

Behr, Andreas; Pötter, Ulrich

Working Paper

Analysing wage differences between the USA and Germany using proportional hazards models

Beiträge zur angewandten Wirtschaftsforschung, No. 16

Provided in Cooperation with:

University of Münster, Münster Center for Economic Policy (MEP)

Suggested Citation: Behr, Andreas; Pötter, Ulrich (2006) : Analysing wage differences between the USA and Germany using proportional hazards models, Beiträge zur angewandten Wirtschaftsforschung, No. 16, Westfälische Wilhelms-Universität Münster, Centrum für Angewandte Wirtschaftsforschung (CAWM), Münster

This Version is available at:

<https://hdl.handle.net/10419/51419>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Analysing wage differences between the USA and Germany using proportional hazards models

Andreas Behr
Institute for Econometrics
University of Münster
Am Stadtgraben 9
48143 Münster, Germany
email: andreas.behr@wiwi.uni-muenster.de

Ulrich Pötter
Faculty of Social Science
University of Bochum
Universitätsstrasse 150
44801 Bochum, Germany
email: ulrich.poetter@ruhr-uni-bochum.de

JEL-Classification J31, C14

March 2006

Abstract

We analyse differences between the wage distributions in the USA and Germany in 2001 both for women and men. The empirical analysis is based on the decomposition of differences using Cox's marginal (partial) likelihood. The approach based on rank invariant estimators such as Cox's is borrowed from the literature on failure time data. Donald et al. (2000) pioneered this approach. However, they did not use the full power of the semi-parametric approach. Instead, they argued for using a piecewise constant hazard rate model. We improve on their work by showing that the semi-parametric features of Cox's marginal likelihood are as appropriate for the analysis of wage decompositions and as easy to interpret. Moreover, we extend their approach by allowing for nonlinear regression effects. We will show empirically that this formulation will both increase the flexibility of their approach and improve the discriminatory power between wage regimes.

1 Introduction

In this paper, we analyse the differences in wage distributions between the USA and Germany in 2001 for women and men. The data used for the analysis are harmonised files of two national longitudinal surveys, the Panel Study of Income Dynamics (PSID) and the German Socio Economic Panel (GSOEP).

In recent research on labour markets, international comparative studies have been fruitfully exploited to highlight variations in labour market institutions, skill endowments and wage distributions (Blau and Kahn 1996 and 2003). Differences in wage distributions between the US and Canada have been studied by DiNardo and Lemieux (1997) and by Donald, Green and Paarsch (2000). Beaudry and Green (2003) analysed differences between the US and Germany focusing on changes in relative capital endowments.

We suggest a new way of using semi-parametric regression techniques to analyse differences in wage distributions. Fortin and Lemieux (1998) argued for the use of rank invariant regression estimators when decomposing wage differences. Their approach as well as that of DiNardo et al. (1996) relies on an inefficient reduction of wages into K ordinal wage levels. Later, Donald et al. (2000) in a pioneering work used ideas from the literature on duration models to suggest a parametric proportional hazards model, namely a piecewise constant hazard rate model, to estimate the underlying distribution. Their use of a parametric model forced them to rely on some arbitrary grouping of the observations as well.

We follow the lead of Donald et al. and base the decomposition of wage differences on a proportional hazards model for the wages. In contrast to their estimation strategy, our approach is based on Cox's marginal likelihood. This allows us to dispense with the arbitrary grouping of observations used by Donald et al., Fortin and Lemieux, and DiNardo et al. Moreover, we use a general additive model for the effects of covariates. It can capture any nonlinearities in the covariate effects and is generally more stable than the traditional use of polynomials.

Our approach can also be compared to both Martins' and Pereira's (2004) as well as Machado and Mata's (2005) using quantile regression. They use a linear parametrisation of the quantile function. In contrast, the proportional hazards model we propose implies a somewhat more complicated dependence on covariates and on a non-parametric component that nevertheless can be easily estimated. Thus, we can compute effects at the quantiles of the wage distribution as well as at the conditional quantiles. While the use of a linear quantile function might not be as restrictive as may appear (Angrist et al. 2006), alternative techniques build on semi-parametric regression models will shed new light on the findings from quantile regressions.

The empirical analysis shows that German skill distributions are located slightly right to the US distributions but that US return-to-skills functions are considerably steeper. This leads to negative wage differentials for US workers compared to their German counterparts except at the rightmost part of the distribution, where US workers gain from the steepness of the return-to-skills functions. The decomposition reveals that only the smallest share of wage differentials can be attributed to differences in years of education, working experience and occupational structure. The price effect is found to be favourable to US men, but this effect was out-weighted by the return-to-skills function effect which favours German employees, especially those in the lower part of the wage distribution. For women we found a small effect of

skills favourable to US women, a mixed effect of the return-to-skills function and a price effect strongly favourable to German women leading to considerable negative wage differentials except at the highest quantiles of the wage distribution. These findings are in accordance with the results of Beaudry and Green (2003) who find parallel developments of skills over the last two decades in the USA and Germany and attribute the flatter relation between skills and wages in Germany to a relative over-accumulation of capital.

The paper is structured as follows: In Section 2, we outline the decomposition approach based on the proportional hazards model as well as the estimation method. The data base is described briefly in Section 3. Section 4 contains the empirical results. Section 5 concludes.

2 A proportional hazards model for wage data

Following Donald et al. (2000), we apply a proportional hazards model to wage data. To introduce the model, let wages W have absolutely continuous distribution function $F(w | x) := \Pr(W \leq w | X = x)$ conditional on a vector of covariates X . The probability that a person has at least wage w is given by the complementary distribution function

$$S(w | x) := \Pr(W > w | x) =: 1 - F(w | x)$$

where $F(w | x)$ denotes the conditional cdf of W . A convenient model for the influence of covariates x on the complementary distribution function is

$$S(w | x; \beta) = S_0(w)^{r(x; \beta)}, \quad 0 < r(x; \beta) < \infty, \quad r(0; \beta) = 1 \quad (1)$$

For an arbitrary, fixed complementary distribution function $S_0(\cdot)$, this family of distributions is referred to as Lehmann family or proportional hazards model. $S_0(w)$ is the baseline complementary distribution function, i.e. $S(w | x; \beta)$ evaluated at $x' = (0, 0, \dots, 0)$. Note that without the normalisation $r(0, \beta) = 1$, $r(x; \beta)$ is determined only up to scale since a fixed scale factor can be absorbed in the baseline complementary distribution function. $r(x; \beta)$ is often called the relative risk function. A suitable choice of $r(x; \beta)$ is the exponential form

$$r(x; \beta) = e^{x' \beta} \quad (2)$$

which ensures $r(x; \beta) > 0$. The normalisation $r(0; \beta) = 1$ is achieved by excluding a constant term.

Let $\Lambda(w | x; \beta)$ denote the negative logarithm of the complementary distribution function

$$\Lambda(w | x; \beta) := -\ln S(w | x; \beta)$$

In the context of duration analysis, the function $\Lambda(w | x; \beta)$ is called the integrated hazard function. The complementary distribution function can be expressed as

$$S(w | x; \beta) = e^{-\Lambda(w | x; \beta)}$$

and with $\Lambda_0(w) := -\ln S_0(w)$ we have

$$S(w | x; \beta) = \exp(-\Lambda(w | x; \beta)) = S_0(w)^{r(x; \beta)} = \exp(-\Lambda_0(w)r(x; \beta))$$

and $\Lambda(w | x; \beta) = r(x; \beta)\Lambda_0(w)$. The effect of covariates is to scale the integrated hazard function $\Lambda_0(\cdot)$, the baseline integrated hazard function corresponding to the baseline complementary distribution function $S_0(\cdot)$. Hence the name “proportional hazards model”.

While it is convenient to introduce the proportional hazards model in terms of the conditional distributions it sometimes is preferable and possibly more customary to express it in terms of random variables as well. If W is a random variable with conditional complementary distribution function $S_0(w)^{r(x; \beta)}$, then

$$\ln \Lambda_0(W) \simeq_d -\ln r(x; \beta) + \epsilon \quad (3)$$

where \simeq_d denotes equality in distribution and ϵ follows the extreme value distribution with complementary distribution function given by

$$\Pr(\epsilon > u) = e^{-e^u}$$

With the special choice of $r(x; \beta) = \exp(x'\beta)$ one arrives at the familiar linear model

$$\ln \Lambda_0(W) \simeq_d -x'\beta + \epsilon$$

so that some monotone increasing transform of the wage follows a linear regression with a fixed error distribution. Note the minus sign for the linear predictor: An increase in $x'\beta$ decreases expected (transformed) wages but increases integrated hazards.

To see the equivalence between (1) and (3) one may start with the latter and compute:

$$\begin{aligned} \Pr(W > w | x; \beta) &= \Pr(\ln \Lambda_0(W) > \ln \Lambda_0(w) | x; \beta) \\ &= \Pr(-\ln r(x; \beta) + \epsilon > \ln \Lambda_0(w)) \\ &= \Pr(\epsilon > \ln r(x; \beta) + \ln \Lambda_0(w)) \\ &= \exp\left(-e^{\ln(r(x; \beta)\Lambda_0(w))}\right) = \exp(-\Lambda_0(w)r(x; \beta)) \\ &= S_0(w)^{r(x; \beta)} \end{aligned}$$

The quantile function implied by the proportional hazards model is

$$Q(p | x; \beta) = S^{-1}(1 - p | x; \beta) = \Lambda_0^{-1}\left(\frac{-\ln(1 - p)}{r(x; \beta)}\right)$$

Thus, for two distinct values of p

$$\frac{\Lambda_0(Q(p_1 | x; \beta))}{\Lambda_0(Q(p_2 | x; \beta))} = \frac{\ln(1 - p_1)}{\ln(1 - p_2)}$$

and for distinct values of x

$$\frac{\Lambda_0(Q(p | x_1; \beta))}{\Lambda_0(Q(p | x_2; \beta))} = \frac{r(x_2; \beta)}{r(x_1; \beta)}$$

While Machado and Mata (2005) suggest a linear quantile function $Q(p|x;\beta) = x'\beta(p)$ with $\beta(p)$ depending on the quantile, the proportional hazards model implies that a transform of the quantile function depends on a fixed parameter β and an unknown monotone transform Λ_0 of the conditional quantile function:

$$\ln \Lambda_0(Q(p|x;\beta)) = \ln(-\ln(1-p)) - \ln(r(x,\beta))$$

The local effect of the k -th covariate on the conditional quantile p at the mean of the covariates $\mu := \mathbb{E}(X)$ is

$$b_k(p) := \left. \frac{\partial}{\partial x_k} Q(p|x;\beta) \right|_{x=\mu} = \left. \frac{\partial r(x;\beta)}{\partial x_k} \right|_{x=\mu} \frac{1}{\lambda_0(Q(p|\mu;\beta))} \frac{\ln(1-p)}{r(\mu;\beta)^2} \quad (4)$$

where $\lambda_0(\cdot)$ is the hazard function, the derivative of the integrated hazard function Λ_0 . When the parametrisation (2) is used, the vector of local effects is

$$b(p) := -\beta \frac{1}{\lambda_0(Q(p|\mu;\beta))} \frac{-\ln(1-p)}{\exp(\mu'\beta)} \quad (5)$$

so that the local effect, which can be interpreted as rates of return of market skills at different points of the conditional wage distribution, is proportional to the regression parameters β .¹

2.1 Observed wages, skill index, and monotone transformations

Juhn et al. (1991) and Fortin and Lemieux (1998) have argued persuasively that since changes in wage structure tend to have the same impact on all workers earning the same wage, measures of wage structure effects should only depend on the position of a worker in a given wage distribution. Moreover, the position within a wage distribution may be assumed to depend on the workers skill level. This would follow from a human capital model in which wages (in equilibrium) are equal to marginal productivity that reflects skill levels.

Thus observed wages result from skills by means of a return-to-skills function

$$W = \Lambda^{-1}(r^*)$$

which should be strictly monotone in the amount of skills r^* . Hence, persons having higher skills will receive a higher wage compared to persons having lower skills and persons with lower skills will receive a lower wage than better skilled persons will irrespective of the wage structure.

A change of wage structure should only be reflected in a change of the function Λ^{-1} while a change in the distribution of skills should only be reflected in a change in the distribution of r^* . Disentangling the effect of wage structure from that of amounts of skill thus means to distinguish effects on the function Λ^{-1} from those on the distribution of r^* , the amount of skills.

Proportional hazards models are an obvious choice to model this distinction between wage structure and amount of skill structure since they naturally allow to specify

¹Dabrowska (2005) discusses general methods to estimate $b(p)$ directly.

Λ^{-1} and r^* separately. Setting $r^* := \exp(\epsilon)/r(x; \beta)$ and $\Lambda \equiv \Lambda_0$ one recovers the proportional hazards model. In fact, the class of proportional hazards models is invariant under the group of strictly monotone transformations. If g is any strictly monotone transformation from \mathbb{R}_+ onto \mathbb{R}_+ , then

$$\Pr(g(W) > w \mid x, \beta) = \Pr(W > g^{-1}(w) \mid x, \beta) = S_0(g^{-1}(w))^{r(x, \beta)}$$

which is once again a proportional hazards model with the same relative risk function $r(x; \beta)$ as the distribution of W . The amount of skills r^* may stay fixed even when the wage structure changes.

The proportional hazards model can be extended by allowing the parameters β to depend on the wages. This was suggested by Donald et al. (2000) who argued that wage dependent parameters may reflect effects like minimum wage regulations. It should be noted, however, that proportional hazards models extended in this way pose no restrictions on the underlying conditional distributions, thus undermining the aim of a clear separation of effects. In particular, the invariance of the relative risk function under monotone transformations is lost when β is made dependent on w . In consequence, the distinction between wage structure and skills can no longer be based on the distinction between relative risk function and baseline hazard function.

Having distinguished between “wage structure” and “amount of skill”, the latter can now be further decomposed by considering changes in the distribution of X (the distribution of endowments with skills) and changes in the magnitude of the parameters β (the relative prices of skills) both across time and country.

Another consequence of the adoption of the proportional hazards model is that a further decomposition of the residual term as in Juhn et al. (1993) is not possible. Any change in their “price of unobserved skills” will be absorbed in the return-to-skills function. In particular, problems of interpretation as indicated by Suen (1997) cannot arise.

2.2 Estimating the regression function based on ranks

Since the relative risk function does not change when data are recorded on a monotonely transformed scale it seems natural to base estimation of the relative risk function on those functions of the data that stay constant under monotone transformations as well. Any other functions of the data on which estimation could be based will extract useful information on the relative risk function only if some a priori knowledge of the baseline integrated hazard function $\Lambda_0(\cdot)$ is presupposed.

Suppose the data are $D = ((w_1, x_1)', \dots, (w_n, x_n)')'$. The data can be recovered from the order statistic $O = ((w_{(1)}, x_{(1)})', \dots, (w_{(n)}, x_{(n)})')'$ with $w_{(1)} \leq w_{(2)} \leq \dots \leq w_{(n)}$, the ordered values of the wages, together with the anti-ranks of the wages $R = ((1), (2), \dots, (n))'$. For example, if the observed wages are $D = (1500, 2000, 1700)'$, then the order statistic is $O = (1500, 1700, 2000)'$ and the anti-ranks are $R = (1, 3, 2)'$. The original data are recovered from (O, R) by rearranging the order statistic in the order given by the anti-ranks. Now the anti-ranks R are invariant under strictly monotone transformations. On the other hand, the order statistic can be transformed into any vector of increasing values of wages by a monotone transformation.

Thus it seems reasonable that estimation of the regression parameters β should be based on the marginal distribution of the anti-ranks R . Their distribution is

$$\begin{aligned} \Pr(R = ((1), (2), \dots, (n))' \mid x_{(1)}, \dots, x_{(n)}; \beta) \\ = \int_0^\infty \int_{w_1}^\infty \dots \int_{w_{n-1}}^\infty \prod_{i=1}^n f(w_i \mid x_{(i)}; \beta) dw_n \dots dw_1 \end{aligned}$$

where $f(\cdot)$ is the conditional density. The innermost integral

$$\int_{w_{n-1}}^\infty f(w_n \mid x_{(n)}; \beta) dw_n = S_0(w_{n-1})^{r(x_{(n)}; \beta)}$$

is just the complementary conditional distribution function. The next integral is

$$\begin{aligned} & \int_{w_{n-2}}^\infty f(w_{n-1} \mid x_{(n-1)}; \beta) S_0(w_{n-1})^{r(x_{(n)}; \beta)} dw_{n-1} \\ &= \int_{w_{n-2}}^\infty \lambda_0(w_{n-1}) r(x_{(n-1)}; \beta) S_0(w_{n-1})^{r(x_{(n)}; \beta) + r(x_{(n-1)}; \beta)} dw_{n-1} \\ &= \frac{r(x_{(n-1)}; \beta)}{r(x_{(n)}; \beta) + r(x_{(n-1)}; \beta)} \\ & \quad \times \int_{w_{n-2}}^\infty \lambda_0(w_{n-1}) ((r(x_{(n)}; \beta) + r(x_{(n-1)}; \beta)) S_0(w_{n-1})^{r(x_{(n)}; \beta) + r(x_{(n-1)}; \beta)} dw_{n-1} \\ &= \frac{r(x_{(n-1)}; \beta)}{r(x_{(n)}; \beta) + r(x_{(n-1)}; \beta)} S_0(w_{n-2})^{r(x_{(n)}; \beta) + r(x_{(n-1)}; \beta)} \end{aligned}$$

Continuing this way, the probability of ranks reduces to

$$\prod_{i=1}^n \frac{r(x_{(i)}; \beta)}{\sum_{l \in \{(i), (i+1), \dots, (n)\}} r(x_l; \beta)}$$

which can be used as a marginal likelihood for β (see Fleming and Harrington 1991, chap. 4.3 for extensions).

The marginal likelihood depends on the observed wages only through their ranks. It is thus invariant under strictly monotone transformations of the wages. Note that this contrasts with classical rank regression which is based on the ranks of the residuals from some regression model (Hettmansperger 1984, chap. 5). Note also how this contrasts with quantile regression which is equivariant under monotone transformations: the quantiles of transformed wages are equal to the transformed quantiles of the wages (Koenker 2005).

The derivation above assumes that the ranks are uniquely defined. But survey data will necessarily contain some ties, i.e. several observations having the same value. In this case it is reasonable to maximise the marginal likelihood over all rankings of the data compatible with the observations. While this would be prohibitively costly, there are easily computed approximations. We use an approximation called after Efron (see Kalbfleisch and Prentice 2002, chap. 4.2.3 for details).

To allow for non-linear effects of covariates, we have chosen to model the risk function as

$$r(x; \beta) := \exp \left(\sum_{j=1}^k g_j(x_j; \beta_j) + \sum_{j=k+1}^m x_j \beta_j \right)$$

where x_1, \dots, x_k are continuous covariates and x_{k+1}, \dots, x_m are discrete. The g_j are smooth functions that we approximate by natural cubic splines. These are composed of several cubic polynomials pieced together as smoothly as possible. They are further restricted by the requirement that the function is linear beyond the range of the covariate. The knots, the places where the different polynomials are pieced together, are chosen from appropriate quantiles of the distribution of the covariate. The advantages of splines over the traditional use of polynomials is that they are able to accommodate functional forms that are difficult to approximate by low degree polynomials, that they are less dependent on effects far apart in the covariate space and that collinearity problems are less relevant. Moreover, the number of free parameters is comparable to the ones in polynomial regression and the fitting of natural cubic spline models with knots determined from the covariate distribution can be accomplished with the estimation methods previously described.

2.3 Estimating the return-to-skills function

Having obtained an estimate for β , the non-parametric likelihood for the baseline complementary distribution function S_0 is

$$L(S_0|x; \hat{\beta}) := \prod_{i=1}^n S_0(w_{i-})^{r(x_i; \hat{\beta})} - S_0(w_i)^{r(x_i; \hat{\beta})}$$

where $S_0(w_-) := \lim_{u \uparrow x} S_0(u)$ is the left limit of $S_0(\cdot)$ at w . The likelihood is non-zero only if $S_0(\cdot)$ has jumps at all the observations w_i . Thus, the non-parametric maximum likelihood estimator with $\hat{\beta}$ plugged in will be a step function with jumps at the observed wages. Since this corresponds to a purely discrete distribution, the baseline integrated hazard function is

$$\hat{\Lambda}_0(w) = \sum_{w_k \leq w} \frac{\hat{S}_0(w_{k-}) - \hat{S}_0(w_k)}{\hat{S}_0(w_{k-})}$$

where w_k runs through the set of jumps of \hat{S}_0 .

2.4 Counterfactual wage distributions

Throughout the analysis, we use the subscripts U for the USA and G for Germany, in graphics we use GER and USA, respectively. Whether we are referring to men or women is denoted by the subscripts m and f .

When using counterfactual components of $\hat{\Lambda}, \hat{\beta}, \hat{F}$ in combination with fitted components, counterfactual wage distributions can be constructed to highlight special features of interest. Here, F denotes the distribution function of the covariates.

For example, applying the estimation procedure to the USA sample of men results in the estimated marginal wage distribution function

$$\begin{aligned} \hat{S}_{U,m}(w) &:= \hat{S}(w; \hat{\Lambda}_{0,U,m}, \hat{\beta}_{U,m}, \hat{F}_{U,m}) \\ &:= \int \exp \left(-\hat{\Lambda}_{0,U,m}(w) r(x; \hat{\beta}_{U,M}) \right) d\hat{F}_{U,m}(x) \end{aligned}$$

This is the marginal distribution implied by the model. We will write $S_{U,m}$ etc. for the empirical complementary distribution function.

Using the distribution of male characteristics in Germany, $\hat{F}_{G,m}$, instead of $\hat{F}_{U,m}$ leads to

$$\tilde{S}_{U,m,\hat{F}_{G,m}}(w) := \tilde{S}(w; \hat{\Lambda}_{0,U,m}, \hat{\beta}_{U,m}, \hat{F}_{G,m})$$

This represents the hypothetical wage distribution function for men in the USA which would have been observed if the transformation function and prices of skills would have been applied to the skill distribution of the German male sample. The difference in wage distributions

$$S_{U,m} - \tilde{S}_{U,m,\hat{F}_{G,m}}$$

therefore isolates the effect of differences in the skill distributions between the USA and Germany.

We will be interested in three contrafactual contrasts.²

- i) The contrast due to differences in skills (X-effect):

$$S_{U,m} - \tilde{S}_{U,m,\hat{F}_{G,m}} = S_{U,m} - \tilde{S}(\hat{\Lambda}_{0,U,m}, \hat{\beta}_{U,m}, \hat{F}_{G,m})$$

- ii) The contrast due to differences in relative prices for skills (β -effect):

$$S_{U,m} - \tilde{S}_{U,m,\hat{\beta}_{G,m}} = S_{U,m} - \tilde{S}(\hat{\Lambda}_{0,U,m}, \hat{\beta}_{G,m}, \hat{F}_{U,m})$$

- iii) The contrast due to differences in the return-to-skills function (Λ_0 -effect):

$$S_{U,m} - \tilde{S}_{U,m,\hat{\Lambda}_{0,G,m}} = S_{U,m} - \tilde{S}(\hat{\Lambda}_{0,G,m}, \hat{\beta}_{U,m}, \hat{F}_{U,m})$$

We will use functionals of the distribution functions to characterise these three “first order” effects. Thus, differences between functionals of the form $\gamma(S_{U,m}) - \gamma(\tilde{S}_{U,m,\hat{\Lambda}_{0,G,m}})$ are used to highlight important contributions to the overall comparison.

3 The Data source: Cross National Equivalent Files

Our analysis is based on two national longitudinal surveys, the PSID for the USA and the GSOEP for Germany. The CNEF are ex-post harmonised versions of the national studies constructed to facilitate international comparative studies.

²One might be tempted to try a complete decomposition using an algebraic identity like:

$$\begin{aligned} S_{U,m} - S_{G,m} &= (S_{U,m} - \hat{S}_{U,m}) + (\hat{S}_{U,m} - \tilde{S}_{U,m,\hat{F}_{G,m}}) + (\tilde{S}_{U,m,\hat{F}_{G,m}} - \tilde{S}_{U,m,\hat{F}_{G,m},\hat{\beta}_{G,m}}) \\ &\quad + (\tilde{S}_{U,m,\hat{F}_{G,m},\hat{\beta}_{G,m}} - \hat{S}_{G,m}) + (\hat{S}_{G,m} - S_{G,m}) \end{aligned}$$

There are 6 logically possible such identities that result from changing the order by which contrafactual terms are introduced. But we do not see how such an exercise might contribute to an understanding of differences between wage distributions across countries.

In our analysis we include employed men and women working at least 800 hours per year in the age between 20 and 60 and having obtained at least four years of education. Our wage variable is based on reported gross yearly wages. As the social systems as well as payment contracts differ considerably between the two countries, these might restrict comparability of the reported wages. Especially problematic is the payment of Christmas and holiday gratifications as well as contributions to the social security system paid by the employer (about 50% of total contributions) in Germany.³ But because our main interest is to analyse the rewards for skills in the labour market, we consider gross wages more adequate than net earnings. To take into account different working hours we scale reported wages by reported total working hours to obtain hourly wage rates. To allow for comparability, the German wage figures are expressed in US Dollar using 2001 purchasing power parities. Note that the choice of the exchange rate has considerable influence on the wage levels. In 2001 the Euro has been undervalued against the US Dollar in the foreign exchange markets relative to the purchasing power parity. The average nominal exchange rate was 0.8956 while 1.2686 Dollars were needed to obtain the purchasing power of 1 Euro (based on German baskets). In 2005 this relative undervaluation (2001 : 29.4%) of the Euro relative to purchasing power parities has almost disappeared (2005 : 6.2%): the average nominal exchange rate was 1.2441 and the purchasing power parity was 1.3267.⁴ In accordance with the literature we use logarithmic wages throughout.⁵

Covariates included are measures of education, experience and occupation. Due to data problems, the occupation variables from the two data sets are transformed into 10 rather broad categories. Detailed information on occupational categories is given in the appendix.⁶ Schooling is given in years of schooling needed to obtain the attained level of education. Working experience is approximated by age minus 6 minus years of schooling.

To prevent outliers from unduly influencing the empirical results, we drop the 1% highest and lowest wages in both countries. Additionally we exclude altogether 26 observations that had an extreme influence on model fit and parameter estimates. Outlier detection is based on the one hand on the residuals $\hat{M}_i := 1 - r(x_i; \hat{\beta})\hat{\Lambda}_0(w_i)$. This has empirical mean 0. Due to the fitting procedure, the residuals are slightly

³For the PSID, the definition is: “Labor earnings include wages and salary from all employment including self-employment (farming, business, market gardening, and roomers and boarders), professional practice or trade, and bonuses, overtime and commissions. The original variable reported in the PSID for labor income includes only the labor part of self-employment income. The variable reported here also includes the asset part of self-employment income. Individual labor income can have negative values because the asset part of self-employment income can take on negative values.” For the GSOEP, “Labor earnings include wages and salary from all employment including training, primary and secondary jobs, and self-employment, plus income from bonuses, overtime, and profit-sharing. Specifically labor earnings is the sum of income from primary job, secondary job, self-employment, 13th month pay, 14th month pay, Christmas bonus pay, holiday bonus pay, miscellaneous bonus pay, and profit-sharing income.”

⁴Deutsche Bundesbank (2006), Statistisches Bundesamt (2002,2006).

⁵While the logarithmic transform could create some technical problems when applying standard techniques from survival data analysis, these can be avoided by assuming a positive lower bound on wages. In any case, neither the estimation techniques nor the inference is invalidated and the results are easier to compare with classical estimates.

⁶While occupation might be affected by problems of endogeneity, we regard this problem as much less severe for occupations, which relates at least partly to vocational and other training, than for other variables like sector which are sometimes used in wage decompositions.

negatively correlated, but this correlation is only of order $O(1/n)$.⁷ The second component of \hat{M}_i should be unit exponentially distributed when the model fits. We excluded observations if $\hat{M}_i < -9$. This event has probability below $5 \cdot 10^{-5}$ in the exponential model. On the other hand, we used the empirical influence functions to check for influences on the estimated parameters. Individual observations were excluded if the influence on spline parameters exceeded 0.2 and on occupational dummies 0.15, both in absolute terms.

4 Decomposing the differences in wage distributions between the USA and Germany

In this section we analyse in some detail, the difference in the wage distributions between the USA and Germany using the decomposition method described above.

4.1 Log-wage distributions and characteristics

Figure 1 displays the densities of logarithmic hourly gross wages for men and women in US Dollars, using 2001 purchasing power parity.

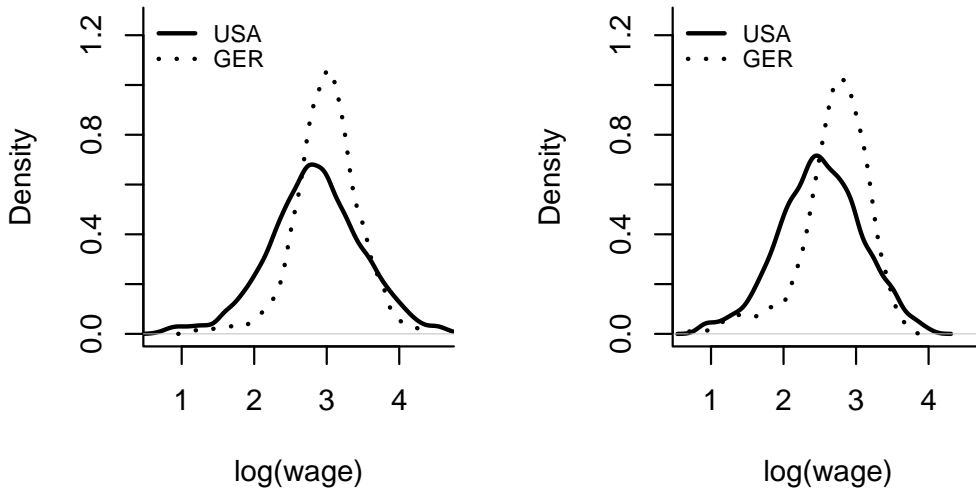


Figure 1: Density estimates of log-wage distribution (left men, right women)

For the kernel density estimates, we select the bandwidth according to the Silverman rule adjusted by the factor of 0.9. We find the log-wage distributions for the USA to be located slightly left of the German distributions and to be platykurtic, having much smaller densities in the middle range. The empirical wage distribution functions are shown in Figure 2. The higher densities for average log-wages in Germany are reflected in steeper distribution functions in the middle range. At low quantiles the distribution functions of the USA are above and at higher quantiles below the German reflecting the higher wage inequality in the USA.

⁷Therneau and Grambsch (2000, chap. 4) discuss further properties of these and related residuals.

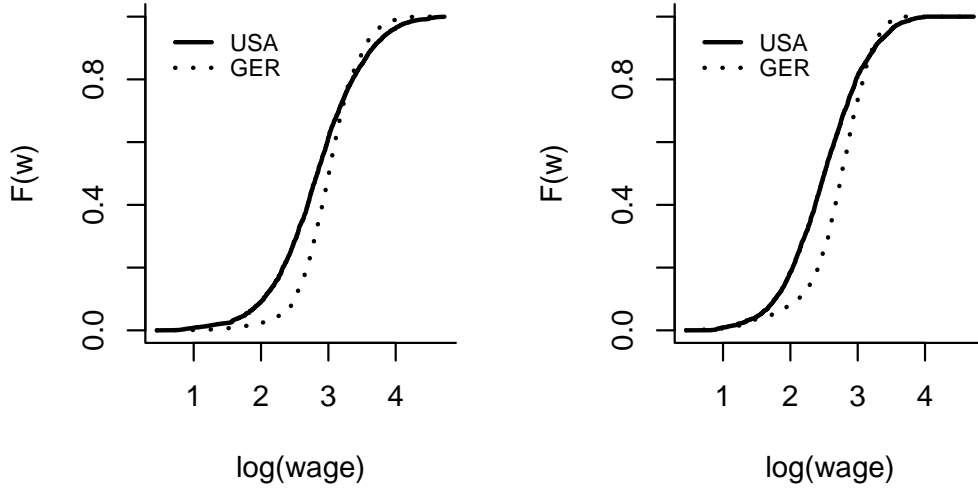


Figure 2: Empirical wage distribution functions (left men, right women)

Table 1: Log-Wage by country, men and women

	USA, men	GER, men	USA, women	GER, women
mean	2.836	2.997	2.505	2.701
median	2.839	3.009	2.506	2.761
sd	0.646	0.435	0.574	0.480
q75-q25	0.809	0.509	0.762	0.534
q90-q10	1.620	1.033	1.479	1.085
n	2,914	1,497	2,895	974

Table 1 contains some descriptive measures of the log-wage distributions depicted in Figure 1. As measures capturing wage inequality, we use standard deviations (*sd*), interquartile ranges and $90^{th} - 10^{th}$ percentile differences. We find that average wages for both men and women to be higher in Germany.⁸ While in the USA wage inequality is higher among men compared to women according to all measures, the opposite holds for Germany.

Means and standard deviations for years of education and experience are given in Table 2. We find only small differences between the groups. Men and women in the USA tend to have slightly more general years of education and slightly less general working experience compared to their German counterparts.

Table 2: Education and experience by country, men and women

	USA, men	GER, men	USA, women	GER, women
eduyears	13.31 (2.1)	12.46 (2.68)	13.12 (1.92)	12.22 (2.52)
experience	20.57 (10.14)	22.26 (9.43)	20.02 (10.16)	22.27 (10.82)

⁸Note that this would be reversed if using nominal exchange rates instead of purchasing power parities.

4.2 Results from the proportional hazards model

Table 3 contains the estimated coefficients and their standard errors of the occupation dummies for country specific proportional hazards models estimated separately for US and German men and women. The reference category is the one referring to elementary occupations including soldiers. Most of the coefficients are quite large documenting considerable dispersion of wages across even broad occupation categories. The line “R-sq.” gives R^2 measures for the models including all covariates based on the marginal likelihoods. They are rescaled so that their theoretical maximum is 1. All the models exhibit a reasonable fit based on this statistic.

Table 3: Effects of occupations in proportional hazards model

	USA, men	GER, men	USA, women	GER, women
Missing	0.244 (0.071)	-0.421 (0.129)	-0.025 (0.097)	-0.349 (0.224)
Technicians	-0.546 (0.081)	-1.123 (0.115)	-0.594 (0.099)	-0.68 (0.236)
Professionals	-0.184 (0.095)	-0.665 (0.142)	-0.44 (0.101)	-0.431 (0.228)
Office worker	-0.528 (0.07)	-0.918 (0.136)	-0.718 (0.103)	-0.702 (0.268)
Clerks	0.13 (0.089)	-0.881 (0.112)	-0.258 (0.09)	-0.464 (0.2)
Sales	-0.476 (0.084)	-0.622 (0.133)	-0.388 (0.116)	-0.205 (0.213)
Service	0.192 (0.085)	-0.289 (0.114)	0.207 (0.1)	-0.046 (0.201)
Production worker	-0.128 (0.084)	-0.281 (0.138)	-0.078 (0.134)	0.398 (0.27)
Craft	-0.213 (0.069)	-0.58 (0.099)	-0.119 (0.157)	0.095 (0.263)
R-sq.	0.34	0.37	0.28	0.31
n	2,914	1,497	2,895	974

While often years of education and general working experience are ad hoc assumed to exert a quadratic influence on log wages, we used natural cubic splines with three degrees of freedom (three interior knots in the range of the variables).

The effect of years of education and experience is visualised in Figures 3 and 4.⁹ The effect of years of education is found to be almost linear for US men and reveals slight concavity for German men. For women we observe for Germany decreasing returns to education while the relationship for the US is somewhat sigmoid, containing an interval of decreasing as well as an interval of increasing returns. The overall effect of years of education is very strong in all cases, while there is barely evidence for non-linearities. This contrasts somewhat with Trostel’s (2005) findings.

⁹The detailed results for the estimation of the coefficients of the natural splines for years of education and experience are given in the appendix.

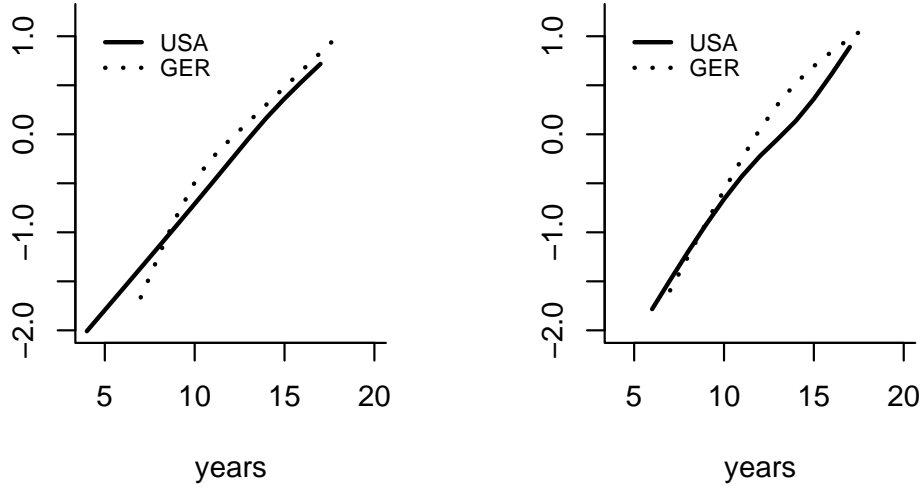


Figure 3: Natural spline for years of education (left men, right women)

For general working experience we find a steep curve until about twenty years of experience. For US employees the relationship is steeper for both genders compared to their German counterparts. For more than twenty years of experience the relationship is only slightly increasing for men and almost horizontal for women. Once again, effects are quite large, with strong indications of a form of non-linearity not easily accommodated by quadratics.

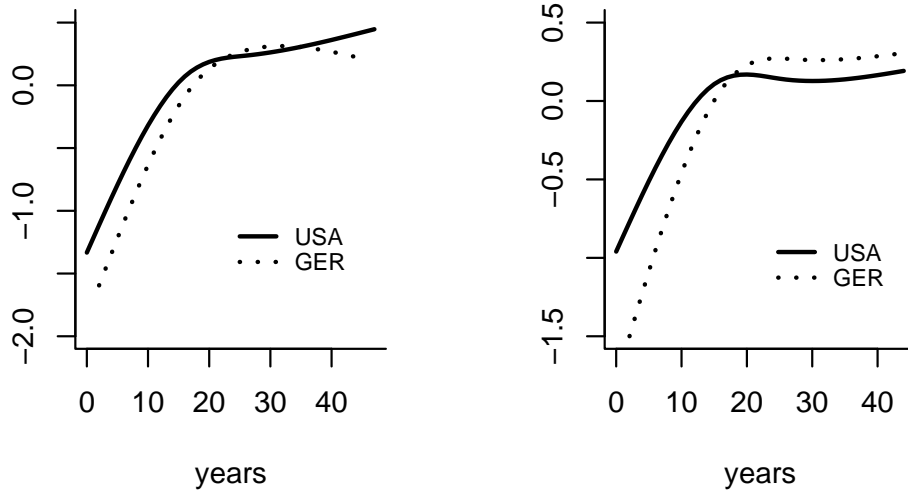


Figure 4: Natural spline for general working experience (left men, right women)

Estimates of individual skill indices $\hat{r}^* = 1/r(x; \hat{\beta})$ are obtained using the observed characteristics and estimated coefficients. Figure 5 shows the distributions of the estimated skill indices. We observe less variation in the female distributions which are located left to the male distributions. For both genders the US distribution is located slightly to the left of the German distribution and shows considerably higher excess kurtosis.

The return-to-skills functions transform the estimated individual amount of skills calculated according to the estimated coefficients $\hat{\beta}$ of the skill function into log-wages. In Figure 6, we show the estimated return-to-skills functions.

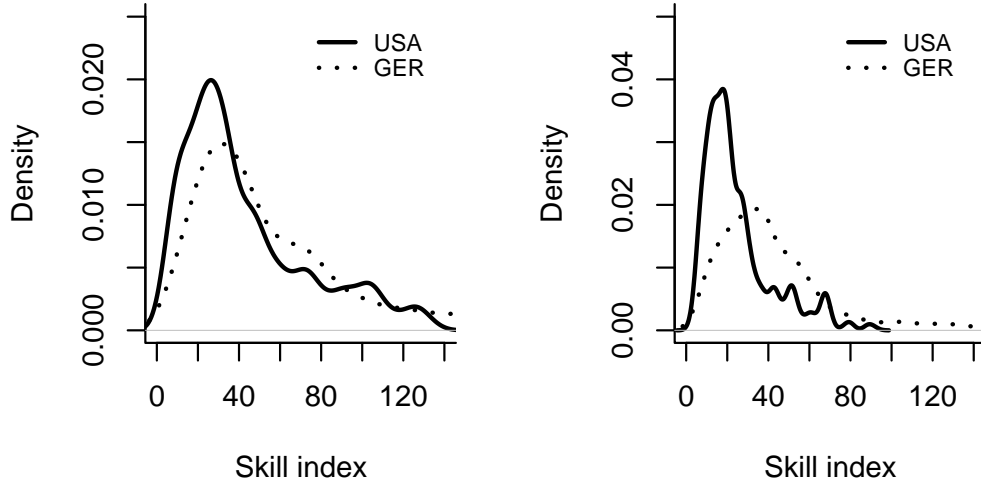


Figure 5: Estimated skill indices (left men, right women)

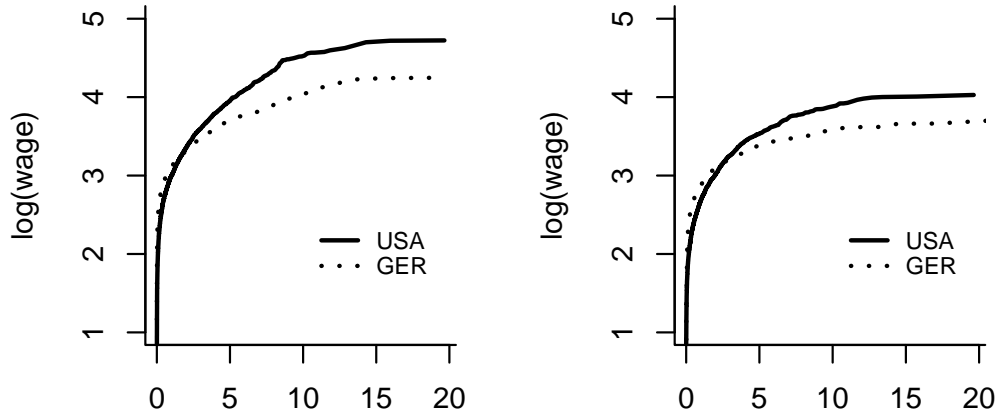


Figure 6: Return-to-skills functions (left men, right women)

By displaying quantiles of the estimated skill indices at the horizontal and the corresponding quantiles of the estimated distribution of log-wages on the y -axis we provide an illustrative presentation of the average relation (Fig. 7). To ensure comparability, the values \hat{r}^* have been normalised.

For both genders the relation between skills and wages are considerably steeper in the USA and cross the German functions at high skill levels. The relationships for Germany reveal a much smaller increase in returns to skills.

In Figure 8 we compare the empirical wage distribution function to the distribution obtained from the proportional hazard model for the USA.

The corresponding distribution functions for Germany are given in Figure 9. Because the distributions implied by the model have a slightly smaller variance than observed wages, the estimated distribution functions are somewhat steeper in the interval of middle wages. The fit is slightly superior for Germany than for the USA.

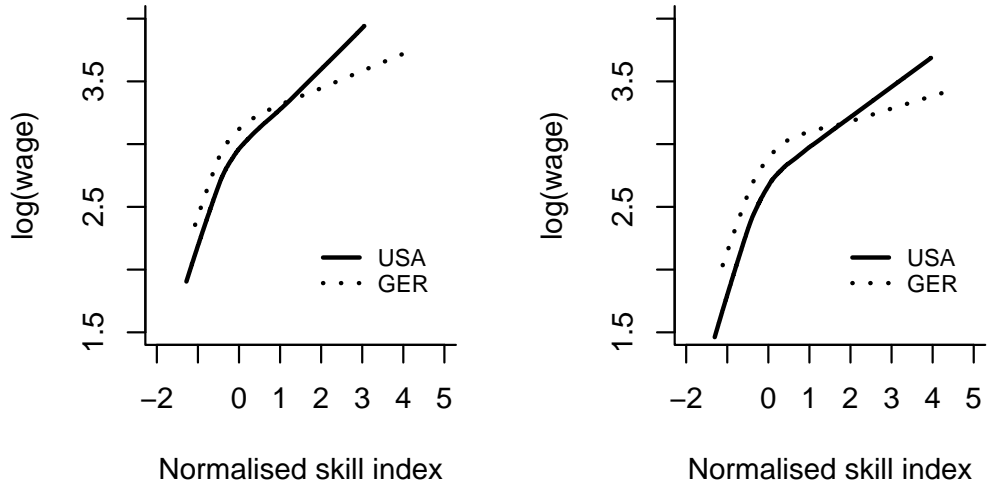


Figure 7: Estimated skills and wages (left men, right women)

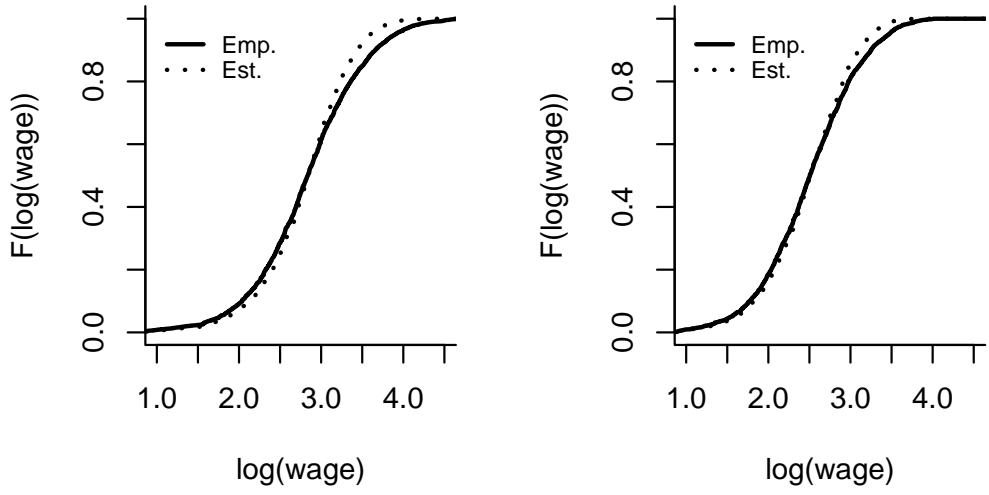


Figure 8: Estimated and empirical distributions, USA (left men, right women)

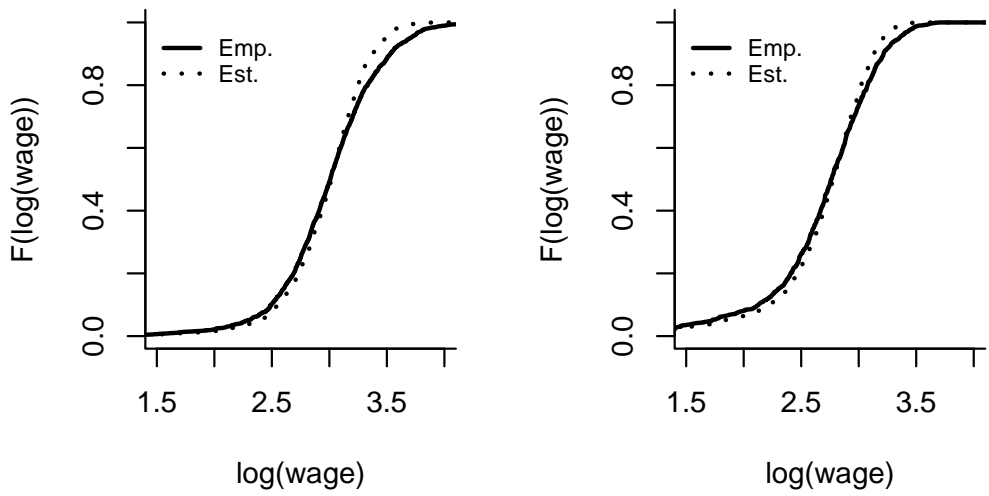


Figure 9: Estimated distribution, Germany (left men, right women)

In Table 4 we compare wage differences between the USA and Germany using the empirical distributions and the distributions implied by the models at selected quantiles. In general, the empirical and the estimated differences at various quantiles are alike, only at the higher quantiles the model slightly underpredicts the empirical wage differences.

Table 4: Wage differences at quantiles

	S, men	\hat{S}, men	$S, women$	$\hat{S}, women$
10%	-0.456	-0.428	-0.360	-0.400
25%	-0.310	-0.289	-0.356	-0.387
50%	-0.170	-0.170	-0.255	-0.263
75%	-0.010	-0.049	-0.128	-0.136
90%	0.130	0.063	0.034	-0.028

Lastly, we present the relative local effects of covariates on conditional quantiles. According to equation (5), the derivative of the conditional quantiles with respect to the covariates is proportional to the regression coefficients when the proportional hazards model holds. In general, the computation of the proportionality factor is difficult since it depends crucially on a pointwise estimate of the derivative of the integrated hazard function. However, the proportionality factor cancels if ratio of regression coefficients are considered. The ratios approximate the change in one variable that is needed to offset a unit change in the other variable. Note that this is constant across all quantiles. Table 5 gives the ratios of local effects on conditional quantiles where the coefficients of the “sales” category are used as denominator. The effects for education and experience are computed using the crossproduct of the respective spline coefficients with the derivatives of the spline basis at the mean of education and experience. At the mean, US men would need roughly 14 years of additional experience to offset a further year of education, it is only 6 years for German men. For women, however, experience has even smaller effect. It is nearly 0 for US women and relatively small for German women.

Table 5: Standardised covariate effects on conditional quantiles

	USA, men	GER, men	USA, women	GER, women
Education	0.454	0.284	0.448	1.313
Experience	0.031	0.044	-0.002	0.042
Missing	-0.513	0.677	0.063	1.705
Technicians	1.146	1.805	1.530	3.323
Professionals	0.387	1.069	1.133	2.104
Office worker	1.110	1.477	1.850	3.426
Clerks	-0.272	1.417	0.665	2.265
Sales	1.000	1.000	1.000	1.000
Service	-0.403	0.465	-0.534	0.223
Production worker	0.270	0.451	0.202	-1.945
Craft	0.448	0.932	0.307	-0.465

4.3 Counterfactual wage distributions based on the components model

To provide insight into different aspects contributing to the overall differences between wage distributions, counterfactual estimates of the USA distribution function for men are given in Figure 10.

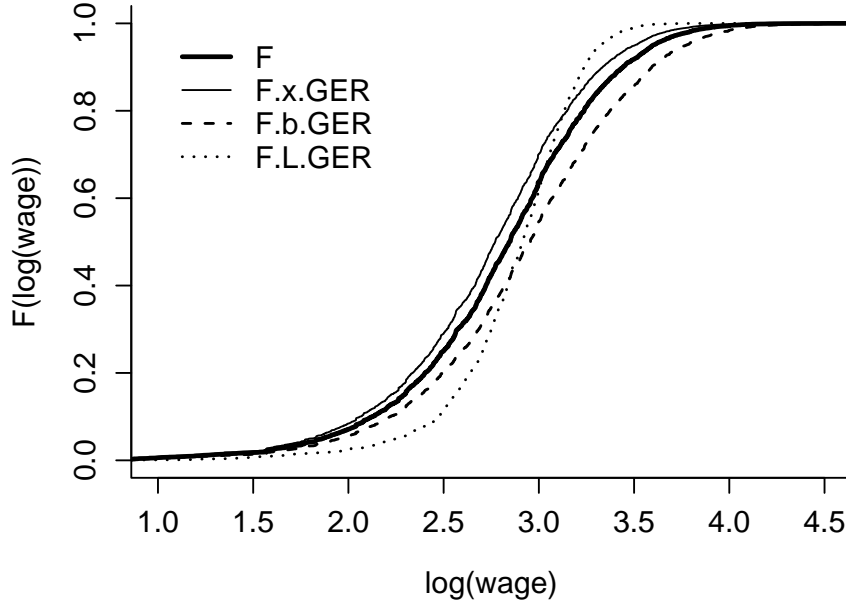


Figure 10: Comparison of contrafactual distributions, men

The distribution function that results from using the German individual characteristics lies above the original estimate throughout. Hence, men in the USA gain from their favourable characteristics. Using German prices shifts the distribution function for US men downwards. Hence, US men would be worse off if their years of education, experience and occupation would be priced at German prices. The largest effect appears when counterfactually using the German return-to-skills function. This shifts the distribution considerably to the right at lower incomes and to the left at higher incomes, leading to a steeper distribution function. Hence, this would make US men considerably better off at lower parts of the income distribution but worse off at higher positions. Therefore, the isolated effect of the return-to-skills function works towards a strong increase in income inequality in the United States.

The analogous estimates for women, given in Figure 11, reveal similar effects of characteristics and of the return-to-skills function. US women would be slightly worse off having the characteristics of their German counterparts and the difference in the return-to-skills functions increases the income inequality in the United States. But by far the strongest effect is the price effect. Weighting the observed characteristics of US women with German prices and pricing with the US return-to-skills function would cause a huge gain for US workers. Hence, the isolated effect of price differences causes the main share of the observed negative income differentials between US and German women. Only at the highest income positions the negative price effect is out-weighted by the positive effects of characteristics, return-to-skills functions and residuals.

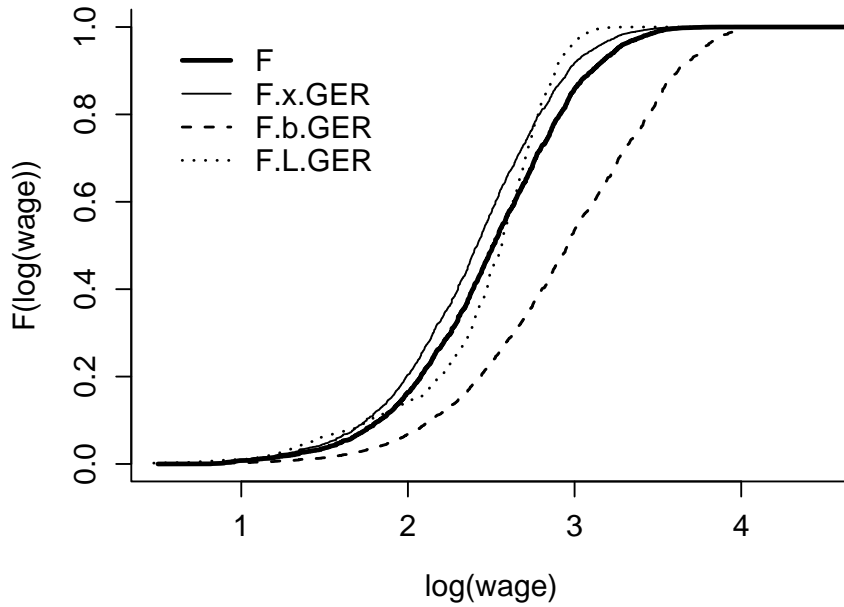


Figure 11: Comparison of contrafactual distributions, women

Therefore, it can be said that while there are almost no differences in measured individual characteristics for both men and women, characteristics are relatively higher weighted for women in Germany. The effect of the return-to-skills function is found positive for high earning women and men in the USA but strongly favourable to German women and men at lower income positions.

4.4 Decomposing the wage differences at quantiles

Based on the empirical and estimated factual and counterfactual wage distribution functions, we now decompose the observed wage differentials at quantiles into underlying sources separately for men and women. Wage differentials are always calculated as USA-wages minus German wages. We find that the three “first order” effects behave quite differently at different quantiles of the wage distribution. This makes evident that traditional Oaxaca-Blinder (Oaxaca 1973) decompositions which focus on single points of the distribution, conceal important differences in wage distributions.

Table 6: Decomposition of wage difference, USA-GER, men

	total	total est.	x-effect	β -effect	Λ -effect
10%	-0.456	-0.428	0.065	-0.094	-0.348
25%	-0.310	-0.289	0.063	-0.095	-0.206
50%	-0.170	-0.170	0.076	-0.109	-0.068
75%	-0.010	-0.049	0.093	-0.128	0.065
90%	0.130	0.063	0.107	-0.151	0.202

For men the column of total differences reveals that there is a substantial wage differential at the lower half of the wage distribution, which vanishes at the 75% quantile and has reversed sign at the 90% quantile. Men earning less than 25%

of their fellow citizens earn -30% less than their counterparts in Germany. This differences decreases strongly at higher quantiles. Men at the 90% quantile in their wage distribution earn 13% more then German men at this quantile.

The 'x-effect' measures the difference that would be observed, if wage differences would solely result from differences in measured characteristics. We find that for men the 'x-effect' has the smallest size of all effects and would lead to a positive differential throughout the distribution. The differences in prices for skills are considerably greater and lead to a negative wage differential. The effect of the return-to-skills function (Λ -effect) is very strong. It increases steeply with the quantiles and changes sign somewhere between the 50% and 75% quantile. Thus, this effect has a strong relative inequality increasing effect on US wages. Therefore, the German labour market favours very strongly the relatively low skilled men compared to the US labour market.

Table 7: Decomposition of wage difference, USA-GER, women

	total	total est.	x-effect	β -effect	Λ -effect
10%	-0.360	-0.400	0.080	-0.308	0.057
25%	-0.356	-0.387	0.080	-0.379	-0.130
50%	-0.255	-0.263	0.097	-0.444	-0.051
75%	-0.128	-0.136	0.123	-0.517	0.102
90%	0.034	-0.028	0.143	-0.515	0.222

For women we find the wage differentials to be smaller compared to men. But the overall pattern is similar to that for men. US women with low wages do much worse relative to their German counterparts, but women positioned in the highest region of the wage distribution earn more than their German counterparts. Nevertheless, the decomposition reveals that here is only a minor isolated effect of differences in skills and opposite to the finding for men, the differences in the return-to-skills functions are positive in the lowest part of the wage distribution. At the median this effect is almost negligible, but at higher quantiles this effect is strongly positive for US women. The most highly skilled are much higher rewarded in the USA than in Germany. The strongest effect is the isolated price effect which is negative throughout the wage distribution and relatively disadvantageous for US women.

5 Conclusion

To analyse empirically observed wage differentials between the USA and Germany we used a new estimation framework. The approach based on rank invariant estimators is borrowed from the literature on failure time data. By showing that the semi-parametric features of a marginal likelihood are appropriate for the analysis of wage decompositions and easy to interpret we improved the approach of Donald et al. (2000). Moreover, we extended their approach by allowing for non-linear regression effects. At least the effect of general experience was found to influence log-wages non-linearly. USA men were found to have negative wage-differentials almost throughout the wage distribution compared to their German counterparts except at highest income positions. The decomposition revealed that the wage

differentials for men can be mostly attributed to differences in the return-to-skills functions which are strongly favourable to low skilled men in Germany compared to the USA. For women this return-to-skills function effect also is more favourable to high paid women but was offset by a negative price effect. Altogether, the evident differences in wage distributions, especially the much higher wage inequality in the USA can be attributed mainly to the steepness of the return-to-skills functions or the overall wage structure despite rather similar underlying skill distributions in both countries.

References

- Angrist, Joshua/Chernozhukov, Victor/Fernández-Val, Iván (2006): Quantile regression under misspecification, with an application to the U.S. wage structure, *Econometrica*, 74, 539–563.
- Beaudry, Paul/Green, David A. (2003): Wages and employment in the United States and Germany: What explains the difference?, *American Economic Review*, 93, 573–602.
- Blau, Francine D./Kahn, Lawrence M. (1996): Wage structure and gender earnings differentials: An international comparison, *Economica*, 63, S29–S62.
- Blau, Francine D./Kahn, Lawrence M. (2003): Understanding international differences in the gender pay gap, *Journal of Labor Economics*, 21, 106–144.
- Dabrowska, Dorota M. (2005): Quantile regression in transformation models, *Sankhyā*, 67, 153–186.
- Deutsche Bundesbank (2006), Devisenkursstatistik, Februar, Statistisches Beiheft zum Monatsbericht 5.
- DiNardo, John/Lemieux, Thomas (1997): Diverging male wage inequality in the United States and Canada, 1981–1988: Do institutions explain the difference?, *Industrial and Labor Relations Review*, 50, 629–651.
- DiNardo, John/Fortin, Nicole M./Lemieux, Thomas (1996): Labor market institutions and the distribution of wages, 1973–1992: A semiparametric approach, *Econometrica*, 64, 1001–1044.
- Donald, Stephen G./Green, David A./Paarsch, Harry J. (2000): Differences in wage distributions between Canada and the United States: An application of a flexible estimator of distribution functions on the presence of covariates, *Review of Economic Studies*, 67, 609–633.
- Fleming, Thomas R./Harrington, David P. (1991): *Counting Processes & Survival Analysis*. Wiley.
- Fortin, Nicole M./Lemieux, Thomas (1998): Rank regressions, wage distributions, and the gender gap, *The Journal of Human Resources*, 33, no. 3, 610–643.
- Hettmansperger, Thomas P. (1984): *Statistical Inference based on Ranks*. Wiley.

- Juhn, Chinhui/Murphy, Kevin M./Brooks, Pierce (1991): Accounting for the Slowdown in Black-White Convergence, in Kosters, Marvin H. (ed.) *Workers and Their Wages*, Washington DC, 107–143.
- Juhn, Chinhui/Murphy, Kevin M./Brooks, Pierce (1993): Wage inequality and the rise in returns to skill, *Journal of Political Economy*, 101, 410–442.
- Kalbfleisch, John D./Prentice, Ross L. (2002²): *Statistical Analysis of Failure Time Data*, Wiley.
- Koenker, Roger (2005): *Quantile Regression*. Cambridge University Press.
- Martins, Pedro S./Pereira, Pedro T. (2004): Does education reduce wage inequality? Quantile regression evidence from 16 countries, *Labour Economics*, 11, 355–371.
- Machado, Jose A. F. / Mata, Jose (2005): Counterfactual decomposition of changes in wage distributions using quantile regression, *Journal of Applied Econometrics*, 20, 445–465.
- Oaxaca, Ronald (1973): Male-female wage differentials in urban labor markets, *International Economic Review*, 14, 693–709.
- Statistisches Bundesamt (2002, 2006): Fachserie 17, Reihe 10, Preise, Internationaler Vergleich der Verbraucherpreise, Februar.
- Suen, Wing (1997): Decomposing wage residuals: Unmeasured skill or statistical artifact? *Journal of Labor Economics*, 15, 555–566.
- Therneau, Terry M./Grambsch, Patricia M. (2000): *Modeling Survival Data. Extending the Cox Model*. Springer.
- Trostel, Philip A. (2005): Nonlinearity in the return to education, *Journal of Applied Economics*, 8, 191–202.

6 Appendix

Table 8: Occupations

Group	Description
0	Missing
1	Technicians and associate professionals
2	Professionals
3	Office worker
4	Clerks
5	Sales
6	Service
7	Agricultural worker, production worker
8	Craft
9	Elementary occupations, soldier

Table 9: Effects of occupations in cox model

	USA, men	GER, men	USA, women	GER, women
$ns(Eduyears)$ 1	-1.84 (0.199)	-1.398 (0.187)	-1.482 (0.164)	-1.735 (0.209)
$ns(Eduyears)$ 2	-3.694 (0.662)	-3.407 (0.595)	-3.615 (0.542)	-3.378 (0.668)
$ns(Eduyears)$ 3	-2.014 (0.145)	-2.069 (0.172)	-1.914 (0.127)	-2.195 (0.22)
$ns(Experience)$ 1	-1.024 (0.087)	-1.413 (0.122)	-0.62 (0.082)	-1.28 (0.147)
$ns(Experience)$ 2	-3.03 (0.241)	-3.318 (0.349)	-2.172 (0.218)	-3.778 (0.402)
$ns(Experience)$ 3	-0.983 (0.161)	-0.946 (0.155)	-0.509 (0.119)	-0.92 (0.167)
n	2,914	1,497	2,895	974

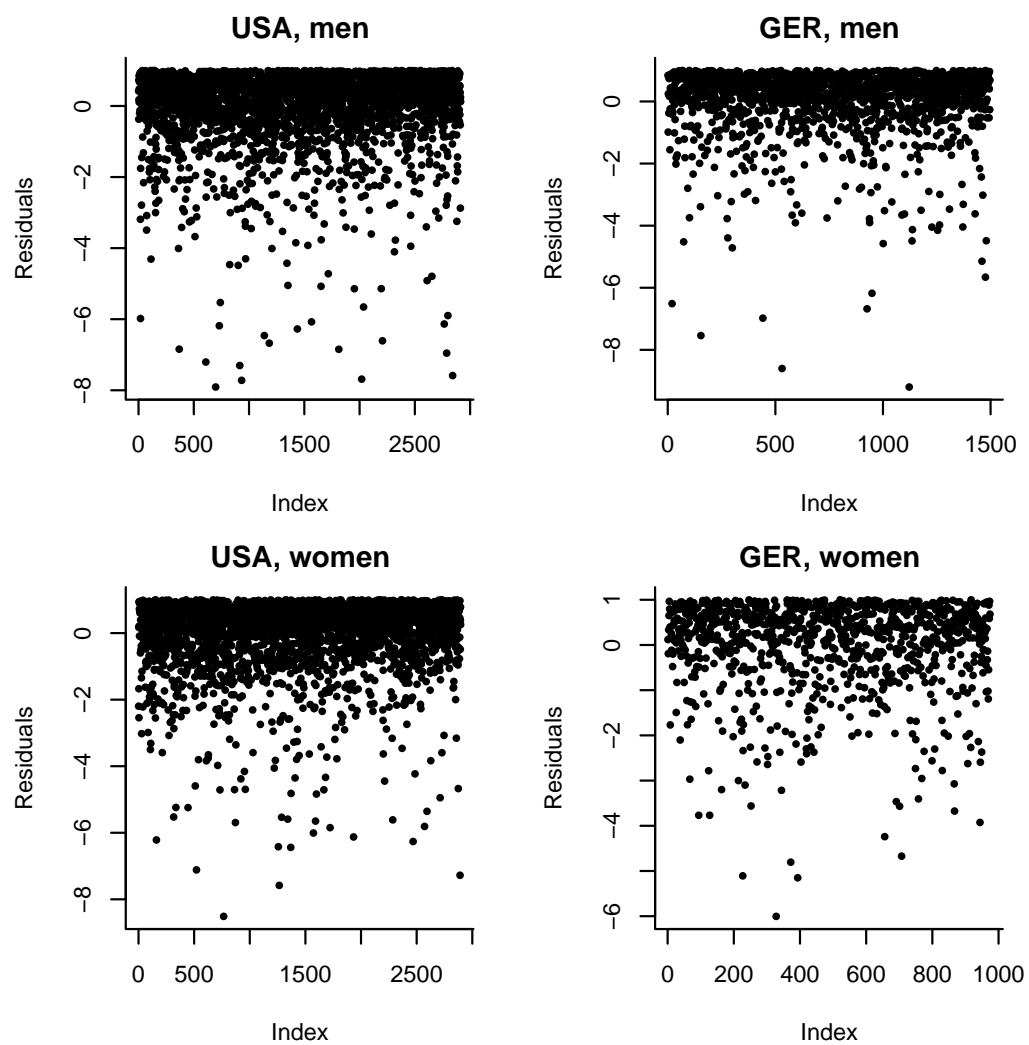


Figure 12: Plot of residuals