0.1568	0.1980	0.1791	0.1919	0.2464
manufacturing II (reference)	0.4993	0.4078	0.4190	0.3997	0.4873	0.4411	0.5342	0.4158	0.5569	0.3745
construction	0.0354	0.0157	0.0569	0.0179	0.0388	0.0155	0.0262	0.0163	0.0196	0.0129
wholesale and retail trade	0.0548	0.1368	0.0926	0.1740	0.0340	0.1744	0.0399	0.0870	0.0526	0.1116
transport and communication	0.0682	0.0464	0.0620	0.0309	0.1005	0.0410	0.0757	0.0692	0.0345	0.0447
financial intermediation	0.0013	0.0010	0.0001	0.0002	0.0001	0.0002	0.0009	0.0005	0.0040	0.0029
real state, renting and business activities	0.0527	0.0704	0.0651	0.0907	0.0315	0.0461	0.0514	0.0619	0.0630	0.0830
education	0.0030	0.0058	0.0033	0.0066	0.0021	0.0054	0.0029	0.0056	0.0037	0.0056
other service activities	0.0307	0.1167	0.0377	0.1496	0.0297	0.1061	0.0285	0.1359	0.0268	0.0754
Berlin-West	0.0439	0.0586	0.0411	0.0509	0.0422	0.0532	0.0569	0.0647	0.0352	0.0657
Schleswig Holstein	0.0511	0.0614	0.0655	0.0758	0.0540	0.0682	0.0449	0.0612	0.0402	0.0402
Hamburg	0.0593	0.0517	0.0430	0.0286	0.0448	0.0345	0.0805	0.0552	0.0690	0.0884
Niedersachsen	0.0842	0.0752	0.1277	0.1171	0.0947	0.0746	0.0700	0.0654	0.0442	0.0436
Bremen	0.0301	0.0353	0.0270	0.0380	0.0285	0.0295	0.0297	0.0364	0.0351	0.0374
North Rhine-Westphalia	0.1964	0.1630	0.1600	0.1124	0.2076	0.1635	0.1929	0.1675	0.2253	0.2086
Hesse	0.1318	0.1332	0.1263	0.1234	0.1234	0.1212	0.1380	0.1296	0.1395	0.1586
Rhineland-Palatinate	0.0472	0.0552	0.0575	0.0800	0.0528	0.0560	0.0415	0.0500	0.0370	0.0349
Baden-Wurttemberg	0.1301	0.1569	0.0879	0.1299	0.1167	0.1477	0.1379	0.1880	0.1780	0.1621
Bavaria	0.1671	0.1714	0.1963	0.1923	0.1616	0.2081	0.1448	0.1479	0.1656	0.1374
Saarland	0.0588	0.0381	0.0676	0.0515	0.0737	0.0436	0.0629	0.0341	0.0310	0.0231
Observations	430,269	113,466	107,569	28,368	107,555	28,363	107,578	28,364	107,567	28,371