Blinder–Oaxaca decomposition for Tobit models

Bauer, Thomas K.; Sinning, Mathias

Postprint / Postprint
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
www.peerproject.eu

Empfohlene Zitierung / Suggested Citation:
Bauer, T. K., & Sinning, M. (2010). Blinder–Oaxaca decomposition for Tobit models. *Applied Economics*, *42*(12), 1569-1575. https://doi.org/10.1080/00036840701721612

Blinder-Oaxaca Decomposition for Tobit Models

Journal:	*Applied Economics*
Manuscript ID:	APE-07-0083.R1
Journal Selection:	Applied Economics
Date Submitted by the Author:	22-May-2007
Complete List of Authors:	Bauer, Thomas; RWI Essen; Ruhr-University Bochum; IZA Bonn; CEPR London Sinning, Mathias; RWI Essen; IZA Bonn
JEL Code:	C24 - Truncated and Censored Models < C2 - Econometric Methods: Single Equation Models < C - Mathematical and Quantitative Methods, J31 - Wage Level, Structure; Differentials by Skill, Occupation, etc. < J3 - Wages, Compensation, and Labor Costs < J - Labor and Demographic Economics
Keywords:	Blinder-Oaxaca decomposition, Tobit model, gender wage gap

Blinder-Oaxaca Decomposition for Tobit Models

Thomas K. Bauer
RWI Essen, Ruhr-University Bochum, IZA Bonn and CEPR London

Mathias Sinning
RWI Essen and IZA Bonn

May 2007

Abstract. In this paper, a decomposition method for Tobit-models is derived, which allows the differences in observed outcome variables between two groups to be decomposed into a part that is explained by differences in observed characteristics and a part attributable to differences in the estimated coefficients. Monte Carlo simulations demonstrate that in the case of censored dependent variables this decomposition method produces more reliable results than the conventional Blinder-Oaxaca decomposition for linear regression models. Finally, our method is applied to a decomposition of the gender wage gap using German data.

JEL-Classification: C24, J31

Keywords: Blinder-Oaxaca decomposition, Tobit model, gender wage gap

* The authors are grateful to Michael Fertig, John Haisken-DeNew and Harald Tauchmann for helpful comments on an earlier draft of this paper. All correspondence to Mathias Sinning, Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI) Essen, Hohenzollernstr. 1-3, 45128 Essen, Germany, Fax: +49-201-8149-284, Email: sinning@rwi-essen.de.

1 Introduction

The decomposition method developed by Blinder (1973) and Oaxaca (1973) and generalized by Juhn, Murphy, and Pierce (1991), Neumark (1988), and Oaxaca and Ransom (1988, 1994) is a very popular descriptive tool, since it permits the decomposition of the difference in an outcome variable between two groups into a part that is explained by differences in the observed characteristics of these groups and a part that is due to differences in the estimated coefficients.

So far, the Blinder-Oaxaca-decomposition and its various generalizations have mainly been used in linear regression models. A decomposition method for models with binary dependent variables has been developed by Fairlie (1999, 2003). In many cases, however, censoring requires the estimation of a Tobit-model because OLS yields inconsistent parameter estimates and in turn misleading decomposition results. In this paper, a Blinder-Oaxaca type decomposition method for Tobit-models is derived. To compare the results obtained from this method to the results obtained when the conventional Blinder-Oaxaca decomposition method for linear regression models is used despite a censored dependent variable, a Monte Carlo simulation is carried out. Finally, our method is applied to a decomposition of the gender wage gap using data from the German *Socio-Economic Panel* (SOEP).

2 Blinder-Oaxaca Decomposition for Tobit Models

Consider the following linear regression model, which is estimated separately for the groups $g = m,f$

$$Y_{ig} = X_{ig}\beta_g + \varepsilon_{ig}, \tag{1}$$

for $i = 1,\dots,N_g$, and $\Sigma_g N_g = N$. For these models, Blinder (1973) and Oaxaca (1973) propose the decomposition

$$\begin{aligned}\bar{Y}_m - \bar{Y}_f = \Delta^{OLS} &= [E_{\beta m}(Y_{im} \mid X_{im}) - E_{\beta m}(Y_{if} \mid X_{if})] \\ &\quad + [E_{\beta m}(Y_{if} \mid X_{if}) - E_{\beta_f}(Y_{if} \mid X_{if})] \\ &= (\bar{X}_m - \bar{X}_f)\hat{\beta}_m + \bar{X}_f(\hat{\beta}_m - \bar{\beta}_f),\end{aligned} \tag{2}$$

where $\bar{Y}_g = N_g^{-1}\Sigma_{i=1}^{N_g} Y_{ig}$ and $\bar{X}_g = N_g^{-1}\Sigma_{i=1}^{N_g} X_{ig}$. $E_{\beta_g}(Y_{ig} \mid X_{ig})$ refers to the conditional expectation of Y_{ig} evaluated at the parameter vector β_g. The first term on the right hand side of equation (2) displays the difference in the outcome variable between the two groups due to differences in observable characteristics. The second term shows the differential that is due to differences in coefficient estimates.[1] [2]

Given X_{ig}, the linear model is a good approximation of the expected value of the outcome variable $E(Y_{ig}|X_{ig})$ for values of X_{ig} close to the mean. If the outcome variable Y_{ig} is censored, however, the use of OLS may lead to biased estimates of the parameter vector and hence misleading results of the decomposition. To illustrate the Blinder-Oaxaca decomposition in the presence of censoring, we consider a Tobit model, where the dependent variable takes on the values a_1 and a_2 with positive probability and represents a continuous random variable over values between a_1 and a_2, i.e.

$$\begin{aligned}
&Y_{ig}^* = X_{ig}\beta_g + \varepsilon_{ig}, \\
&Y_{ig} = a_1 \quad \text{if} \quad Y_{ig}^* \le a_1 \\
&Y_{ig} = a_2 \quad \text{if} \quad Y_{ig}^* \ge a_2 \\
&Y_{ig} = Y_{ig}^* = X_{ig}\beta_g + \varepsilon_{ig} \quad \text{if} \quad a_1 < Y_{ig}^* < a_2, \\
&\varepsilon_{ig} \sim N(0, \sigma_g^2).
\end{aligned} \tag{3}$$

Given that we are interested in the marginal effects of the latent censored outcome variable, $\partial E(Y_{ig}^* \mid X_{ig}) / \partial X_{ig} = \beta_g$, a solution to this problem would be the use of the Tobit estimator in the standard Blinder-Oaxaca decomposition of equation (2). However, the conventional decomposition method leads to erroneous predictions of the components of the decomposition equation if we aim at analyzing the observed outcome variable Y_i instead of the unobserved counterpart Y_i^*, i.e. the parameters of interest are given by the marginal effects of the observed outcome

---

[1] Based on this decomposition, Fairlie (1999, 2003) develops the following decomposition equation for models with binary dependent variables: $\overline{Y}_m - \overline{Y}_f = [F(X_{im}\beta_m) - F(X_{if}\beta_m)] + [F(X_{if}\beta_m) - F(X_{if}\beta_f)]$, where $F(\cdot)$ is the cumulative distribution function (a logistic distribution is assumed for a Logit model and a standard normal distribution for a Probit model).

[2] Oaxaca and Ransom (1994) demonstrate that this decomposition may be considered as a special case of the decomposition $\overline{Y}_m - \overline{Y}_f = (\overline{X}_m - \overline{X}_f)\hat{\beta}^* + \overline{X}_m(\hat{\beta}_m - \hat{\beta}^*) + \overline{X}_f(\hat{\beta}^* - \hat{\beta}_f)$, where $\hat{\beta}^* = \Omega_\beta \hat{\beta}_m + (I - \Omega_\beta)\beta_f$ and Ω_β is equal to the identity matrix *I*. A consideration of equation (2), however, is sufficient for the derivation of the decomposition method for Tobit-models. The generalization of the decomposition method to the equation proposed by Oaxaca and Ransom (1994) is straightforward.

variable, $\partial E(Y_{ig} \mid X_{ig}) / \partial X_{ig} = \Phi(\beta_g, X_g, \sigma_g)\beta_g$,where $\Phi(\cdot)$ represents the cumulative standard normal density function. In this case, an alternative decomposition method must be applied.

Assuming homoscedastic and normally distributed error terms ε_{ig} ,the unconditional expectation of Y_{ig} given X_{ig} consists of the conditional expectations of Y_{ig} weighted with the respective probabilities of observing a_1, a_2, or a value between a_1 and a_2:

$$\begin{aligned} E(Y_{ig} \mid X_{ig}) &= a_1\Phi_1(\beta_g, X_g, \sigma_g) + a_2\Phi_2(\beta_g, X_g, \sigma_g) \\ &+\Lambda(\beta_g, X_g, \sigma_g)\left[X_{ig}\beta_g + \sigma\frac{\lambda(\beta_g, X_g, \sigma_g)}{\Lambda(\beta_g, X_g, \sigma_g)}\right], \end{aligned} \tag{4}$$

where $\Phi_1(\beta_g, X_g, \sigma_g) = \Phi[\sigma_g^{-1}(a_1 - X_{ig}\beta_g)]$, $\Phi_2(\beta_g, X_g, \sigma_g) = 1 - \Phi[\sigma_g^{-1}(a_2 - X_{ig}\beta_g)]$, $\Lambda(\cdot) = \Phi_2(\cdot) - \Phi_1(\cdot)$ and $\lambda(\beta_g, X_g, \sigma_g) = \varnothing[\sigma_g^{-1}(a_1 - X_{ig}\beta_g)] - \varnothing[\sigma_g^{-1}(a_2 - X_{ig}\beta_g)]$; $\varnothing(\cdot)$ represents the standard normal density function.

Equation (4) shows that a decomposition of the outcome variable similar to equation (2) is not appropriate for the observed outcome variable of the Tobit-model, because the conditional expectations $E(Y_{ig}|X_{ig})$ in the Tobit-model depend on the standard error σ_g. Even though the ancillary parameter σ_g does not affect the sign of the marginal effects, it affects their magnitudes and therefore becomes important for the decomposition. Depending on which σ_g is used in the counterfactual parts of the decomposition equation, several possibilities of decomposing the mean difference of Y_i between the two groups can be derived. Two possibilities are

$$\begin{aligned} \Delta_f^{Tobit} &= [E_{\beta_m,\sigma_m}(Y_{im} \mid X_{im}) - E_{\beta_m,\sigma_f}(Y_{if} \mid X_{if})] \\ &+[E_{\beta_m,\sigma_f}(Y_{if} \mid X_{if}) - E_{\beta_f,\sigma_f}(Y_{if} \mid X_{if})], \end{aligned} \tag{5}$$

$$\begin{aligned} \Delta_m^{Tobit} &= [E_{\beta_m,\sigma_m}(Y_{im} \mid X_{im}) - E_{\beta_m,\sigma_m}(Y_{if} \mid X_{if})] \\ &+[E_{\beta_m,\sigma_m}(Y_{if} \mid X_{if}) - E_{\beta_f,\sigma_f}(Y_{if} \mid X_{if})], \end{aligned} \tag{6}$$

where $E_{\beta_o,\sigma_o}(Y_{ig} \mid X_{ig})$ now refers to the conditional expectation of Y_{ig} evaluated at the parameter vector β_g and the standard error σ_g. In both equations, the first term on the right hand side displays the part of the differential in the outcome variable between the two groups that is due to differences in the covariates X_{ig}, and the second term the part of the differential in Y_{ig} that is due to differences in coefficients.

The two versions of the decomposition equation may differ from each other, if large differences in the variance of the error term between the two groups exist. Note however, that the decomposition using σ_f to calculate the counterfactual parts, as in equation (5), is more comparable to the OLS decomposition described in equation (2), since the counterfactual parts differ from $E_{\beta_f,\sigma_f}(Y_{if} \mid X_{if})$ only by using the parameter vector for group *m*, β_m, rather than by using the parameter vector *and* the standard error for group *m* in the alternative decomposition described in equation (6).

Defining the sample counterpart $S(\cdot)$ of equation (4),

$$S(\hat{\beta}_g, X_{ig}, \hat{\sigma}_g) \equiv N^{-1} \sum_{i=1}^{N} \Big\{ a_1 \Phi_1(\hat{\beta}_g, X_{ig}, \hat{\sigma}_g) + a_2 \Phi_2(\hat{\beta}_g, X_{ig}, \hat{\sigma}_g) + \Lambda(\hat{\beta}_g, X_{ig}, \hat{\sigma}_g) \left[X_{ig}\hat{\beta}_g + \hat{\sigma}_g \frac{\lambda(\hat{\beta}_g, X_{ig}, \hat{\sigma}_g)}{\Lambda(\hat{\beta}_g, X_{ig}, \hat{\sigma}_g)} \right] \Big\}$$

equation (5) can be estimated by

$$\begin{aligned} \Delta_f^{Tobit} &= [S(\hat{\beta}_m, X_{im}, \hat{\sigma}_m) - S(\hat{\beta}_m, X_{if}, \hat{\sigma}_f)] \\ &\quad + [S(\hat{\beta}_m, X_{if}, \hat{\sigma}_f) - S(\hat{\beta}_f, X_{if}, \hat{\sigma}_f)]. \end{aligned} \tag{7}$$

Similarly, equation (6) can be estimated by

$$\begin{aligned} \Delta_m^{Tobit} &= [S(\hat{\beta}_m, X_{im}, \hat{\sigma}_m) - S(\hat{\beta}_m, X_{if}, \hat{\sigma}_m)] \\ &\quad + [S(\hat{\beta}_m, X_{if}, \hat{\sigma}_m) - S(\hat{\beta}_f, X_{if}, \hat{\sigma}_f)]. \end{aligned} \tag{8}$$

If the dependent variable is not censored, i.e. if $a_1 \to -\infty$ and $a_2 \to \infty$, both equations reduce to the original Blinder-Oaxaca decomposition described in equation (2).

3 Monte Carlo Simulation

To examine the differences between the OLS and the Tobit decomposition, a Monte Carlo simulation with 1,000 replications is applied, departing from the following linear regression model for the two groups g = 1,2:

$$y_{ig} = a_g + b_g v_g x_{ig} + w_g e_{ig},$$

with i = 1,..., N_1 ,N_1 + 1,..., N_1 + N_2 and N_1 = 500, N_2 = 500. The explanatory variable x and the error term e represent normally distributed random variables. The model parameters a and b are assumed to be equal for the two groups: $a_1 = a_2 = 1$, $b_1 = b_2 = 5$. In order to ensure that $y_1 > y_2$, the random variables are weighted using the following weights: $v_1 = 1.5$, $v_2 = 1$, $w_1 = 4$, $w_2 = 1$.[3]

Figure 1 displays the deviations of the explained and unexplained parts of the OLS and the Tobit decomposition when the dependent variable is censored from below and/or above at a certain percentile. While the gap between the original decomposition and the values estimated by the two models is zero if the dependent variable is uncensored, the decomposition estimates deviate from the original decomposition as soon as the dependent variable is censored. Moreover, the gap widens if the degree of censoring increases. Overall, the results of the Monte Carlo simulation indicate that the Tobit decomposition produces more reliable results than the decomposition based on OLS estimates. However, the differences between the components of the two decomposition methods are not statistically significant.

4 Empirical Illustration: The Gender Wage Gap in Germany

In order to apply the Blinder-Oaxaca decomposition for Tobit-models empirically, we analyze the gender wage gap using data from the German *Socio-Economic Panel* (SOEP) for the year 2004. We estimate the following wage equation separately for males *(m)* and females *(f)*:

$$\ln(w_{ig}) = X_{ig}\beta_g + \varepsilon_{ig}\,, \tag{9}$$

for $g = m, f$, where w_{ig} refers to the gross hourly wage rate of individual i in group g. The explanatory variables X_i include the years of completed schooling, potential labor market experience (calculated as *Age – Years of Schooling – 6)* and potential labor market experience squared, the number of children, and dummy variables for married individuals, part-time workers, immigrants, and persons residing in East-Germany.[4] We restrict our sample to working individuals aged 16 to 65 and

[3] Note that all parameters of the model were chosen arbitrarily. Alternative simulations were also carried out using different parameters. In addition, different choices of the error term σ were made. The choice of model parameters and error terms, however, did not change our results significantly.

[4] Descriptive statistics are given in the Appendix.

eliminated all observations with missing values for at least one of the variables used in the analysis. This procedure results in a sample of 3,610 observations for men and 2,465 observations for women.

The wage information in the SOEP is uncensored. Therefore, we first apply the original Blinder-Oaxaca decomposition described in equation (2), using the OLS-estimates of the regression model (9). In a second step, we generate an artificial outcome variable by censoring the distribution of gross hourly wages at the lower and upper 10th percentile. We estimate equation (9) by OLS using the transformed wage information as dependent variable to show the potential bias in the estimation results and wage decomposition when ignoring the censoring of the dependent variable. In a final step, we use the transformed wage variable and estimate equation (9) using a Tobit model and apply the Tobit-Blinder-Oaxaca decompositions described in equations (7) and (8). To be able to test the different decomposition results against each other, we obtain standard errors for the decomposition parts by bootstrapping with 1000 replications.

Table 1 shows the estimates of the OLS and Tobit-models. In all cases, the estimated coefficients have the expected signs and are statistically significant at conventional levels. When using the artificially censored outcome variable, the Tobit estimates perform slightly better than the OLS-estimates in the sense that they are closer to the respective estimation results when using the original uncensored wage information.

Based on the uncensored wage information, the estimates of the decomposition analysis reported in column 1 of Table 2 (which do not differ between the OLS and the various Tobit-decomposition methods) indicate that 67.6% of the wage differential between men and women is attributable to differences in observable characteristics.

When using the original Blinder-Oaxaca decomposition *(Δ^{OLS})*, censoring the dependent variable from below or from both sides of the wage distribution increases the unexplained part of the wage differential, while the decomposition results do not change very much when wages are censored just from above. Furthermore, for left-censoring and censoring from both sides of the wage distribution the Tobit decomposition methods perform better than the original Blinder-Oaxaca decomposition. In our example, t-tests demonstrate, however, that the differences in the decomposition results between the uncensored and the three censored estimations are not statistically significant.

5 Conclusion

In this paper, a decomposition method for Tobit-models is developed, which allows the decomposition of the difference in a censored outcome variable between two groups into a part that is explained by differences in the observed characteristics and a part attributable to differences in the estimated coefficients of these characteristics. Monte Carlo simulations reveal that this decomposition method produces more reliable results than applying the conventional Blinder-Oaxaca decomposition for linear regression models in the case of censored dependent variables. Using data of the SOEP, we find that the major part of the wage differential between men and women is attributable to differences in observable characteristics. The results further show that applying the Tobit decomposition method produces better results than the original Blinder-Oaxaca decomposition when wages are censored from below and from both sides of the wage distribution. However, in our example the differences between the various decomposition methods are not statistically significant, confirming the results of the Monte Carlo simulation.

6 Tables and Figures

6.1 Monte Carlo Simulation

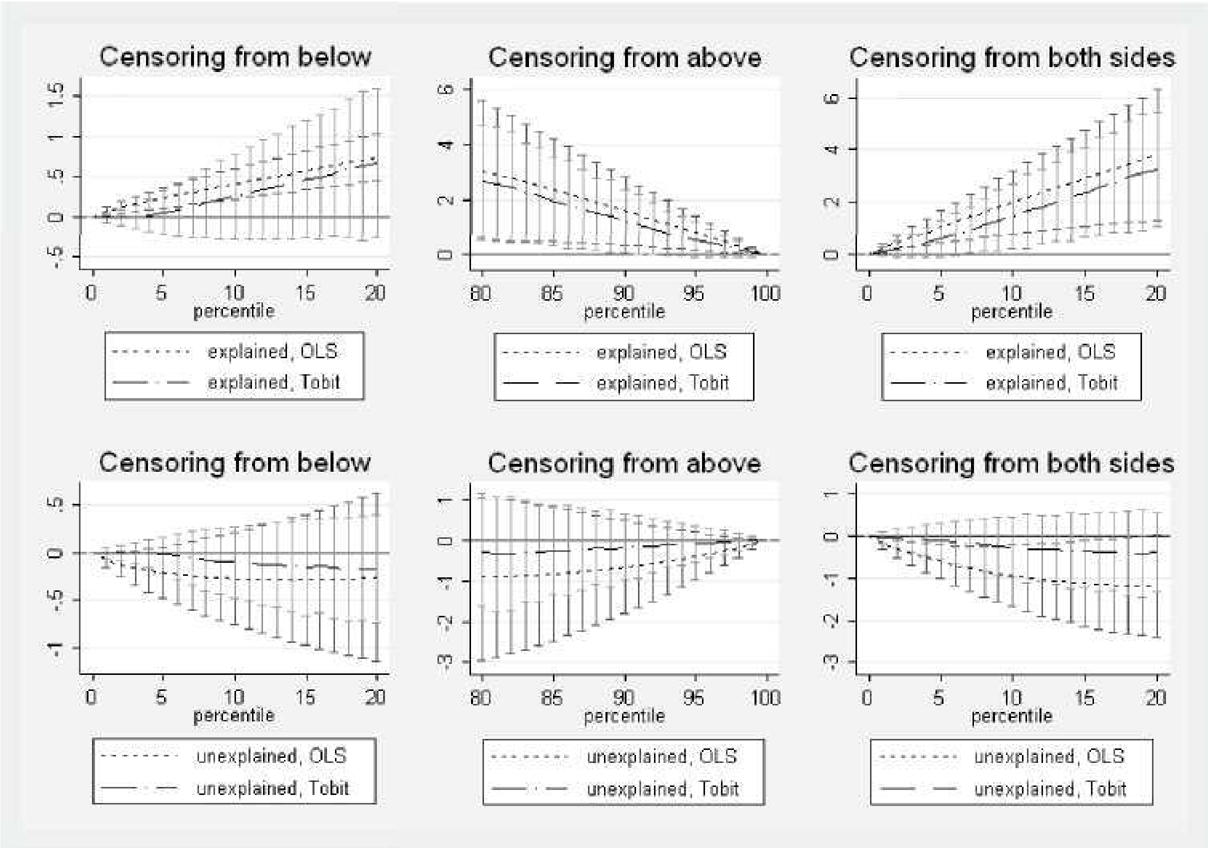

Figure 1: Monte Carlo simulation, OLS and Tobit estimates, 1,000 replications

6.2 Empirical Application

Table 1: Estimation Results

	A: OLS estimates		**B: Tobit estimates**	
	Men	**Women**	**Men**	**Women**
		uncensored		
Education (Yrs.)	0.085***	0.077***	0.085***	0.077***
	(0.003)	(0.004)	(0.003)	(0.004)
Experience	0.027***	0.035***	0.027***	0.035***
	(0.003)	(0.004)	(0.003)	(0.004)
Experience^2/100	-0.031***	-0.062***	-0.031***	-0.062***
	(0.007)	(0.008)	(0.007)	(0.008)
Constant	1.223***	1.136***	1.223***	1.136***
	(0.054)	(0.064)	(0.054)	(0.064)
R^2/McFadden Pseudo R^2	0.38	0.26	0.27	0.19
		left-censored		
Education (Yrs.)	0.082***	0.074***	0.086***	0.083***
	(0.003)	(0.003)	(0.003)	(0.004)
Experience	0.024***	0.029***	0.026***	0.035***
	(0.003)	(0.003)	(0.003)	(0.004)
Experience^2/100	-0.027***	-0.048***	-0.029***	-0.059***
	(0.006)	(0.006)	(0.007)	(0.007)
Constant	1.306***	1.247***	1.230***	1.230***
	(0.048)	(0.052)	(0.052)	(0.061)
R^2/McFadden Pseudo R^2	0.40	0.30	0.29	0.22
		right-censored		
Education (Yrs.)	0.065***	0.070***	0.085***	0.085***
	(0.002)	(0.004)	(0.003)	(0.004)
Experience	0.024***	0.033***	0.027***	0.035***
	(0.003)	(0.003)	(0.003)	(0.004)
Experience^2/100	-0.030***	-0.059***	-0.033***	-0.061***
	(0.006)	(0.007)	(0.007)	(0.008)
Constant	1.482***	1.238***	1.215***	1.166***
	(0.046)	(0.060)	(0.054)	(0.062)
R^2/McFadden Pseudo R^2	0.36	0.25	0.27	0.19
		left/right-censored		
Education (Yrs.)	0.064***	0.068***	0.085***	0.085***
	(0.002)	(0.003)	(0.003)	(0.004)
Experience	0.021***	0.027***	0.025***	0.034***
	(0.003)	(0.003)	(0.003)	(0.003)
Experience^2/100	-0.026***	-0.046***	-0.030***	-0.058***
	(0.005)	(0.006)	(0.006)	(0.007)
Constant	1.540***	1.339***	1.242***	1.108***
	(0.040)	(0.048)	(0.051)	(0.059)
R^2/McFadden Pseudo R^2 R^2	0.41	0.31	0.29	0.23

Notes: 3,610 observations for men and 2,465 observations for women. Standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Additional variables used: number of children and dummy-variables for marital status, part-time employment, immigrants and East-Germany.

Table 2: Decomposition Results

	(1)	(2)	(3)	(4)
	uncensored	**left-censored**	**right-censored**	**left/right-censored**
$\hat{\Delta}^{OLS}$	0.326***	0.301***	0.288***	0.288***
	(0.014)	(0.012)	(0.013)	(0.012)
Explained Part	0.220***	0.173***	0.198***	0.153***
	(0.025)	(0.019)	(0.023)	(0.017)
in % of $\hat{\Delta}^{OLS}$	67.6***	57.3***	68.8***	57.1***
	(7.7)	(6.1)	(8.3)	(6.2)
Unexplained Part	0.105***	0.128***	0.089***	0.115***
	(0.026)	(0.019)	(0.025)	(0.017)
in % of $\hat{\Delta}^{OLS}$	32.3***	42.6***	31.1***	42.8***
	(7.7)	(6.1)	(8.3)	(6.2)
$\hat{\Delta}_f^{Tobit}$	0.326***	0.301***	0.293***	0.270***
	(0.014)	(0.013)	(0.013)	(0.013)
Explained Part	0.220***	0.189***	0.194***	0.164***
	(0.025)	(0.019)	(0.024)	(0.017)
in % of $\hat{\Delta}^{Tobit}$	67.6***	62.7***	66.3***	60.6***
	(7.7)	(6.1)	(8.2)	(6.4)
Unexplained Part	0.105***	0.112***	0.098***	0.106***
	(0.026)	(0.019)	(0.025)	(0.018)
in % of $\hat{\Delta}^{Tobit}$	32.3***	37.2***	33.6***	39.3***
	(7.7)	(6.1)	(8.2)	(6.4)
	0.453	0.420	0.439	0.399
$\hat{\Delta}_m^{Tobit}$	0.326***	0.301***	0.293***	0.270***
	(0.014)	(0.013)	(0.013)	(0.013)
Explained Part	0.220***	0.187***	0.195***	0.163***
	(0.025)	(0.018)	(0.024)	(0.017)
in % of $\hat{\Delta}^{Tobit}$	67.6***	62.1***	66.5***	60.3***
	(7.7)	(6.0)	(8.3)	(6.3)
Unexplained Part	0.105***	0.114***	0.098***	0.107***
	(0.026)	(0.019)	(0.025)	(0.018)
in % of $\hat{\Delta}^{Tobit}$	32.3***	37.8***	33.4***	39.6***
	(7.7)	(6.0)	(8.3)	(6.3)
$\hat{\delta}^m$	0.455	0.431	0.444	0.412

Notes: Decomposition results based on the regression results in Table 1. Bootstrapped (1,000 replications) standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

References

BLINDER, A. S. (1973): „Wage Discrimination: Reduced Form and Structural Estimates," *Journal of Human Resources,* 8, 436-455. FAIRLIE, R. W. (1999): „The Absence of the African-American Owned Business: An Analysis of the Dynamics of Self-Employment," *Journal of Labor Economics,* 17, 80-108.

----------- (2003): „An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models," *Yale University Economic Growth Center Discussion Paper No. 873,* pp. 1-11.

JUHN, C., K. M. MURPHY, AND B. PIERCE (1991): „Accounting for the Slowdown in Black-White Wage Convergence," in *Workers and Their Wages: Changing Patterns in the United States,* ed. by M. H. Kosters. American Enterprise Institute, Washington.

NEUMARK, D. (1988): „Employers' Discriminatory Behavior and the Estimation of Wage Discrimination," *Journal of Human Resources,* 23, 279-295.

OAXACA, R. L. (1973): „Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review,* 14, 693-709.

OAXACA, R. L., AND M. RANSOM (1988): „Searching for the Effect of Unionism on the Wages of Union and Nonunion Workers," *Journal of Labor Research,* 9, 139-148.

----------- (1994): „On Discrimination and the Decomposition of Wage Differentials," *Journal of Econometrics,* 61, 5-21.

Appendix

Table: Descriptive Statistics

	Uncensored		Left-censored		Right-censored		Left/Right-censored	
	Men	**Women**	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**
Wage	15.879	11.654	15.977	11.836	14.847	11.360	14.945	11.542
	(0.239)	(0.234)	(0.236)	(0.230)	(0.164)	(0.164)	(0.160)	(0.159)
Education	12.398	12.263						
	(0.070)	(0.082)						
Experience	23.707	22.664						
	(0.268)	(0.379)						
Married	0.580	0.521						
	(0.013)	(0.016)						
Regular part-time employment	0.036	0.391						
	(0.005)	(0.015)						
Number of children in household	0.679	0.503						
	(0.026)	(0.025)						
Immigrant to Germany since 1948	0.061	0.059						
	(0.005)	(0.007)						
East Germany	0.173	0.192						
	(0.008)	(0.010)						

Notes: 3,610 observations for men and 2,465 observations for women. Standard deviations in parentheses.

The following tables are not intended to be published

Table: Estimation Results

	A: OLS estimates		B: Tobit estimates	
	Men	**Women**	**Men**	**Women**
		uncensored		
Education (Yrs.)	0.085***	0.077***	0.085***	0.077***
	(0.003)	(0.004)	(0.003)	(0.004)
Experience	0.027***	0.035***	0.027***	0.035***
	(0.003)	(0.004)	(0.003)	(0.004)
Experience^2/100	-0.031***	-0.062***	-0.031***	-0.062***
	(0.007)	(0.008)	(0.007)	(0.008)
Widowed	0.091***	0.022	0.091***	0.022
	(0.020)	(0.022)	(0.020)	(0.022)
Regular part-time employment	-0.432***	-0.160***	-0.432***	-0.160***
	(0.046)	(0.021)	(0.046)	(0.021)
Number of children in household	0.025***	-0.025**	0.025***	-0.025**
	(0.009)	(0.013)	(0.009)	(0.013)
Immigrant to Germany since 1948	-0.175***	-0.129***	-0.175***	-0.129***
	(0.030)	(0.038)	(0.030)	(0.038)
East Germany	-0.499***	-0.392***	-0.499***	-0.392***
	(0.018)	(0.021)	(0.018)	(0.021)
Constant	1.223***	1.136***	1.223***	1.136***
	(0.054)	(0.064)	(0.054)	(0.064)
R^2/McFadden Pseudo R^2	0.38	0.26	0.27	0.19
		left-censored		
Education (Yrs.)	0.082***	0.074***	0.086***	0.083***
	(0.003)	(0.003)	(0.003)	(0.004)
Experience	0.024***	0.029***	0.026***	0.035***
	(0.003)	(0.003)	(0.003)	(0.004)
Experience^2/100	-0.027***	-0.048***	-0.029***	-0.059***
	(0.006)	(0.006)	(0.007)	(0.007)
Widowed	0.089***	0.022	0.094	0.027
	(0.018)	(0.018)	(0.019)***	(0.021)
Regular part-time employment	-0.314***	-0.138***	-0.390***	-0.161***
	(0.042)	(0.017)	(0.046)	(0.017)
Number Of Children In Household	0.026***	-0.020*	0.027***	-0.025**
	(0.008)	(0.010)	(0.008)	(0.012)
Immigrant to Germany since 1948	-0.184***	-0.135***	-0.182***	-0.135***
	(0.027)	(0.031)	(0.029)	(0.036)
East Germany	-0.471***	-0.344***	-0.502***	-0.408***
	(0.016)	(0.017)	(0.017)	(0.020)
Constant	1.306***	1.247***	1.230***	1.230***
	(0.048)	(0.052)	(0.052)	(0.061)
R^2/McFadden Pseudo R^2	0.40	0.30	0.29	0.22

Notes: 3,610 observations for men and 2,465 observations for women. Standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table: Estimation Results

	A: OLS	estimates	B: Tobit	estimates
	Men	Women	Men	Women
		right-censored		
Education (Yrs.)	0.065***	0.070***	0.085***	0.085***
	(0.002)	(0.004)	(0.003)	(0.004)
Experience	0.024***	0.033***	0.027***	0.035***
	(0.003)	(0.003)	(0.003)	(0.004)
Experience^2/100	-0.030***	-0.059***	-0.033***	-0.061***
	(0.006)	(0.007)	(0.007)	(0.008)
Widowed	0.086***	00.0	0.097***	0.021
	(0.017)	(0.021)	(0.020)	(0.021)
Regular part-time employment	-0.398***	-0.155***	-0.430***	-0.155***
	(0.040)	(0.019)	(0.045)	(0.020)
Number Of Children In Household	0.020***	-0.028**	0.028**0.028*	-0.027**
	(0.008)	(0.012)	(0.009)	(0.012)
Immigrant to Germany since 1948	-0.140***	-0.124***	-0.179***	-0.128***
	(0.026)	(0.036)	(0.030)	(0.037)
East Germany	-0.445***	-0.378***	-0.499***	-0.390***
	(0.016)	(0.020)	(0.018)	(0.018)
Constant	1.482***	1.238***	1.215***	1.166***
	(0.046)	(0.060)	(0.054)	(0.062)
R^2/McFadden Pseudo R^2	0.36	0.25	0.27	0.19
		left/right-censored		
Education (Yrs.)	0.064***	0.068***	0.085***	0.085***
	(0.002)	(0.003)	(0.003)	(0.004)
Experience	0.021***	0.027***	0.025***	0.034***
	(0.003)	(0.003)	(0.003)	(0.003)
Experience^2/100	-0.026***	-0.046***	-0.030***	-0.058***
	(0.005)	(0.006)	(0.006)	(0.007)
Widowed	0.086***	00.0	0.099***	0.025
	(0.015)	(0.016)	(0.018)	(0.020)
Regular part-time employment	-0.283***	-0.128***	-0.381***	-0.154***
	(0.035)	(0.016)	(0.044)	(0.019)
Number Of Children In Household	0.022***	-0.023**	0.029***	-0.027**
	(0.007)	(0.010)	(0.008)	(0.012)
Immigrant to Germany since 1948	-0.153***	-0.130***	-0.183***	-0.183***
	(0.023)	(0.029)	(0.028)	(0.034)
East Germany	-0.423***	-0.332***	-0.497***	-0.332***
	(0.014)	(0.016)	(0.017)	(0.019)
Constant	1.540***	1.339***	1.242***	1.108***
	(0.040)	(0.048)	(0.051)	(0.059)
R^2/McFadden Pseudo R^2	0.41	0.31	0.29	0.23

Notes: 3,610 observations for men and 2,465 observations for women. Standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.