Heinze, Anja; Beninger, Denis; Beblo, Miriam; Laisney, François

Working Paper

Measuring Selectivity-Corrected Gender Wage Gaps in the EU

ZEW Discussion Papers, No. 03-74

Discussion Paper No. 03-74

Measuring Selectivity-Corrected Gender Wage Gaps in the EU

Miriam Beblo, Denis Beninger,
Anja Heinze and François Laisney

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 03-74

Measuring Selectivity-Corrected Gender Wage Gaps in the EU

Miriam Beblo, Denis Beninger,
Anja Heinze and François Laisney

Non technical summary

In the European Union the raw wage gap between women and men amounts to 16% on average. In order to assess a gender wage gap adjusted for male and female characteristics, particularly two main methodological issues have to be considered. First, male and female wage equations have to be estimated consistently. This requires proper treatment of self-selection into the labour market. In other words, those working may not form a random subgroup of the (sampled) population but differ systematically, in unobservable aspects, from those not employed. The second issue concerns the appropriate decomposition of the gender pay gap, that allows meaningful interpretation of its components.

In this study we propose different techniques to assess the gender pay gap while exploring different estimation methods for the wage equations, and different decomposition approaches for the wage gap. We concentrate on the estimation methods most often used in the gender gap literature (OLS and Heckman), and also use a recently developed estimator (Lewbel 2002). These procedures differ in the treatment of self-selection. We decompose the pay gap both at the mean, following Oaxaca (1973) and Blinder (1973), and across the wage distribution, as proposed by Juhn, Murphy and Pierce (1993).

The empirical application, based on the European Community Household Panel (ECHP) for the five largest European countries (France, Germany, Italy, Spain and the United Kingdom), shows that at most half of the difference in earnings between the sexes can be attributed to differences in characteristics. This confirms the findings of other studies, such as the report Employment in Europe 2002 (European Commission 2002a). However, the size of this endowment effect differs considerably between countries and depends on the choice of estimator. Our results suggest that correcting for self-selection has a significant impact on both the wage estimates and the pay gap decomposition. Furthermore, the results are sensitive to the choice of estimator, that is to the way self-selection in estimation is treated. We recommend the Lewbel approach because it imposes fewer arbitrary restrictions on the model than the Heckman approach, and performs better in terms of predictions with our data.

Another main result of the study is derived from the pay gap decomposition over quantiles of the wage distribution. Remarkable differences are revealed within as well as between countries. A further recommendation derived from our analysis would therefore be to pay careful attention to differences over the wage distribution when drawing policy conclusions. Focusing only on the mean pay gap may conceal politically relevant aspects of the problem.

Measuring Selectivity-Corrected Gender Wage Gaps in the EU

Miriam Beblo*, Denis Beninger†,
Anja Heinze‡ and François Laisney§

November 21, 2003

Abstract

We investigate different techniques to assess the gender pay gap in five EU countries (France, Germany, Italy, Spain and United Kingdom), focusing on self-selection into market work. Results show that selectivity correction has an impact on both wage estimates and wage gap decomposition. If there is a positive correlation between the wage and the propensity to participate, the estimated pay gap understates the true difference in earnings when self-selection is ignored. The estimated pay gap differs considerably at different quantiles of the wage distribution, and is sensitive to the choice of estimator.

JEL Classification: J16, J31

Acknowledgements: This paper builds on the final report of the project *Methodological Issues Related to the Analysis of Gender Gaps in Employment, Earnings and Career Progression*, carried out for and financially supported by the European Commission, Employment and Social Affairs DG. We thank Eurostat for providing us with data from the European Community Household Panel (ECHP UDB). We thank Frank Siebern-Thomas, Mette Deding, Anna d'Addio, Björn Döhring, Monika Oels, Elke Wolf, and participants of several meetings at the European Commission for helpful comments. Doris Weichselbaumer advised us on the literature. Franziska Mientus and Susanne Steffes provided competent research assistance. Views expressed represent exclusively the positions of the authors and do not necessarily correspond to those of the European Commission. All errors remain our own.

*ZEW, Mannheim. Beblo@zew.de
†ZEW, Mannheim. Beninger@zew.de
‡ZEW, Mannheim. Heinze@zew.de
§Université Louis Pasteur, Strasbourg, and ZEW, Mannheim. Laisney@zew.de

1 Introduction

Gender gaps in employment, earnings and career progression have been on the European Union's policy agenda for several decades. Despite repeated commitments to promote gender equality, e.g. through the European Employment Strategy and national governmental measures, women are still less favoured on the labour market than men. In 1998 women's average gross hourly earnings amounted to 76 % - 94 % of men's earnings throughout the European Union (European Commission 2002a). Women encounter more difficulties in their career progression and are less likely to be promoted (OECD Employment Outlook 2002). They are under-represented in managerial occupations and in jobs with a supervisory role.

Although the observed or raw difference in wages between men and women provides an overall picture of the actual gender pay gap, encompassing both differences in endowments and differences in remuneration, it is worthwhile to analyse the factors related to the raw wage gap in more detail. On average, men and women have different human capital endowments. To start with, men and women differ in their education levels and tracks. Second, women and men have quite distinct work histories. Traditionally, women are more likely to interrupt gainful employment for family reasons, leading to less work experience compared with men and loss of human capital. The quasi totality of part-time jobs are occupied by women, and their overall participation rate remains steadily below that of men. Finally, the segregation of men and women with respect to occupation or industry may explain a part of the wage gap. Empirical studies consider these differences by estimating an adjusted pay gap that controls for individual and job characteristics. However, many published studies fail to take account of important methodological issues.

The aim of our study is to explain the difference in earnings by gender and improve the analysis of the factors related to the gender pay gap in the European Union. Our focus is a methodological one. The typical procedure when estimating pay gaps is to first investigate the determinants of women's and men's wages and then use the empirical results in order to draw conclusions on differences in the wage formation processes of women and men. Thereby, two issues have to be dealt with. First, male and female wage equations have to be estimated consistently. This requires proper treatment of methodological problems such as, for instance, self-selection into participation. This problem arises if the working individuals do not form a random subgroup of the sample population but differ systematically, in unobservable aspects of preferences, opportunities, and productivity, from those not employed. The second methodological issue discussed here concerns the appropriate decomposition of the gender pay gap in the presence of self-selection.

As regards obtaining consistent estimates for the returns to endowments, other important issues do arise, of course. A prominent one concerns the potential endogeneity of several of the variables typically included in the list of endowments, and unobserved heterogeneity also deserves a mention. While acknowledging these methodological issues, our focus here is deliberately on the treatment of self-selection both in estimation and in wage gap decomposition.[1]

[1] We nevertheless document attempts to control for unobserved heterogeneity and for endogeneity in our report (Beblo et al. 2003). However, due to the very low variation over time of the meaningful variables, panel data techniques controlling also for self-selectivity, Wooldridge

We use a new microeconometric approach to estimate wages for women and men that takes account of the selection process into the labour market. The Lewbel estimator (Lewbel 2002), allows us to derive consistent wage estimates for women and men under less stringent assumptions than those required for consistencey of Heckman's estimator. We also provide new empirical evidence on the gender pay gap in Europe by using different decomposition methods while drawing on comparable data for all EU countries. Our study therefore complements previous work published in Employment in Europe 2002, Employment Outlook 2002 and by the Group of Experts on Gender and Employment (European Commission 2002b), while stressing the self-selection issue.

The outline of the paper is as follows. Section 2 presents stylised facts on the unadjusted pay gap in EU countries, introduces the estimation methods we implement for the wage equations, and describes the decompositions of the wage gap we use. Thereby, the main focus is on proper treatment of self-selection in the data. Section 3 describes the data, and Section 4 provides empirical evidence. The variety of econometric procedures, including OLS, Heckman two-step and Lewbel estimators, and the different wage gap decomposition issues, are illustrated with German data. In addition, we provide estimation and decomposition results based on cross-section Lewbel estimates for the other four largest EU member states (France, Italy, Spain, and the United Kingdom). The pay gap decomposition results presented for all countries are based on the methods proposed by Oaxaca (1973) and Blinder (1973), and Juhn, Murphy and Pierce (1993). Section 5 concludes.

2 Analysing gender pay gaps

This section first presents stylised facts on gender earnings in EU countries. It introduces the estimation methods we implement for the wage equations. Finally we describe the decompositions of the wage gap used.

2.1 The unadjusted pay gap in EU countries

We start off by taking a look at the raw or unadjusted gender gap in gross earnings across the EU (details on the data are given in Section 4). According to the ECHP and as listed in Table 1, the average hourly wage gap between women and men in the EU varies between 6.5 % in Portugal and 26.5 % in the United Kingdom.

While these aggregate numbers give us a first idea of the gender pay gap in the EU countries, they conceal the prevalence of wage differences for different groups within each country. The distribution of the gender gap for each country is given in Figure 1. The raw wage gap is displayed by percentiles according to the wage rankings of all women and men in the respective country sample. The graphs highlight the fact that the wage distributions of women and men are not typically congruent, that is, male wages are usually more spread.

(1995), Kyriazidou (1997) and Lewbel (2002), failed to produce meaningful results. The lack of suitable instruments in the ECHP thwarted our attempt to account satisfactorily for sectoral information (in particular employment in the public sector). Results obtained with the available instruments were not significant at all.

Table 1: Average raw wage gap between men and women in the EU countries

Country	absolute wage gap (euro)	relative wage gap (%)
Austria	2.13	20.1
Belgium	0.98	7.7
Denmark	1.72	11.7
France	1.44	13.0
Finland	1.81	17.7
Germany	2.68	21.0
Greece	0.72	9.5
Ireland	2.48	20.2
Italy	0.67	6.8
The Netherlands	3.70	22.1
Portugal	0.38	6.5
Spain	1.37	14.9
United Kingdom	3.38	26.5
EU	1.82	16.3

Data source: ECHP, country files 1998. Sample of 25-55 year old women and men, who are employed at least 8 hours per week. Data for Sweden and Luxembourg are not available.
Note: Absolute wage gap = Male minus female average gross hourly wages in euro. Relative wage gap = Absolute wage gap / average male wage rate.
We use the individual and country weights (ratio of sample size and size of the population 16 years and above) provided in the ECHP for all graphs and descriptive statistics in the paper.

We can distinguish three main groups of patterns: a decreasing, an increasing and a U-shaped distribution of the raw wage gap. While the Scandinavian countries Denmark and Finland have increasing wage differences, Southern European countries, e.g. Italy and Portugal, and Ireland show a decreasing pattern. The Netherlands provide an extreme example of a U-shape. Portugal is the only country where the wage gap curve drops below zero (for higher earnings). There is a small glass ceiling effect for some countries, for example Ireland and the Netherlands, in the sense that the wealthiest part of the (wage-earning) population is composed largely of men.

The unadjusted wage gap provides an aggregate measure of the earnings inequality between men and women. It compounds differences in characteristics and differences in the remuneration of these characteristics (as well as potential direct and indirect discrimination). Investigating the components of the gender pay gap separately will be helpful, in particular if the aim is to better target policy measures at reducing the earnings gap.

2.2 The adjusted pay gap: methodological issues

The ways in which the factors related to a pay gap are analysed in the literature are diverse. The crudest approach consists in including a sex dummy in a single wage regression for women and men. The underlying assumption here is that female and male wages differ by a fixed amount (shift parameter), but that human capital characteristics and other explanatory variables have the same impact on women's and men's wages.

A more flexible approach to investigate the earnings gap has been derived from human capital theory (Mincer 1958, 1974, Becker 1964), where an individual's wage rate reflects the productivity potential based on various human capital characteristics. According to Oaxaca (1973) and Blinder (1973), any wage differential between two groups of people (defined by gender, race, ethnicity, etc.) can therefore be decomposed into two parts. The first is explained by differences in observable human capital endowments and other job-related variables between both groups (*endowment effect*), the second reflects differences in the values that are assigned to women's and men's characteristics, that is the prices or remuneration of these endowments (*remuneration effect*). This latter part of the wage differential is often interpreted as an estimate of wage discrimination.[2] Most empirical studies estimate an adjusted pay gap that controls for these variables and thereby accounts for the observed differences in personal and job characteristics of women and men.

The wage decomposition suggested by Blinder (1973) and Oacaxa (1973) has been subject to criticism on two points at least. First, their method is based on the endowment prices of one of the sexes (the male in most applications), thereby introducing a potential dissymmetry in the effects depending on which gender is considered as the reference. To overcome this problem, Reimers (1983) and

[2] There is an ongoing debate on the interpretation of the gender wage gap or parts of it as discrimination. On the one hand, the different endowments of women and men may already be the result of discrimination because of feedback effects (see e.g. European Commission 2002c). On the other hand, the residual ("unexplained") part of the gap may still consist of unobserved differences in human capital characteristics. Due to these ambiguities we prefer not to speak of discrimination, and neither of an "explained" or "unexplained" part of the gap, but use the terms endowment and remuneration effect instead (as well as other effects to be introduced later on).

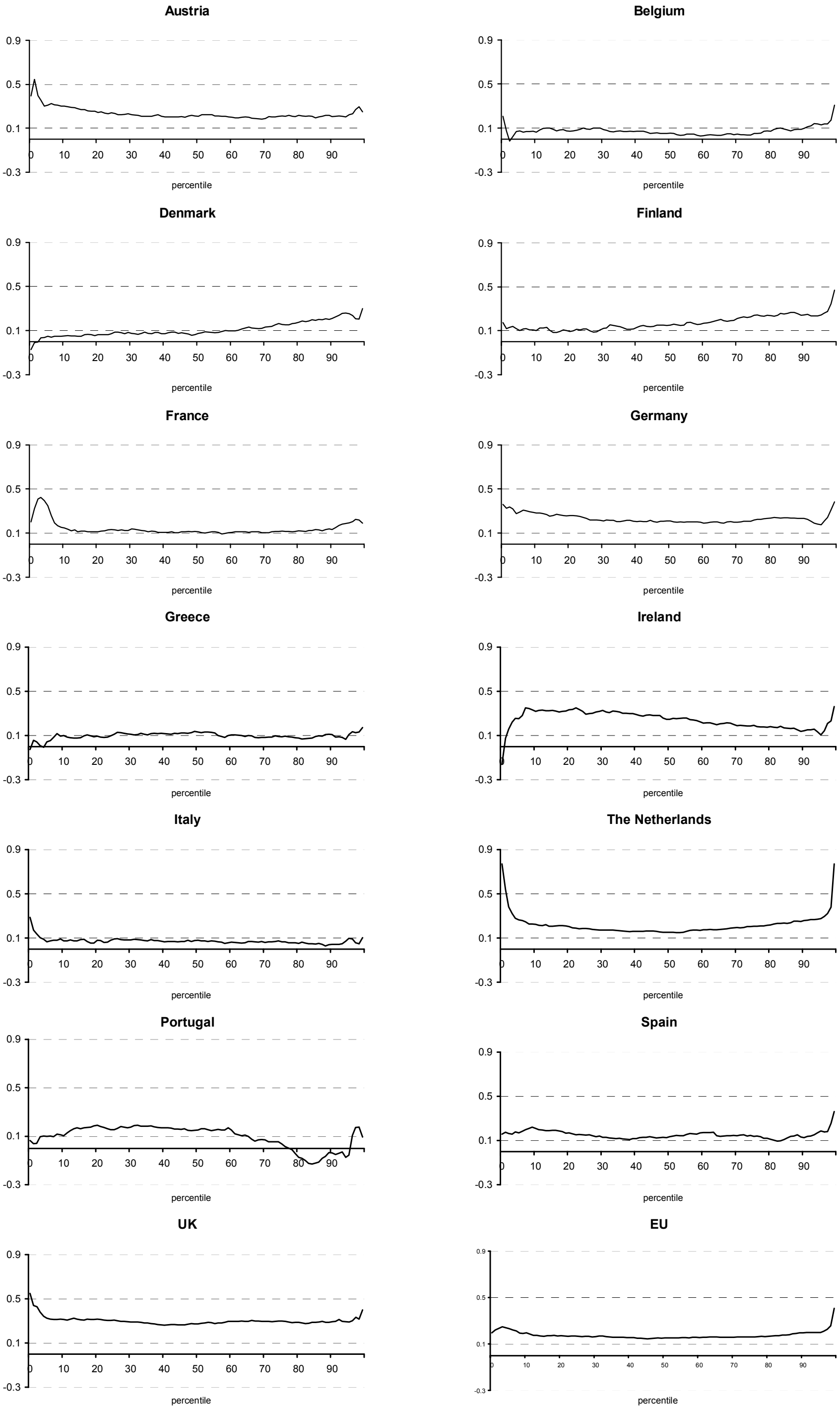

Figure 1: Distribution of the raw wage gap in the EU countries
Data source: ECHP, country files 1998. Sample of 25-55 year old women and men, who are employed at least 8 hours per week. Data for Sweden and Luxembourg are not available.
Note: Male minus female gross hourly wage rate per percentile of the wage distribution.

Oaxaca and Ransom (1994) use a matrix combination of both female and male prices in decomposing the wages. A more important problem, in our view, is that the Oaxaca and Blinder proposal considers only wage decomposition at the mean, thus occulting potential variations of the different effects over the wage distribution. Juhn, Murphy and Pierce (1993) have developed a method that sheds light on the profile of the wage gap across the whole wage distribution.

Since each of these gender pay gap decompositions is based on the estimation of wage equations, it is important to use a consistent, and possibly efficient, estimation technique. The better the wage determination process can be identified, the more knowledge about the factors related to the gender pay gap can be gained, and the better policy measures can be targeted. The following section therefore introduces different methods for both wage estimation and wage decomposition.

The remainder of this section is organised in three parts. First, different wage regression models are presented, notably those handling self selection. In particular, we describe a new estimator proposed by Lewbel (2002). Second, different decomposition techniques of the gender wage gap are discussed. A final subsection discusses how self selection may be treated when decomposing the wage gap. For more details on the methodological aspects and a review of the literature on gender wage gaps we refer to (Beblo et al. 2003).

2.2.1 Estimation of wage equations

We assume the following log-linear wage regression model:

$$\ln W_i^J = X_i^J \beta^J + \varepsilon_i^J, \qquad J = M, F, \tag{1}$$

where $\ln W_i$ is the logarithmic wage. Vector X_i contains the set of Mincerian explanatory variables augmented of job attributes, labour market features and demographic characteristics.[3] ε_i is an i.i.d. idiosyncratic error term with mean zero and constant variance σ_ε^2. i indexes individuals within the male (M) and female (F) samples. For simplicity the index on gender ($J = M, F$) will be skipped hereafter, when not explicitly needed.

Most studies on gender pay gaps use a simple OLS regression for the wage estimation, arguing that only working individuals are considered. However, if a significant part of the sample is not working, endogenous selection does arise if unobservables in the wage and in the participation equations are correlated, and this requires proper treatment.

Heckman (1979) proposed two estimation techniques to overcome the self-selection problem, one consisting in maximum likelihood (ML) estimation of a selection model assuming bivariate normality of the error terms in the wage and participation equations. The other method proceeds in two steps, ML probit estimation of the participation equation, and OLS (or GLS) estimation of the wage equation using participants only and the normal hazard $\hat{\lambda}$ estimated from

[3] Heckman et al. (2003) provide a profound critique of this approach, underlining the important differences that arise between cohort-based and cross-sectional estimates of returns to schooling, as well as the crucial roles of expectation formation and sequential resolution of uncertainty.

the first step as additional regressor. For further reference, note that this is a decreasing function of the participation probability.[4]

Lest identification be obtained by arbitrary functional form restrictions, both procedures require the availability of valid instruments, i.e. variables which contribute to determining the propensity to work but are not related to wages. In practice, such exclusion restrictions are difficult to find and collinearity problems are likely to prevail. Furthermore, the consistency of the two-step estimator hinges on the correct specification of the normal hazard as additional regressor. In order to circumvent this latter restriction, other correction terms have been suggested in the literature, such as the propensity of participating (Olsen 1980). In the latter approach the propensity score instead of the normal hazard is included in the wage equation. An empirical survey about various correction terms is provided in Vella (1998).[5]

To obviate the self selection problem in a less restrictive way than the Heckman correction, we propose to use the endogenous sample selection model introduced by Lewbel (2002) which, to our knowledge, has not yet been applied to the estimation of gender wage gaps. It provides a two stage least squares estimator of the coefficients, where regressors are allowed to be endogenous, mismeasured, or otherwise correlated with the model errors. Unlike the Heckman two-step estimator, no structure is imposed on the distribution of the error terms, permitting a more general form of unknown heteroscedasticity. The estimator does not even require an estimation of the selection equation, in other words the participation decision. Yet, the latter may be estimated separately in order to gain information on the factors influencing the labour market activity of women and men in Europe.

In the Lewbel procedure, identification is obtained by observing a "special" variable S. S is supposed to be continuously distributed with large support. The selection process is assumed monotonic in S. The special variable affects participation in otherwise unspecified ways but does not affect the dependent variable. Hence S can be interpreted as an instrument for participation. Here, an appropriate choice for S may be the unearned income of the individual. Define

$$W = \frac{I}{f(S|U)}, \tag{2}$$

where I is the participation indicator, $f(S|U)$ is the conditional probability density function of S given U, where U is a subset of the variables in X. If X contains endogenous variables these should be instrumented.

In a first stage, $f(S|U)$ is estimated either parametrically or non-parametrically. Lewbel shows that, by using $f(S|U)$ as a weighting function in the wage

[4] There are two reasons to prefer two-stage estimation to the direct ML estimation of the Heckman model. First, ML relies on joint normality of the errors in the selection and level equations. Its advantages and drawbacks are twofold: if none of the equations is misspecified, simultaneous estimation yields efficiency gains. On the other hand, misspecification of one equation may contaminate the other, resulting in inconsistency. By contrast, the two-step estimator only relies on conditional moments which, although derived under joint normality, may hold for a wider class of distributions, at least approximately. Second, using OLS in the second stage has the advantage that the average of the residual is zero, which does not hold for the ML Heckman estimator.

[5] In Beblo et al. (2003), we used both the predicted linear index and the normal hazard (separately), and found very similar results with both approaches.

estimation of the second stage, one can obtain a consistent estimator for the β-parameters of the wage equation, either by least squares or by instrumental variable estimation if some of the regressors are endogenous.

In our application, we compare the cross-section estimates obtained from the endogenous sample selection model by Lewbel, using the non-earned income of the household as the special variable S, with those resulting from OLS estimation and Heckman two-step estimation. We choose the German sample as a reference sample. For the remaining four countries we use the Lewbel procedure only. Concentrating on the cross-section Lewbel model, we are able to disentangle the components of the gender pay gap while considering at least the most prevailing of the methodological problems discussed above, that is, taking account of the selection effect in a non-restrictive way.

2.2.2 Decomposition of the gender wage gap

The general procedure of all decomposition methods is that, first of all, wage equations are estimated using individual and other characteristics, such as firm characteristics, as explanatory variables. The estimated price vector $\widehat{\beta}$ and the average human capital and job characteristics for males and females are used to compute weighted differences in mean characteristics. This part of the wage gap is assumed to reflect productivity differences. We call it the *endowment effect*. The adjusted pay gap is then measured as the difference between the total wage differential observed and the fraction explained by differences in human capital endowments of women and men. This remaining part measures differences in the remuneration of the characteristics. We call this second term the *remuneration effect*.[6]

Suppose wages are estimated without bias, i.e.:

$$\overline{\ln W}^J = \overline{X}^J \widehat{\beta}^J, \qquad \forall\ J = M, F, \tag{3}$$

where $\overline{X}$ is the sample mean of X. Oaxaca (1973) and Blinder (1973) propose to decompose the raw wage gap as follows

$$\begin{aligned}\underbrace{\left\{\overline{\ln W}^M - \overline{\ln W}^F\right\}}_{\text{raw wage gap}} &= \left\{\overline{\ln W}^M - \overline{\ln W^*}^F\right\} + \left\{\overline{\ln W^*}^F - \overline{\ln W}^F\right\} \\ &= \underbrace{\left(\overline{X}^M - \overline{X}^F\right)\widehat{\beta}^M}_{\text{endowment effect}} + \underbrace{\overline{X}^F\left(\widehat{\beta}^M - \widehat{\beta}^F\right)}_{\text{remuneration effect}},\end{aligned} \tag{4}$$

[6] It is important to stress that the decomposition into endowment and remuneration effects is conditioned by the list of explanatory variables included in the wage regression. Beyond the endogeneity problems already touched upon, a problem of variable selection thus arises. An extreme view is that "all relevant variables measuring individuals' productive endowments" should be included (Berndt 1991, p. 184). A more pragmatic strategy when elaborating policy recommendations could be to involve the policy makers in the definition of the explanotory variables to be used. A prevailing opinion, already expressed by Oaxaca (1973), seems to be that the greater the number of control variables is, the smaller the endowment effect will be. While the accumulated empirical evidence tends to support that opinion we would like to stress that it is by no means guaranteed that including more regressors has that effect. An extreme example is given by the Italian case: the endowment effect is negative (see Section 4).

where $\overline{\ln W^{*}}^{F} = \overline{X}^{F}\widehat{\beta}^{M}$.[7] Here $\widehat{\beta}^{M}$ is taken as the non-discriminating wage structure. The first term $\overline{\ln W}^{M} - \overline{\ln W^{*}}^{F}$ indicates the hypothetical wage differential if women had the same wage structure as men. The second term $\overline{\ln W^{*}}^{F} - \overline{\ln W}^{F}$ shows the distance between the hypothetical wage rate for women and their actual mean wage. When using female prices $\widehat{\beta}^{F}$ or a weighted price vector as reference, the wage decomposition results may differ. These variations and the empirical consequences of using them are discussed in detail in the study of Oaxaca and Ransom (1994).

Going beyond the Oaxaca-Blinder decomposition at the mean, and taking a closer look at the wage differential by quantiles of the wage distribution provides additional information on the nature of gender earnings inequality. This is done in the method developed by Juhn, Murphy and Pierce (1993), which also takes the residual wage distribution into account. The main feature of the Juhn-Murphy-Pierce approach is the decomposition of the raw gap at different points of the wage distribution. Consequently this decomposition method allows an analysis at the quantile level:

$$\underbrace{\Delta\overline{\ln W_q}}_{\text{raw wage gap}} = \underbrace{(\overline{X}_q^{M} - \overline{X}_q^{F})\widehat{\beta}^{M}}_{\text{endowment effect}} + \underbrace{\overline{X}_q^{F}(\widehat{\beta}^{M} - \widehat{\beta}^{F})}_{\text{remuneration effect}} + \underbrace{(\overline{\varepsilon}_q^{M} - \overline{\varepsilon}_q^{F})}_{\text{unobservable effect}} \quad , \qquad (5)$$

where $\overline{X}_q$ represents the sample mean of X for quantile q. The Juhn-Murphy-Pierce decomposition has been applied mostly in studies analysing the wage gap between two groups of workers over time or across countries.

In our application we compute the Oaxaca-Blinder and Juhn-Murphy-Pierce decompositions based on the wage estimations selected. We use the estimates from the male wage regressions as the reference remuneration, that is, as the non-discriminatory salary structure. As Ginther and Hayes (2003) point out, men are the usual comparison group in legal proceedings concerning gender discrimination. Hence a pooled approach, obtained from a (matrix) weighted average of the male and female wage structures, is not likely to be used in legal cases concerned with equal opportunities for women and men.

2.2.3 Treatment of the sample selection correction

This section discusses how to handle the selectivity bias correction within the decomposition of the raw wage gap. In the following, the decomposition method is applied for the case where a selection bias correction is performed for both sexes. If a random sample for men is assumed, the correction term is set to zero.

We first consider the Oaxaca-Blinder decomposition and Heckman two-step estimates. When applying the Heckman two-step regression technique, we are able to distinguish the endowment and remuneration effects from a selection effect. This gives us an idea of how the wage distribution of women, and hence the wage gap, would look like in the absence of sample selection. The selection

[7] Note that the fact that the unweighted average of estimation residuals is zero does not necessarily imply the same property for the corresponding weighted average. However, empirically, we have found the weighted averages to be negligible, and have omitted them in the presentation for simplicity.

correction terms enter the wage decomposition as follows:

$$\underbrace{\Delta\overline{\ln W}}_{\text{raw wage gap}} = \underbrace{(\overline{X}^M - \overline{X}^F)\widehat{\beta}^M}_{\text{endowment effect}} + \underbrace{\overline{X}^F(\widehat{\beta}^M - \widehat{\beta}^F)}_{\text{remuneration effect}} + \underbrace{(\overline{\widehat{\lambda}}^M\widehat{\theta}^M - \overline{\widehat{\lambda}}^F\widehat{\theta}^F)}_{\text{selection effect}}, \quad (6)$$

where $\widehat{\theta}$ is an estimate of $\rho\sigma_\varepsilon$. The first two terms of the right-hand side in equation (6) are the familiar endowment and remuneration components.

However, it is not obvious how the last term in equation (6) should be treated in the overall decomposition scheme, that is, whether it should be attributed to differences in endowments or included in the remuneration effect. Several variants are found in the literature.

In most studies, the last term on the right-hand side of the equation (6) is subtracted from the observed wage gap on the left-hand side. In this form the left-hand side provides a measure of the difference in potential or offered wages, in contrast to observed wages realized only by those participating in the labour market (see among others Oglobin 1999).[8] The studies which proceed in this way find that the existence of a sample selection bias implies that the "offered wage gap" greatly exceeds the observed wage gap. However, this empirical result is obtained with $\widehat{\theta}^M = 0$ and $\widehat{\theta}^F > 0$ (no selection for men, positive selection for women, that is, positive correlation between unobservables in the wage and participation equations) and is therefore by no means general. Note also that with other selectivity correction approaches, the relative magnitude of the offered and observed wage gap may not relate directly to the sign of a coefficient. With Olsen's (1980) approach, for instance, the additional regressor designed to correct for selection is the predicted linear index of the participation equation. Given that this is inversely related to the normal hazard included in the Heckman correction, one expects a negative coefficient for this regressor. However, the sign of the mean of the participation index can differ depending on the participation rate.

Thus the impact on the remuneration and endowment effects of taking the correction of sample selection into account, is ambiguous. For instance in the investigation of Oglobin (1999) both the remuneration and the endowment effects decline in comparison with the results of an OLS regression. But both effects increase in the study of Miller (1987). Miller und Rummery (1991) show that the effects may point in opposite directions. In their study the endowment effect declines and the remuneration effect increases.

Dolton and Makepeace (1986) treat the correction term as a regular explanatory variable. The difference in $\overline{\widehat{\lambda}}$ between men and women weighted with $\widehat{\theta}^M$ enters into the endowment effect. The distance between the estimated coefficients for the correction term add to the remuneration effect. In contrast to this, Dolton and Kidd (1994) and Choudhury (1993) choose a decomposition similar to equation (6) and interpret the third term as representing the appropriate correction for non-random sampling.

Summarising, sample selection may be taken into account in various ways. It is not possible to point out any of these as the right one, since the appropriate procedure depends on the specific empirical problem and the data at hand, as discussed by Neuman and Oaxaca (2001). However, empirical studies largely

[8] The offered wage coincides with the observed wage for participants, but is not observed for non-participants.

support the significance of the selection bias correction, particularly for women. Following Dolton and Kidd (1994) and Choudhury (1993), we will treat the selectivity bias as an additional effect to the endowment, remuneration (and unobservables) effects. Equation (6) can be straightforwardly extended to the Juhn-Murphy-Pierce decomposition.

Contrary to OLS or Heckman two-step estimators (at least the OLS version of the latter), the sample mean of the unobservables ($\bar{\varepsilon}$) is not zero for the Lewbel estimator. Thus, even a wage decomposition at the mean yields an unobservables effect if we use Lewbel estimates. This effect cannot be interpreted straightforwardly. However, it may capture part of the selection effect.

3 Data

Our empirical analyses are based on data from the user data base of the European Community Household Panel (ECHP UDB, version December 2002). In order to meet the requirement of more in-depth knowledge and greater compatibility of data on social and economic conditions in the European Union, the ECHP was launched as a closely coordinated component of a system of household surveys. The ECHP is a standardized survey conducted in member states of the European Union under the auspices of the Statistical Office of the European Communities.[9] It involves annual interviews of a representative panel of households and individuals in each country, covering a wide range of topics on living conditions. This includes comparable information across member states, on income, work and employment, poverty and exclusion, housing, health, and many other social indicators. The key feature of the ECHP is harmonisation of its methodology, specifically through the creation of a centralised questionnaire which serves as a point of departure for all national surveys. The ECHP is thus a rich data set, providing a wide range of information for investigating the distribution of wage income within and across European countries. Despite its merits, the ECHP involves restrictions when analyzing employment and wages in many ways. For example, sectors, education and children information as well the income sources are very aggregated. Nevertheless, the comparable nature of the ECHP data permits a cross-country analysis for the EU member states on the highest compatibility level available.

We select our estimation samples according to the following criteria. We include respondents from all nationalities, aged 25 to 55, who are presently employed or out of the labour force. The quite restrictive selection on age is made to prevent the results from being excessively affected by education and early retirement decisions that may influence participation behaviour and differ between countries. We exclude the self-employed and people working in family businesses because of the difficulty to obtain credible information on earned income for these categories. Unemployed workers, pensioners, students, those in special training programs or national service (military or civil) as well as people with a disability are excluded, too. These restrictions are justified by the aim to form a fairly homogeneous sample of employed persons and individuals who are “voluntarily” out of the labour force.

[9] For a detailed description of the ECHP methodology and questionnaires see Eurostat (1996).

Table 2: Sample sizes of the ECHP country files

Country	# women	% working	# men	% working
Austria	1,343	70	1,270	100
Belgium	1,144	79	1,076	99
Denmark	963	97	985	100
France	2,260	70	1,935	97
Finland	1,431	92	1,310	100
Germany	2,608	77	2,576	99
Greece	1,680	47	1,206	100
Ireland	1,407	58	969	99
Italy	3,235	53	2,506	99
The Netherlands	2,069	82	2,147	99
Portugal	1,943	72	1,778	100
Spain	2,616	49	2,176	99
United Kingdom	2,166	82	1,740	99
EU	24,865	69	21,677	99

Data source: ECHP, country files 1998. Samples of 25-55 year old women and men, who are employed or out of the labour force.

Furthermore, we restrain wage earners to have positive earnings and work for at least 8 hours per week. This last restriction aims at minimising measurement errors connected with the wage measure. Wages are not observed directly, but obtained by dividing current monthly total gross earnings by the current total number of hours worked per week in main and additional jobs, the latter multiplied by 4.3. It turns out that outliers in the resulting wage distribution correspond to individuals reporting very low hours of work. Observations with missing information on household net income – required to model participation – do not enter the analysis. The remaining sample sizes for all ECHP country files are listed in Table 2. Total sample sizes range from more than 3000 women and 2500 men in the Italian sample to less than 1000 women and men for Denmark. In these samples, the female participation rates differ considerably between countries, while male participation rates are near to 100 % in almost every country. The female participation rate is particularly low in the South of Europe (Spain, Italy and Greece), while it is remarkably high in the Scandinavian countries Denmark and Finland. Also remarkable are the unbalanced sizes of the selected samples between the sexes in Southern Europe. One explanation is that men are more often self-employed in rural activities or in services and are therefore excluded. On the other hand, women, especially mothers, are more likely not to participate and thus are kept in our sample. As mentioned earlier, Luxembourg and Sweden are not included, due to missing information on earnings.

Based on both, sample and population size, we choose to focus our further investigation on the gender wage gaps in the five largest EU countries, Germany, France, Italy, Spain and the United Kingdom. From the ECHP, we choose the 1998 cross section to facilitate comparison with the results published in the report Employment in Europe 2002 (European Commission 2002a).

Table 3: Goodness of fit for female wages

Estimation model	No. of obs.	Mean wage	Std. dev.
Actual wage	2,015	9.80	4.57
OLS	1,985	9.82	2.73
Heckman	1,982	9.82	2.75
Lewbel	1,884	9.83	2.84

Data source: ECHP, German data file 1998. Sample of 25-55 year old women, who are employed at least 8 hours per week.

4 Results

We investigate various estimation techniques and the resulting decomposition effects. For the 1998 German sample of the ECHP data set, we provide results for both the Oaxaca-Blinder and the Juhn-Murphy-Pierce wage gap decompositions, based on three different estimators for the wage equations: OLS, two-stage Heckman, and Lewbel. For the latter, we only report results from the simplest version (with parametric estimation of $f(S|U)$, see 2), since performing a non-parametric estimation for the first stage had only a small impact on the estimated coefficients. For the other four countries we present only the wage gap decompositions based on the Lewbel estimates.

4.1 Wage equation estimation for Germany

As the male participation rate is 99 %, the wage equation for men is estimated by OLS. For decompositions based on the Lewbel method, both wage equations, for males and females, result from the Lewbel procedure. As explanatory variables we include individual characteristics (such as age, education, tenure, information on household composition and regional information) and job characteristics (firm size, sector, occupational group). The participation equation for the Heckman two-step procedure is set up with age, family status, children, non-earned income and various interaction terms as independent variables. The selection correction variable turns out to be statistically significant, with a negative coefficient, and thus a negative correlation between unobservables in the participation equation and in the wage equation. This result underlines the necessity to correct for sample selection, since the parameters would not be estimated consistently otherwise.

As the special variable in the Lewbel model, we use annual non-work net household income. In addition, we experimented with the respective non-work individual income, as well as total household income, excluding only the individual's labour earnings. The first measure performed best with regard to the predicted wages. We opt for the non-earned household income for theoretical reasons also, since this income measure is less likely to be correlated with the individual wage rate than the other candidates examined.

Table 3 summarises the predictive performance of the competing wage equation models. The variance of the predicted wages is largest in the Lewbel model. The Lewbel predictions therefore represent best the spread of the actual wage distribution.

The estimation results for the German sample are listed in Appendix 1. The results for the other countries can be found in Beblo et al. (2003).

4.2 Wage decomposition

We discuss in turn the Oaxaca-Blinder and Juhn-Murphy-Pierce decomposition results.

4.2.1 Oaxaca-Blinder decomposition

As illustrated in Figure 2 the magnitude of the various terms in the raw wage gap decomposition differ remarkably depending on the wage regression model used. According to the OLS estimation, a little bit more than 50 % of the German wage differential can be explained by different endowments of women and men. The remaining half is due to differences in the remuneration of these endowments. The results of the Lewbel procedure reflect an endowment effect of similar size. The selection effect from the Heckman regression corresponds to a negative selection on unobservables (negative correlation between the unobservables in both equations). Note, however, that this means that a woman with a relatively low predicted probability of participation is predicted to earn less conditional on participation than a woman with the same productivity endowment X but a higher predicted probability of participation. The Heckman findings for Germany reveal that the potential wage gap between women and men, in the absence of selection, would be lower than observed. No selection effect is displayed for the Lewbel estimation since selection is not estimated explicitly, but it shows off indirectly in the unobervables whereas the unobservable effect is restricted to zero on average for OLS and Heckman estimates. The Lewbel estimate tells us that unobserved characteristics of women lead to an offered wage gap which exceeds the observed wage gap, contrary to the findings obtained with the Heckman estimates.

In levels, the offered gender wage gaps estimated with the Lewbel approach in the other country samples differ from the German sample in both directions. However, according to Figure 3, the fraction explained by different endowments of women and men is generally smaller in the other countries. Consequently, the remuneration effect, that is, the relative effect of differences in remuneration, is larger. Italy and Spain have a negative endowment effect. This means that if Italian and Spanish women had the same characteristics as Italian and Spanish men they would receive even lower wages than observed. In Italy the gap in potential wages is almost 50 % higher than that in observed wages.

4.2.2 Juhn-Murphy-Pierce decomposition

Juhn-Murphy-Pierce decompositions of the gender wage gaps within countries are displayed in Figures 4 to 10 in Appendix 1. Recall that the advantage of this approach is that it allows an investigation of the wage gap over the whole wage distribution. It thus relates to the Oaxaca-Blinder decomposition by disentangling the pay difference at the overall sample mean into quantile gaps. To illustrate this feature let us take a closer look at the German example. In the Figures 4 to 6 the raw wage gap is displayed according to the wage rankings of all women and men in the sample. As in the figures on the raw

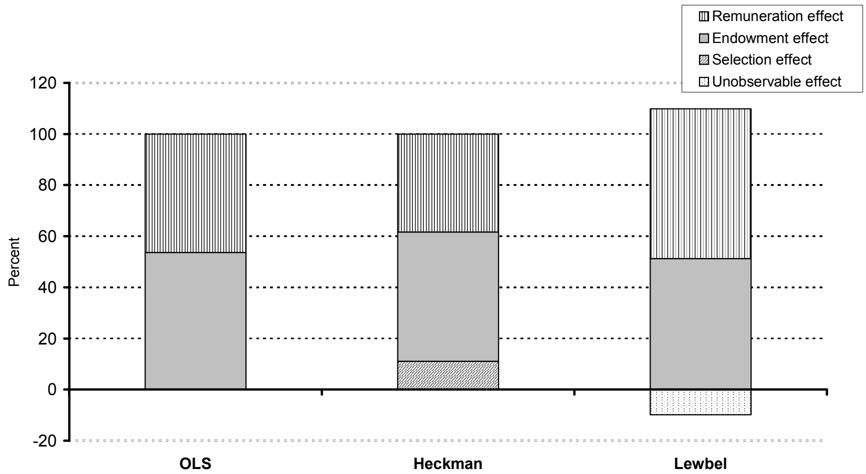

Figure 2: Oaxaca-Blinder decomposition of the gender wage gap in Germany.
Data source: ECHP, German data file 1998. Sample of 25-55 year old women and men, who are employed at least 8 hours per week.

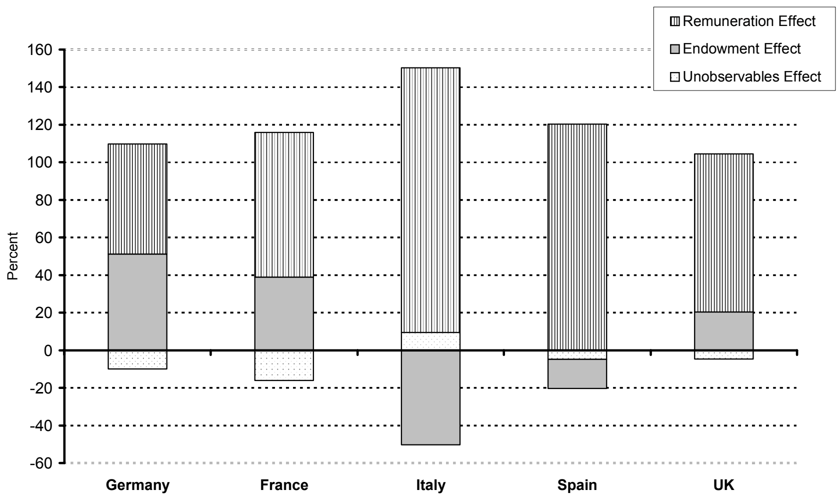

Figure 3: Oaxaca-Blinder decomposition of the gender wage gap in selected EU countries
Data source: ECHP, country files 1998. Sample of 25-55 year old women and men, who are employed at least 8 hours per week.
Note: Decomposition based on the Lewbel wage regression.

pay gap, women and men are ranked separately by their wage rate and then compared pairwise in each decile of the distribution.[10] That is, the mean wage of the 10 % lowest paid women is now compared with the mean wage of the 10 % lowest paid men. This gap is displayed as the first observation to the left of the panel. As the wage distributions of women and men are not congruent typically, the decile gaps also reflect the extent to which male and female wage dispersions differ.

In the German case lowest-wage earners face a gender wage gap of about 31 %. The raw gap first decreases and then increases when moving up the income distribution. For middle-wage earners it is as low as 20 % whereas for high-wage women it amounts to 28 %. As in the Oaxaca-Blinder decomposition, the raw gap can be decomposed into an endowment and a remuneration effect. Furthermore, since we are no longer confined to the sample means, the effect of unobservables can also be illustrated. The fraction of the wage gap explained by different endowments of women and men increases over the wage distribution. While taking up only about 20 % at the lower end of the wage distribution, it accounts to more than 80 % of the gap between high-wage women and their male counterparts. The remuneration effect has the reverse pattern. Different remuneration of characteristics seems to affect mostly women at the lower end of the wage scale. Unobservables take up some of the wage gap for low income receivers but their effect is close to zero for the rest.

In contrast to Figure 4, Figure 5 is based on selectivity corrected wage estimations for the female sample. As seen in the Oaxaca-Blinder decomposition above, the selection effect corresponds to a negative selection on unobservables. It takes up part of the observed wage gap. This means that the offered wage gap is smaller than the observed wage gap. Remarkably enough, this applies to the whole of the wage distribution to more or less the same extent. As was already the case for the Oaxaca-Blinder representation, the Lewbel estimates differ quite a bit from the Heckman estimates when differentiated over deciles (see Figure 6). While the increasing pattern of the endowment effect with rising wage is confirmed, the remuneration effect decreases from the second decile on, for the Heckman estimates, but its profile is more or less flat according to the Lewbel estimates.

In France the U-shape of the raw gender wage gap is quite pronounced (see Figure 7). Starting with some 25 % in the lowest decile, the gap diminishes to little more than 10 % for the rest of the wage distribution, except for the highest earners where it amounts to 18 %. In contrast to the German case, the explanatory power of the endowment effect varies a lot over the income distribution. Also the pattern of unobservables shows more volatility for France due to, first, the better specification of the wage equation model in Germany, and second, the smaller sample size for France. As a consequence, the quasi-totality of the raw wage gap in France is attributed to a different remuneration of women and men. The UK graph reveals a similar pattern, although the endowment effect takes up a larger fraction of the pay gap, and so does the Spanish. In Italy only the wage difference of the highest earners can be explained by different endowments of women and men.

[10] Note that the the raw wage gap was displayed by percentile.

5 Conclusion

Based on the European Community Household Panel for five European countries (France, Germany, Italy, Spain and the United Kingdom) we investigate different techniques to assess the gender pay gap in the European Union. These different techniques concern the estimation of wage equations as well as the decomposition of the estimated gaps. Our comparative study shows that at most 50 % of the difference in payment between the sexes can be attributed to differences in characteristics. However, the size of the endowment effect differs considerably between countries. It depends on the information used and on the estimation model and decomposition method applied.

Regarding the wage estimation, we find that a selectivity correction may have a significant impact both on wage estimates and on the pay gap decomposition: our results suggest that, for Germany, the offered wage gap is smaller than the observed wage gap on the basis of the Heckman estimates, but the reverse using the Lewbel estimates. Therefore, careful attention should be paid to the choice of estimation method. We prefer the Lewbel approach because it is less restrictive on the structure of the data. No structure is imposed on the distribution of the error terms, permitting a more general form of unknown heteroscedasticity.[11]

Regarding the wage decomposition, another main result of the study is derived from the decomposition over quantiles of the wage distribution (Juhn, Murphy and Pierce method). Remarkable differences are revealed within as well as between countries. A further recommendation derived from our analysis would therefore be to pay careful attention to differences over the wage distribution when drawing policy conclusions.

Future research may include applications of further wage estimation techniques (e.g. quantile regression) and gap decomposition methods (e.g. Machado and Mata 2003). Furthermore, confidence bands for the adusted pay gap should be investigated, as well as for the different effects in the wage decomposition.

References

[1] Beblo, M., D. Beninger, A. Heinze and F. Laisney (2003), *Methodological Issues Related to the Analysis of Gender Gaps in Employment, Earnings and Career Progression*, Final report, European Commission, Employment and Social Affairs DG.

[2] Becker, G. (1964), *Human Capital - A Theoretical and Empirical Analysis with Special Reference to Education*, Columbia University Press, Chicago.

[3] Berndt, E.R. (1991), *The Practice of Econometrics: Classic and Contemporary*, Addison-Wesley Publishing Company.

[4] Blinder, A. S. (1973), Wage Discrimination: Reduced Form and Structural Estimates, *The Journal of Human Resources,* 8(4), 436-455.

[11] From a meta-analysis on the empirical literature on gender wage differentials, Weichselbaumer and Winter-Ebmer (2002) find that sample selection techniques à la Heckman do not seem to matter in terms of their impact on the residual gender pay gap. Though various factors may play a role in such an international and, hence, heterogenous set-up, part of the explanation may be that the distribution assumptions have simply not been fulfilled.

[5] Choudhury, S. (1993), Reassessing the Male-Female Wage Differential: A Fixed Effects Approach, *Southern Economic Journal,* 60 (2), 327-341.

[6] Dolton, P.J. and G. H. Makepeace (1986), Sample Selection and Male-Female Earnings Differentials in the Graduate Labour Market, *Oxford Economic Papers,* 38, 317-341.

[7] Dolton, P. J. and M. P. Kidd (1994), Occupational Access and Wage Discrimination, *Oxford Bulletin of Economics and Statistics,* 56 (4), 457-474.

[8] European Commission (2002a), *Employment in Europe 2002*, Luxembourg.

[9] European Commission (2002b), *The Gender Pay Gap and Gender Mainstreaming Pay Policy: Synthesis Report of Gender Pay Equality in EU Member States*, Report prepared by the Group of Experts on Gender and Employment commissioned by DG Employment and Social Affairs, November 2002, Brussels.

[10] European Commission (2002c), *The Adjusted Gender Pay Gap: A Critical Appraisal of Standard Decomposition Techniques*, Report prepared by D. Grimshaw and J. Rubery as part of the work by the co-ordinating team of the Group of Experts on Gender and Employment commissioned by DG Employment and Social Affairs, March 2002, Brussels.

[11] Eurostat (1996), *The European Community Household Panel (ECHP): Volume 1- Survey Methodology and Implementation and the European Community Household Panel (ECHP): Volume 1 - Survey Questionnaires: Wave 1-3 - Theme 3, Series E*, Eurostat, OPOCE, Luxembourg.

[12] Ginther, D.K. and K.J. Hayes (2003), Gender Differences in Salary and Promotion for Faculty in the Humanities 1977-95, *Journal of Human Resources,* 38(1), 34-73.

[13] Heckman, J. (1979), Sample Selection Bias as a Specification Error, *Econometrica,* 47, 153-163.

[14] Heckman, J.J., L.J. Lochner and P.E. Hold (2003), Fifty Years of Mincer Earnings Regression, IZA Discussion Paper 775, Bonn.

[15] Juhn, C., K. M. Murphy and B. Pierce (1993), Wage Inequality and the Rise in Returns to Skill, *Journal of Political Economy,* 101 (3), 410-442.

[16] Kyriazidou, E. (1997), Estimation of a Panel Data Sample Selection Model, *Econometrica,* 65, 1335-1364.

[17] Lewbel, A. (2002), Selection Model and Conditional Treatment Effects, Including Endogenous Regressors, *mimeo*, Boston College.

[18] Machado, J., and J. Mata (2003), Counterfactual Decomposition of Changes in Wage Distribution Using Quantile Regression, *mimeo,* Universidade de Lisboa.

[19] Miller, P. W. (1987), Gender Differences in Observed and Offered Wages in Canada, 1980, *The Canadian Journal of Economics*, 20 (2), 225-244.

[20] Miller, P. and S. Rummery (1991), Male-Female Wage Differentials in Australia: A Reassessment, *Australian Economic Papers*, 30 (56), 50-69.

[21] Mincer, J. (1958), Investment in Human Capital and Personal Income Distribution, *Journal of Political Economy*, 66(4), 281-302.

[22] Mincer, J. (1974), *Schooling, Experience and Earnings*, Columbia University Press, New York.

[23] Neuman, S. and R. L. Oaxaca (2001), Estimating Labor Market Discrimination With Selectivity-Corrected Wage Equations: Methodological Considerations and an Illustration From Israel, *mimeo.*

[24] Oaxaca, R. L. (1973), Male-Female Wage Differentials in Urban Labor Markets, *International Economic Review,* 14 (3), 693-709.

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

[26] OECD (2002), Women at Work: Who are They and How are They Faring?, *OECD Employment Outlook*, Chapter 2, 61-125.

[27] Oglobin, C. G. (1999), The Gender Earnings Differential in the Russian Transition Economy, *Industrial and Labor Relations Review,* Vol. 52 (4), 602-627.

[28] Olsen, R. J. (1980), A Least Squares Correction for Selectivity Bias, *Econometrica,* 48 (7), 1815-1820.

[29] Reimers, C. (1983), Labor Market Discrimination Against Hispanics and Black Men, *Review of Economics and Statistics,* 65 (4), 570-579.

[30] Rice, P. (1999), Gender Earnings Differentials: The European Experience, *The World Bank Development Research Group,* WP No.8.

[31] Vella, F. (1998), Estimating Models with Sample Selection Bias: A Survey, *The Journal of Human Resources,* 33 (1), 127-169.

[32] Weichselbaumer, D. and R. Winter-Ebmer (2002), The Impact of Markets, Politics and Society on the Gender Wage Gap: A Meta Analysis, *Working Paper,* University of Linz.

[33] Wooldridge, J. M. (1995), Selection Corrections for Panel Data Models Under Conditional Mean Independence Assumptions, *Journal of Econometrics*, 68, 115-132.

6 Appendix 1: Juhn-Murphy-Pierce figures

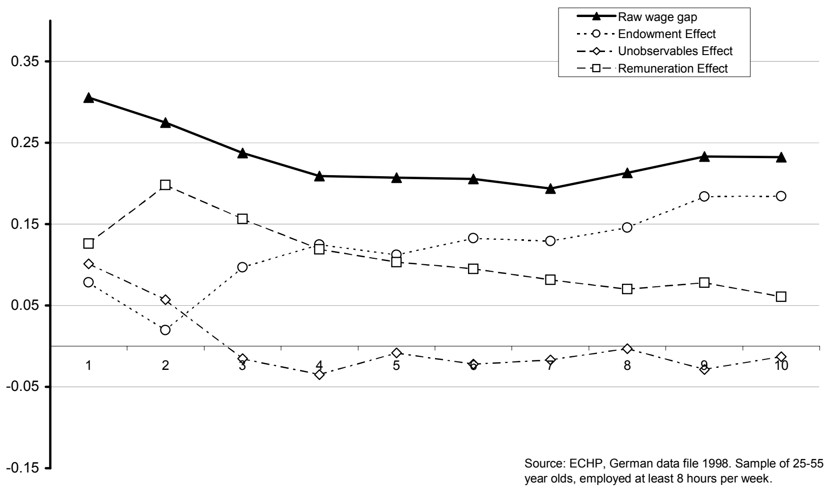

Figure 4: Juhn-Murphy-Pierce decomposition of the gender wage gap in Germany (OLS wage estimation)

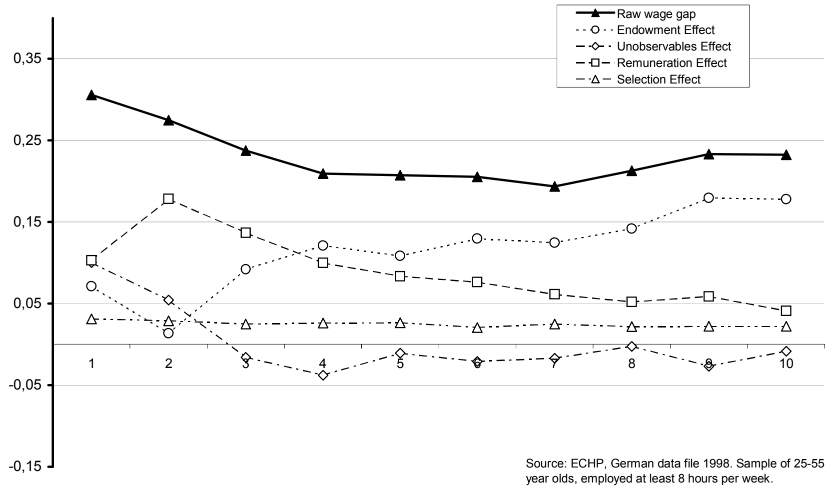

Figure 5: Juhn-Murphy-Pierce decomposition of the gender wage gap in Germany (Heckman two-step wage estimation)

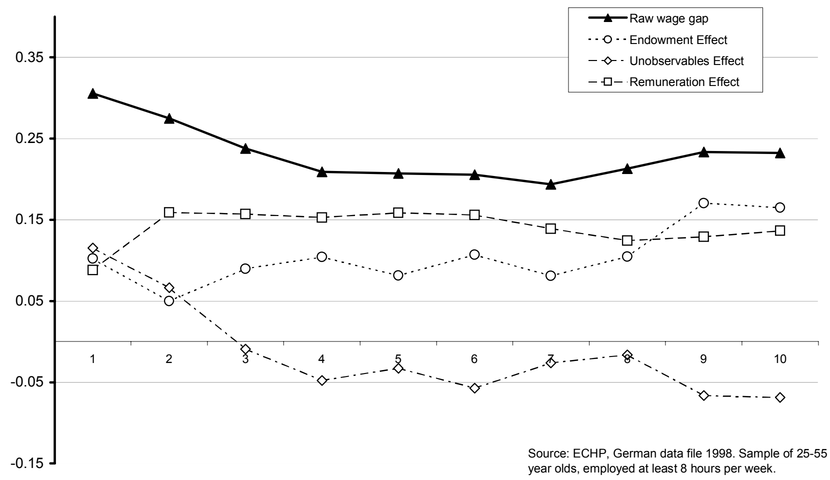

Figure 6: Juhn-Murphy-Pierce decomposition of the gender wage gap in Germany (Lewbel two-step wage estimation)

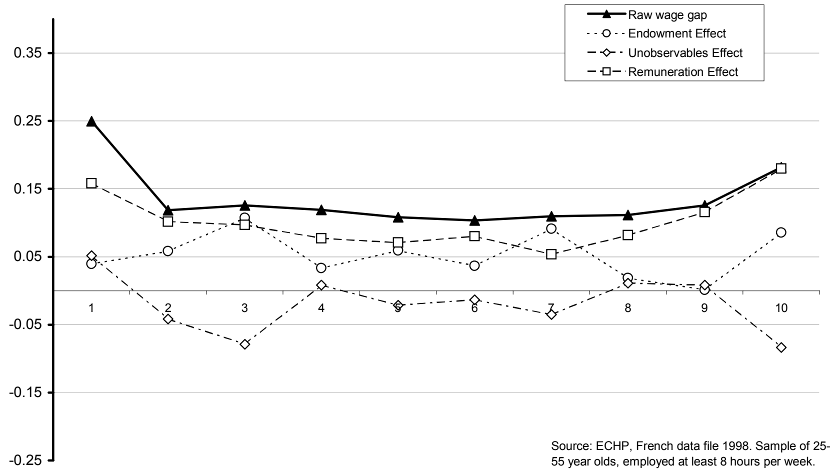

Figure 7: Juhn-Murphy-Pierce decomposition of the gender wage gap in France (Lewbel two-step wage estimation)

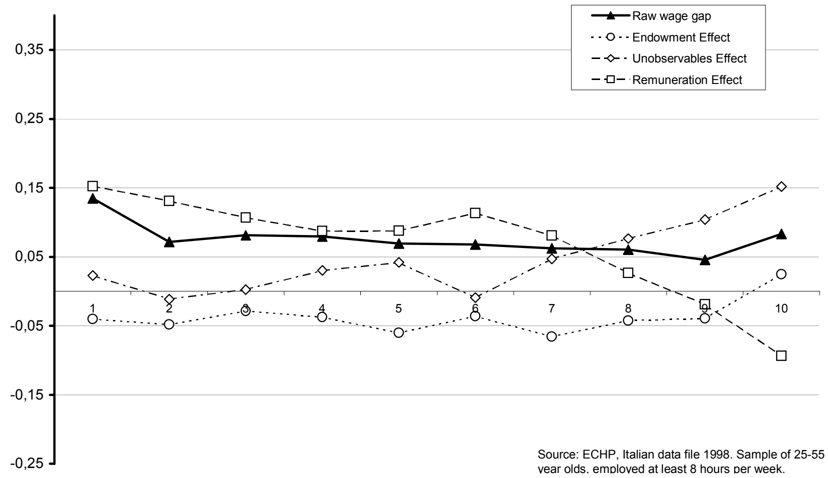

Figure 8: Juhn-Murphy-Pierce decomposition of the gender wage gap in Italy (Lewbel two-step wage estimation)

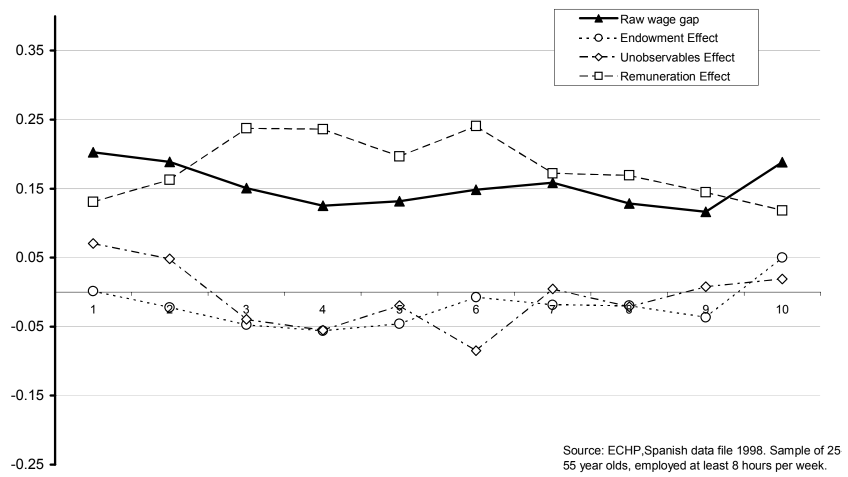

Figure 9: Juhn-Murphy-Pierce decomposition of the gender wage gap in Spain (Lewbel two-step wage estimation)

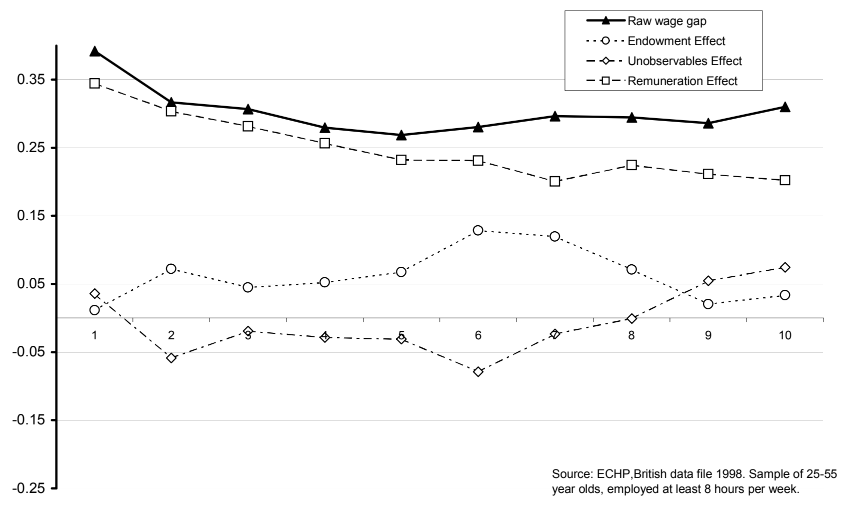

Figure 10: Juhn-Murphy-Pierce decomposition of the gender wage gap in the UK (Lewbel two-step wage estimation)

7 Appendix 2: Descriptive statistics and wage estimation results for Germany

Table 4: Descriptive statistics for Germany (whole sample)

		Men		Women	
Variable name	Variable	Mean	Std. Dev.	Mean	Std. Dev.
a_y	Dummy: age<30	.19	.40	.20	.40
a_m	Dummy: age 30-45	.53	.50	.51	.50
age	Age	38.73	8.38	38.99	8.42
mar	Dummy: married	.72	.45	.74	.44
mar_coh	Dummy: married or cohabiting	.82	.38	.84	.37
educ_3	Dummy: higher education	.27	.44	.20	.40
educ_2	Dummy: secondary education	.57	.50	.58	.49
sizehh_015	Number of children <16	.86	1.00	.84	1.01
sizehh_1415	Number of Children 14-15	.14	.37	.17	.41
sizehh	Household size	3.19	1.26	3.16	1.25
child_0	Dummy: newborn	.06	.23	.03	.17
child_011	Dummy: child<12	.39	.49	.35	.48
child_111	Dummy: child 1-11	.34	.47	.32	.47
income_ne	Annual hh income − individual earnings	11807.86	9580.88	20202.43	13194.17
east	Dummy: living in east Germany	.24	.43	.23	.42
citizen_other	Dummy: non EU citizen	.09	.29	.11	.31
citizen_EU	Dummy: EU citizen	.07	.25	.05	.23
ihhu	Annual nonearned hh income	1401.35	4906.59	1685.38	5349.64
job_part	Dummy: participation	.99	.10	.77	.42
# obs.		2562		2577	

Data source: ECHP, German data file 1998. Samples of 25-55 year old women and men, who are employed or out of the labour force.

Table 5: Descriptive statistics for Germany (only working individuals)

		Men		Women	
Variable name	Variable	Mean	Std. Dev.	Mean	Std. Dev.
job_magr	Dummy: legislators. senior officials and managers	.09	.28	.03	.07
job_prof	Dummy: professionals	.12	.32	.10	.30
job_cler	Dummy: clerks	.07	.26	.24	.43
job_sale	Dummy: service and sales workers	.04	.20	.14	.34
job_wagr	Dummy: skilled agricultural and fishery workers	.01	.10	.01	.10
job_wser	Dummy: craft and related trades workers	.33	.47	.04	.19
job_wqua	Dummy: plant and machine operators and assemblers	.15	.36	.19	.22
job_welm	Dummy: elementary occupations	.07	.25	.11	.31
job_sizelow	Dummy: small size firm	.19	.40	.24	.43
job_sizemid	Dummy: middle size firm	.55	.50	.43	.50
job_public	Dummy: public sector	.20	.40	.36	.48
job_tenure	Firm tenure (years)	7.39	6.40	5.95	5.90
job_pc	Dummy: permanent contract	.94	.23	.87	.34
job_break	Dummy: job interruption	.13	.34	.17	.38
hours_pt	Dummy: part time employment	.00	.06	.16	.37
wage_grossm	Hourly gross wage	12.30	6.50	9.79	4.59
# obs.		2388		1747	

Data source: ECHP, German data file 1998. Samples of 25-55 year old women and men, who are employed at least 8 hours per week.

Table 6: OLS wage estimation for German women 1998

log wage_grossm	Coef.	T-stat.
job_magr	.1836	3.94
job_prof	.2526	8.38
job_wagr	-.4525	-6.00
job_cler	-.0722	-3.41
job_sale	-.2790	-10.89
job_wser	-.1897	-4.47
job_wqua	-.2249	-5.92
job_welm	-.3590	-12.55
job_public	.0736	4.21
job_sizelow	-.2256	-10.50
job_sizemid	-.0615	-3.31
educ_3	.0992	3.32
educ_2	.0634	2.86
job_tenure	.0259	5.61
job_tenure^2	-.0006	-2.43
east	-.2417	-13.65
constant	2.2417	74.20
obs	1985	
R-squared	.4108	
Adj R-squared	.4060	

Data source: ECHP, German data file 1998. Samples of 25-55 year old women, who are employed at least 8 hours per week.

Table 7: OLS wage estimation for German men 1998

log wage_grossm	Coef.	T-stat.
log age	3.9321	3.63
$(\log\ \text{age})^2$	-.5282	-3.53
log (age × educ_3)	.3369	4.82
job_magr	.2060	7.21
job_prof	.1548	5.68
job_wagr	-.2204	-3.23
job_cler	-.0235	-0.79
job_sale	-.1949	-5.54
job_wser	-.0913	-4.16
job_wqua	-.1256	-4.95
job_welm	-.1864	-5.95
job_public	-.0631	-3.71
job_break	-.1264	-6.45
sizehh_015	.0222	3.27
hours_pt	-.3935	-4.04
job_sizelow	-.2976	-14.78
job_sizemid	-.1280	-8.34
educ_3	-1.1194	-4.36
educ_2	.0421	2.23
job_tenure	.0052	4.23
east	-.3442	-22.23
constant	-4.6979	-2.41
obs.	2388	
R-squared	0.4513	
Adj R-squared	0.4464	

Data source: ECHP, German data file 1998. Samples of 25-55 year old men, who are employed at least 8 hours per week.

Table 8: Probit participation estimation for German women 1998

job_part	Coef.	T-stat.
a_y	.5354	4.35
a_m	.3982	4.64
lage_ch0	3.5495	3.57
lage_ch111	1.4292	3.91
mar	-.7473	-5.24
mar_coh	.5489	3.10
log income_ne	-.3555	-2.65
child_0	-14.3697	-4.12
child_111	-6.0151	-4.60
educ_3	.8160	6.91
educ_2	.4454	6.08
east	.7322	7.56
constant	4.0094	3.15
obs	2608	
Pseudo R2	.2719	
Log pseudo-likelihood	-1017.9915	

Data source: ECHP, German data file 1998. Samples of 25-55 year old women, who are employed or out of the labour force.

Table 9: Heckman wage estimation for German women 1998

log wage_grossm	Coef.	T-stat.
job_magr	.1866	4.01
job_prof	.2496	8.29
job_wagr	-.4650	-6.20
job_cler	-.0702	-3.32
job_sale	-.2759	-10.78
job_wser	-.1852	-4.37
job_wqua	-.2190	-5.77
job_welm	-.3412	-11.82
job_public	.0738	4.22
job_pc	.0676	2.78
job_sizelow	-.2362	-10.78
job_sizemid	-.0733	-3.86
educ_3	.0782	2.52
educ_2	.0474	2.06
job_tenure	.0138	10.16
east	-.2550	-13.72
lmbd	-.0886	-2.75
constant	2.2603	58.91
obs	1982	
R-squared	.4143	
Adj R-squared	.4092	

Data source: ECHP, German data file 1998. Samples of 25-55 year old women, who are employed at least 8 hours per week.

Table 10: Lewbel regression for Germany 1998, 1. stage: density function for women

ihhu	Coef.	T-stat.
log age	-15.1919	-2.08
$(\log\ age)^2$	2.2318	2.21
educ_3	1.1174	7.45
educ_2	.6172	5.15
east	-.4282	-3.97
child_0	.6529	2.57
citizen_other	-1.221	-7.89
cititzen_EU	-.6343	-3.24
sizehh_015	-.0592	-1.21
constant	30.9817	2.36

Data source: ECHP, German data file 1998. Samples of 25-55 year old women, who are employed at least 8 hours.

Table 11: Lewbel regression for Germany 1998, 2. stage: selectivity-corrected wage estimation for women

log wage_grossm	Coef.	T-stat.
log age	8.6983	2.08
(log age)^2	-1.1965	-2.09
job_magr	.1838	.81
job_prof	.3042	3.14
job_wagr	-.0463	-.57
job_cler	-.3233	-3.75
job_sale	-.5803	-3.70
job_wser	.1484	1.14
job_wqua	-.3357	-2.58
job_welm	-.5323	-5.63
job_sizelow	-.2767	-3.57
job_sizemid	-.1141	-1.61
educ_3	-.1357	-1.57
educ_2	-.0648	-.82
job_tenure	.0171	3.32
east	-.2005	-3.49
constant	-13.3174	-1.75

Data source: ECHP, German data file 1998. Samples of 25-55 year old women, who are employed at least 8 hours per week.

Table 12: Lewbel regression for Germany 1998, 1. stage: density function for men

income_ne	coef.	t-stat.
mar_coh	-1.96	-2.46
mar_coh× log age	.649	3.03
sizehh_015	-.149	-3.42
educ_3	1.48	10.6
educ_2	.661	5.27
east	-.542	-5.57
citizen_other	-.869	-5.69
cititzen_EU	-.514	-2.98
constant	4.84	32.5

Data source: ECHP, German data file 1998. Samples of 25-55 year old men, who are employed or out of the labour force.

Table 13: Lewbel regression for Germany 1998, 2. stage: selectivity-corrected wage estimation for men

log wage_grossm	coef.	t-stat.
job_magr	.404	4.92
job_prof	.068	.90
job_cler	-.013	-.16
job_sale	-.287	-3.41
job_wagr	-.190	-1.34
job_twser	-.132	-1.90
job_wqua	-.188	-2.27
job_welme	-.216	-2.54
job_sizelow	-.177	-2.80
job_sizemid	-.033	-.69
job_break	-.097	-1.81
sizehh_015	.054	2.89
educ_3	.216	3.35
educ_2	.110	2.21
job_tenure	.011	3.45
east	-.366	-6.52
hours_pt	-.322	-2.96
constant	2.39	25.4

Data source: ECHP, German data file 1998. Samples of 25-55 year old men, who are employed at least 8 hours per week.