Occupational Segregation within Establishments and the Gender Wage Gap

preliminary and incomplete

Johannes Ludsteck, IAB Nuremberg

May 18, 2010

Abstract

We investigate the impact of the occupational segregation within establishments and the establishment level proportion of females on the gender wage gap. Segregation explains about 10 percent of the discriminatory part of the wage gap in levels regressions. The effect becomes insignificant, however, in fixed establishment effects (difference) regressions. The proportion of female which explains a similar share of the discriminatory wage gap remains relevant even in difference models.

Keywords: Occupational segregation, gender segregation, gender wage gap
JEL classifications: J16, J71, M52

1 Introduction

1.1 Motivation

A huge number of empirical studies (see e.g. Weichselbaumer & Winter-Ebmer, 2005) suggests that gender wage discrimination appears to prevail despite the declaration of anti discrimination laws. A possible explanation for the imperfect effectiveness of legal bans may be based on fuzziness in the formulation of job tasks and measurement of output. It were simple for female assembly line workers to sue their employer if they were remunerated differently from male colleagues despite of identical (measurable) performance. But things become much more involved if tasks are heterogenous and non-routine. If performance measurement is difficult and output cannot be attributed unambiguously to individual workers, employers have discretion to discriminate regardless of whether they follow discriminatory preferences (in line with Becker's 1971 theory) or try to maximize their profits (as in the case of monopsonistic discrimination, see Robinson, 1969, Schlicht, 1982 or Schlicht, 2010). These thoughts give rise to the hypothesis that discriminatory remuneration of female is easier in occupationally segregated working environments. A female secretary cannot compare her tasks and output with her chief but a female electrician may be able to do so if she works together in a team with male electricians. If occupations characterise narrow ranges of qualifications and tasks, a reduction of the occupational segregation within establishments should be accompanied with diminishing wage discrimination. This logic applies to all forms of discrimination, irrespective of whether the are based on taste for discrimination or economic motives as e.g. monopsonistic or statistical discrimination or on selection wages (Schlicht, 2010). Note that this segregation effect may work even in absence of legal bans through effects on workers' motivation.

To test this hypothesis empirically, we select 10,000 sufficiently large establishments from the employment register 1985-2005 of the German Federal Employment Services. Then we compute an appropriately adjusted version of the Duncan segregation index at the establishment level and use it (together with a rich set of control variables) to explain wages in levels and fixed effects regressions of (log) wages. The regression results are feed into the well know Oaxaca-Blinder (Oaxaca, 1973; Blinder, 1973) decomposition to analyse the gender wage gap at the establishment level.

The remainder of the paper is organised as follows. After some brief remarks on related studies, we describe the most important data preprocessing steps (definition/extraction of the sample) and the computation of the required establishment level indicators. Then we present the estimation model and the results. The paper concludes with some further interpretations, disclaimers and prospects on future work.

1.2 A Scetchy Review of Related Work

To the best of our knowledge, the analysis of within-establishment segregation on discrimination is new. But there exist several studies employing matched employer-employee data sets to investigate the effects of other establishment characteristics on discrimination. Gartner and Hinz (2009, 2005) use the LIAB[1] to focus on the wage gap within narrow job cells. Their

[1]This data set is generated by matching the German employment register with the IAB establishment panel. The LIAB contains more information on establishments than our data set but is restricted to a considerably shorter period and a subsample of all establishments. We decided to dispense with the additional establishment info in favour

results regarding the within-occupation wage gap are similar to ours and can be taken as an important validation since they are based on somewhat different methods. Another study based on the LIAB is Heinze and Wolf (2009). They focus on the effects of product market competition (proxied by establishment size) and other institutional aspects, mainly work councils and collective agreements on the gender wage gap. Their approach differs in several respects from ours. They compute wage gaps within establishments in regressions conducted for every establishment separately and regress these gaps in a second step on establishment level characteristics. Their main findings are that competitive pressure as well as works councils and collective agreements reduce the gap. Regarding the proportion of female they obtain positive but insignificant effects on a measure of the wage gap which accounts for differences in human capital endowment (education). This deviation from our results may be explained by noteworthy differences in the empirical approach and the sample.[2]

The study most similar to ours (but related to Portugese data) is Vieira, Cardoso, and Portela (2005).[3] They employ a large an representative matched employer-employee data set for Portugal to investigate (under others) the effects of female participation at the establishment level (i.e. proportion of female in the establishment work force) on the wage gap. They find even somewhat larger contributions of the female proportion to the unexplained part (valuation effect) of the wage gap than we.

of the longer time period and the full sample (including enough large establishments) since many additional variables from the establishment panel are either almost time-invariant and would be wiped out in the fixed effects estimates or they appear to be of minor importance in our application.

[2]Their approach which is based on separate regressions for every establishment to compute the wage gap, does not allow to control appropriately for occupation effects. Furthermore the analysis of works councils and wage bargaining effects restricts the application of fixed effects models (since variation of these variables in the time dimension is small and prone to reporting errors). Besides that the don't conduct Oaxaca-Blinder type decompositions but include a female to estimate the gap. Note that the differences between their and our results can *not* be explained by omission of firm level characteristics as works councils and collective wage agreements from our estimates since they are practically time-invariant and therefore irrelevant for the fixed effects results.

[3]Note, however, that they use the term 'gender segregation' in a different meaning from ours. They focus on segregation *across establishments* which ist represented by the female proportions in the regressions.

2 Data Source, Preprocessing and Basic Descriptive Statistics

Data Source

Our main data source is the employment register 1985-2005 of the German Federal Employment Services. It is well suited for our analysis as it covers nearly 80 percent of the German workforce, excluding only the self-employed, civil servants, individuals in (compulsory) military service, and individuals in so-called 'marginal part-time jobs' (jobs with no more than 15 hours per week or temporary jobs that last no longer than 6 weeks). It contains important personal characteristics (sex, age, education, job status) as well as information on occupation, industry, establishment identifiers, wages and regional identifiers at the municipality level. The employment register contains complete biographies in spell data form. To simplify data processing, we extract spells at cut-off dates (30.6.) in every year.

Definition of the Segregation Measure and Implied Sample Selection

We employ the Duncan segregation index[4] as measure of occupational segregation within establishments. It is defined for every establishment e in period t as

$$D_{et} = (1/2) \sum_{o} \left| \frac{s^f_{eot}}{S^f_{et}} - \frac{s^m_{eot}}{S^m_{et}} \right| .$$

where $s^f_{eot}, (s^m_{eot})$ denote the proportion of female (male) workers in establishment e, occupation o and year t and the $S^f_{et}, (S^m_{et})$ denote the respective proportoins of female (male) workers in the establishment work force. The index (with range $[0, 1]$) can be interpreted as the share of workers which has to be reallocated to eliminate segregation completely. Usage of the index at the establishment level implies an important sample restriction. Firstly, the index can be computed only for establishments with male and female proportions strictly above zero and below unity. And secondly, the work force must comprise at least two occupations. Note, however, that this selection is harmless since occupational segregation is defined e.g. only for establishments employing both sexes in at least two occupations. Consequently other

[4] Another frequently employed segregation measure is the Gini index. Though its statistical properties are even more favourable, we restrict our analysis in this preliminary version of the paper to the Duncan index since several studies show high correlations between the both measures.

establishments are not relevant for the analysis. To be on the save side, we employ somewhat narrower restrictions. To enter our estimation sample, an establishment must have at at least 50 full-time employees in every one of the years 1985, 1990, 1995, 2000 and 2005. Furthermore it must employ at least 5 female and at least 5 male in at least two different occupations.

Inspection of the statistical properties of the index reveals that it cannot be taken as a sensible measure for small units since then considerable segregation is generated even by purely random allocation of workers. We follow Carrington and Troske (1997) to correct this by computing the index measure generated by purely random allocation and using this to rescale the index. Formally the corrected index has the form

$$\tilde{D} = \begin{cases} 1 + 0.5\,\frac{D-D^*}{1-D^*} & \text{if } D \geq D^* \\ 1 + 0.5\,\frac{D-D^*}{D^*} & \text{if } D < D^* \end{cases} \tag{1}$$

where D^* denotes the index obtained under random allocation.[5]

Aggregation of Occupational Classification

The occupation classification in our data comprises more that 300 occupations. In order to simplify data processing, to reduce effects of coding errors and to avoid artificial segregation,[6] occupations are aggregated by joining two classes whenever the annual bilateral between-occupation transition rate exceeds 0.3 percent. This reduces the number of occupations to 119.

Treatment of Wage Censoring

About 12-16 percent of male wages and 2-6 percent of female wages in our estimation sample are right-censored at the social insurance contribution limit. This could in principle be handled by using Tobit estimators. Since this would imply possibly severe computational and statistical problems for fixed establishment effects estimation, we choose a considerably simpler and appropriate imputation procedure. To retain establishment level heterogeneity in the data, the imputation is performed for each establishment and each year individually. Predictions for the censored wages are obtained using tobit

[5]The random allocation indes is computed via Monte Carlo simulation based on draws from the binomial distribution. See Carrington and Troske (1997) for a detailed descprition of the procedure.

[6]The classification is too fine if workers performing very similar tasks and workers who typically work together in teams are sorted into different occupations. This occurs e.g. for the occupations blacksmith, locksmith and building fitter. Simple measures of proximity are transition rates.

estimates based on the same regressors as the final analysis. Furthermore, a pseudo random variable from a appropriately truncated normal distribution is added to the conditional mean predictions to preserve the statistical properties of the conditional wage distribution in the censored (imputed) range.[7]

Definition of Final Estimation Sample

As explained above, the computation of the Duncan measure restricts the sample to establishments with at least 50 full-time workers, at least 5 full time male and at least 5 full time female workers in every year. To avoid possible problems due to training spells and early retirement schemes in Germany, we drop apprentices from the sample and restrictict it to full-time prime age workers aged 20-54 years. Possible misreporting of working time or wages is handled by dropping wages below the lower social contribution limit[8] The base sample is restricted to West-German establishments to obtain a sufficiently large period. This appears to be important since the fixed effects models identify the segregation effects based on variation of the duncan measure within establishments over time which is small for short time periods. Finally, we exclude working spells lasting less than one month to avoid bias due to exceptional working relations.

To keep the data manageable, the final regressions and the corresponding descriptive statistics are based on a 50 percent sample of persons (sampling is performed for every establishment × year × gender cell to preserve the proportions of male and female within establishments).

Basic Descriptive Statistics

Let us consider the most important stylized facts on the proportions of female and the (corrected) gender segregation measure at the establishment level before we step into the details of the analysis. Table 1 shows that the proportion of female remains almost constant over the estimation period. This differs considerably from the base population of full-time employees in the Employment register where the proportion of female increased from about 40 in 1985 to about 48 percent in 2005. Note, however, that the differ-

[7]This procedure could be improved in principle using multiple imputation techniques. In our application, however, this seems not worth the required additional effort, since the significance is beyond doubt for all relevant coefficients and from our experience with multiple imputation in other cases we see no reason to expect a noteworthy increase of standard errors in a multiple imputation context.

[8]less than one percent of the sample is affected by this in every year.

ence can be explained by our necessary establishment size restriction. The corrected Duncan index decreased from about 0.55 to 0.51, indicating that the occupational segregation decreased in this period.[9]

Table 1: Summary statistics of the Proportion of Female and the Segregation Index

Year	Mean	Lower Quartile	Upper Quartile
	Proportion of Female		
1985	0.296	0.114	0.449
1990	0.301	0.117	0.451
1995	0.309	0.124	0.462
2000	0.309	0.129	0.457
2005	0.306	0.132	0.450
Corrected Duncan Segregation Index			
1985	0.554	0.424	0.700
1990	0.536	0.403	0.678
1995	0.532	0.402	0.671
2000	0.538	0.453	0.663
2005	0.515	0.396	0.647

Legend: The table contains means together with lower and upper quartiles of the Proportion of Females in the establishment work force and the corrected Duncan Segregation index. All computations are based on the

3 Specification of the Empirical Model and Results

Empirical Model

Following Vieira et al. (2005) we employ the Oaxaca-Blinder decomposition. The decomposition is based on regressions of the form

$$w_{iget} = X_{iget}\, b_g + \epsilon_{iget} \tag{2}$$

[9]As expected, the Duncan Index is considerably lower before correction. Its mean is about 0.34 for our estimation sample. A more detailed comparison can be obtained from the author upon request.

for worker i of gender g in establishment e and year t. Where X_{iget} comprises individual and establishment level characteristics. A version including fixed establishment effects is obtained by adding the coefficients η_e (one for each establisment). All regressions contain a rich set of individual an firm level controls (see detailed regression output tables in the appendix) and 118 dummies for occupations as well as 118 proportions of occupations in the establishment work force.

Our decomposition of the gender wage gap into endowment and valuation effects is based on the formula

$$\bar{w}_m - \bar{w}_f = (\bar{X}_m - \bar{X}_f)\,\hat{b}_m + \bar{X}_f\,(\hat{b}_m - \hat{b}_f) \tag{3}$$

where the bars denote (estimation) sample means. The first term on the right hand side can be interpreted as the wage difference due to differences in endowments and the second term as the wage difference due to different valuation ('pricing') of these endowments.[10] To answer the main question of this paper we have to split the decomposition further into contributions of single regressors or dummy variable groups. The 'pricing' contribution of the share of female (PF) in the establishment work force and the occupational segregation measure $\tilde{D}$ are obtained from the terms $\bar{PF}_f \times (\hat{b}_m^{PF} - \hat{b}_f^{PF})$ and $\bar{\tilde{D}}_f \times (\hat{b}_m^{\tilde{D}} - \hat{b}_f^{\tilde{D}})$. They are presented in rows with headings 'PROP. FEMALE' and 'ADJ. Duncan' in table 2.

Results

Table 2 tells us that male and female wages differ on average by roughly 30 percent (28.2 log points). In the OLS regressions, about one third of the gross differential (roughly 10 percent) are explained by differences in observable characteristics. The rest (18 log points, amounting to roughly 20 percent) is due to different valuations of the characteristics. The table also shows the contributions of some variables or variable groups which are of importance in our analysis.[11] A sizeable contribution (9.6 log points) to the valuation

[10] Several papers are devoted to the discussion of the appropriateness of the implicit weighting by $\hat{b}_m$ and $\bar{X}_f$. Cotton (1988) and Neumark (1988) propose to obtain a reference price $\hat{b}^*$ as weighted average of $\hat{b}_m$ and $\hat{b}_f$ using estimation sample proportions of male and female as weights. We disregard this issue here as it is of second order importance in our context. (Note that the interpretation of $\hat{b}_m - \hat{b}_f$ is clear and unambiguous. We have to be careful to abstain from narrow *structural* interpretations, however, i.e. we should not consider $\hat{b}_m$ as true marginal productivity of men. But also $\hat{b}^*$ cannot be interpreted in this way.

[11] Note that the reported effects do not add to the total valuation effect since other variables which are not relevant here, are omitted (they are, however, available from the author upon request.)

effects comes from the occupation dummies. This means that female wages are roughly 10 percent lower than the wages of their male colleagues within narrow occupation groups, conditional on all other (observed) characteristics.[12] Also the contributions of our main variables are of considerable size. Wage gaps of roughly 3.5 percent are due to the proportion of female within the establishment and about 2 percent are due to the (adjusted) Duncan segregation measure, i.e. 'explained' by occupational segregation. Note that 2 log points amount to about 10 percent ($0.02/0.18 \times 100\%$) of the total valuation effect.

Things are quite different for the fixed effects models. Whereas the Occupation dummies remain as important as before, the share of female shrinks somewhat and the segregation measure coefficient becomes insignificant.

To understand the female proportions and segregation effects in more detail, we have to look at the respective coefficients which are reproduced from the appendix tables in table 3 below. From the OLS estimates we find that *both* male and female earn *(cet. par.)* less in establishments with a higher proportion of female workers but male loose less than female. A one percentage point increase of the proportion of female reduces male wages by 0.123 percent which is less than half of the corresponding loss of 0.267 percent for female. Note that the last sentence is somewhat imprecise since the OLS estimates explain levels rather than changes. The interpretation of changes is more appropriate for the fixed effects estimates as they exploit only deviations from establishment means of the respective variables. They tell us that the male workers are not affected by an increase of the proportions of female (the coefficient 0.004 is highly insignificant) whereas female loose 0.123 percent due to a one percent increase of their proportion.

The differences between OLS and fixed establishments effects estimates regarding the segregation measure can be explained by unobserved establishment characteristics using the standard formula on omitted variables. With the fixed effects model as a benchmark, OLS can be interpreted as a misspecified model neglecting the fixed establishment dummies. According to the standard omitted variables bias formula, included variables capture the effect of the omitted ones. Thus the increase of the segregation coefficients (cf. table 3) after omission of the fixed effects points to a positive correlation

[12]Gartner and Hinz (2009) obtain a similar estimate (12 percent) in an analysis focussing on differentials within narrowly defined job-cells. We have to add the disclaimer that valuation effect for dummy regressor groups depend on the choice of the base category. This arbitrariness could in principle be avoided by reparametrising the dummies such that the respective coefficients measure deviations from the sample mean over all dummy categories. We postpone this issue to a later version of the paper since it appears to be of minor significance in our application.

Table 2: Oaxaca-Blinder Decomposition of male-female (log) wage differences.

	OLS		Establ. FE	
	point estim.	std. err.	point estim.	std. err.
Log wage difference (total diff.)	0.282	0.004	0.282	0.000
Endowment effects (explained)	0.102	0.004	0.059	0.007
Valuation effects ('prices')	0.180	0.003	0.223	0.007
qualification	-0.002	0.003	-0.001	0.003
occup. dummies	0.096	0.011	0.099	0.012
OCCUP. SHARES	0.026	0.028	-0.023	0.085
PROP. FEMALE	0.035	0.004	0.031	0.006
ADJ. DUNCAN	0.019	0.005	0.005	0.005

Notes: The decompositions are based on OLS (columns 2 and 3) and fixed establishment effects estimates (columns 4 and 5). The respective coefficients and further sample information can be found in the appendix tables 4, 5 and 6. Regressions are based on observations from 2,552,431 male and 1,150,075 female in 10.000 West-German establishments with at least 50 employees and relate to the years 1985,1990,1995,2000,2005. Heteroscedasticity-robust standard errors in column 3 account for clustering at the establishment level.

Legend: Establishment level variables are printed in capital letters. A complete description of variable names can be found in appendix table 7. Note that only some important components of the valuation effects are contained in the table, implying that they do not add to the total valuation effect. Results for other variables are available from the author on request.

between the segregation measure and the fixed effects. Segregation is greater in establishment paying higher wages to both, men and women. Table 3 tells us further that the correlation appears to be even stronger for men as their segregation coefficients increase by more after dropping the the fixed effects.

Of course, the fixed establishment effects appear to be the more trustworthy ones for several good reasons. Nevertheless they could be questioned by two objections. Firstly, they are still biased and cannot be interpreted as causal effects in presence of endogeneity or reverse causality. Employers pursuing discriminatory policies may adjust their hiring decisions such as to keep the occupational segregation on an high level. Secondly the within-establishment variation of the segregation measure in the time dimension may be too small or measurement error may be important. In presence of

Table 3: Coefficients for proportions of female and the segregation measure

Sample	male		female	
	Coeff.	S.E.	Coeff.	S.E.
	OLS estimates			
PROP. OF FEMALE	-0.123	0.0014	-0.267	0.0018
ADJ. DUNCAN	0.050	0.0009	0.014	0.0015
	Fixed establishment effect estimates			
PROP. OF FEMALE	0.004	0.0198	-0.123	0.0171
ADJ. DUNCAN	-0.000	0.0067	-0.010	0.0062

Note: coefficients reproduced from appendix tables 4 and 5.

measurement error, the fixed effects estimates may suffer from noteworthy downward bias. After a glance at the variances this appears to be irrelevant. The within and between variances of the adjusted Duncan measure are 0.008 and 0.031, respectively. They have to be compared with the corresponding values for the within-establishment proportions of female which are 0.002 and 0.077. Though the within variance of the proportion of female is considerably smaller, its fixed effects coefficients are nevertheless stable and significant.[13]

4 Conclusion and Prospect on Future Work

This paper focuses on the impact of the establishment level proportion of female and the occupational within-establishment gender segregation to check the hypothesis whether comparability of tasks and communication within teams reduces employer's discretion to discriminate female workers. The OLS results show that discrimination is more pronounced *cet. par* in establishments with greater proportions of female workers and in establishments characterised by greater segregation. Whereas the segregation effects vanish after controlling for unobserved establishment heterogeneity, the female proportion effects remain quite robust. Because of endogeneity of both variables at the firm level, we should, however, be cautious to give our results a strict causal interpretation.[14]

[13] It is still possible that the aggregation of occupations generates measurement error for the segregation measure. We will have to check that in future versions.

[14] Hiring decisions and wage setting are directly interrelated e.g. in theories of monopsonistic discrimination, c.f. Manning (2005) or Schlicht (1982).

A preliminary version of a paper should not end without prospects for work and robustness checks to be amended in future versions. Firstly, we use potential experience since true experience cannot be computed directly in our base data set.[15] This is not an optimum choice since employment interruptions are more frequent for female (mainly due to maternity leave or child care periods) and renders the male-female comparison fuzzy. Gartner and Hinz (2009) compare gender wage gaps based on potential and true experience and find that the bias induced by using potential experience vanishes in regressions controlling for occupations. Though this results could be taken as an 'all-clear' for our applications and the inclusion of interaction terms between pot. experience and the qualification dummies leaves the decomposition results unchanged, we should check this issue in a future version of the paper to be on the save side.

Secondly, since the partitioning of jobs into occupation classes is important for the measurement of segregation, our decomposition results may depend on the aggregation level of the occupation scheme through the well known error-in-variables problem. We will check this by varying the aggregation thresholds and the number of resulting occupation classes in future versions.

References

Becker, G. (1971). *The Economics of Discrimination.* Chicago: University of Chicago Press.

Blinder, A. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources*, *8*, 436–455.

Carrington, W., & Troske, K. (1997). On measuring segregation in samples with small units. *Journal of Business and Economic Statistics*, *15* (4), 402-409.

Cotton, J. (1988). On the decomposition of wage differentials. *Review of Economic Studies*, *70* (2), 236–243.

Gartner, H., & Hinz, T. (2005). Geschlechtsspezifische Lohnunterschiede in Branchen, Berufen und Betrieben. *Zeitschrift fr Soziologie*, *34* (4), 22–39.

Gartner, H., & Hinz, T. (2009). Geschlechtsspezifische Lohnungleichheit in Betrieben, Berufen und Jobzellen (19932006). *Berliner Journal für Soziologie*, *19* (4), 557-575.

[15]Note that the the observations in our data set relate to the 30. of June in year year. But tell us not for all workers whether they were employed the whole year.

Heinze, A., & Wolf, E. (2009). The intra-firm gender wage gap, a new view on wage differentials based on linked employer-emploee data. *Journal of Population Economics*, *forthcoming*.

Manning, A. (2005). *Monopsony in Motion:Imperfect Competition in Labor Markets*. Princeton: Princeton University Press.

Neumark, D. (1988). Employers' Discriminatory Behaviour and the Estimation of Wage Discrmination. *Journal of Human Ressources*, *23*(3), 279–295.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, *14*, 693–709.

Robinson, J. (1969). *The Economics of Imperfect Competition*. London: Macmillan.

Schlicht, E. (1982). A Robinsonian Approach to Discrimination. *Journal of Institutional and Theoretical Economics*, *138*(1), 64–83.

Schlicht, E. (2010). Selection Wages and Discrimination. *Economics: The Open-Access, Open-Assessment E-Journal*, *4*(2010-6), 1–30.

Vieira, J. A. C., Cardoso, A. R., & Portela, M. (2005). Gender Segregation and the Wage Gap in Portugal: an Analysis at the Establishment Level. *Journal of Economic Inequality*, *3*, 145-168.

Weichselbaumer, D., & Winter-Ebmer, R. (2005). A Meta-analysis of the International Gender Wage Gap. *Journal of Economic Surveys*, *19*(3), 479–551.

Table 4: Coefficient estimates from OLS models

Regressor	male		female	
	Coeff.	S.E.	Coeff.	S.E.
Dummy 1990	0.113	0.0005	0.104	0.0008
Dummy 1995	0.135	0.0005	0.157	0.0008
Dummy 2000	0.151	0.0005	0.193	0.0008
Dummy 2005	0.144	0.0006	0.181	0.0009
ln(ESTABL. SIZE)	0.067	0.0010	0.111	0.0018
ln(ESTABL. SIZE)^2	-0.003	0.0001	-0.006	0.0001
NUMB. OF OCCUP. IN ESTABL.	-0.002	0.0000	-0.001	0.0001
pot. exper.	0.027	0.0001	0.019	0.0001
pot. exper.^{2}/100	-0.048	0.0002	-0.039	0.0003
tenure	0.015	0.0001	0.019	0.0001
Dummy(tenure $\geq$ 10)	0.110	0.0005	0.175	0.0008
PROP. APPRENTICES	-0.096	0.0021	-0.079	0.0024
PROP. FOREIGN	0.056	0.0020	-0.028	0.0032
PROP. FOREIGN ADV.	0.018	0.0056	0.061	0.0091
PROP. PART TIME $<$ 50%	-0.304	0.0040	-0.471	0.0046
PROP. PART TIME $\geq$ 50%	-0.111	0.0023	0.138	0.0027
Dummy medium qualif.	0.149	0.0005	0.153	0.0007
Dummy high qualif.	0.384	0.0008	0.379	0.0014
Dummy foreigner	-0.033	0.0007	-0.015	0.0010
Dummy foreigner adv.	0.059	0.0016	0.042	0.0026
PROP. OF FEMALE	-0.123	0.0014	-0.267	0.0018
ADJ. DUNCAN	0.050	0.0009	0.014	0.0015
constant	3.287	0.0051	3.168	0.0084
+ occupation dummies				
+ PROP. OF OCCUP.				
number of observations	2,552,431		1,150,075	
number of establishments	10,000		10,000	
adj. R^2	0.521		0.483	

Notes: To keep the data set manageable, the final estimations are based on a 50 percent sample of persons. (Proportions of male and female within establishments are preserved through sampling by establishment and sex and year). See table 7 for definitions of the variable names.

Table 5: Coefficient estimates from fixed establishment effects models

Regressor	male		female	
	Coeff.	S.E.	Coeff.	S.E.
Dummy 1990	0.111	0.0015	0.112	0.0014
Dummy 1995	0.133	0.0021	0.166	0.0018
Dummy 2000	0.143	0.0025	0.190	0.0025
Dummy 2005	0.132	0.0035	0.179	0.0036
ln(ESTABL. SIZE)	-0.047	0.0148	0.024	0.0191
ln(ESTABL. SIZE)^2	0.003	0.0013	-0.002	0.0018
NUMB. OF OCCUP.	-0.002	0.0004	-0.002	0.0004
pot. exper.	0.026	0.0005	0.018	0.0003
pot. exper.^{2}/100	-0.045	0.0009	-0.036	0.0006
tenure	0.013	0.0002	0.018	0.0002
Dummy(tenure $\geq$ 10)	0.106	0.0022	0.167	0.0026
PROP. APPRENTICES	-0.007	0.0103	0.006	0.0076
PROP. FOREIGN	-0.064	0.0287	0.012	0.0274
PROP. FOREIGN ADV.	0.166	0.0776	0.061	0.0718
PROP. PART TIME ($<$ 50%)	-0.105	0.0177	-0.064	0.0190
PROP. PART TIME ($\geq$ 50%)	0.073	0.0192	0.123	0.0169
Dummy medium qualif.	0.134	0.0021	0.137	0.0024
Dummy high qualif.	0.348	0.0037	0.332	0.0044
Dummy foreigner	-0.036	0.0020	-0.027	0.0022
Dummy foreigner adv.	0.063	0.0046	0.054	0.0052
PROP. OF FEMALE	0.004	0.0198	-0.123	0.0171
ADJ. DUNCAN	-0.000	0.0067	-0.010	0.0062
constant	3.868	0.0842	3.574	0.0711
+ occupation dummies				
+ PROP. OF OCCUP.				
number of observations	2,552,431		1,150,075	
number of establishments	10,000		10,000	
R^2	0.456		0.371	

Notes: To keep the data set manageable, the final estimations are based on a 50 percent sample of persons. (Proportions of male and female within establishments are preserved through sampling by establishment, sex and year). See table 7 for definitions of the variable names.

Table 6: Summary statistics of variables used in the estimation sample

Regressor	male		female	
	Mean	S.E.	Mean	S.E.
log wage (imputed)	4.502	0.370	4.220	0.367
Dummy 1990	0.234	0.423	0.233	0.423
Dummy 1995	0.209	0.407	0.216	0.411
Dummy 2000	0.194	0.395	0.198	0.399
Dummy 2005	0.159	0.365	0.158	0.365
ln(ESTABL. SIZE)	6.212	1.390	5.904	1.223
NUMB. OF OCCUP.	16.759	10.161	14.970	9.237
pot. exper.	19.741	9.261	17.325	10.141
Tenure[a]	2.227	2.805	2.356	2.724
Dummy(tenure $\geq$ 10)	0.335	0.472	0.235	0.424
PROP. APPRENTICES	0.053	0.082	0.067	0.112
PROP. FOREIGN	0.095	0.105	0.084	0.101
PROP. FOREIGN ADV.	0.013	0.033	0.012	0.032
PROP. PART TIME (< 50%)	0.019	0.046	0.031	0.062
PROP. PART TIME ($\geq$ 50%)	0.067	0.092	0.123	0.119
Dummy medium qualif.	0.677	0.468	0.672	0.470
Dummy high qualif.	0.123	0.328	0.065	0.246
Dummy foreigner	0.097	0.296	0.087	0.282
Dummy foreigner adv.	0.013	0.115	0.012	0.109
PROP. OF FEMALE	0.239	0.178	0.451	0.212
ADJ. DUNCAN	0.534	0.204	0.537	0.191
number of observations	2,552,431		1,150,075	
number of establishments	10,000		10,000	

Notes: To keep the data set manageable, the final estimations are based on a 50 percent sample of persons. (Proportions of male and female within establishments are preserved through sampling by establishment, sex and year). See table 7 for definitions of the variable names.

[a] Tenure is interacted with the Dummy(tenure $\geq$ 10) to handle right-censoring of the tenure at 10 year. The statistics in the table relate to the censored tenure.

Table 7: Definitions of variable names

log wage (imputed)	log daily wage (censored wages imputed in regressions conducted separately for each establishment, sex and year
Dummy 1990 ...	Dummy variable for year 1990 (omitted base is 1985)
ln(ESTABL. SIZE)	log establishment size
NUMB. OF OCCUP.	Number of different occupations within establishment
Pot. exper.	Potential experience of worker in years
Tenure	Tenure of worker in years, interacted with Dummy(tenure $\geq$ 10) since tenure is censored from above at 10 years
Dummy(tenure $\geq$ 10)	Dummy for (tenure $\geq$ 10), see explanation of Tenure above
PROP. APPRENTICES	Proportion of apprentices in the establishment work force
PROP. FOREIGN	Proportion of foreign workers in establishment work force
PROP. FOREIGN ADV.	Proportion of foreign workers from advanced industrial countries in establishment work force
PROP. PART TIME (< 50%)	Proportion of part time workers with less than 50 percent of regular full time hours in establishment work force
PROP. PART TIME ($\geq$ 50%)	Proportion of part time workers with at least 50 percent of regular full time hours in establishment work force
Dummy medium qualif.	Dummy for medium qualification (completed apppprenticeship training)
Dummy high qualif.	Dummy for high qualification (technical college or college)
Dummy foreigner	Dummy for foreign workers
Dummy foreigner adv.	Dummy for foreign workers from advanced industrial countries
PROP. OF FEMALE	Proportion of female in establishment work force
ADJ. DUNCAN	Adjusted Duncan Measure

Note: Establishment level variables (i.e. variables taking on the same value for all workers in the establishment) are printed in capital letters.