ISSN 1977-0375

eurostat
Methodologies and Working papers

Development of econometric methods to evaluate the Gender pay gap using Structure of Earnings Survey data

2009 edition

eurostat
EUROPEAN COMMISSION

eurostat
Methodologies and
Working papers

Development of econometric methods to evaluate the Gender pay gap using Structure of Earnings Survey data

2009 edition

eurostat
EUROPEAN COMMISSION

Europe Direct is a service to help you find answers to your questions about the European Union

Freephone number (*):

00 800 6 7 8 9 10 11

(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

More information on the European Union is available on the Internet (http://europa.eu).

Luxembourg: Office for Official Publications of the European Communities, 2009

ISBN 978-92-79-13181-3
DOI 10.2785/31247
Cat. No. KS-RA-09-011-EN-N

Theme: Population and social conditions
Collection: Methodologies and working papers

© European Communities, 2009

Development of econometric methods to evaluate the Gender pay gap using Structure of Earnings Survey data

A study carried out by the Research Center for Education and the Labour Market, Maastricht University

Foreword

The EU has a long-standing commitment to promoting gender equality, and the goal of eliminating inequalities and promoting equality between women and men is set out in Article 3(2) of the EC Treaty. One of the most important areas in this regard is the equal treatment of men and women in the labour market, particularly with respect to wages and salaries. In July 2007, the European Commission adopted a communication on the issue,[1] recalling the need to implement a series of measures in order to tackle the wage differences between women and men, including through the provision of better statistics and analysis of the factors influencing the wage differences.

In this framework and on request of the Indicators Group of the Employment Committee of the Council and in coordination with the Directorate-General Employment and Social Affairs of the European Commission, Eurostat collects the Structural Indicator "Gender pay gap (GPG) in unadjusted form" on an annual basis. The unadjusted GPG is the relative difference between the average gross hourly earnings of women and men within the economy.[2]

As an unadjusted indicator, the GPG gives an overall picture of gender inequalities in terms of pay and measures a concept which is broader than the concept underlying the principle of equal pay for equal work. In addition, the overall GPG figure does not take into account differences in individual characteristics of employed men and women, nor can it give an indication of the incidence and level of discrimination or segregation in the labour market.

From the reference year 2006 onwards, this indicator is based on the Structure of Earnings Survey (SES), a rich employer-employee matched data set.[3] However, the SES

[1] "Tackling the pay gap between women and men", COM(2007) 424 final

[2] $$\text{GPG} = \frac{\text{gross hourly earnings of male paid employees - gross hourly earnings of female paid employees}}{\text{gross hourly earnings of male paid employees}} \%$$

[3] The SES is set up by Council Regulation (EC) No 530/1999 of 9 March 1999 concerning structural statistics on earnings and on labour costs and Commission Regulation 1738/2005 amending Regulation (EC) No 1916/2000 as regards the definition and transmission of information on the structure of earnings.

has also some limitations, e.g., SES data do not cover employees in the public sector, in enterprises with less than 10 employees or self-employed. Moreover, it does not collect information related to personal characteristics such as marital status, number of children or work history and does not cover inactive or unemployed, which might be relevant for women's decisions to participate in education or the labour market. However, controlling for available variables would provide a first decomposition of the pay gap and result in methodological elements for the calculation of an adjusted GPG in Europe.

Hence, in order to investigate such a possible measurement of an adjusted GPG (based on the SES) that can be better interpreted and compared between countries, as well as to recommend (in the SES context) the most appropriate methods to measure the extent of the pay gaps, Eurostat launched in 2008 a study on the "Development of econometric methods to evaluate the Gender pay gap using Structure of Earnings Survey data". More specifically, the goal was to evaluate the SES data in the light of its above shortcomings and to propose a framework in which econometric analyses of the GPG using SES data can meaningfully be interpreted.

This report presents the results of this study that was completed in April 2009.

Eurostat Unit F/2 – Labour Market Statistics

Acknowledgements

This study was prepared by Arnaud Dupuy, Didier Fouarge and Bianca Buligescu, Research Centre for Education and the Labour Market (ROA), Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands.

The views expressed in this publication are those of the authors and do not necessarily reflect the opinion of the European Commission.

The SES includes very detailed information on wages (e.g. payments made for overtime or shift work, bonuses, days of paid leave) and firm data (e.g., bargaining regime and economic and financial control of the firm).

Summary

Gender equality has a high priority on the policy agenda of the European Union (EU), and the Gender pay gap is a key indicator of gender equality. The aim of this research is to develop a methodology for measuring the Gender pay gap on the Structure of Earnings Survey (SES) data, a new and rich database at Eurostat. The SES, however, has two drawbacks that must be accounted for when calculating the Gender pay gap: 1) it suffers from a potential self-selection problem as it only includes individuals in paid employment, and 2) it suffers from a potential sample selection problem as it does not include employees in small firms, in agriculture and the self-employed. The aim of the research project is twofold: 1) to assess the bias resulting from the two above-mentioned selection mechanisms, 2) to propose a methodology to correct for such selection that can be applied to the SES data. Applying econometric methods on another data source, the European Community Household Panel (ECHP), we quantify the bias from both selection mechanisms. We conclude that self-selection into paid employment is an issue in a number of EU Member States, and that this process affects the measurement of the Gender pay gap. Omission of employees in small firms, however, has only a small effect on the measurement of the Gender pay gap. To identify the self-selection into paid employment, literature suggests that variables such as marital status and the number and age of children should be used. While such variables are available in the ECHP, they are not available in the SES data. We show that using a simplified approach in which self-selection is captured only by variables such as gender, educational level and age (variables that are available in the SES) is not satisfactory, and that it can lead to misleading conclusions. We conclude that the Gender pay gap that accounts for observable differences in background characteristics of female and male workers is a better measurement than the simple comparison of difference in average wage of females and males. We also conclude that controlling for self-selection into employment is not feasible in the current SES because it lacks the variables that are key in explaining such selection processes.

1 Introduction

Although the difference in pay between males and females has decreased sharply over the past decades in a number of European Countries, the pay differential is still significant (Plantega and Remery, 2006). This Gender pay gap represents one of the areas targeted by the European Union policies in order to promote gender equality. The Gender pay gap is an important issue on the European agenda (see Section 2) as it is desired that the pay gap between men and women – once it is corrected for differences in socio-economic characteristics – be eliminated. To this end, the European Commission has engaged in a monitoring process of the magnitude of the Gender pay gap in EU Member States, and of its evolution over time.

An accurate monitoring to the Gender pay gap is of utmost importance. Indeed, policy makers use existing measures of the Gender pay gap to evaluate progress towards the objective of gender equality, and to promote new strategies and legislation in this area. An accurate monitoring to the Gender pay gap also requires the use of high quality data. It is the aim of this research to develop a methodology for measuring the Gender pay gap on the basis of Structure of Earnings Survey (SES) data, a new and rich database at Eurostat. The SES, however, has a number of drawbacks that must be accounted for when calculating the Gender pay gap. Firstly, the SES suffers from a potential self-selection problem as it only includes individuals in paid employment. Secondly, SES suffers from a potential sample selection problem as employees in small firms, in agriculture and the self-employed are not included in the data.

The aim of the research project is twofold: 1) to assess the bias resulting from the two above-mentioned selection mechanisms; 2) to propose a methodology to correct for such selection that can be applied to the SES data.

This report is structured as follows. In Section 2, we give a brief account of the policy context of gender pay inequality. In Section 3, we review the literature on gender pay inequality in order to give an overview of explanations that have been given for the Gender pay gap, and the methods that have been used in empirical research to assess the

Gender pay gap. Section 4 reviews the methodological issues in estimating the Gender pay gap. In Section 5, we estimate the Gender pay gap from an alternative data source (the European Community Household Panel; ECHP), and quantify the effect of self-selection and sample selection. Section 6 reports on our application of the correction mechanisms for self-selection that we developed for the ECHP on the SES data, and the effect it has on the measure of the Gender pay gap. In Section 7 we conclude with recommendations on how to measure the Gender pay gap in data such as the SES.

2 The European policy context

The principle of equal pay for men and women for work of equal value has been instated in the European Community primary legislation since its very beginning (Article 119 of the EEC Treaty/1957). This principle has been brought into practice by Directive 75/117EEC on the approximation of the laws of the Member States stipulating equal pay for equal work and work of equal value. Furthermore, the principle of Article 119 has become directly effective in national courts through an extensive case law of the European Court of Justice.

The European Employment Strategy (1997) drew attention to quantitative aspects (i.e. employment rate of females), as well as on qualitative aspects of labour market participation (i.e. the issue of equal opportunities). These inherent tensions between the quality and the quantity of employment were then addressed at the Nice European Council summit in 2000 which incorporated employment quality as a policy goal. Later, at the Laeken summit, concrete indicators to measure the new goal were set. "More and better jobs" became a strategic objective at the EU level.

Although over the last ten years progress has been made in female employment in terms of quantity, further efforts were needed in order to improve the quality aspect. In this sense, the European Employment Strategy guidelines have been revised in 2003 in order to include the targets and goals agreed at Lisbon. Specifically employment guideline six referring to gender equality specified as a target to reduce the Gender pay gap by 2010 with a view to its elimination. A multi-faceted approach addressing the underlying factors of the Gender pay gap – including sectoral and occupational segregation, education and training, job classifications and pay systems, awareness-raising and transparency – is to be used in view of reducing the Gender pay gap.

The European Commission set as one of the priorities the reduction in Gender pay gap (as set out in guideline 17) in the 'Roadmap for Gender Equality'. This document states the six priority areas for EU action on gender equality for the period 2006-2010 and represents the Commission's commitment to driving the gender equality agenda forward, reinforcing partnership with Member States, and other actors. The 2006 European Commission's Report on equality between women and men shows that women in the EU earn on average 15% less than men, and that progress has been slow in closing the Gender pay gap (Directorate-General for Employment, Social Affairs and Equal Opportunities, 2007). Eliminating the Gender pay gap becomes a priority, since its persistence results from direct discrimination against women and structural inequalities, such as segregation in sectors, occupations and work patterns, access to education and training, biased evaluation and pay systems, and stereotypes (Commission of the European Communities, 2006).

Despite engaged commitment to promote gender equality women are still less favoured on the labour market than men. The 2008 report on the Equality between women and men acknowledges that the Gender pay gap has remained steady at 15% since 2003, and has narrowed by only one point since 2000 (Commission of the European Communities, 2008). Sectoral and occupational segregation by gender is not diminishing, and is even increasing in certain countries. This is a sign that women who have recently joined the labour market have gone into sectors and occupations already dominated by women.

3 Explanations for the Gender pay gap: A review of the literature

There is an abundant economic literature on the Gender pay gap. The available research focuses on the level of the Gender pay gap, its measurement, its evolution over time, the causes of the Gender pay gap, and the effectiveness of policies to reduce it. In this short review, we go into the possible explanations for the Gender pay gap, and the methodological issues involved in measuring it.

In the literature, the Gender pay gap has been attributed to a number of factors including human capital endowments and career interruptions of women, discrimination by employers, job characteristics, occupational self-selection and labour market institutions. Here, we review each of these explanations.

A first type of explanations for gender pay differentials focuses on the fact that men and women have different individual characteristics, labour market histories and job characteristics. Men and women have different patterns of human capital accumulation due to different educational tracks, and therefore have different endowments in skills. Women have more career interruptions and intermittent employment histories due to coping with childbirth and care (Manzoni et al., 2008, Fouarge and Muffels, 2008). They tend to interrupt their career more frequently for childcare reasons leading to less experience and depreciation of skills (Paull, 2006; Fouarge and Muffels, 2009). Moreover, women tend to adjust working hours by taking part-time jobs that have lower returns and less training opportunities (Manning and Petrongolo, 2008). These differentials in labour market histories make women to accumulate lower work experience compared to men and therefore to have lower earnings (Buligescu et al., 2009).

Job characteristics are important too in the wage differential. For example, many women are employed in the public sector where the returns are higher for the low skilled but lower for the highly skilled compared to the private sector. Moreover, in the public sector the penalty for career interruption is lower but the potential wage growth is also lower.

Typically, the Gender pay gap refers to the differences between the wages earned by women and by men that are not explained by different patterns of human capital accumulation and depreciation or by differences in individual and job characteristics (occupation, sector, industry, firm, workplace and job characteristics). The Gender pay gap is therefore a residual that cannot be explained by differences in endowments between men and women. It is interpreted as the result of differences in the relative prices paid by employers. The difference in pay that is unexplained is generally assumed to be the outcome of gender discrimination by employers.

Gender segregation implies that women are concentrated in specific segments of the labour market in low graded jobs, in service work, in the public sector and in part-time jobs (Bardasi and Gornick, 2008). Pay policies and practices that impact differently by sector, by job level, contract or type of work will have gender effects. However, other researchers argue that the chief way in which women are being discriminated against in the labour market is not the pay rates but the opportunities of being promoted to better positions. Women are being relegated to lower positions on the job ladder than one could expect on the basis of their qualifications, and therefore they cannot reach the higher strata in their occupation. Oaxaca and Ransom (2003) find that the expected seniority is greater for women in order to move up the career ladder in the U.S. They also find a pattern of intrafirm mobility and initial job assignment that generally penalises women, even after taking account of individuals' characteristics. Apart from promotion discrimination, there are other studies arguing that women are discriminated in employment, as employers would rather hire men than women. This is because women are expected to interrupt their career as a result of child bearing.

Alternatively, other explanations for the gender wage differentials suggest that the distribution of men and women differs across work places. The latter can be described by the occupation, sector, industry and job status. Women tend to select themselves in certain occupations. This creates a crowding effect as women concentrate in those occupations thus decreasing the returns. Occupational selectivity bias affects wage differentials as occupations differ in average wage rates (even after controlling for workers' characteristics) and barriers to entrance of the subordinate group create another source of discrimination. The implications of this hypothesis would be the existence of a wage gap across occupations but the absence of gender differentials within occupation. Selection into occupations by women could be explained by the fact that those occupations have higher non-pecuniary rewards compared to those chosen by men[4] or by the fact that women choose occupations in order to mitigate the effects of a more intermittent labour attachment.

Polachek's (1981) segregation theory relies on the fact that women's employment is intermittent because of domestic responsibilities. The human capital of women out of the labour market is depreciating. Therefore, it is rational for women who are planning to spend a lot of time out of the labour market to choose jobs with low penalties for intermittent employment (Görlich and de Grip, 2009). Even in the absence of discrimination by employers, occupational segregation would still occur as a result of the minimization of costs incurred by women. "If life cycle behaviour labour force participation differs across individuals, and if the costs of varying degrees of labour force intermittency vary across occupations, then individuals will choose those occupations with the smallest penalty for the life cycle participation" (Polachek, 1981). Therefore, women who plan an intermittent employment will prefer occupations with lower penalties for depreciation, whereas men or women who plan a continuous employment will have no reason to avoid jobs with high depreciation risks. Therefore, the occupations preferred by women are those with low rewards for experience. High appreciation indexes an occupation's penalty for intermittence only when 1) the rate of appreciation is not high enough to make up for the rate of depreciation in the occupation, and 2) the starting wages in the occupation are low enough that they more than offset the effect of appreciation on the projected lifetime earnings of women planning intermittent employment (England, 1982). For this reason, occupational assignment may be the primary source of the gap and the rates of return to endowments, collective bargaining, personal and market characteristics differ by occupation. This approach emphasises that earnings should be estimated for each occupation.

However, not only market forces, discrimination by employers and the human capital explanations matter for wage differentials. Wage differentials are also due to different

[4] This idea implies that in fact the Gender pay gap would be smaller if fringe benefits would be taken into account.

wage setting mechanisms and different wage structures.[5] Hence, institutions play a role in the size of the wage gap. A more compressed wage structure, for example, results in a lower Gender pay gap. Similarly, greater collective bargaining leads to lower wage dispersion and is negatively related to the Gender pay gap. This is because centralisation and unionisation tend to increase the minimum wage floors and thereby the relative position of women in the bottom of the wage distribution. Therefore, the level of bargaining and centralisation in the wage setting of minimum wage floors is an important factor affecting the gender wage gap. Whatever the level of the minimum wage, a higher share of women are affected than men (Rubery et al. 2005). Card and Di Nardo (2002) emphasise on the fact that trends in minimum wage, and other factors such as a decline in unionisation and the reallocation of labour offer a better explanation to the rise of the wage inequality in the 1980s than technological change.

Other institutions like family friendly policies could have an effect on the gender wage gap too (Dupuy and Fernandez-Kranz, 2007). On one hand, family friendly policies can lead to a widening of the Gender pay gap by providing incentives for women to take a longer time out of the labour market for childcare reasons. On the other hand, such policies could also lead to a decrease in the Gender pay gap by preserving the employment relationship during maternal leave which avoids loss of firm specific capital and encouraging human capital investments that lead to higher female wages. Arulampalam et al. (2007) suggest that countries with more "generous" work-family policies, such as the Netherlands and Denmark, have a lower wage gap at the bottom of the wages distribution and a wider gap at the top.

It seems that where datasets have information on demand-side variables, these often prove to be more informative or have more explanatory power than differences in personal characteristics for the gender wage gap (Rubery et al. 2005).

[5] Traditional econometric analyses of the Gender pay gap start from a presumption that wage structures reflect market factors or productivity differences, except for gender discrimination. Even the incorporation of institutional factors into the analysis is done by removing the impact of institutional factors or the noise in the search for an underlying rational structure of wages, related to productivity. However, as Rubery et al. (2005) note wages are not only a reflection of productivity and discrimination they could also be used as a managerial tool to increase productivity and motivate workers.

In Table A1 in Appendix 1, we provide an overview of recent studies on the level of the Gender pay gap. In general, women are found to earn less than men do, although the size of the wage differential differs depending on the method used.

4 Issues in estimating the Gender pay gap

Due to the fact that all decompositions rely on consistent estimates from wage regression by gender and that biased estimates affect the size of the wage gap found, it is important to estimate an adjusted Gender pay gap which takes into account problems like heterogeneity, endogeneity, omitted variable bias and non-random sample selection.

Biased estimates of the wage rate might occur due to heterogeneity if the wage rate is related to unobserved individual characteristics such as motivation, ability or risk aversion (Dohmen et al. 2005) which are in turn correlated with regressors. Heterogeneity is usually handled by using panel data methods dealing with the individual effect.

Endogeneity is a problem if explanatory variables like experience, time out of the labour market, education, working in the public sector are not given exogenously but subject to an individual's decision and are at risk of being correlated with unobserved factors affecting the wage rate (Beblo et al. 2003a; Kunze 2007). The impact on the wage rate of these factors is likely to be estimated with bias and the estimates of the price vector, say β, are not going to be consistent. Indeed, endogeneity is removed by using instrumental variables like the number of children, age, region, the education of parents and occupation (Beblo et al. 2003b) and (structural) control functions (Heckman, 1979).[6]

If omitted productivity variables are correlated with gender then it is possible that the gender variable will serve in part as a proxy for omitted variables. Omitted variables and measurement errors are a source of bias in the decomposition approaches. Usually this

[6] The issue of the gender wage gap has also been dealt with in the literature using a decomposition approach (e.g. Blinder-Oaxaca decomposition, the Juhn-Murphy-Pierce decomposition, and the Brown-Moon-Zoloth decomposition). However, such approaches are less suitable in the case of the SES data for they do not allow to take account of self-selection and sample selection issues. Dealing with these two selection mechanisms is the core of this research project.

issue is solved by including job characteristics and fringe benefits as explanatory variables (Solberg and Laughlin, 1995).

Sample selection biases the estimates if the individuals observed in a sample (working for instance) differ systematically in attributes and characteristics from those not observed in the sample (those not working), and therefore inference about the population is not possible based on a non-random observed subgroup (self-selection). Selection into employment can be higher in the Southern countries where women labour force participation remains low thus affecting the estimation of the wage rate and of the Gender pay gap. A second type of non-random sample selection results from data truncation, i.e. when a non-random part of the employed population (e.g. people employed in small firms) is not observed in the data. In this report, we focus exclusively on these two types sample selection. Heterogeneity, endogeneity and omitted variable bias fall beyond the scope of this project and cannot adequately be dealt with using the SES data. The next section is looking more into detail in the type of biases one might expect as a result of ignoring selectivity and which approach we suggest to use in order to cope with this problem.

Although it is difficult to correct for data truncation sensitivity on alternative data sources can help us understand the impact of it (i.e. to answer questions of the type: to what extent does the Gender pay gap change when small firms are excluded from the sample?).

For the selection into employment the models used in the literature usually make use of variables such as gender, age, educational level, marital status, number of children and their age, health status, partner's labour supply, other household member's income and regional employment characteristics (Beblo et al., 2003a; Albrecht et al., 2004, Buligescu et al., 2008). Especially variables pertaining to the presence and the age of children turn out to be strong predictors of the participation choice of females. This is also true for the presence of a (working) partner and the other household member's income. Our estimations of the selection process into employment on the ECHP will include such variables. However, as we discuss later (Section 5), not all such variables are available in the SES. This implies that our correction for self-selection in the SES will make use of a restricted set of covariates.

Traditionally, scientific research aiming at assessing the level of the Gender pay gap focuses on prime ages workers. Arulampalam et al. (2007) for example focus on the age group 23-54, while Beblo et al. (2003b) focus on the age group 25-55 (see also Albrecht et al., 2004) and Bardasi and Gornick (2008) restrict their analyses to individuals in the age group 25-59. In this report, we will follow this procedure and we select individuals in the age group 22-55. We do this in order to exclude young labour market entrants as well as older workers whose participation decision and wage level could colour our measurement of the Gender pay gap. This is especially important in international comparative research owing to the large country differences in youth unemployment rate and participation rates among older workers.

5 Estimating the Gender pay gap with alternative data

5.1 Approach for the quantification of the Gender pay gap

Monitoring of the Gender pay gap requires the use of high quality data. The Gender pay gap indicator that is currently being published by Eurostat 1) is based on a mix of several national data sources and EU-SILC data, 2) is unadjusted for composition effects, and 3) is unadjusted for selection effects. Point 1) implies that there is a lack of comparability between sources that makes it difficult to compare Gender pay gap indicators across countries. To tackle this problem, a harmonised dataset should be used to measure the Gender pay gap in all countries. The proposed dataset is the SES, which offers a unique opportunity to derive comparable Gender pay gap measures across countries. This data is attractive as it contains detailed information about variables known to explain wages that are therefore potential candidates to explain the Gender pay gap, i.e. bargaining regime, regular versus overtime pay or bonuses, education, sector of industry (private sector only), occupation etc. Moreover, the dataset contains information about both workers and their employers. This enables one to take into account firm effects and work place characteristics. Henceforth, the SES makes it possible to account for observable differences in composition of the male and female workforce (point 2 above).[7]

Concerning point 3), the SES has two drawbacks that must be taken into account when estimating the level of the Gender pay gap. First, the SES data suffer from a potential self-selection problem as it only includes individuals in paid employment. Second, the data also suffers from a sample selection problem because employees in small firms and the self-employed are not included in the data. Because the income generating process of the self-employed is much different from that of employees, and because of lack of international comparability of the status of self-employed itself and their income, we do not discuss the issue of the non-inclusion of this group in the SES data any further.

The main purpose of this section is to assess the bias resulting from the two above-mentioned selection mechanisms. We do this on data from the ECHP, bearing in mind the possibility to apply the suggested correction procedure on the SES data.

Self-selection into paid employment

As discussed in Sections 3 and 4, the literature has pointed to the existence of selection bias when estimating the wage of female workers. The most appropriate approach is to first deal with the self-selection issue, and then with the sample selection problem that is due to the non-inclusion of small firms and workers in agriculture in the SES. The intuition behind the correction for self-selection is explained in the textbox below.

[7] In the meantime, Eurostat also releases the unadjusted Gender pay gap based on SES data, see http://epp.eurostat.ec.europa.eu/portal/page/portal/labour_market/earnings

Correcting for self-selection: intuition

By definition, the wage from work is only observed for people engaged in paid employment. This means that only people who choose to enter the labour market have a positive wage. If males and females differ in their preference for paid employment, the process by which they select themselves into paid employment will be different, and this will affect the measurement of the Gender pay gap. Consider the following example: in a particular year all males participate in paid employment, but among females, one group does participate while the other group does not. Suppose that the decision whether or not to work is made on the basis of both observed (e.g. education, age) and unobserved (e.g. motivation, preference for paid labour) characteristics of individuals. If the group of working women is a random draw from all women, then there is no selection issue, and we can compare the wages of males and females. Likewise, if the group of working women is selected on the basis of a number of observed characteristics, then we can compare the wage of males and females, provided we control for the appropriate observed characteristics when estimating the wage differential. However, if the group of working women is selected on the basis of some unobservable characteristics that also affect the wage they get, then there is a problem of self-selection. This is because the decision whether or not to work is based on unobserved characteristics, such as preferences for paid employment, that correlate with observed characteristics in the wage equation, such as the respondent being a female. In this case the Gender pay gap will be biased, unless we include information in the wage regression that captures the effect of selection on the unobservables (see also Vella, 1998: 128-129). Using information in the data, we control for this potential difference in unobserved 'preference' for labour market participation. To achieve this, it is crucial to be able to identify the process of selection into paid employment. In the case of the males-females comparison, variables such as the educational level, the marital status, the presence and the age of children are crucial.

Self-selection into paid employment could affect the Gender pay gap in a positive or negative way: its sign cannot be determined beforehand as it differs across countries. We propose to model self-selection into paid employment on the basis of observable endowments (gender, human capital, age, marital status, presence and number of children, age of the youngest child, …) based on data other than the SES.[8] Estimation on other data is necessary because the SES does not include people who are not engaged in paid employment. In principle, a model for self-selection could be estimated from the Labour Force Survey (LFS). This has the advantage that the selection process can be modelled for all countries included in the SES for which LFS data are also available. However, here we used the European Community Household Panel (ECHP) data for eleven countries, Denmark, The Netherlands, Ireland, Italy, Greece, Spain, Portugal, Austria, Finland, Germany and the United Kingdom. The advantage of the ECHP over LFS is that it allows us to estimate the selection bias, while at the same time assessing its effect on the Gender pay gap. Moreover, the ECHP allows us to measure the impact on the Gender pay gap of leaving out small firms and workers in agriculture from the observation set. This would not have been possible on the LFS because it does not contain wage information.

The following model for the decision whether or not to work (w_i, with $w_i = 1$ when in paid employment, and 0 otherwise) can be estimated for each country separately:

$$\Pr(w_i) = \beta Z_i + \varepsilon_i \qquad [1]$$

where Z represents a vector of observable endowments of individual i (such as gender, age, marital status, presence and number of children, …), β is a parameter to be estimated and ε and error term. The estimation of the model allows us to retrieve the returns to

[8] There is an additional problem that must be mentioned. It is likely to be the case that people self-select into different types and levels of education, with those who expect higher returns investing more in schooling. This issue can technically be dealt with but it requires rich data that can capture this self-selection process through correction for individual differences in expectations, ambition and effort. In the light of the available information concerning SES (data documentation for the Structure of Earning Survey 2006) we do not think it is possible to adequately deal with this issue with the data at hand.

endowments (β) on the basis of which we can calculate the nonselection hazard, i.e. the so-called inverse Mills ratio (Heckman, 1979):

$$\lambda_i = \phi(\hat{\beta}Z_i) \big/ \Phi(\hat{\beta}Z_i) \qquad [2]$$

where Φ is the cumulative distribution function of the standard normal distribution and ϕ is the corresponding density.

The vector of endowments (Z) must include variables that have been shown to determine the participation choice in the literature (see Section 4). The restriction, however, is that the variables used must also be included in the SES so that [2] can be calculated in the SES data.

Once [1] has been estimated, the inverse Mills ratio [2] is included in wage regressions to control for self-selection into employment as follows:

$$y_i = \delta X_i + \kappa\lambda_i + \pi female_i + \tau_i \qquad [3]$$

where y_i represents the log hourly wage of individual i, X is a vector of background characteristics that affect wages. It includes age, age squared, dummies for educational level, dummy for private sector firms, dummy for full-time workers, dummy for fixed terms contracts, dummies for firm size[9], as well as 8 occupational dummies. In the model, λ is the self-selection term, *female* is a dummy variable for females, and δ, κ, π and τ parameters to be estimated. The idea is that, for the purpose of identification, Z includes variables that are omitted in X. The parameter of interest, π, reflects the wage differential between males and females that is not accounted for by observed characteristics or self-selection in paid employment. A test for no self-selection is therefore readily available and consists of testing for κ=0.

[9] In the ECHP, reliable information on firm size is only available for private sector employees. Therefore, the effect of firm size is only modelled for them.

5.2 Data

Both the ECHP and the SES have data limitations. On the one hand, a number of variables are missing in the SES such as information about actual work experience and career breaks (unemployment, parental leave, other career interruptions) which are very important for isolating the effect of gender. On the other hand, certain variables like firm size, wages and industry, are truncated in both data sets (see Table 1).

Table 1: Summary of data limitations in ECHP and SES

Variables	Data limitations in ECHP	Data limitations in SES
Firm size	Firm size for public firms missing unless there is a change in the number of employees	Firms with less than 10 employees not present for all the countries
Wages	Wages of people working less than 15 hours in a week not observed	Unemployed and inactive not observed
Industry	No limitations	- No information for people in agriculture and fishery - No information for public administration and defence servants; compulsory social security servants

In our analysis, we face a number of selection criteria. These are either imposed by us in order to obtain more reliable results, or they are imposed for comparability between sources. Starting with the first type of constraints imposed on the data: following the standard approach in the literature (see Section 4), we focus all our analysis in both datasets on persons in the age group 22-55. Furthermore, we exclude outliers in the SES data because the average wage as a statistical measure would be highly influenced by the very rich in the 99th wage percentile or the very poor in the 1st wage percentile.[10] As to constraints that were imposed for a better comparability of data sources: we treat people working fewer than 15 hours as being out of employment as their wages are not reported in the ECHP. Moreover, in order to understand the effect of missing observations for workers in agriculture, public administration or small firms in the SES, we simulate these restrictions on ECHP. We summarise the different selection filters imposed on the data in

[10] Descriptive statistics on the SES data revealed very large and low hourly wages. We therefore decided to exclude these values. For the analyses on the ECHP, this selection does not need to be performed.

Table 2. In the analyses we will indicate what filters were applied for the computation of the Gender pay gap.

Table 2: Selection criteria applied to the SES / ECHP data

Filter 1.	Age 22-55
Filter 2.	Exclusion of outliers: top 1% percentile and bottom 1% percentile of the hourly-wage distribution
Filter 3.	People working more than 15 hours
Filter 4.	Excluding workers in small firms with less than 10 employees
Filter 5.	Excluding workers in agriculture
Filter 6.	Excluding workers in public administration and defence

5.3 Results

According to the literature (Section 4), the presence and the age of children, the presence of a (working) partner and other household member's income are strong predictors of the labour market participation choice of females. Therefore, such variables should be included in Z for the estimation of the self-selection process. However, such variables are not available in the SES. For the purpose of this research project, we have estimated the selection model [1] using two sets of explanatory variables: a restricted set of variables that are available in the SES, and one richer set of variables that the literature suggests are pertinent for explaining self-selection.

This approach allows us to assess the added value of having a richer dataset than the SES to control for self-selection. Stated otherwise, if the literature suggests that variable z should be included in Z (e.g. marital status, the number of children), and z is available in the alternative data sources used but not in the SES, then we can measure the effect of estimating [1] on a restricted set of variables.

For each of the eleven countries listed above, we have estimated equation [1] twice, using two different specifications. The first specification (<u>specification 1</u>: S1) corresponds to the set of variables Z that are both available in the ECHP and the SES data. This is the restricted model that controls for:

- age,
- age square,
- education,
- gender,
- the interactions of gender with age,
- the interactions of gender with education.

The second specification (specification 2: S2) in addition to the previous set of variables also includes:

- the marital status,
- the number of children,
- the total income of other household members,
- a dummy for children under the age of 12,
- a dummy for children under the age of 1,
- the interaction of the gender dummy with all of the above variables.

These are typically the type of variables that have been included in empirical studies where selection into paid employment is accounted for (see Table A1 for an overview of the literature).

Self-selection and Gender pay gap

Table 3 reports the estimated coefficients for self-selection into paid employment in the wage equation in ECHP (this is coefficient κ in equation [3]). For comparability with the SES, the model was estimated on three different sets of observations. Columns 2 and 3 include all individuals in the age group 22-55 working at least 15 hours per week. Columns 4 and 5 exclude individuals working in small firms, and columns 6 and 7 also exclude employees in agriculture.

Most of the selection parameters appear to be significant according to specification 1. However, rather than being indicative of strong self-selection in the data, this could be due to the fact that the selection model does not sufficiently control for the processes that lead to self-selection. Because this specification does not include variables that are excluded from the wage regression, identification relies purely only the nonlinearity of the inverse Mills ratio function estimated from the probit model.

In fact, it turns out that this nonlinearity is not sufficient to pick-up the selection problem. This can be seen from the density plot of the inverted Mills ratio for Denmark (that we chose for the purpose of illustration) in Figure 1. Identification would require a kink in the plot of the inverse Mills ratio. Such a kink does not show in the left part of Figure 1, and it does not show for the other countries either. In fact, additional analyses have revealed that the linear prediction from the probit model [1], the argument of the inverse Mills ratio, alone already explains more than 90% of the inverse Mills ratio. Moreover, the inverse Mills ratio is highly collinear to the covariates included in the wage regression [3]. This multicollinearity problem means that when estimating [3] the effects of observed covariates could either load on the parameters δ or on the parameter κ: the two measure the same thing. We have no reason to assume that this result would be different when using another dataset such as the Labour Force Survey.

The implication of this lack of nonlinearity in the inverse Mills ratio is that a more extensive set of covariates is to be preferred to the restricted set. Specification 2 better picks-up self-selection (right pane of Figure 1), and there is not collinearity with the individual characteristics entered in model [3]. The extensive set of covariates in specification 2 shows significant self-selection in 4 of the 11 countries (Table 3). The fact that self-selection is not found for all countries is not surprising as other studies using the ECHP have reached similar conclusions (see e.g. Nicodemo, 2008). Note, however, that using a different selection on age could lead to significantly different results. This is because of the international differences in participation rate of the youth and older workers in Europe.

Excluding employees working in small firms or in agriculture from the ECHP (in order to make the data comparable to the SES) reveals that self-selection only affects 3 countries.

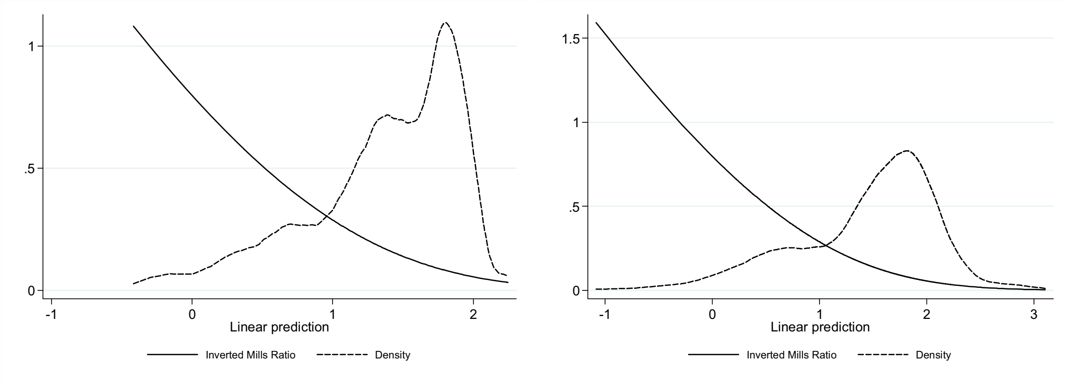
This can be explained by the fact that women are more likely to be employed in small firms.

Table 3: Inverse Mills Ratios (coefficients for κ in equation [3])

	No truncation (uses filters 1 and 3 from Table 2)		**Truncation small firms** (uses filters 1, 3 and 4 from Table 2)		**Truncation small firms and agriculture** (uses filters 1, 3, 4 and 5 from Table 2)	
	OLS with sample selection (S1)	OLS with sample selection (S2)	OLS with sample selection (S1)	OLS with sample selection (S2)	OLS with sample selection (S1)	OLS with sample selection (S2)
1	2	3	4	5	6	7
Denmark	-0.376***	-0.193***	-0.338**	-0.186***	-0.373***	-0.189***
	(0.138)	*(0.066)*	*(0.139)*	*(0.064)*	*(0.142)*	*(0.065)*
The Netherlands	-0.255***	-0.040	-0.254***	-0.010	-0.245***	-0.008
	(0.070)	*(0.042)*	*(0.067)*	*(0.041)*	*(0.066)*	*(0.041)*
Ireland	-0.239*	-0.038	-0.231*	-0.021	-0.219*	-0.018
	(0.125)	*(0.047)*	*(0.120)*	*(0.045)*	*(0.119)*	*(0.045)*
Italy	-0.150***	-0.016	-0.132***	-0.020	-0.123***	-0.019
	(0.036)	*(0.016)*	*(0.037)*	*(0.017)*	*(0.037)*	*(0.017)*
Greece	-0.136***	-0.098***	-0.115**	-0.097***	-0.109**	-0.094***
	(0.046)	*(0.025)*	*(0.049)*	*(0.027)*	*(0.049)*	*(0.028)*
Spain	-0.084**	-0.071***	-0.064	-0.034	-0.068*	-0.019
	(0.041)	*(0.025)*	*(0.039)*	*(0.027)*	*(0.040)*	*(0.027)*
Portugal	-0.447***	-0.033	-0.421***	-0.032	-0.422***	-0.028
	(0.116)	*(0.037)*	*(0.102)*	*(0.041)*	*(0.104)*	*(0.041)*
Austria	0.247	-0.242***	0.233	-0.280***	0.233	-0.281***
	(0.157)	*(0.037)*	*(0.148)*	*(0.038)*	*(0.149)*	*(0.038)*
Finland	0.363***	-0.060	0.338***	-0.073	0.338***	-0.080
	(0.100)	*(0.052)*	*(0.095)*	*(0.050)*	*(0.097)*	*(0.050)*
Germany	-0.204	-0.009	-0.185	0.028	-0.155	0.023
	(0.140)	*(0.037)*	*(0.135)*	*(0.038)*	*(0.138)*	*(0.038)*
United Kingdom	0.175	-0.027	0.182	0.015	0.182	0.015
	(0.230)	*(0.050)*	*(0.203)*	*(0.049)*	*(0.201)*	*(0.049)*

Significance levels * $p<0.10$ ** $p<0.05$ *** $p<0.01$. Standard errors are provided in parentheses.
Source: ECHP, authors' computation.

Figure 1: Density distribution and inverse Mills ratio using specification 1 (left pane) and specification 2 (right pane), Denmark



Source: ECHP , authors' computation.

Table 4 reports estimates of the Gender pay gap for the eleven countries in the ECHP. The raw Gender pay gap is reported in columns 2 and 3. Comparing columns 2 and 3 provides an indication of the effect on the Gender pay gap of excluding workers in small firms and in agriculture from the sample. In two countries, this has no effect on the Gender pay gap: Italy and Greece. For six countries, the resulting Gender pay gap is smaller after exclusion of employees in small firms and in agriculture: the Netherlands, Spain, Portugal, Austria, Germany and the UK. In the other three countries, excluding employees from small firms and in agriculture results in a larger raw Gender pay gap. The unadjusted Gender pay gap in the full sample of ECHP respondents ranges from 5.5% in Italy to 26.9% in Germany.

Apart from the raw Gender pay gap, the table also reports corrected Gender pay gaps (π, in terms of equation [3]) for the whole sample, for employees not employed in small firms, and for employees neither employed in small firms or in agriculture. For each of these three samples, three corrected Gender pay gaps are reported: 1) the OLS Gender pay gap as estimated from equation [3], but without correction for self-selection, 2) the OLS Gender pay gap as estimated from equation [3] with the restricted set of covariates to control for self-selection (specification 1), and 3) the OLS Gender pay gap as estimated from equation [3] with the full set of covariates to control for self-selection (specification 2). First, comparing column 2 (raw Gender pay gap) to column 4 (OLS Gender pay gap), we note that including controls to measure the Gender pay gap already tells a different story than the unadjusted Gender pay gap for most countries. In most countries (Finland, and the Latin rim countries Portugal, Spain and Italy being exceptions), correcting for observed characteristics already reduces the Gender pay gap. Second, columns 5 and 6 give very different results for most countries. This means that the inclusion of additional variables in the self-selection into employment equation plays an important role in the estimation of the Gender pay gap. This is problematic since the information contained in specification 2 is not available in the SES data. However, it should be noted that the results from column 6 are also more comparable to those in column 4. For five countries (Ireland, Italy, Portugal, Germany and the United Kingdom), the differences between both estimates is less than half a percentage point. For the other countries, however, the difference range from 1.4 to 3.0 percentage points. The conclusion so far is that specification 2 should be preferred to specification 1 and the simple OLS estimator.

In all countries, the Gender pay gap is significant and it ranges from 7.1% in Denmark to 25.5% in Germany, according to the model specification 2 for the sample selection equation (column 6 in Table 4).

Data truncation

The effect of truncation on small firms and truncation on specific sectors on the estimator of self-selection into employment seems to be very limited as shown in Table 3.[11] Column 2 compares to columns 4 and 6 and column 3 compares to columns 5 and 7. However, the incidence of the two types of truncation on the various measures of the Gender pay gap is rather limited if not absent (see columns 2 and 12 from Table 4). For Ireland, Austria and Finland, data truncation results in a somewhat larger Gender pay gap than would otherwise have been measured. For Spain data truncation results in a somewhat lower Gender pay gap.

[11] Data truncation was applied by excluding workers in small firms from the data. We also tested an alternative approach in which we redefined their employment status to not-employed, and their wage to be missing. In theory, it is hard to say which of the two approaches is the best, but both lead to similar results.

Table 4: Gender pay gap measured in ECHP, various specifications and data truncations

	Raw GPG		**No truncation** (filters 1 and 3 from Table 2)			**Truncation small firms** (filters 1, 3 and 4 from Table 2)			**Truncation small firms and agriculture** (filters 1, 3, 4 and 5 from Table 2)		
	Whole sample (filter 3 from Table 2)	22-55 y.o. (filter 1 and 3 from Table 2)	OLS	OLS with sample selection (S1)	OLS with sample selection (S2)	OLS	OLS with sample selection (S1)	OLS with sample selection (S2)	OLS	OLS with sample selection (S1)	OLS with sample selection (S2)
1	2	3	4	5	6	7	8	9	10	11	12
Denmark	-0.102***	-0.125***	-0.088***	-0.053**	-0.071***	-0.091***	-0.060***	-0.076***	-0.088***	-0.053**	-0.072***
			(0.019)	*(0.023)*	*(0.020)*	*(0.019)*	*(0.023)*	*(0.020)*	*(0.019)*	*(0.023)*	*(0.020)*
The Netherlands	-0.214***	-0.195***	-0.168***	-0.071**	-0.158***	-0.160***	-0.060*	-0.158***	-0.161***	-0.064**	-0.159***
			(0.017)	*(0.032)*	*(0.020)*	*(0.017)*	*(0.031)*	*(0.020)*	*(0.016)*	*(0.031)*	*(0.019)*
Ireland	-0.139***	-0.183***	-0.128***	-0.054	-0.127***	-0.140***	-0.067	-0.142***	-0.141***	-0.071	-0.143***
			(0.023)	*(0.045)*	*(0.024)*	*(0.022)*	*(0.044)*	*(0.023)*	*(0.022)*	*(0.044)*	*(0.023)*
Italy	-0.055***	-0.050***	-0.114***	-0.041**	-0.112***	-0.107***	-0.037*	-0.105***	-0.108***	-0.043*	-0.105***
			(0.010)	*(0.020)*	*(0.011)*	*(0.011)*	*(0.023)*	*(0.011)*	*(0.011)*	*(0.022)*	*(0.012)*
Greece	-0.174***	-0.176***	-0.146***	-0.070**	-0.129***	-0.151***	-0.080**	-0.133***	-0.152***	-0.085**	-0.135***
			(0.015)	*(0.030)*	*(0.016)*	*(0.017)*	*(0.034)*	*(0.018)*	*(0.017)*	*(0.034)*	*(0.018)*
Spain	-0.183***	-0.146***	-0.205***	-0.168***	-0.190***	-0.178***	-0.146***	-0.169***	-0.181***	-0.147***	-0.176***
			(0.014)	*(0.023)*	*(0.015)*	*(0.015)*	*(0.025)*	*(0.016)*	*(0.015)*	*(0.025)*	*(0.016)*
Portugal	-0.146***	-0.129***	-0.229***	-0.125***	-0.226***	-0.235***	-0.121***	-0.231***	-0.237***	-0.122***	-0.234***
			(0.014)	*(0.030)*	*(0.014)*	*(0.015)*	*(0.032)*	*(0.016)*	*(0.015)*	*(0.032)*	*(0.016)*
Austria	-0.218***	-0.209***	-0.199***	-0.281***	-0.168***	-0.215***	-0.301***	-0.175***	-0.216***	-0.302***	-0.177***
			(0.018)	*(0.055)*	*(0.019)*	*(0.019)*	*(0.058)*	*(0.020)*	*(0.019)*	*(0.058)*	*(0.020)*
Finland	-0.145***	-0.164***	-0.162***	-0.211***	-0.127***	-0.178***	-0.226***	-0.145***	-0.178***	-0.226***	-0.145***
			(0.014)	*(0.019)*	*(0.016)*	*(0.014)*	*(0.019)*	*(0.017)*	*(0.014)*	*(0.020)*	*(0.017)*
Germany	-0.269***	-0.256***	-0.256***	-0.195***	-0.255***	-0.253***	-0.194***	-0.257***	-0.251***	-0.202***	-0.255***
			(0.016)	*(0.045)*	*(0.017)*	*(0.016)*	*(0.046)*	*(0.017)*	*(0.016)*	*(0.047)*	*(0.017)*
United Kingdom	-0.248***	-0.228***	-0.187***	-0.230***	-0.184***	-0.176***	-0.235***	-0.179***	-0.176***	-0.234***	-0.179***
			(0.016)	*(0.058)*	*(0.017)*	*(0.018)*	*(0.067)*	*(0.020)*	*(0.018)*	*(0.067)*	*(0.020)*

Significance levels * p<0.10 ** p<0.05 *** p<0.01. Standard errors are provided in parentheses.
Source: ECHP, authors' computation.

6 Gender pay gap in SES

Having performed the analyses reported in Section 5, we now turn to estimating the Gender pay gap on the SES. First, we report on the raw Gender pay gap, i.e. the average difference in female to male gross hourly wage without correction for background characteristics and self-selection in employment. Then we report on our implementation of the correction for self-selection in the SES and its effect on the GPG.

6.1 Raw Gender pay gap

Table 5 reports the raw Gender pay gap as calculated from the SES for 11 European countries. Column 2 is the Gender pay gap as computed using the methodology of Eurostat (it includes employed people from all ages, but excludes workers in small firms, in agriculture, and in public administration and defence).[12] Column 3 is the Gender pay gap calculated including public administration and defence. Column 4 reports the same pay gap, but excluding people working in small jobs (with less that 15 hours) and excluding people younger than 22 or older than 55 years old. We do this for comparability with the ECHP. Column 5 also excludes outliers in the distribution of wages: it excludes those people whose wage level is in the bottom or top 1 percentiles of the wage distribution. We do this in order to obtain a measurement of the Gender pay gap that is less sensitive to extreme wage values.

[12] http://epp.eurostat.ec.europa.eu/portal/page/portal/labour_market/earnings

Table 5: Raw Gender pay gap in SES

	Raw Gender pay gap using Eurostat methodology (filters 4, 5 and 6 from Table 2)	**Raw Gender pay gap including workers in the public administration and defence** (filters 4 and 5 from Table 2)	**Raw Gender pay gap excluding small jobs of less than 15 hours per week** (filters 3, 4 and 5 from Table 2)	**Raw Gender pay gap excluding small jobs of less than 15 hours per week, age selection 22-55** (filters 1, 3, 4 and 5 from Table 2)	**Raw Gender pay gap excluding small jobs of less than 15 hours per week, age selection 22-55, and excluding outliers** (filters 1, 2, 3, 4 and 5 from Table 2)
1	2	3	4	5	6
Denmark	-0.213	-0.187***	-0.188***	-0.194***	-0.156***
		(0.005)	*(0.005)*	*(0.005)*	*(0.005)*
The Netherlands	-0.234	-0.205***	-0.187***	-0.180***	-0.158***
		(0.009)	*(0.009)*	*(0.008)*	*(0.007)*
Ireland	-0.178	-0.166***	-0.161***	-0.151***	-0.116***
		(0.008)	*(0.009)*	*(0.009)*	*(0.008)*
Italy	-0.043	-0.068***	-0.066***	-0.064***	-0.039***
		(0.008)	*(0.008)*	*(0.007)*	*(0.007)*
Greece	-0.217	-0.176***	-0.176***	-0.168***	-0.143***
		(0.008)	*(0.008)*	*(0.008)*	*(0.008)*
Spain	-0.184	-0.211***	-0.206***	-0.192***	-0.163***
		(0.007)	*(0.007)*	*(0.008)*	*(0.007)*
Portugal	-0.081	-0.018*	-0.017*	-0.020**	0.019**
		(0.010)	*(0.010)*	*(0.010)*	*(0.010)*
Austria	-0.269	-0.218***	-0.218***	-0.235***	-0.209***
		(0.007)	*(0.008)*	*(0.007)*	*(0.006)*
Finland	-0.216	-0.229***	-0.231***	-0.222***	-0.194***
		(0.005)	*(0.005)*	*(0.006)*	*(0.005)*
Germany	-0.270	-0.127***	-0.125***	-0.126***	-0.096***
		(0.008)	*(0.008)*	*(0.007)*	*(0.007)*
United Kingdom	-0.253	-0.225***	-0.212***	-0.223***	-0.184***
		(0.008)	*(0.008)*	*(0.009)*	*(0.008)*

Standard errors in parentheses. Significance level: * $p<0.10$ ** $p<0.05$ *** $p<0.01$.
Note: filters 4 and 5 cannot be relaxed as that the SES excludes workers in small firms with less than 10 employees and workers in agriculture and fishery.
Source: SES, authors' computation.

The data in column 2 of Table 5 is comparable to the official Gender pay gap as published by Eurostat. The only exception is Denmark for which the Gender pay gap we calculated is larger (-21.3%) than the official number (17.6%). In all countries but Spain and, to a lesser extent Italy and Finland, including workers from the public administration and defence sector in the computation results in a lower Gender pay gap (compare column 3 to column 2). Excluding employees in small jobs of less than 15 hours from the computation tends to result in a lower Gender pay gap (compare column 4 to columns 3). Imposing the age restriction of 22-55 years old on the SES has mitigated effect on the Gender pay gap: it increases it in fives countries, reduces in five countries an leaves it more or less unchanged in one country (compare column 5 and column 4). Finally, excluding extremely low and extremely high wage from the computation reduces the level of the Gender pay gap in all countries (column 6). Compared to the other constraints imposed on the data, this one has the largest effect on the level of the Gender pay gap.

6.2 Corrected Gender pay gap in SES

The starting point in estimating equation [1] is that the set of controls (*Z*) should include variables that are also measured in the SES. The selected set of variables from specification 1 (S1) fulfils this requirement. Using this model specification, we have applied the estimates obtained from the ECHP to the SES data. The resulting Gender pay gaps are reported in Table 6, and the coefficients from the self-selection term are reported in Table 7.

On the basis of the SES, we performed OLS regression analyses in order to obtain a measurement of the Gender pay gap, after correction for observed differences in individual background characteristics. Apart from a gender dummy, the regression model used includes control variables that are readily available in the SES: age, age squared, dummies for educational level, dummy for private sector firms, dummy for full-time workers, dummy for fixed terms contracts, dummies for firm size, as well as 8 occupational dummies (columns 2 and 3 in Table 6). Because the ECHP data do not allow us to compute the hourly wage rate for people employed for less than 15 hours, we replicated the SES regressions excluding such workers in "small" jobs (columns 4 and 5 in Table 6).[13] In order to assess the differences in the measurement of the Gender pay gap

[13] For the sake of comparability with the results from ECHP, firm size is only modelled for private sector employees.

between the two datasets, one can compare the last column in Table 6 to the 11th column of Table 4 (OLS results from the truncation occupations in agriculture column).

Table 6: Gender pay gap in SES, adjusted for observed characteristics and corrected for selection bias

	Gender pay gap (filters 1, 2, 4 and 5 from Table 2)		**Gender pay gap, excluding small jobs** (filters 1, 2, 3, 4 and 5 from Table 2)	
	OLS	OLS with sample selection (S1)	OLS	OLS with sample selection (S1)
1	2	3	4	5
Denmark	-0.120***	-0.123***	-0.119***	-0.120***
	(0.004)	*(0.005)*	*(0.004)*	*(0.005)*
The Netherlands	-0.121***	-0.051***	-0.120***	-0.054***
	(0.007)	*(0.012)*	*(0.007)*	*(0.012)*
Ireland	-0.157***	-0.077***	-0.153***	-0.078***
	(0.007)	*(0.012)*	*(0.007)*	*(0.012)*
Italy	-0.134***	-0.090***	-0.126***	-0.081***
	(0.005)	*(0.010)*	*(0.005)*	*(0.010)*
Greece	-0.110***	-0.057***	-0.110***	-0.056***
	(0.006)	*(0.012)*	*(0.006)*	*(0.012)*
Spain	-0.194***	-0.162***	-0.194***	-0.164***
	(0.006)	*(0.011)*	*(0.006)*	*(0.011)*
Portugal	-0.202***	-0.157***	-0.208***	-0.161***
	(0.006)	*(0.013)*	*(0.006)*	*(0.013)*
Austria	-0.169***	-0.172***	-0.183***	-0.186***
	(0.006)	*(0.016)*	*(0.006)*	*(0.017)*
Finland	-0.158***	-0.194***	-0.160***	-0.190***
	(0.004)	*(0.006)*	*(0.004)*	*(0.005)*
Germany	-0.151***	-0.162***	-0.154***	-0.168***
	(0.006)	*(0.014)*	*(0.006)*	*(0.014)*
United Kingdom	-0.135***	-0.161***	-0.133***	-0.174***
	(0.006)	*(0.033)*	*(0.007)*	*(0.034)*

Standard errors in parentheses. Significance level: * p<0.10 ** p<0.05 *** p<0.01.
Source: SES, authors' computation.

What do we learn from Table 6? First, after correcting for observable differences across individuals, a significant Gender pay gap remains (column 2). Second, omitting "small" jobs of less than 15 hours has only a small effect on the Gender pay gap (compare columns 2 and 4). The largest, but still small, effects are found for Austria and Italy. Third, it turns out that the Gender pay gap corrected for observed individual characteristics (column 4 in Table 6) is smaller than the raw Gender pay gap (reported in column 6 of Table 5) for six countries. However, for the Latin rim countries (Italy, Spain, and Portugal) – which is in line with the findings from ECHP –, and for Ireland and Germany, the OLS corrected Gender pay gap is larger than the raw Gender pay gap.

What do we learn from the comparison with the adjusted Gender pay gap in ECHP?[14] Although the OLS Gender pay gap is significant on the basis of both datasets, the corrected Gender pay gap calculated on the SES appears to be lower in eight of the countries. In the other three countries – Ireland, Italy and Spain – the Gender pay gap calculated on the SES is larger.

Although the estimations on the ECHP suggest that implementing selection model S1 is not likely to be a fruitful option, we do report the outcomes from the regression model in columns 3 and 5 of Table 6. The parameters for the self-selection term are reported in Table 7. In most of the countries, the non-selection terms are significant. Exceptions are Germany, United Kingdom, Denmark and Austria. For Denmark and Austria, the estimated Gender pay gap in the model with sample selection is similar to that in the model without sample selection, but in Germany and the United Kingdom, the Gender pay gap in the model that controls for self-selection is larger. This is also the case in Finland. In all other countries, the Gender pay gap in the model with controls for self-selection is found to be smaller. However, as argued above, we cannot be sure that these are correct point estimates since the model for selection only includes a limited set of variables and does not seem to do a good job at identifying selection into paid employment.

[14] Compare column 4 from Table 6 to column 10 from Table 4.

Table 7: Inverse Mills Ratios (coefficients for κ in equation [3]) from SES

	Inverse Mills ratio (filters 1, 2, 4 and 5 from Table 2) OLS with sample selection (S1)	**Inverse Mills ratio, excluding small jobs** (filters 1, 2, 3, 4 and 5 from Table 2) OLS with sample selection (S1)
1	2	3
Denmark	0.024	0.010
	(0.031)	*(0.031)*
Netherlands	-0.165***	-0.157***
	(0.024)	*(0.024)*
Ireland	-0.296***	-0.280***
	(0.037)	*(0.037)*
Italy	-0.088***	-0.090***
	(0.017)	*(0.017)*
Greece	-0.094***	-0.095***
	(0.017)	*(0.017)*
Spain	-0.061***	-0.056***
	(0.017)	*(0.017)*
Portugal	-0.192***	-0.201***
	(0.048)	*(0.047)*
Austria	0.011	0.011
	(0.046)	*(0.046)*
Finland	0.266***	0.235***
	(0.028)	*(0.028)*
Germany	0.021	0.042
	(0.046)	*(0.046)*
United Kingdom	0.110	0.174
	(0.138)	*(0.142)*

Standard errors in parentheses. Significance level: * $p<0.10$ ** $p<0.05$ *** $p<0.01$.
Note: the model used to calculate IMR is Selection 1 model with age and education interacted with gender, and age square. The IMR is calculated based on the estimates from ECHP.
Source: SES authors' computation.

6.3 Effect of data truncation in SES

As discussed in Section 5.2, one drawback of the SES is that it excludes workers employed in small firms and workers employed in the sectors public administration and defence. For three countries included in the SES (Ireland, Spain and Germany), we were able to assess the effect that excluding workers in small firms has on the Gender pay gap. This was possible because the raw SES data delivered to Eurostat by these three countries did include wage information for employees in small firms. Henceforth, we could replicate the analyses in Section 6.1 and 6.2, while including workers in small firms. In terms of the selection criteria summarised in Table 2, the data allows us to drop 'filter 4'.

The results for the raw Gender pay gap are reported in Table 8, the Gender pay gaps from the econometric analyses are reported in Table 9, and Table 10 reports the coefficients from the self-selection term.

The raw Gender pay gaps as well as the regression based Gender pay gaps for the three countries are similar to the Gender pay gaps reported in Table 5 and 6. This suggests that excluding employees in small firms from the observation set has no major impact on the measurement of the Gender pay gap in these three countries. Although we cannot say in what way excluding small firms affects the measurement of the Gender pay gap in other countries, results from the ECHP suggest that the effect is small or zero (see Section 5.3).

Table 8: Raw Gender pay gap in SES including workers in small firms (<10 employees)

	Raw Gender pay gap using Eurostat methodology (filters 5 and 6 from Table 2)	**Raw Gender pay gap including workers in public administration and defence** (filter 5 from Table 2)	**Raw Gender pay gap excluding small jobs of less than 15 hours per week** (filters 3 and 5 from Table 2)	**Raw Gender pay gap excluding small jobs of less than 15 hours per week, age selection 22-55** (filters 1, 3 and 5 from Table 2)	**Raw Gender pay gap excluding small jobs of less than 15 hours per week, age selection 22-55, and excluding outliers** (filters 1, 2, 3 and 5 from Table 2)
1	2	3	4	5	6
Ireland	-0.177	-0.170***	-0.166***	-0.159***	-0.127***
		(0.008)	*(0.008)*	*(0.008)*	*(0.008)*
Spain	-0.179	-0.207***	-0.202***	-0.189***	-0.161***
		(0.007)	*(0.007)*	*(0.007)*	*(0.007)*
Germany	-0.269	-0.126***	-0.125***	-0.126***	-0.096***
		(0.008)	*(0.008)*	*(0.007)*	*(0.007)*

Standard errors in parentheses. Significance level: * p<0.10 ** p<0.05 *** p<0.01.
Source: SES, authors' computation.

Table 9: Gender pay gap in SES, adjusted for observed characteristics and corrected for selection bias, including workers in small firms

	Gender pay gap (filters 1, 2 and 5 from Table 2)		**Gender pay gap, excluding small jobs** (filters 1, 2, 3 and 5 from Table 2)	
	OLS	OLS with sample selection (S1)	OLS	OLS with sample selection (S1)
1	2	3	4	5
Ireland	-0.162***	-0.078***	-0.156***	-0.082***
	(0.007)	*(0.012)*	*(0.006)*	*(0.012)*
Spain	-0.192***	-0.161***	-0.191***	-0.162***
	(0.006)	*(0.010)*	*(0.006)*	*(0.010)*
Germany	-0.151***	-0.163***	-0.154***	-0.168***
	(0.006)	*(0.014)*	*(0.006)*	*(0.014)*

Standard errors in parentheses. Significance level: * p<0.10 ** p<0.05 *** p<0.01.
Note: the calculations have been done for countries that include information about workers in small firms with less than 10 employees. Workers in agriculture and fishery are still excluded.
Source: SES, authors' computation.

Table 10: Inverse Mills Ratios (coefficients for κ in equation [3]) from SES, including small firms workers

	Inverse Mills ratio (filters 1, 2 and 5 from Table 2) OLS with sample selection (S1)	**Inverse Mills ratio, excluding small jobs** (filters 1, 2, 3 and 5 from Table 2) OLS with sample selection (S1)
1	2	3
Ireland	-0.309***	-0.274***
	(0.036)	*(0.035)*
Spain	-0.058***	-0.054***
	(0.016)	*(0.016)*
Germany	0.023	0.044
	(0.046)	*(0.046)*

Standard errors in parentheses. Significance level: * p<0.10 ** p<0.05 *** p<0.01.
Note: the calculations have been done for countries that include information about workers in small firms with less than 10 employees. Workers in agriculture and fishery are still excluded.
Source: SES, authors' computation.

7 Recommendations

Based on the results presented above, in this section we make recommendations to use the SES data to generate measurements of the Gender pay gap. Our general impression of the SES data is reported in Appendix 3.

Our analyses suggest that data truncation in the SES, such as the non-coverage of employees in small firms, does not affect the measurement of the Gender pay gap in a significant way.

Analyses using ECHP data have revealed that modelling selection on the restricted set of variables is tenuous. The identification of the selection process relies essentially on the nonlinearity of the inverse Mills ratio but as it turns out, this nonlinearity is limited in most of the countries. This generates multicollinearity problems that bias our estimates of the Gender pay gap. Modelling the selection mechanism using the richer set of variables available in ECHP seems to be more reliable. The multicollinearity problem is solved, and identification now hinges on exclusion restrictions (information included in these additional variables that are used to model the participation choice but that are excluded from the wage equation). However, as it turns out, the selection process modelled this way does not seem to play a significant role in the age group 22-55. In only 4 countries out of 11 is the coefficient for self-selection found to be statistically significant. This could mean either that the selection in most countries is not very stringent or that our way of modelling selection is not capable of picking the true selection. In fact, the Gender pay gaps measured from the selection model with extended set of variables are very similar to that obtained from simple OLS.

Based on our findings, we have rank ordered four alternative approaches (from the most preferred to the least preferred) to use the SES to generate measures of the Gender pay gap.

Approach 1: *Extent the variables included in SES*

The first approach we suggest is to enrich the SES data by gathering information about variables that are known to be important in the participation decision:
- the marital status,
- the number of children,
- the age of the children,

- the total income of other household members.

These variables and their interaction with gender are typically the type of variables that have been included in empirical studies where selection into paid employment is accounted for.

With these variables available in the SES, we could use specification S2 above to model selection in the SES (using the estimated coefficient from ECHP applied to the variables in the SES), and then derive the Gender pay gap using this correction for selection in the wage regression.

Approach 2: Construct the missing variables from existing variables

As it might not be possible to enrich the data – or if it is possible, it will probably take time to gather this new information – an alternative approach as long as these variables are not available would be to construct the missing variables. The approach here relies on the idea that there exists a reduced form expression of the missing variables as a function of the existing variables Z_1, i.e. $Z_2 = Z_1*f+e$ where f are parameters and e is an error term. Since we observe both sets of variables in ECHP, we could estimate the parameters f in ECHP and predict Z_2 in the SES based on these estimates, and the variables Z_1 available in the SES. The inverse Mills ratio could then be calculated in the SES using the predicted set of variables $Z_2 = Z_1*f$ instead of the true but unavailable Z_2.

The Z_2 variables are the marital status, the presence of children, and the other household member's income, and their interaction with gender. These variables need to be predicted using only the limited set of variables age, education, and gender (Z_1). It is likely that the prediction will be imprecise (very low R-square) and hence provide poor results when plugged into the inverse Mills ratio. To some extent, this problem could be overcome by using interactions and higher orders (age, age square, age cubic, etc.) to gain flexibility. Various specifications could be tested using ECHP. For the sake of the argument, we have tested this method using one specification with age, age square and age cubic, education and the interaction of gender with all these variables. The results are reported in Appendix 2. Comparison of specification 2 presented above and this imputed version of specification 2 indicates strong divergence between the two specifications except for Spain and Germany. The coefficients of the inverse Mills ratio in the wage regressions differ significantly as well as the Gender pay gap measures except for Spain and Germany where both specifications are very much in line. These differences are due to the fact that the R-square of the reduced form regressions of the exclusion restrictions on age, age square, age cubic, education and their interaction with gender is very low in most countries (between 0.02 and 0.15 depending on the dependent variable) except for Spain and Germany (between 0.15 and 0.30). It seems that, even though there is some scope for improvement for most countries, results for Spain and Germany are encouraging and suggest that testing other specifications – especially with additional interactions and high orders – could deliver satisfactory results for other countries too.

Approach 3: Use the OLS results

Even though we are able to model selection quite accurately using specification 2, for some of the countries in our sample, the selection does not appear to be stringent, i.e. the coefficient of the IMR is not significant, and the OLS Gender pay gap and the Gender pay gap from specification 2 are very much in line. This means that the OLS estimates of the Gender pay gap might in fact be a good proxy for the true Gender pay gap, at least for these countries where selection is weak.

Approach 4: Re-scale the OLS Gender pay gap

Another way to approach the problem could be to re-scale the OLS Gender pay gap measured in the SES proportionally to the difference of the OLS Gender pay gap to the Gender pay gap from specification 2 measured in ECHP. So for instance for Denmark, the OLS Gender pay gap is -0.088 while the Gender pay gap from specification 2 is -0.071. One could therefore re-scale the OLS Gender pay gap in the SES data, i.e. -0.137, by decreasing this value by 0.017 and obtain a Gender pay gap of -0.120. The problem with this approach is that the correction to be applied would need to be different for each country.

Cautious note about Approach 3 and 4

Although it is possible in theory that the selection process modelled in specification 2 changes over time as labour market participation changes or as the incentives and disincentives to work (policy interventions mainly) induce different persons to participate over time, we believe the specification is flexible enough to pick up most changes. Hence, the parameters estimates of the selection model in specification 2 should be expected to remain constant over time. However, the impact of the selection process on wages probably changes over time. This implies that the bias induced when using OLS techniques will also change over time. This means that the difference between the OLS Gender pay gap and the Gender pay gap derived from specification 2 will change over time. Approach 3 and 4, when chosen should be updated regularly to account for the changes in the selection into labour participation. Unfortunately, the ECHP data set could not be used anymore since the data covers the years 1994-2001 only.

Extension: Estimating the self-selection from LFS

In this research, we have estimated the self-selection process from the ECHP. We did this because this made it possible to measure the effect of correcting for self-selection on the Gender pay gap. However, estimating the self-selection from another dataset like LFS would be straightforward. All that is required is to rename the variables in the datasets. The estimation procedures developed by us can then be applied. This, however, would only have added value when the SES does include the variables that are relevant to explain self-selection (such as marital status, number and age of children and their interaction with gender).

References

Albrecht, J., van Vuuren A., Vroman, S. (2004), "Decomposing the Gender Wage Gap in the Netherlands with Sample Selection Adjustments", *IZA DP*, 1400, p.1-50.

Arulampalam, W., booth, A.L., Bryan, M.L. (2007), "Is there a glass ceiling over Europe? Exploring the Gender pay gap across the Wage Distribution", *Industrial and Labor Relations Review*, 60(2), p.163-186.

Bardasi, E., Gornick. J. (2008), "Working for less? Women's part-time wage penalties across countries". *Feminist Economics*, 14(1), 37-72.

Beblo, M., Beninger D., Heinze, A., Laisney., F., (2003a), "Methodological Issues Related to the Analysis of Gender Gaps in Employment, Earnings and Career Progression", Project Prepared for the European Commission Employment and Social Affairs DG, p.1-136.

Beblo M., Beninger D., Heinze, A., Laisney., F., (2003b), "Measuring Selectivity-Corrected Gender Wage Gaps in the EU", *ZEW DP* 03-74, p. 1-33.

Blau, F., Kahn, L. (1996), "International differences in male wage inequality: Institutions versus market forces", *Journal of Political Economy*, 104 (4), p. 791-837.

Blinder, A. (1973), "Wage discrimination: reduced form and structural estimates", *Journal of Human Resources*, 8(4), p.436– 455.

Buchinsky, M. (1998). "Recent advances in quantile regression models", *Journal of Human Resources*, 33, 88-126.

Buligescu, B., de Crombrugghe, D., Menteşoğlu, G., Montizaan, R. (2009), "Panel estimates of the wage penalty for maternal leave", *Oxford Economics Papers*, 61 (1), i35-i55.

Card, D., Di Nardo, J. (2002), "Skill-biased technological change and rising wage inequality: some problems and puzzles", *Journal of Labor Economics*, 20(4), p. 733-783.

Commission of the European Communities (2006), "Communication from the Commission to the Council, The European Parliament, The European Economic and Social Committee and the Committee of the Regions. A Roadmap for equality between women and men 2006-2010", p.1-22.

Commission of the European Communities (2008), "Communication from the Commission to the Council, The European Parliament, The European Economic and Social Committee and the Committee of the Regions. Equality between women and men-2008", p. 1-27.

Directorate-General for Employment, Social Affairs and Equal Opportunities, Unit EMPL/G/2, (2007), "Legal Aspects of the Gender pay gap Report by the Commission's Network of legal experts in the fields of employment, social affairs and equality between men and women", p. 1-98.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G. (2005), "Individual Risk Attitudes : New Evidence from a Large, Representative, Experimentally-Validated Survey", Discussion Papers of DIW Berlin 511.

Dupuy, A., Fernandez-Kranz, D., (2007), "International differences in the family gap in pay: the role of labour market institutions". *IZA Discussion Paper*, N. 2719, p. 1-36.

England, P. (1982), "The Failure of Human Capital Theory To Explain Occupational Sex Segregation", *Journal of Human Resources* 17(3), p. 358-370.

Fouarge, D., Muffels, R. (2008), "Part-time Work and Childbirth in Europe: Scarring the Career or Meeting Working-time Preferences?". In: R, Muffels, *Flexibility and* *employment security in Europe. Labour markets in transition*, Cheltenham: Edward Elgar (pp. 223-254).

Fouarge, D., Muffels, R. (2009), "Working part-time in The British, German and Dutch Labour Market: Scarring for the Wage Career?", *Journal of Applied Social Science Studies*, forthcoming.

Görlich, D., de Grip, A. (2009), "Human capital depreciation during hometime", *Oxford Economic Papers*, 61 (1), 98-121.

Heckman, J. (1979), "Sample selection bias as a specification error", *Econometrica*, 47, 153–61.

Heinze, A. (2007), "Going behind the gender wage gap: Are women less educated or are they in worse firms?", Paper presented at EALE Conference, p.1-28.

Henneberger, F., Sousa-Poza, A. (1998), "Estimating wage functions and wage discrimination using data from the 1995 Swiss labour force survey: a double-selectivity approach", *International Journal of Manpower*, 19(7), p. 486-506.

Kim, M-K, Polachek, S. (1994), "Panel estimates of male–female earnings functions" *Journal of Human Resources,* 29(2), p. 406–428.

Kunze, A. (2007), "Gender wage gap studies: consistency and decomposition", *Empirical Economics*, s00181-007-0143-4, p. 1-14.

Machado, J., Mata, J. (2005), "Counterfactual decomposition of changes in wage distribution using quantile regression", *Journal of Applied Econometrics*, 20, p. 445–465.

Manning, A., Petrongolo, B. (2008), "The part-time pay penalty for women in Britain". *Economic Journal*, 118 (526), 28-51.

Manzoni, A., Fouarge, D., Luijkx, R., Muffels, R. (2008), "Childbirth and mother's labour supply: comparative evidence from German, Dutch and UK life-history data", RC28 Conference on Social Stratification and Insiders/Outsiders, Florence, 15-17.05.2008.

Melly, B. (2005), "Public-private sector wage differentials in Germany: Evidence from quantile regression", *Empirical Economics*, 30, p.505-520.

Meng, X., Meurs, D. (2004), "The gender earnings gap: effects of institutions and firms—a comparative study of French and Australian private firms", *Oxford Economic Papers,* 56, p. 189–208.

Neuman, S., Oaxaca, R. (1998), "Estimating Labour Market Discrimination with Selectivity-Corrected Wage Equations: Methodological Considerations and An Illustration from Israel", CEPR Discussion Paper No. 1915.

Nicodemo, C. (2008), "Participation and Wage Equations for Married Women in European Countries", European Population Conference 2008, Barcelona.

Oaxaca, R., Ransom, M. (1994), "Using Econometric Models for Intrafirm Equity Salary Adjustments", *Journal of Economic Inequality*, vol. 1(3), p. 221-249.

Paull, G. (2006), "The impact of children on women's paid work", *Fiscal Studies*, 27(4), 473-512.

Plantega, J., Remery, C. (2006), *The Gender pay gap. Origin and policy responses. A comparative review of 30 European countries*. Luxembourg: Office for Official Publications of the European Communities.

Polachek, S.W., (1981) "Occupational self-selection: a human capital approach to sex differences in occupational structure", *The Review of Economics and Statistics* , 63(1), p. 60–69.

Rubery, J., Grimshaw, D., Figueiredo, H. (2005), "How to close the Gender pay gap in Europe: towards gender mainstreaming of pay policy", *Industrial Relations Journal*, 36(3), p. 184-213.

Rycx, F., Tojerow I. (2004), "Rent sharing and the gender wage gap in Belgium", *International Journal of Manpower*, 25(3-4), p. 279-299.

Stanley, T., Jarell, S. (1998), "Gender Wage Discrimination Bias? A Meta-Regression Analysis", *The Journal of Human Resources*, 33(4), p. 947-973.

Solberg, E., Laughlin, T. (1995), "The Gender pay gap, Fringe Benefits, and Occupational Crowding", *Industrial and Labor Relations Review*, 48(4), pp. 692-708.

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

Appendix 1: Overview of Gender pay gap literature

Table A1: Overview of findings from the Gender pay gap literature

Authors	Estimation method	Selection	Heterogeneity	Findings
Machado and Mata 2005	**Method**: Quantile regression Counterfactual Simulation by comparing the marginal distributions implied by different distributions for the covariates. **Data:** Quadro de Pesoal **Country**: Portugal			Changes in individual attributes and in returns to these attributes contribute to increase the wage inequality by the same amount in Portugal. Education is driving the wage inequality. The distribution of the education returns became more spread out: higher education returns for the top wage distribution and constant returns for the bottom contributes to an increase in inequality. Further the observed increase in the level of education in Portugal also contributed towards a less equal wage distribution.
Beblo et al 2003a, Beblo et al 2003b	**Method**: OLS **Data**: ECHP and GSOEP **Countries**: France, Germany, Italy, Spain, the United Kingdom	Heckman and Lewbel for Germany only, significant impact on Gender pay gap	Low variation over time therefore panel data methods not successful in accounting for heterogeneity	Their results show that at most half of the difference is attributed to differences in characteristics but that the size of the endowment differs between countries and depends on the choice of the estimator. Correcting for sample selection has a significant impact on both the wage estimates and the pay gap decomposition. When using quantile regression decomposition the authors find a remarkable difference between as well as within countries and recommend not to focus only on the mean pay gap as differences over the wage distribution can be informative for policy conclusions.
Albrecht et al 2004	**Method**: Quantile regression **Data**: OSA Labour Supply Panel **Country**: Netherlands,	Buchinsky (1998) correction for sample selection into full time work, only for women		Taking first the population of women working full time as given, they find a wage gap of 20% which increases as we move up the distribution, that is there is a glass ceiling effect. Three quarters of the wage gap can be attributed to gender differences in rewards. However there is a strong sample selection since most

				women in the Netherlands work part-time. There is a positive selection of women into full-time work in the Netherlands; i.e., women who get the greatest return to working full time do work full time. About two thirds of this selection is due to observables such as education and experience with the remainder due to unobservables. Once selection into full time work is corrected for, differences in labour market characteristics between men and women play a larger role than before in explaining the gender gap however, most of the gender gap across the distribution continues to be accounted for by differences in how men and women are rewarded. The decompositions show that the majority of the gender log wage gap is due to differences between men and women in returns to labour market characteristics rather than to differences in the characteristics.
Melly (2005)	**Method:** Quantile regression **Data**: GSOEP 1984-2001 **Country**: Germany			Oaxaca decomposition results suggest that conditional wages are higher in the public sector for women but lower for men. Using the quantile regression decomposition technique proposed by Machado and Mata (2004), they find that the conditional distribution of wages is more compressed in the public sector. At the low end of wages, differences in characteristics explain less than the raw wage gap when it is the opposite at high wages. Separate analyses by work experience and educational groups reveal that the most experienced employees and those with basic schooling do best in the public sector. All results are stable over the 80s

				and 90s.
Solberg and Laughlin (1995)	**Method**: Pooled OLS with gender dummy by occupation, Occupational decomposition, **Data:** 1991 National Longitudinal Survey for Youth, labour force participants aged 26-34 years, **Country:** The USA,	Heckman correction insignificant and labour force participation equation did not manage to predict well the probability of working since most participants were working. Therefore selection not used in final version		Two sets of equations were estimated one of the log wage and one with the total compensation index (which includes fringe benefits). Fringe benefits pay an important role in the determination of total compensation and the measurement of the wage gap. However when the index of total compensation is used in regressions by occupational category the gender coefficient is no longer significant in six out of seven categories. The only occupational category where the gender variable is still significant is the operatives which can be explained by job and contract heterogeneity. Therefore occupational assignment is the primary determinant of the gender differences. The absence of gap for female dominated occupations and male dominated occupations suggest that the crowding hypothesis and occupational segregation may be the reason of gender wage differentials.
Henneberger and Sousa-Poza, 1998	**Method:** OLS **Data**: Swiss Labour Force Survey **Country**: Switzerland	Heckman two step Behrman *et al.* (1981) double selectivity Tunali (1986) double selectivity		Selectivity correction only makes sense when a certain correlation between the estimate of λ and the other regressors in the wage equation exists.
Meng and Meurs (2004)	**Method**: Firm Fixed effects **Data**: employer-employee linked data French Labour Cost and Wage Structure Survey 1992 Australian Workplace Industrial Relations Survey 1993 , data only on medium and large private enterprises			The degree to which firm wage policies can influence the gender earnings gap may be affected by labour market institutions. For example, countries with a more centralized wage bargaining system may leave less room for the effect of firm wage policies than countries with a more decentralized wage bargaining system. The authors

	Countries: Australia and France			find that firm wage policies in Australia play a much larger role in narrowing the gender earnings gaps than in France. This is mainly due to the fact that Australia has a more decentralized wage bargaining system and that such system is operated under an environment where there is a strong union presence. In Australia firms which face stronger market competition are more likely to pay a firm wage premium to workers that is equal across gender groups. In addition, although firms with enterprise level wage bargaining are more likely to pay higher premia to male than female employees, if the enterprise level wage bargaining is conducted by a union delegate rather than individual employees it benefits women.
Heinze A. (2007)	**Method**: Quantile regression **Data**: employer-employee data IAB **Country**: Germany			The unconditional gender gap is not constant across the wage distribution. It is sharply increasing within the first quartile and the decrease decelerates until the 70th percentile after that the gap is increasing again. Female employees are better educated then men in the lower tail of the wage distribution but they work in the inferior firms. In the upper tail of the distribution men and women work in similar firms but the female employees have less human capital.
Blinder (1973)	**Method**: OLS **Data**: Panel Data of Income Dynamics **Country**: the USA			70 percent of the overall race differential and 68 percent of the sex differential are ultimately attributable o discrimination of various sorts. Age-wage profiles are different between white men and white women. Women have a flat wage-age profile compared to men whose earnings increase over the life cycle. Thus the failure of some women within

				the same education-occupation category to rise on the economic ladder over their working lives is seen to be the single largest cause for male-female wage differentials among whites.
Arulampalam et al. 2007	**Method**: Quantile regression **Data**: ECHP **Countries**: 11 EU countries			They estimate the wage gaps for each sector with and without the industry and occupational controls. The author split the sample by public and private sector and find evidence in some countries of glass ceilings (Gender pay gaps are larger at the top of the distribution) and in others of sticky floors (Gender pay gaps are larger at the bottom of the wage distribution).
Kim and Polachek (1994)	**Method**: Fixed effects and random effects with intercept and slope specific effect **Data**: PSID **Country**: USA	Endogeneity: Experience, time out of work	Panel methods FE, with intercept and slope specific effects	About 50% of unexplained male-female wage gap can be attributed to unmeasured individual differences. These results emerge both from individual-specific intercept and individual-specific slope models, with individual-specific slope models resulting in a slightly smaller unexplained male-female wage gap.
Stanley and Jarell (1998)				The estimated gender gap has been steadily declining however large biases are likely when researchers omit experience or fail to correct for selection bias. The wage gap is reduced by .197 or nearly 18% for the USA when selection bias is not controlled for.
Neuman and Oaxaca (1998)	**Method**: bivariate probit **Data**: 20% sample of the 1995 Census of Population and Housing **Country**: Israel	Selection into work and selection into occupations, two stage Heckman procedure		They estimate gender (or ethnic) wage differentials within a given occupation – Professionals accounting for occupational and sample selection. The overall results are the following. Gender wage differentials (at the mean points) are larger than ethnic wage differentials. Not only can the magnitudes of the discrimination estimates be greatly affected but even the

				direction of discrimination is affected when correcting for selection. In all cases the Inverse Mills Ratio was statistically significant, indicating the presence of selection bias in professional employment.
Polacheck (1981)	**Method**: regression, logit, simultaneous equation models **Data**: National Longitudinal Survey of Women, 30-44 years of age **Country**: the US	Endogeneity: Home time		It is hypothesized that, at least for females, duration of time in and out of the labour force is related to occupation. This implication stems from a model that utilizes a hedonic approach to embed the occupational choice decision into the human capital framework. Empirically this hypothesis is tested by measuring the effect of home-time on occupational choice. To answer the remaining question of how important intermittent labour force behaviour really is in explaining occupational segregation, male-female occupational dissimilarity can be compared before and after adjustments are made for differences in lifetime labour force participation. An occupational probability density function is obtained for each woman Aggregation of all individual probabilities yields a projected population-wide occupational distribution. From this comparison, it can be seen that differences in labour force commitment alone account for much of the difference in professional and menial employment. If women were to have a full commitment to the labour force, the number of women professionals would increase by 35%, the number of women in managerial professions would more than double, and women in menial occupations would decrease by more than 25%. These results hold even when using simultaneous equation models and

				endogenising home time.
Rycx and Tojerow (2004)	**Method**: OLS, adjusting for group effects the covariance matrix , control for industry wage differentials and a wide range of observable individual and firm characteristics. **Data**: Belgian Structure of Earnings Survey and Structure of Business Survey supplemented with financial information for the firm from 1995 SBS firm-level survey **Country**: Belgium	Endogeneity: wage-profit elasticity		Empirical findings show that individual gross hourly wages are significantly and positively related to firm profits-per-employee even when controlling for group effects in the residuals, individual and firm characteristics, industry wage differentials and endogeneity of profits. Of the overall gender wage gap (on average women earn 23.7 per cent less than men), results show that around 14 per cent can be explained by the fact that on average women are employed in firms where profits-per-employee are lower. Thus, findings suggest that a substantial part of the gender wage gap is attributable to the segregation of women in less profitable firms.
Blau and Kahn (1996)	**Method:** **Data:** ISSP Austria (1985-87 and 1989), West Germany (1985-88), Hungary (1986-88), Switzerland (1987), Britain (1985- 89), the US (1985-89), and Norway (1989). Complemented by Class Structure and Class Consciousness Survey for Sweden and Norway, Income Distribution Survey for Australia and Bank of Italy Survey for Italy			The greater overall U.S. wage dispersion primarily reflects substantially more compression at the bottom of the wage distribution in the other countries. While differences in the distribution of measured characteristics help to explain some aspects of the international differences, higher U.S. prices (i.e., rewards to skills and rents) are an important factor. Labour market institutions, chiefly the relatively decentralized wage-setting mechanisms in the United States, provide the most persuasive explanation for these patterns.

Appendix 2: Additional results from ECHP

The table below reports the outcome of our exercise consisting of simulating the exclusion restrictions from S2 with the limited set of variables included in S1.

Table A2: Actual and simulated exclusion restrictions[1)]

	Actual exclusion restrictions S2[2)]		**Predicted exclusion restrictions[3)]**	
	Gender pay gap, OLS with sample selection (S2) (column 3 of Table 4)	IMR (column 3 of Table 3)	Gender pay gap, OLS with sample selection (S2) with predicted exclusion restrictions	IMR, from predicted exclusion restriction
1	2	3	4	5
Denmark	-0.071***	-0.193***	-0.043**	-0.470***
	(0.020)	*(0.066)*	*(0.022)*	*(0.114)*
The Netherlands	-0.158***	-0.040	-0.050	-0.307***
	(0.020)	*(0.042)*	*(0.031)*	*(0.068)*
Ireland	-0.127***	-0.038	-0.043	-0.270***
	(0.024)	*(0.047)*	*(0.039)*	*(0.100)*
Italy	-0.112***	-0.016	-0.039*	-0.155***
	(0.011)	*(0.016)*	*(0.020)*	*(0.036)*
Greece	-0.129***	-0.098***	-0.081***	-0.115***
	(0.016)	*(0.025)*	*(0.029)*	*(0.044)*
Spain	-0.190***	-0.071***	-0.187***	-0.039
	(0.015)	*(0.025)*	*(0.023)*	*(0.039)*
Portugal	-0.226***	-0.033	-0.149***	-0.344***
	(0.014)	*(0.037)*	*(0.029)*	*(0.110)*
Austria	-0.168***	-0.242***	-0.284***	0.258*
	(0.019)	*(0.037)*	*(0.054)*	*(0.153)*
Finland	-0.127***	-0.060	-0.207***	0.332***
	(0.016)	*(0.052)*	*(0.019)*	*(0.092)*
Germany	-0.255***	-0.009	-0.269***	0.045
	(0.017)	*(0.037)*	*(0.029)*	*(0.082)*
United Kingdom	-0.184***	-0.027	-0.214***	0.110
	(0.017)	*(0.050)*	*(0.051)*	*(0.197)*

Significance levels: * p<0.10 ** p<0.05 *** p<0.01. Standard errors are provided in parentheses
Notes:
1) All analyses use filters 1 and 3 from Table 2.
2) The exclusion restrictions are: household income, number of children, dummy for children under 1 year, dummy for children under 12, dummy for bad health, married or cohabiting.
3) The exclusion restrictions are predicted based on: education, age, age square and age cube and their interactions with gender.
Source: ECHP, authors' computation.

Appendix 3: Impression of the SES data

It is evident from this project so far that both data files used have their own shortcomings. Both ECHP and the SES have truncation problems. ECHP does not include information on the number of hours worked for people in small jobs with less than 15 hours. Moreover it does not include reliable information on the firm size for workers employed in the public sector.

SES has four main problems:

1) it does not include workers in small firms with less than 10 employees or working in agriculture;

2) it does not include information of work experience making it difficult to get a good measurement of the Gender pay gap;

3) it does not include data on non-employed persons, making it impossible to investigate self-selection into employment straight from the data;

4) it includes very few variables that can be used to account for possible self-selection. It is clear that the lack of variables to explain selection into employment is a serious drawback of the data. Our implementation of the selection processes as computed on the ECHP into the SES is limited due to this data limitation. Essentially, variables such as family structure, marital status, age and number of children, and employment status of the partner have been used in the literature to indentify self-selection. The data – and the approach we suggest – would be greatly improved if such information was embodied into the SES. Our measurement of the Gender pay gap would also be improved if we did have information on work experience of male and female workers.

On the bright side, the SES is a very large dataset making it possible to study specific groups of workers in great detail. For example, the data can make it possible to investigate the wage structure and the wage differentials at a detailed sector and occupational level.

European Commission

Development of econometric methods to evaluate the Gender pay gap using Structure of Earnings Survey data

Luxembourg: Office for Official Publications of the European Communities

2009 — 51 pp. — 21 x 29.7 cm

ISBN 978-92-79-13181-3
ISSN 1977-0375

How to obtain EU publications

Our priced publications are available from EU Bookshop (http://bookshop.europa.eu), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents. You can obtain their contact details by sending a fax to (352) 29 29-42758.

KS-RA-09-011-EN-N

Publications Office