Unequal pay or unequal employment? A cross-country analysis of gender gaps.*

Claudia Olivetti
Boston University

Barbara Petrongolo
London School of Economics
CEP, CEPR and IZA

First draft, May 2005

Abstract

There is substantial international variation in gender pay gaps, from 25-30% in the US and the UK, to 10-20% in a number of central and northern EU countries, down to an average of 10% in southern EU. We argue that non-random selection of women into work across countries may explain part of such variation. This ides is supported by the observed variation in employment gaps, from 10% in the US, UK and Scandinavian countries, to 15-25% in northern and central EU, up to 30-40% in southern EU and Ireland. If women who are employed tend to have relatively high-wage characteristics, low female employment rates may become consistent with low gender wage gaps simply because low-wage women would not feature in the observed wage distribution.

We explore this idea across the US and EU countries estimating gender gaps in potential wages. In order to do this, we recover information on wages for those not in work in a given year by simply making assumptions on the position of the imputed wage observations with respect to the median, not on the actual level. Imputation is based on wage observations from nearest available waves in the sample and/or observable characteristics of the nonemployed. We estimate median wage gaps on the resulting imputed wage distributions. Our estimates for 1999 deliver higher median wage gaps on imputed rather than actual wage distributions for most countries in the sample, meaning that, as one would have expected, women tend on average to be more positively selected into work than men. However, this difference is tiny or virtually zero in the US and northern and central EU countries (except Ireland), and becomes sizeable in Ireland, France and southern EU, all countries in which gender employment gaps are high. In particular, in Spain, Portugal and Greece the median wage gap on the imputed wage distribution reaches 20 log points, a closely comparable level to that of the UK and other central and northern EU countries.

*We wish to thank seminar participants at Boston University, IFAU Uppsala, Ente Einaudi, CEP/LSE, Toulouse and CEMFI for very useful comments.

1 Introduction

There is substantial international variation in gender pay gaps, from 25-30 log points in the US and the UK, to 10-20 log points in a number of central and northern European countries, down to an average of 10 log points in southern Europe. International differences in overall wage dispersion are typically found to play a role in explaining differences in gender wage gaps (Blau and Kahn 1996, 2003). The idea is that a given level of dissimilarities between the characteristics of working men and women translates into a higher gender wage gap the higher the overall level of wage inequality. However, OECD (2002, chart 2.7) shows that, while differences in the wage structure do explain an important portion of the international variation in gender wage gaps, the inequality-adjusted wage gap in southern Europe remains lower than in the rest of Europe and the US.

In this paper we argue that, besides differences in wage inequality and therefore in the returns associated to characteristics of working men and women, a significant portion of the international variation in gender wage gaps may be explained by differences in characteristics themselves, whether observed or unobserved. This idea is supported by the striking international variation in employment gaps, ranging from 10 percentage points in the US, UK and Scandinavian countries, to 15-25 points in northern and central Europe, up to 30-40 points in southern Europe and Ireland. If selection into employment is non-random, then it makes sense to worry about the way in which selection may affect the resulting gender wage gap. In particular, if women who are employed tend to have relatively high-wage characteristics, low female employment rates may become consistent with low gender wage gaps simply because low-wage women would not feature in the observed wage distribution.

Although there exist substantial literatures on gender wage gaps on one hand, and gender employment, unemployment and participation gaps on the other hand,[1] to our knowledge the variation in both quantities and prices has not been simultaneously exploited to understand important differences in gender gaps across countries. In this paper we claim that the international variation in gender employment gaps can indeed shed some light on well-known stylized facts of international gender wage gaps. In particular, we explore this view by estimating selection-corrected wage gaps.

The existing literature contains a number of country-level studies that estimate selection-corrected wage gaps, based on alternative methodologies and applied to both race and gender gaps. Neal (2004) estimates wage gaps between black and white women in the US and finds that "the black-white gap in log-potential wages among young adult women in 1990 was at least 60 percent larger than the gap implied by reported earnings and hours worked". The

[1] See Altonji and Blank (1999) for an overall survey on both employment and gender gaps for the US, Blau and Kahn (2003) for international comparisons of gender wage gaps and Azmat, Güell and Manning (2004) for international comparisons of unemployment gaps).

gap in log-potential wages is in turn estimated by fitting median regressions on imputed wage distributions, using alternative methods of wage imputation for women non employed in 1990. Using a similar approach, Chandra (2003) finds that the wage gap between black and white US males was also understated, due to selective withdrawal of black men from the labor force during the 1970s and 1980s.

Turning to gender wage gaps, Blau and Kahn (2004) study changes in the US gender wage gap between 1979 and 1998 and find that sample selection implies that the 1980s gains in women's relative wage offers were overstated, and that it may also explain part of the slowdown in convergence between male and female wages in the 1990s. Mulligan and Rubinstein (2004) also argue that the narrowing of the gender wage gap in the US during 1964-2002 may be a direct impact of progressive selection into employment of high-wage women, in turn attracted by widening within-gender wage dispersion. This idea follows the implications of the Roy's (1951) model, as applied to the choice between market and non-market work in the presence of rising dispersion in the returns to market work. Related work on European countries includes Blundell, Gosling, Ichimura and Meghir (2004), Albrecht, van Vuuren and Vroman (2003) and Beblo, Beninger, Heinze and Laisney (2003). Blundell et al. examine changes in the distribution of wages in the UK during 1978-2000, using bounds to the distribution of potential wages, in order to allow for the impact of non-random selection into work. Bounds are first constructed based on the worst case scenario and then progressively tightened using restrictions motivated by economic theory. Features of the resulting wage distribution are then analyzed, including overall wage inequality, returns to education, and gender wage gaps. Albrecht et al. estimate gender wage gaps in the Netherlands having corrected for selection of women into market work according to the Buchinsky's (1998) semi-parametric method for quantile regressions. They find evidence of strong positive selection into full-time employment: were all Dutch women working full-time, the gender wage gap would be much higher. Finally, Beblo et al. show selection corrected wage gaps for Germany using both the Heckman (1979) and the Lewbel (2002) two-stage selection models. They find that correction for selection has an ambiguous impact on gender wage gaps in Germany, depending on the method used.

Interestingly, most of the studies cited find that correction for selection has sizeable consequences for our assessment of gender wage gaps. At the same time, none of these studies use data from southern European countries, where employment rates of women are lowest, and thus the selection issue should be most relevant. In this paper we use data for the US and for a representative group of European countries to investigate how non-random selection into work may have affected the gender wage gap. In doing this, we use panel data sets that are as comparable as possible across countries, namely the Panel Study of Income Dynamics (PSID) for the US and the European Community Household Panel

Survey (ECHPS) for Europe. Our analysis is based on the period 1994-2001, the longest time span for which data are available for all countries.

In our empirical analysis we aim at recovering the counterfactual wage distribution that would prevail in the absence of non-random selection into work - or at least some of its characteristics, and we then estimate gender gaps in potential wages. In order to do this, we recover information on wages for those not in work in a given year following the approach of Johnson, Kitamura and Neal (2000) and Neal (2004), which is based on wage imputation for the nonemployed. This approach simply requires assumptions on the position of the imputed wage observations with respect to the median. Importantly, it does not require assumptions on the actual level of missing wages, as typically required in the matching approach, nor it requires arbitrary exclusion restrictions often invoked in two-stage Heckman sample selection correction models.

We then estimate median wage gaps on the resulting imputed wage distributions. The impact of selection into work is assessed by comparing estimated wage gaps on the base sample with those obtained on a sample enlarged with wage imputation. The attractive feature of median regressions is that, if missing wage observations fall completely on one or the other side of the median regression line, the results would in this case only be affected by the position of wage observations with respect to the median, and not by specific values of imputed wages. One can therefore make assumptions motivated by economic theory on whether an individual who is not in work should have a wage observation below or above median wages for their gender.

Imputation can be performed in several ways. First, we exploit the panel nature of our data sets and, for all those not in work in some base year, we search backwards and forwards to recover hourly wage observations from the nearest wave in the sample. This is equivalent to assuming that an individual's position with respect to the base-year median can be recovered by the ranking of her wage in the nearest wave in the base-year distribution. As the position with respect to the median is determined using levels of wages in other waves in the sample, we are in practice allowing for selection on unobservables.

While imputation based on this procedure arguably uses the minimum set of potentially arbitrary assumptions, it has the disadvantage of not providing any wage information on individuals who never worked during the sample period. To recover wage observations also for those never observed in work during the whole sample period, we make assumptions about whether they place above or below the median wage offer, based on their observable characteristics, specifically education, experience and spouse income. In this case we are allowing for selection on observables only.

Our estimates for 1999 deliver higher median wage gaps on imputed rather than actual wage distributions for most countries in the sample, meaning that, as one would have expected, women tend on average to be more positively selected into work than men. However, this difference is tiny or virtually zero in the US, northern European countries (except Ireland) and most central European countries, it becomes sizeable in Ireland, France and southern Europe, i.e. countries in which the gender employment gap is highest. In particular, in Spain, Portugal and Greece the median wage gap on the imputed wage distribution reaches nearly 20 log points, a closely comparable level to that of the UK and of other central and northern European countries. In other words, correcting for selection into employment

We also estimate wage gaps adjusted for characteristics on both actual and imputed wage distributions and perform Oaxaca-Blinder decompositions of wage gaps. Countries whose gender wage gap is not seriously affected by sample inclusion rules also have a roughly unchanged gap decomposition across specifications. In countries where wage imputation indeed affects the estimated wage gap, it is both characteristics and returns components that matter. In other words, in Ireland and southern Europe, women with lower labor market attachment have a higher wage penalty with respect to men because both they have relatively poorer characteristics than women with higher labor market attachment and because they receive a lower remuneration for a given set of characteristics.

Finally, in order to relate our findings to those of the existing literature on cross-country differences in gender wage gaps, we use the methodology proposed by Juhn, Murphy and Pierce (1991) and Blau and Kahn (1996) to decompose such differences into differences in characteristics, both observed and unobserved, and differences in (male) returns to these characteristics. This decomposition is used in the literature in order to quantify the contribution of cross-country differences in the wage structure to the explanation of the variation in the gender wage gap. We perform this decomposition on both the actual and the imputed wage distribution. Overall we find that the contribution of characteristics relative to that of the wage structure is much stronger in southern Europe than elsewhere. This effect is attenuated on the imputed wage distribution.

The paper is organized as follows. Section 2 describes the data sets used and the specification of our wage equations. Section 3 describes our methodology. Section 4 estimates median gender wage gaps on actual and imputed wage distributions, to illustrate how alternative sample selection rules affect the estimated gaps. Section 5 decomposes international differences in gender wage gaps into a component explained by differences in characteristics, observed and unobserved, and another component explained by differences in the wage structure. Section 6 concludes.

2 Data

2.1 The PSID

Our analysis for the US is based on the Michigan Panel Study of Income Dynamics (PSID). This is a longitudinal survey of a representative sample of US individuals and the families in which they reside. It has been ongoing since 1968. The data were collected annually through 1997 and biennially starting in 1999. In order to be consistent with the data for Europe, we consider five waves from the PSID, from 1994 to 2001. We restrict our attention to employed workers aged 16-64 and we exclude the self-employed, full-time students, and individuals in the armed forces.

The wage concept that we use throughout the analysis is the gross hourly wage. Our basic wage equation specification reflects a simple human capital model and includes controls for an individual's education, work experience, industry and occupation. We also include 51 state dummies and year dummies. The results of our wage equations were not sensitive to the inclusion of a dummy variable for ethnic origin. The results reported below are based on specifications that do not control for ethnic origin, for consistency with the specifications used for the EU. Below we briefly describe how we construct the main variables of interest for the US sample.

For the education variable, we group individuals according to three broad educational categories: less than high school, high school completed, and college completed. We construct three education dummies based on this categorization. The dummy EDU1 is equal to one if an individual has completed less than twelve years of schooling and is equal to zero otherwise. EDU2 is a dummy variable that is equal to one if he or she has completed between twelve and fifteen years of schooling and is equal to zero otherwise. Finally, EDU3 is equal to one if an individual has completed at least sixteen years of education and equals zero otherwise. We have chosen this particular categorization to be consistent with the education variable available in the European data set, which is only available by recognized qualifications. We include 12 dummy variables for occupation, based on the 3-digits occupation codes from the 1970 Census of the Population. We also include 12 dummy variables for the industry an individual worked in during the previous year. Detailed occupation and industry categories are described in Table A1.

Following Blau and Kahn (2004) we construct the variable representing actual years of experience according to the following methodology. In 1976 and in 1985 a retrospective question about an individual's years of actual working experience since he or she was 18 years old was asked, in the PSID questionnaire, to all heads and wives irrespective on the year they had joined the sample. The answer to this question in 1985 provide the basis from which we build our variable for an individual's actual work experience. After 1976 the question was asked to all head and wives when they join the panel. Once we have the initial values for this variable in 1985, we use the entire work history file from the PSID to compute the actual experience of the individuals in the years of interest. For example, in order to know the years of actual experience in 1994 for an individual who was in the survey in 1985, we take the number of years of actual experience he or she had in 1985 and we add one for each year between 1985 and 1994 in which the individual has worked a positive number of hours.[2] If the individual has worked 0 hours in any given year we add a zero to the initial value of the variable. As discussed in Blau and Kahn (2004), this procedure allows one to construct the full work experience of an individual for every year he or she has been in the survey except for the last two waves. This is because the PSID has started collecting information biannually since 1999 (of course, this is relevant only for those individuals who were already in the sample in 1997. The information on years of actual experience is available from the PSID for new entrants in 1999 and in 2001). In order to solve this problem we follow the methodology developed by Blau and Kahn (2004) and estimate the experience for the missing years (that is, the year between 1997 and the 1999, and the year between 1999 and 2001) by averaging the two predicted values from, gender-specific, logit regression for the two adjacent years. The explanatory variables in the regressions include race, schooling, experience a marital status indicator and variables for the number of children aged 0-2, 3-5, 6-10, and 11-15 who are living in the household at the time of the interview. We use the experience variable constructed according to this procedure in all the regressions.

2.2 The ECHPS

Data for European countries are drawn from the European Community Household Panel Survey. This is an unbalanced household-based panel survey, containing annual information on a few thousands households per country during the period 1994-2001.[3] The ECHPS has the advantage that it asks a consistent set of questions across the 15 members states of the pre-enlargement EU. The Employment section of the survey contains information on the jobs held by members of selected households, including wages and hours of work. The household section allows to obtain information on the family composition of respondents.

As for the US, we restrict our analysis of wages to employed workers aged 16-64 as of the survey date, and exclude the self-employed, those in full-time education and the military.

[2] The measure of actual experience used here includes both full-time and part-time work experience, as this is better comparable to the measure of experience available from the ECHPS.

[3] The initial sample sizes are as follows. Austria: 3,380; Belgium: 3490; Denmark: 3,482; Finland: 4,139; France: 7,344; Germany: 11,175; Greece: 5,523; Ireland: 4,048; Italy: 7,115; Luxembourg: 1,011; Netherlands: 5,187; Portugal: 4,881; Spain: 7,206; Sweden: 5,891; U.K.: 10,905. These figures are the number of household included in the first wave for each country, which corresponds to 1995 for Austria, 1996 for Finland, 1997 for Sweden, and 1994 for all other countries.

The specification of our wage equations on EU data is as similar as possible to that used for the US, subject to slight data differences. The EU education categories are: less than upper secondary high school, upper secondary school completed, and higher education. These correspond to ISCED 0-2, 3, and 5-7, respectively. We consider 9 broad occupational groups: although this is not the finest occupational disaggregation available in the ECHPS, it is the one that allows the best match with the occupational classification available in the PSID. We also consider 18 industries. (See Table A1 for definition of categories).

The main differences with respect to the specification of the wage equation concern the race and experience variables. No information on ethnicity is available in the ECHPS, nor a measure of actual experience. Our wage equations for the EU thus do not control for race, and control for a measure of potential experience computed as the current age of an individual, minus the age at which she started her working life. We also control for region of residence at the NUT1 level, meaning 11 regions for the UK, 1 for Finland and Denmark, 15 for Germany, 1 for the Netherlands, 3 for Belgium and Austria, 2 for Ireland, 8 for France, 12 for Italy, 7 for Spain, 2 for Portugal and 4 for Greece.

All descriptive statistics for both the US and the EU samples are reported in Table A1.

2.3 Descriptive evidence on gender gaps

Table 1 reports raw gender gaps in log gross hourly wages and employment rates for all countries in our sample. All these are computed for the population aged 16-64. At the risk of some oversimplification, one can classify countries in three broad categories according to their levels of gender wage gaps. In the US and the UK men's hourly wages are 25 to 30 log points higher than women's hourly wages. Next, in northern and central Europe the gender wage gap in hourly wages is between 10 and 20 log points, from a minimum of 11 log points in Denmark, to a maximum of 24 log points in the Netherlands. Finally, in southern European countries the gender wage gap is on average 10 log points, from 6.3 in Italy to 13.4 in Spain. Such gaps in hourly wages display a roughly negative correlation with gaps in employment to population rates. Employment gaps range from 10 percentage points in the US, the UK and Scandinavia, to 15-25 points in northern and central Europe, up to 30-40 points in southern Europe and Ireland. The relationship between wage and employment gaps is represented in Figure 1. The coefficient of correlation between them is -0.497 and is significant at the 7% level.

Such negative correlation between wage and employment gaps may reveal significant sample selection effects in observed wage distributions. If the probability of an individual being at work is positively affected by the level of her potential wage offers, and this mechanism is stronger for women than for men, then high gender employment gaps become consistent with relatively low gender wage gaps simply because low wage women are relatively less likely than men to feature in observed wage distributions.

Table 1 also reports wage and employment gaps by education. Employment gaps everywhere decline with educational levels, if anything more strongly in southern Europe than elsewhere. On the other hand, the relationship between gender wage gaps and education varies across countries. While the wage gap is either flat or rises slightly with education in most countries, it falls sharply with education in Ireland and southern Europe. In particular, if one looks at the low-education group, the wage gap in southern Europe is closely comparable to that of other countries - while being much lower for the high-education group. However, the fact that the low-education group has the lowest weight in employment makes the overall wage gap substantially lower in southern Europe.

Interestingly, in the four southern European countries, the overall wage gap is smaller than each of the education-specific gaps, and thus lower than their weighted average. One can think of this difference in terms of an omitted variable bias. The overall gap is simply the coefficient on the male dummy in a wage equation that only controls for gender. The weighted average of the three education-specific gaps would be the coefficient on the male dummy in a wage equation that controls for both gender and education. Education would thus be an omitted variable in the first regression, and the induced bias has the sign of the correlation between education and the male dummy, given that the correlation between education and the error term is always positive. While the overall correlation between education and the male dummy tends to be positive in all countries, such correlation becomes negative and fairly strong among the employed in southern Europe, lowering the overall wage gap below each of the education-specific wage gaps. The fact that, if employed, southern European women tend to be more educated than men may be itself interpreted as a signal of selection into employment based on high-wage characteristics.

In Table 1A we report similar gaps for the population aged 25-54, as international differences in schooling and/or retirement systems may have affected relevant gaps for the 16-64 sample. However, when comparing the figures of Table 1 and 2, we do not find evidence of important discrepancies between the gender gaps computed for those aged 16-64 and those aged 25-54. The rest of out analysis therefore uses the population sample aged 16-64.

3 Methodology

We are interested in measuring the gender wage gap:

$$D = E\left(w|X, \text{male}\right) - E\left(w|X, \text{female}\right), \tag{1}$$

where D denotes the gender gap in mean log wages, w denotes log wages and X is a vector of observable characteristics. Average wages for each gender are given by

$$E(w|X,g) = E(w|X,g,I=1)\Pr(I=1|X,g) + E(w|X,g,I=0)\left[1-\Pr(I=1|X,g)\right], \quad (2)$$

where I is an indicator function that equals 1 if an individual is employed and zero otherwise and g =male, female. Wage gaps estimated on observed wage distributions are based on $E(w|X,g,I=1)$ alone. If there are systematic differences between $E(w|X,g,I=1)$ and $E(w|X,g,I=0)$, cross-country variation in $\Pr(I=1|X,g)$ may translate into misleading inferences concerning the international variation in the gender wage gap. This problem typically affects estimates of female wage equations; even more so when one is interested in cross-country comparisons of gender wage gaps, given the evidence described in the previous section on the cross-country variation in $\Pr(I=1|X,\text{male}) - \Pr(I=1|X,\text{female})$. Our goal is to retrieve gender gaps in potential (offer) wages, i.e. we seek a measure for (1), where $E(w|X,g)$ is given by (2). For this purpose, the data provide information on both $E(w|X,g,I=1)$ and $\Pr(I=1|X,g)$, but clearly not on $E(w|X,g,I=0)$, as wages are only observed for those who are in work.

A number of approaches can be used to correct for non-random sample selection in wage equations and/or recover the distribution in potential wages. The seminal approach suggested by Heckman (1974, 1979) consists in allowing for selection on unobservables, i.e. on variables that do not feature in the wage equation but that are observed in the data.[4] Heckman's two-stage parametric specifications have been used extensively in the literature in order to correct for selectivity bias in female wage equations. More recently, these have been criticized for lack of robustness and distributional assumptions (see Manski 1989). Approaches that circumvent most of the criticism include semi-parametric selection correction models à la Heckman (1980). Nonparametric methods allow in principle to approximate the bias term by a series expansion of propensity scores from the selection equation, with the qualification that the term of order zero in the polynomial is not separately identified from the constant term in the wage equation, unless some additional information is available. Usually, the constant term in the wage regression is identified from a subset of workers for which the probability of work is close to one (Buchinski 1998), but in our case this route is not feasible since for no type of women the probability of working is close to one in all

[4] In this framework, wages of employed and nonemployed would be recovered as

$$\begin{aligned} E(w|X,g,I=1) &= X\beta + E(\varepsilon_1|\varepsilon_0 > -Z\gamma) \\ E(w|X,g,I=0) &= X\beta + E(\varepsilon_1|\varepsilon_0 < -Z\gamma), \end{aligned}$$

respectively, where ε_1 and ε_0 are the error terms in the wage and the selection equation, and Z is the set of covariates used in the selection equation.

countries.[5]

Selection on observed characteristics is instead exploited in the matching approach, which consists in imputing wages for the nonemployed by assigning them the observed wages of the employed with matching characteristics (see Blau and Beller 1992 and Juhn 1992, 2003 for parametric and non parametric applications, respectively, to race gaps).

In this paper we follow the approach of Johnson et al. (2000) and Neal (2004), which is also based on some form of wage imputation for the nonemployed, and simply requires assumptions on the position of the imputed wage observations with respect to the median.[6] We then estimate median wage gaps on the resulting imputed wage distributions. The attractive feature of median regressions is that, if missing wage observations fall completely on one or the other side of the median regression line, the results would in this case only be affected by the position of wage observations with respect to the median, and not by specific values of imputed wages, as it would be in the matching approach. One can therefore make assumptions motivated by economic theory on whether an individual who is not in work should have a wage observation below or above median wages for their gender.

More formally, let's consider the linear wage equation

$$w_i = X_i\beta + \varepsilon_i, \tag{3}$$

where w_i denotes (log) wage offers, X_i denotes characteristics, with associated coefficients β, and ε_i is an error term such that $Med\left(\varepsilon_i|X_i\right) = 0$. Let's denote by $\hat{\beta}$ the hypothetical LAD estimator based on true wage offers. However, wage offers w_i are only observed for the employed, and missing for nonemployed. One can then define a transformed dependent variable y_i that is equal to w_i for the employed and to some arbitrarily low imputed value $\widetilde{w}_i$ for the nonemployed. If missing wage offers fall completely below the median regression line, i.e. $w_i < X_i\widehat{\beta}$ for the nonemployed, then the following result holds:

$$\hat{\beta}_{imputed} \equiv \arg\min_{\beta} \sum_{i=1}^{N} |y_i - X_i'\beta| = \hat{\beta} \equiv \arg\min_{\beta} \sum_{i=1}^{N} |w_i - X_i'\beta|. \tag{4}$$

(and viceversa for $w_i > X_i\widehat{\beta}$ for the nonemployed). Condition (4) states that the LAD estimator is not affected by imputation (see Johnson et al. 2000 for details).

It should be noted, however, that in order to use median regressions to evaluate gender wage gaps in (1) one should assume that the mean and the median of the (log) wage distribution coincide, in other words that the (log) wage distribution is symmetric. This is clearly true for the log-normal distribution, which is typically assumed in Mincerian wage equations.

[5] See Vella (1998) for an extensice survey of both parametric and non-parametric sample selection models.

[6] See also Chandra (2003) for a non-parametric application to racial wage gaps among US men.

In what follows we therefore assume that the distribution of offer wages is log-normal.[7]

Having said this, imputation can be performed in several ways. First, we exploit the panel nature of our data sets and, for all those not in work in some base year, we recover hourly wage observations from the nearest wave in the sample. The underlying identifiying assumptiom is that an individual's position with respect to the base-year median can be recovered looking at the level of her wage in the nearest wave. As the position with respect to the median is determined using levels of wages in other waves in the sample, we are allowing for selection on unobservables.

This procedure of imputation makes sense when an individual's position in the latent wage distribution stays on the same side of the median across adjacent waves in the panel. In other words, as we estimate the wage gap at the median, we do not need an assumption of stable rank throughout the whole wage distribution, but only with respect to the median.

While imputation based on this procedure arguably exploits the minimum set of potentially arbitrary assumptions, it has the disadvantage of not providing any wage information on individuals who never worked during the sample period. It is therefore important to understand in which direction this problem may distort, if at all, the resulting median wage gaps. If women are on average less attached to the labor market than men, and if individuals who are less attached have on average lower wage characteristics than the fully attached, then the difference between the median gender wage gap on the imputed and the actual wage distribution tends to be higher the higher the proportion of imputed wage observations in total nonemployment in the base year. Consider for example a country with very persistent employment status: those who do not work in the base year and are therefore less attached are less likely to work at all in the whole sample period. In this case low wage observations for the less attached are less likely to be recovered, and the estimated wage gap is likely to be lower. Proportions of imputed wage observations over the total nonemployed population in 1999 (our base year) are reported in Table A2: the differential between male and female proportions tends to be higher in Germany, Austria, France and southern Europe than elsewhere. Under reasonable assumptions we should therefore expect the difference between the median wage gap on the imputed and the actual wage distribution to be biased downward relatively more in this set of countries. This in turn means that we are being relatively more conservative in assessing the effect of non-random employment selection in these countries than elsewhere.

[7] If one does not impose symmetry of the (log) wage distribution, the equivalent of (2) would be

$$\begin{aligned} Med\,(w|X,g) &= F^{-1}(1/2) \\ &= F^{-1}\left\{F\left[Med\,(w|X,g,I=1)\right]\Pr(I=1|X,g)+F\left[Med\,(w|X,g,I=1)\right]\left[1-\Pr(I=1|X,g)\right]\right\} \end{aligned}$$

Even so, it would of course be preferable to recover wage observations also for those never observed in work during the whole sample period. To do this, we can recover wage observations for the nonemployed by making assumptions about whether they place above or below the median wage offer, based on their observable characteristics, specifically education, experience and spouse income. In this case we are allowing for selection on observables only. Having done this, earlier or later wage observations for those with imputed wages in the base year can shed light on the goodness of our imputation methods.

4 Results

4.1 Raw wage gaps

While wage imputation is supposed to enlarge our sample so as to include those with presumably lower labor market attachment, we also perform the reverse exercise and remove from our base sample of those employed at the time of interview those who were not employed for the whole year. Due to slightly different information available in different sources, full-year employees are identified in the PSID as those who report a number of annual hours worked at least equal to 1500, and in the ECHPS are those who are continuously employed during the 12 months preceding the survey date. Our most restricted sample is thus made by the full-year employed, then it is enlarged to include those employed at time of survey, and finally it is further enlarged to include those for whom we can impute a wage observation according to the two methods explained above.

The results are reported in Table 2. Column 2 reports raw (unadjusted) wage gaps for individuals with hourly wage observations in 1999, which is our base year. These replicate very closely the wage gaps reported in Table 1, with the only difference that mean wage gap for the whole sample period are reported in Table 1, while median wage gaps for 1999 are reported here. As in Table 1, the US and the UK stand out as the countries with the highest wage gaps, followed by central and northern Europe, and finally Scandinavia and Southern Europe. The column to the left restricts the sample to those employed during the full year in 1999, while the two columns to the right extend the sample to all those who have worked at some point during the whole sample period. In particular, in column 3 missing wage observation in 1999 are replaced with the real value of the nearest wage observation in a 2-year window, while in column 3 they are replaced with the real value of the nearest wage observation in the whole sample period, meaning a maximum window of [-4, +1] years for the US and [-5, +2] years for Europe - this last difference being due to different wave availability in the two data sets. Comparing figures in columns 1-4, one can see that the median wage gap remains substantially unaffected or affected very little in the US the UK and a number of European countries down to Austria, and increases substantially in Ireland, France and southern Europe, this latter groups including countries with the highest gender employment gap. While sample selection seems to be fairly neutral in a large number of countries in the sample, or, in other words, selection in market work does not seem to vary systematically with wage characteristics of individuals, it is mostly high-wage individuals who work in catholic countries, and this seems to give a downward biased estimate of the gender wage gap when one does not account for non-random sample selection. As one would expect from these results, controlling for selection removes most of the observed negative correlation between wage and employment gaps. The correlation coefficient between unadjusted median wage gaps and employment gaps is -0.517, and is significantly different from zero at the 6% level. Using the adjusted estimates from column 4 of Table 2, this falls to -0.118, and is not significantly different from zero at standard levels.

The estimates of columns 3 and 4 do not control for aggregate wage growth over time. If aggregate wage growth were homogeneous across genders and countries, then estimated wage gaps based on wage observations for adjacent years would not be not affected. But if there is a gender differential in wage growth, and if such differential varies across countries, then simply using past (future) wage observations would deliver a higher (lower) median wage gap in countries where women have relatively lower wage growth with respect to men.[8] We thus estimate real wage growth by regressing log real hourly wages for each country and gender on a linear trend.[9] The resulting coefficients are reported in Table A3. These are then used to adjust real wage observations outside the base year and re-estimate median wage gaps. The resulting median wage gaps on the imputed wage distribution are reported in column 5 and 6. Despite some differences in real wage growth rates across genders and countries, adjusting estimated median wage gaps does not produce any appreciable change in the results reported in columns 3 and 4, which do not control for real wage growth.

Note that in Table 2 we are (at best) recovering on average 24% of the nonemployed females in the four southern European countries, as opposed to approximately 46% in the rest of countries (see Table A2). For men, the respective proportions are 54% and 60%. Such differences happen because (non)employment status tends to be more persistent in southern Europe than elsewhere, much more so for women than for men. As briefly noted in Section 3, given that we recover relatively fewer less-attached women in southern Europe, we are being relatively more conservative in assessing the effect of non-random employment selection in southern Europe than elsewhere. For this reason is important to try to recover wage observations also for those never observed in work in any wave of the sample period,

[8]Note however that, even if real wage growth were homogeneous across genders, imputation based on wage observations from adjacent waves would not be affected only if the proportion of men and women in the sample remained unchanged after imputation.

[9]Of course, for our estimated rates of wage growth to be unbiased, this procedure requires that participation into employment be unaffected by wage growth, which may not be correct.

as explained below.

In Table 3 we estimate median wage gaps on imputed wage distributions, making assumptions on whether nonemployed individuals in 1999 had potential wage offers above or below the median for their gender. Colum1 reports for reference the median wage gap on the base sample, which is the same as the one reported in column 2 of Table 2. In column 2 we assume that all those not in work would have wage offers below the median for their gender. This is an extreme assumption, and can be taken as a benchmark. This assumption is clearly violated in cases like Italy, Spain and Greece, in which more than a half of the female sample is not in work in 1999, as by definition all missing observations cannot fall below the median. For this reason we do not report estimated gaps for these three countries. However, also for other countries there are reasons to believe that not all nonemployed individuals would have wage offers below their gender mean. Of course, we cannot know exactly what wages these individuals would have received if they had worked in 1999. However, we can form an idea of the goodness of this assumption looking again at wage observations (if any) for these individuals in all other waves of the panel. This allows us to compute what proportion of imputed observations had at some point in time wages that were indeed below their gender median. Such proportions are also presented separately for each gender in column 2. They are fairly high for men, but sensibly lower for women, which makes the estimates based on this extreme imputation case a benchmark rather than a plausible measure for the gender wage gap. Having said this, estimated median wage gaps increase substantially for most countries, except the UK and Scandinavia.

We next make imputations based on observed characteristics of nonemployed individuals. In column 3 we impute wage below the median to all those who are unemployed (as opposed to non participants) in 1999. With respect to the base sample, the implied median wage gap stays roughly unchanged everywhere down to Austria, and increases substantially in Ireland, France and southern Europe. Also, the proportion of "correctly" imputed observations, computed as for the previous imputation case, is now much higher. Those who do not work because they are unemployed are thus relatively more likely to be over-represented towards the bottom of the wage distribution. In column 4 we assume that all those with less than upper secondary education and less than 10 years of potential labor market experience have wage observations below the median for their gender. Those with at least higher education and at least 10 years of labor market experience are instead placed above the median. The change in the estimated wage gap is similar as in column 3, and so are the proportions of correctly imputed observations (except for some reason in the UK). The next imputation method is implicitly based on the assumption of assortative mating and consists in assigning wages below the median to those whose partner has total income in the bottom quartile of the gender-specific distribution of total income. The results are broadly similar to those in column 3: the wage gap is mostly affected in Ireland and southern Europe. It would be natural to perform a similar exercise at the top of the distribution, by assigning a wage above the median to those whose partner has total income in the top quartile. However, in this case the proportion of correctly imputed observations was too low to rely on the assumption used for imputation.

We finally use imputation based on characteristics to recover wage observations only for those who never worked, i.e. we first use wage observations available from other waves, and then we impute the remaining missing observations using education and experience as we did in column 3. The results show again a much higher gender gap in Ireland, France, and southern Europe, and not much of a change elsewhere with respect to the base sample of column 1.

One could argue that less restrictive sample inclusion rules are bound to affect the estimated wage gap less in countries where female employment rates are higher, simply because there is less room to enlarge the sample with imputation methods, and therefore including, say, individuals who are not working in 1999 but have been working at least once in the range of a few years would not substantially affect the sample size. But this is not completely true. Table 4 reports the total number of observations for each gender and country, and the fraction with actual or imputed wages under alternative sample inclusion rules.[10] Comparing columns labeled 1-4, corresponding to the full-year employed in 1999, those employed at the time of survey in 1999, and those employed in time windows of different length, one can see that the fraction of women included increases substantially in southern Europe, and only slightly less in countries like Germany or the UK, where the estimated wage sample is virtually unaffected by the sample inclusion rules. It is thus not simply the lower female employment rate in southern Europe than plays a role, it is also and mostly the fact that in several countries selection into work seems to be less correlated to wage characteristics than in others. This clearly affects our assessment of international variation in gender wage gaps.

To broadly summarize our evidence on unadjusted wage gaps, one could note that whether one corrects for selection unobservables (Table 2) or on unobservables (Table 3), our results were both qualitatively and quantitatively consistent in identifying a clear role of sample selection in Ireland, France and souther Europe. The fact that controlling for unobservables did not greatly change the picture obtained when controlling for a small number of observables alone (education, experience and spouse income) implies that most of the selection role can indeed be captured by a bunch of observable individual characteristics, and possibly some unobservables closely correlated to them.

[10] In column labelled 4 such proportions are generally not equal to 100% because we did not impute wages to those who are employed but have missing information on hourly wages. There is indeed no reason why this group should be placed below the median.

We have performed a number of robustness tests and more disaggregate analyses on the results obtained and reported in Tables 2 and 3. First, we have restricted the estimates of Tables 2 and 3 to individuals aged 25-54 in 1999, and the results were very similar to those obtained on the larger sample. Second, we have repeated our estimates separately for three education groups (less than upper secondary education, upper secondary education, and higher education), and we found that most of the selection occurs across rather than within groups, as median wage gaps disaggregated by education are much less affected by sample inclusion rules than in the aggregate. We will returns to the issue of selection of observables versus unobservable characteristics in the next subsection. Finally, we have repeated our estimates separately for three demographic groups: single individuals without kids in the household, married or cohabiting without kids, and married or cohabiting with kids. We found evidence of a strong selection effect in Ireland, France and southern Europe among those who are married or cohabiting, especially when they have kids, and much less evidence of selection going on among single individuals without kids.

4.2 Adjusted wage gaps

Our discussion so far referred to unadjusted wage gaps. In other words, imputation was based on whether an individual with certain education and experience characteristics should place below or above the median, conditional on gender. While similar imputation methods could in principle be used in estimating adjusted wage gaps, in practice one needs stronger assumptions in order to establish whether a missing wage observation should be placed above or below the median. For example, if the X vector contains, say, a gender dummy and human capital variables, then we should need to assume that those with missing wage and a certain level of education and experience place above or below the median, conditional on their gender and human capital levels. In order to avoid making such stronger assumptions, when estimating adjusted wage gaps we only impute wages based on wage observations in other waves in the sample. We report estimates obtained on three alternative samples: (i) those employed full-year in 1999; (ii) those employed at the time of survey in 1999; (iii) those employed at least once in the sample period. We do not report estimates for those employed at least once in a window of [-2,+2] years, as they do not provide extra relevant information from those based on those employed at least once in the sample period, nor we report estimates corrected for real wage growth, as they do not differ much from those at point (iii).

We estimate separate wage equations for males and females, controlling in each for education (less than upper secondary, upper secondary and higher education) experience (and its square), broad occupation groups (12 categories for the US and 9 categories for Europe), industry (12 categories for the US and 18 categories for Europe), public sector, and state or region dummies. The resulting average gender wage gap can be thus decomposed according to the well known Oaxaca (1973) decomposition into a component represented by gender differences in characteristics and gender differences in the returns to characteristics:

$$\overline{w}^M - \overline{w}^F = \left(\overline{X}^M - \overline{X}^F\right)\widehat{\beta}^M + \overline{X}^F\left(\widehat{\beta}^M - \widehat{\beta}^F\right) \quad (5)$$

where upper bars denote sample averages, hats denote OLS estimates and superscripts denote gender.

A similar decomposition can be also performed at different quantiles of the wage distribution. In this case, the decomposition has an extra term, representing the impact of unobservables, whose mean is non-zero in quantile regressions:

$$\overline{w}_q^M - \overline{w}_q^F = \left(\overline{X}_q^M - \overline{X}_q^F\right)\widehat{\beta}_q^M + \overline{X}_q^F\left(\widehat{\beta}_q^M - \widehat{\beta}_q^F\right) + \left(\overline{\varepsilon}_q^M - \overline{\varepsilon}_q^F\right) \quad (6)$$

where upper bars denote sample averages at quantile q, and hats denote estimates from quantile regressions.

This exercise is performed on both our base sample and a sample in which missing observations are imputed using the real hourly wage in other sample waves. We already know from Table 2 that extending the sample including rules delivers a substantially higher gender wage gap for some countries but not for others. The next set of results are going to tell whether the impact of sample selection (if any) on the gender wage gap is going to come mostly through characteristics or returns, i.e. whether in some countries women with lower labor market attachment have a higher wage penalty with respect to men because they have relatively poorer characteristics or they receive lower returns for a given set of characteristics.

The results of the Oaxaca decomposition are reported in Table 5. Belgium is excluded as the relatively small sample size left us with several empty cells in the estimation of adjusted wage gaps.

The raw wage gaps reported in Table 5 are not necessarily the same as those of Table 2, because of slightly smaller sample size in Table 5, having dropped observations with missing information on any of the right-hand side variables used. In all countries in the sample except the US the contribution of differences in coefficients is much more important than that of differences in characteristics. While this could be in part due to the limited set of X-variables included, we also estimated a specification that controlled for marital status and number of kids in age brackets 0-2, 3-5, 6-10, 11-15, and the split of the raw wage gap into characteristics' and coefficients' components was not greatly affected with respect to figures reported in Table 5.

Another feature to be noticed is that the contribution of characteristics is actually negative in most cases in southern Europe,[11] meaning that working women in these cases have higher wage characteristics than working men (and that differences between male and female coefficients explain more than 100% of the observed wage gap). This is a consequence of very low female employment rates in these countries, in the presence of selective participation into employment. One could also argue that it could be a consequence of the limited set of explanatory variables used, but when we repeated the same kind of Oaxaca decomposition having added marital status and number of dependent kids by age category among the set of explanatory variables, we obtained very similar results to those reported in Table 5.

As a comparison among the three panels of Table 5 shows, countries whose gender wage gap is not seriously affected by sample inclusion rules also have a roughly unchanged gap decomposition. In countries where sample inclusion rules indeed affect the estimated wage gap, it is both components that matter, although the change in the characteristics component seems in general more important than that in the returns component. In other words, in Ireland and southern Europe, women with lower labor market attachment have a higher wage penalty with respect to men mostly because they have relatively poorer characteristics than women with higher labor market attachment. This seems to confirm the importance of selection on observable rather than unobservable characteristics.

Tables 6-8 report the results of such decomposition at the 25th, 50th and 75th percentiles. In general quantile regressions show a more important role of gender differences in characteristics in explaining differences in the gender wage gap under alternative sample inclusion rules, especially at the 75th percentile. In particular, it seems that in Ireland and southern Europe women who are less likely to be employed have poorer characteristics than those who are more attached to the labor market. Concerning other countries, no major differences are found across sample inclusion rules in either the gender wage gap or its components.

5 Sample selection and wage dispersion

We have noticed in the previous sections that nonrandom selection into employment indeed matters for our assessment of the gender wage gap in a set of countries where the gender employment gap is relatively high. In particular, we showed that a number of corrections for sample selection explained part of the international variation in gender wage gaps. To date, the existing literature has mostly related such variation to international differences in overall wage inequality. Blau and Kahn (1996, 2003) argue that institutional differences across countries due, among other factors, to different degrees of unionization or different sizes of public sectors may be responsible for differences in overall levels of wage inequality. Higher wage inequality in turn translates into a higher gender wage gap, given a certain

[11] This is mostly the consequence of gender differences in average educational and occupational levels.

degree of dissimilarity between the characteristics of working men and women.

In order to compare the importance of sample selection versus overall inequality in explaining cross-country differences in the gender wage gap, we analyze such differences using a method initially proposed by Juhn et al. (1991) in order to study the slowdown in the convergence of black and white wages. Such method was first adapted to the study of cross-country differences in the gender wage gap by Blau and Kahn (1996).[12] It consists in decomposing the difference in the gender wage gap between two countries into differences in observed and unobserved characteristics of women compared to men, and differences in their respective returns.

To achieve this decomposition one estimates a male (log) wage equation for each country c:

$$w_{ic} = X_{ic}\beta_c + \theta_{ic}\sigma_c, \tag{7}$$

where θ_{ic} is the standardized male residual and σ_c is the standard deviation of male residuals, i.e. a measure of male residual wage inequality. While X_{ic} and θ_{ic} denote characteristics, observed and unobserved respectively, β_c and σ_c denote the associated prices. The difference in the gender pay gap between country A and country B can be thus decomposed into the following four terms:

$$\begin{aligned} D_A - D_B \;\; = \;\; & (\Delta X_A - \Delta X_B)\beta_A + \Delta X_B(\beta_A - \beta_B) \\ & + (\Delta\theta_A - \Delta\theta_B)\sigma_A + \Delta\theta_B(\sigma_A - \sigma_B), \end{aligned} \tag{8}$$

where $D_c \equiv \overline{w_{iA}} - \overline{w_{iB}}$ and Δ represents the difference in male-female averages in X_{ic} and θ_{ic}. The first term in (8) represents the contribution of country differences in gender differentials in observed characteristics, all evaluated at the male coefficients for country A (thus the reference country). The second term reflects the effect of differences in prices of such observed characteristics. The last two terms represent the impact of differences in unexplained gaps. In particular, the third term reflects country differences in gender differentials in unobserved characteristics. This is known as the "gap effect", and measures the effect of differences in the relative position of males and females in the male residual wage distribution. It is a sort of black-box term, which is supposed to capture the effect of differences in women's unmeasured characteristics with respect to men, but it is also consistent with differences in the extent of pay discrimination against women. Finally, the fourth term represents the impact of international differences in residual (male) wage inequality, given the relative position of men and women in the residual distribution.

Computation of the first two terms is straightforward, simply based on sample averages of included right-hand side variables and coefficients from male regressions. The second

[12] See Blau and Kahn (1997, 2004) for an application to trends in the US gender wage gap.

and third terms could be computed directly using the estimated values of σ_c, and then the sample averages of $\Delta\theta_c$, exploiting the assumption of normal disturbances. However, such an assumption is not necessary, and is typically not used in applications of the Juhn et al (1991) decomposition, if one uses the entire distribution of estimated residuals. Specifically, the $\Delta\theta_c\sigma_c$ are simply equal to minus the average female residual, evaluated at male coefficients (the average male coefficient being zero). For the $\Delta\theta_B\sigma_A$ term one needs to compute what the mean country B female residuals would be if the standard deviation of residuals were that of country A (again for men the mean is zero). Thus we assign each woman in country B a percentile in the country B male residual distribution, based on her residual. Then she is assigned the residual that corresponds to that percentile in country A.

This exercise is similar in spirit to the one performed in Tables 5-8 using the traditional Oaxaca (1973) decomposition. The decomposition in (8) is based on the coefficients obtained from male wage regressions only, implicitly assuming that female coefficients would be the same in the absence of discrimination or misspecification due to non-random selection into work. In principle it has the advantage of separately identifying the contribution of differences in overall wage inequality from that of differences in characteristics (observed and unobserved) in the international variation in gender wage gaps.

We implement decompositions as in (8) for pairs of countries in our sample. As the specification used for the male wage equation has to be identical within each pair, we need to drop the US from the sample, as the industry and occupation classification in the PSID does not exactly mirror the one available in the ECHPS, plus the definition of experience is also somewhat different. We therefore take the UK as our reference country and perform bilateral comparisons between the UK gender wage gap and that of each other EU country. In the notation of (8), country A is the UK. The X vector includes controls for education, experience, occupation, sector, public sector, part time work and temporary contract. Regional dummies are not included here, again for the need of an identical wage equation specification across countries.

We perform our decompositions on two samples. The first is the base-sample, including observed wages in the 1999 wave of the ECHPS. The second also includes imputed wage observations, recovered using wage information from other waves in the sample. Where sample selection matters, we would expect the impact of at least one of the components to decrease, as the total differential $D_{UK} - D_c$ is presumably reduced. Moreoever, we expect this change to be stronger for countries where sample inclusion rules make a bigger difference.

The results of the two decompositions are reported in Tables 10 and 11. Table 10 is based on the base 1999 sample. The first column is always positive, as it reports the difference in the wage gap between the UK and that of each other country. Column 6 reports the contribution of the differences in both observable (column 2) and unobservable (column 4) characteristics, while column 7 reports the contribution of differences in the wage structure, in turn given by the sum of the contribution of differences in observed (column 3) and unobserved (column 5) prices.

The wage structure component is everywhere positive, meaning that the UK has the most unequal wage structure in our sample. Not surprisingly, this term is highest for Scandinavian countries. Wage structure differences by themselves would explain even more than the actual difference in wage gaps $D_{UK} - D_c$ for all northern and central European countries (except Netherlands) than the actual one. Hence, the characteristics component is negative, implying that the average characteristics of working women relative to men are worse in these countries than in the UK. A different pattern emerges in Netherlands and southern European countries. There, the wage structure component is also positive, but the difference with respect to the top set of countries is that the characteristics component becomes positive, implying that the average characteristics of working women relative to men are better in these countries than in the UK. This is not surprising given the descriptive evidence of subsection 2.3 and the results presented in section 4.

Note however that this decomposition is not robust to the specific set of explanatory variables used in the wage regression. In particular, when dummies for part time and temporary work where not included, we found that the decomposition for the Netherlands became similar to that of other northern European countries, i.e. the contribution of the characteristics component became negative. Moreover, the contribution of characteristics in France and Ireland became positive, although with much smaller magnitude than in southern Europe.

Table 11 reports the same decomposition based on the sample that includes imputed wage observations from other waves. The raw wage gap decreases mostly in central and, even more, southern Europe. The characteristics component tends to fall in most countries, with the exception of Scandinavia and Greece. This means that among those with weaker labor market attachment the gender wage gap in characteristics is higher in most countries than in the UK. Second, the wage structure component tends to fall in Scandinavia, France, Spain and especially Greece (being roughly unaffected elsewhere). This means returns to characteristics among low-attached men in these countries tend to be lower than in the UK.

6 Conclusions

In this paper we show the importance of non random selection into work in understanding the observed international variation in gender wage gaps. To do this, we performed wage imputation for those not in work, by simply making assumptions on the position of the imputed wage observations with respect to the median. We then estimated median wage gaps on imputed wage distributions, and assessed the impact of selection into work by comparing estimated wage gaps on the base sample with those obtained on a sample enlarged with wage imputation. Our estimates delivered higher median wage gaps on imputed rather than actual wage distributions for most countries in the sample, meaning that, as one would have expected, women tend on average to be more positively selected into work than men. However, this difference is negligible in the US, northern European countries (except Ireland) and most central European countries, it becomes sizeable in Ireland, France and southern Europe, i.e. countries in which the gender employment gap is highest. In particular, in Spain, Portugal and Greece the median wage gap on the imputed wage distribution reaches nearly 20 log points, a closely comparable level to that of the UK and of other central and northern European countries.

This analysis has identified a number of direction for future work. First, we may investigate alternative methodologies for imputing wages based on observable characteristics and/or self-reported reservation wages. This is important since there is a very large fraction of women who are never observed in work in our sample. Second, we plan to include the self-employed into the main picture of gender wage gaps. There is a large cross-gender and cross-country variation in the share of self-employment, which tends to be higher among men and in southern European economies. How our results would be affected by the inclusion of self-employed work depends on the overall inequality differential between self-employed and dependent work, and on the cross-country variation of gender gaps in self-employment. Finally, as we argued that gender employment gaps are important in understanding cross-country differences in gender wage gaps, one should ultimately assess the importance of demand and supply factor in explaining variation in these gaps.

References

[1] Albrecht, J., A. van Vuuren and S. Vroman (2004), "Decomposing the Gender Wage Gap in the Netherlands with Sample Selection Adjustment", IZA DP No. 1400.

[2] Altonji, J. and R. Blank (1999), "Race and Gender in the Labor Market", in O. Ashenfelter and D. Card (eds.) *Handbook of Labor Economics*, North-Holland, volume 3C: 3141-3259.

[3] Azmat, G., M. Güell and A. Manning (2004), ""Gender Gaps in Unemployment Rates in OECD Countries", CEPR DP No 4307.

[4] Beblo, M., D. Beninger, A. Heinze and F. Laisney (2003), "Measuring Selectivity-corrected Gender Wage Gaps in the EU", ZEW DP No. 03-74, Mannheim.

[5] Blau, F. and A. H. Beller (1992), "Black-White Earnings over the 1970s and 1980s: Gender Differences in Trends", *Review of Economics and Statistics* 72(2): 276-286.

[6] Blau, F. and L. Kahn (1996), "Wage Structure and Gender Earnings Differentials: An International Comparison", *Economica* 63, S29-S62.

[7] Blau, F. and L. Kahn (1997), "Swimming Upstream: Trends in the Gender Wage Differentials in the 1980s", *Journal of Labor Economics* 15, 1-42.

[8] Blau, F. and L. Kahn (2004), "The US gender Pay Gap in the 1990s: Slowing Convergence?", NBER WP No 10853.

[9] Blundell, R., A. Gosling, H. Ichimura and C. Meghir (2004), "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds", CEPR DP No. 4705.

[10] Buchinsky, M. (1998), "The Dynamics of Changes in the Female Wage Distribution in the USA: A Quantile Regression Approach", *Journal of Applied Econometrics* 13, 1-30.

[11] Chandra, A. (2003), "Is the Convergence in the Racial Wage Gap Illusory?", NBER WP No. 9476.

[12] Dinardo, J., N. Fortin and T. Lemieux (1996), "Labor Market Institutions and the Distribution of wages, 1973-1992: A semiparametric Approach", *Econometrica* 64(5): 1001-1044.

[13] Heckman, J. (1974), "Shadow Prices, Gender Differenced and Labor Supply", *Econometrica* 42, 679-694.

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

[15] Heckman, J. (1980), "Addendum to Sample Selection Bias as a Specification Error". In E. Stromsdorfer and G Ferkas (eds.) *Evaluation Studies*. San Francisco: Sage, Volume 5.

[16] Johnson, W., Y. Kitamura and D. Neal (2000), "Evaluating a Simple Method for Estimating Black-White Gaps in Median Wages", *American Economic Review* 90, 339-343.

[17] Juhn, C. (1992), "Decline of Labor Market Participation: The Role of Declining Market Opportunities", Quarterly Journal of Economics 107, 79-122.

[18] Junh, C. (2003), "Labor Market Dropouts and Trends in the Wages of Black and White Men", *Industrial and Labor Relations Review* 56(4), 643-662.

[19] Juhn, C., K. Murphy and B. Pierce (1991), "Accounting for the Slowdown in Black-white Wage Convergence." In *Workers and Their Wages*, by M. Kosters (ed.), 107–43. Washington, DC: AEI Press, 1991.

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

[21] Manski, C. F. (1989), "Anatomy of the Selection Problem", *Journal of Human Resources* 24, 343-360.

[22] Mulligan, C. and Y. Rubinstein (2004), "The Closing of the Gender Gap as a Roy Model Illusion", NBER WP No. 10892.

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

[24] Neal, D. (2004), "The Measured Black-white Wage Gap Among Women is Too Small", *Journal of Political Economy* 112, S1-S28.

[25] OECD (2002), *Employment Outlook*, Paris.

[26] Roy, A.D. (1951), "Some Thoughts on the Distribution of Earnings", *Oxford Economic Papers* 3, 135-146.

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

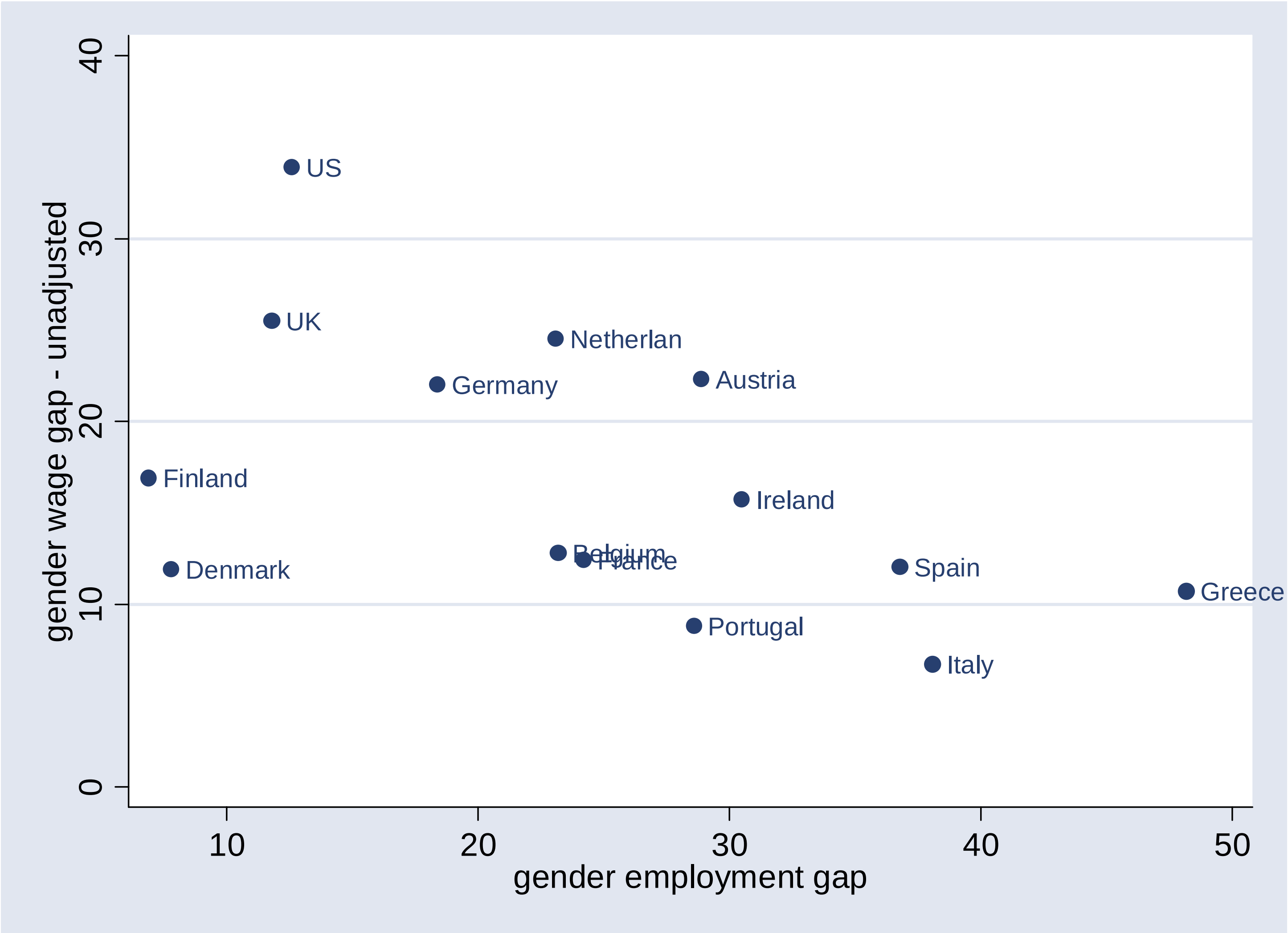

Figure 1: Gender gaps (%) in median hourly wages and in employment gaps (%).

Table 1
Raw (mean) wage and employment gaps, 1994-2001
Aged 16-64

	Wage gaps by highest qualification				Employment gaps by highest qualification			
Country	Total	Low	Medium	High	Total	Low	Medium	High
US	30.2	29.6	31.0	39.4	12.6	22.1	13.8	9.2
UK	27.0	24.5	22.2	25.0	11.8	12.2	10.2	8.5
Sweden	-	-	-	-	5.2	10.8	5.2	3.2
Finland	17.8	17.7	17.5	27.8	6.9	5.8	8.7	8.1
Denmark	10.8	8.0	10.1	16.8	7.8	17.5	6.7	3.0
Germany	23.8	15.5	21.4	25.3	18.4	23.2	17.5	8.5
Netherlands	24.2	23.7	23.5	27.7	23.1	23.2	26.0	12.5
Belgium	12.1	20.1	14.3	15.4	23.2	38.7	26.8	6.7
Austria	22.3	10.4	23.5	26.3	28.9	39.6	24.3	10.5
Ireland	15.1	29.4	15.9	10.4	30.5	36.6	29.8	13.6
France	14.3	17.8	15.7	17.9	24.2	32.3	21.5	11.6
Italy	6.3	15.9	5.6	9.5	38.1	49.8	24.7	14.1
Spain	13.4	24.2	21.2	15.0	36.8	43.8	29.0	16.9
Portugal	9.8	22.7	15.8	8.0	28.6	34.7	9.0	2.0
Greece	12.0	20.9	18.2	12.6	48.2	58.8	42.4	22.1

Notes
1. The sample includes individuals aged 16-64, excluding the self-employed, the military and those in full-time education.
2. Definitions. Low qualification: less than upper secondary education. Medium qualification: upper secondary education. High qualification: higher education.
3. Source: PSID (1994-2001) and ECHPS (1994-2001). Wage data are not available for Sweden.

Table 1A
Raw (mean) wage and employment gaps, 1994-2001
Aged 25-54

	Wage gaps by highest qualification				Employment gaps by highest qualification			
Country	Total	Low	Medium	High	Total	Low	Medium	High
US	31.7	30.9	30.6	35.9	13.4	27.31	14.22	10.16
UK	30.5	30.4	26.8	24.0	13.5	13.8	12.2	9.5
Sweden	-	-	-	-	5.8	13.2	6.0	3.4
Finland	18.4	19.7	17.6	27.0	7.5	4.4	10.1	8.8
Denmark	11.2	12.1	9.6	15.6	7.1	17.4	6.6	2.9
Germany	24.0	28.3	20.3	23.9	18.5	25.1	17.7	9.4
Netherlands	23.9	24.0	22.6	27.0	24.5	24.6	28.1	13.8
Belgium	10.9	20.0	13.7	13.4	20.8	36.3	26.1	6.4
Austria	22.5	25.8	20.9	25.1	26.8	35.7	24.1	11.5
Ireland	17.9	35.2	19.5	5.1	28.9	32.9	31.2	13.2
France	14.2	19.1	15.7	16.9	22.6	29.9	21.7	11.3
Italy	5.7	16.5	5.0	7.1	37.9	51.1	26.4	13.9
Spain	11.6	23.1	21.1	12.4	37.9	46.9	32.5	17.3
Portugal	11.8	26.4	15.4	6.1	26.5	33.0	9.2	2.2
Greece	9.6	21.6	15.3	7.2	46.5	58.6	44.6	20.6

Notes

1. The sample includes individuals aged 25-54, excluding the self-employed, the military and those in full-time education.
2. Definitions. Low qualification: less than upper secondary education. Medium qualification: upper secondary education. High qualification: higher education.
3. Source: PSID (1994-2001) and ECHPS (1994-2001). Wage data are not available for Sweden.

Table 2
Raw (median) wage gaps, 1999, under alternative sample inclusion rules

	1	2	3	4	5	6
US	0.287	0.339	0.359	0.371	0.361	0.374
UK	0.244	0.255	0.252	0.259	0.271	0.276
Finland	0.159	0.169	0.149	0.149	0.158	0.158
Denmark	0.074	0.119	0.095	0.095	0.086	0.086
Germany	0.195	0.220	0.236	0.232	0.247	0.244
Netherlands	-	0.245	0.215	0.220	0.218	0.225
Belgium	0.077	0.128	0.106	0.115	0.105	0.115
Austria	0.213	0.223	0.239	0.238	0.235	0.235
Ireland	0.178	0.157	0.256	0.260	0.272	0.279
France	0.108	0.124	0.144	0.158	0.152	0.168
Italy	0.057	0.067	0.060	0.073	0.070	0.081
Spain	0.070	0.120	0.170	0.184	0.161	0.171
Portugal	0.128	0.088	0.175	0.180	0.183	0.200
Greece	0.135	0.107	0.194	0.212	0.197	0.196

Notes. All estimates are significant at the 1% level. Sample: aged 16-64, excluding the self-employed, the military and those in full-time education. Sweden is excluded as no wage information is available in any wave. Source: PSID and ECHPS.
Sample inclusion rules by columns:

1. Employed full year in 1999
2. Employed at time of survey in 1999
3. Wage imputed from other waves when missing (-2,+1 for US; -2,+2 for EU)
4. Wage imputed from other waves when missing (-4,+1 for US; -5,+2 for EU)
5. Wage imputed from other waves when missing (-2,+1 for US; -2,+2 for EU), adjusted for real wage growth
6. Wage imputed from other waves when missing (-4,+1 for US; -5,+2 for EU), adjusted for real wage growth

Table 3
Raw (median) wage gaps, 1999, under alternative imputation rules

	1	2			3			4			5			6
	Wage gap	Wage gap	Goodness imputation		Wage gap	Goodness imputation		Wage gap	Goodness imputation		Wage gap	Goodness imputation		Wage gap
			M	F		M	F		M	F		M	F	
US	0.339	0.455	0.81	0.71	0.340	1.00	0.90	0.350	0.70	0.78	0.355	0.63	0.86	0.376
UK	0.255	0.354	0.77	0.59	0.221	0.80	0.78	0.214	0.52	0.46	0.248	0.78	0.76	0.249
Finland	0.169	0.163	0.78	0.71	0.120	0.78	0.81	0.126	0.50	0.44	0.147	0.88	0.78	0.149
Denmark	0.119	0.105	0.67	0.75	0.078	0.73	0.75	0.079	0.88	0.59	0.100	0.88	0.63	0.095
Germany	0.220	0.403	0.72	0.47	0.239	0.74	0.67	0.218	0.64	0.65	0.241	0.67	0.77	0.232
Netherlands	0.245	0.422	0.45	0.43	0.257	0.65	0.59	0.202	0.75	0.69	0.216	0.45	0.73	0.217
Belgium	0.128	0.267	0.72	0.66	0.143	0.79	0.75	0.085	0.70	0.50	0.111	0.70	0.94	0.108
Austria	0.223	0.438	0.71	0.48	0.222	0.71	0.74	0.213	1.00	0.76	0.250	0.73	0.75	0.239
Ireland	0.157	0.718	0.82	0.18	0.217	0.86	0.71	0.217	0.84	0.74	0.267	0.70	0.91	0.254
France	0.124	0.442	0.76	0.38	0.140	0.81	0.81	0.073	0.54	0.59	0.123	0.75	0.90	0.145
Italy	0.067	-	0.69	-	0.115	0.73	0.66	0.063	0.92	0.77	0.141	0.70	0.87	0.075
Spain	0.120	-	0.59	-	0.205	0.74	0.60	0.103	0.77	0.68	0.159	0.52	0.90	0.170
Portugal	0.088	0.377	0.59	0.43	0.182	0.59	0.63	0.178	0.81	0.74	0.187	0.63	0.55	0.194
Greece	0.107	-	0.75	-	0.240	0.75	0.66	0.174	0.68	0.79	0.281	0.73	0.61	0.239

Notes. All estimates are significant at the 1% level. In specification 2 no results are reported for Italy, Spain and Greece as more than 50% of women in the sample have missing wages. Sample: aged 16-64, excluding the self-employed, the military and those in full-time education. Sweden is excluded as no wage information is available in any wave; Luxembourg is excluded as no wage information is available after wave 3.
Source: PSID and ECHPS.
Sample inclusion rules by columns:

1. Employed at time of survey in 1999;
2. Impute wage<median(wage|gender) if wage is missing;
3. Impute wage<median(wage|gender) if wage is missing and & individual is unemployed;
4. Impute wage<median(wage|gender) if wage is missing & education<upper secondary & experience<10;
   Impute wage>median(wage|gender) if wage is missing & education >=higher ed. & experience>=10;
5. Impute wage<median(wage|gender) if wage is missing & spouse income in bottom quartile;
6. Wage imputed from other waves when missing (-4,+1 for US; -5,+2 for EU) and (3).

Table 4
Percentage of adult population in samples for Tables 2 and 3:

	No. obs. in 1999		1 (%)		2 (%)		3 (%)		4 (%)		5 (%)		6(%)		7(%)		8(%)		9(%)	
	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F
US	3386	4301	86.5	59.7	94.8	81.8	97.4	90.0	97.7	91.2	100.0	100.0	95.3	82.6	96.2	87.9	96.1	85.8	97.9	92.8
UK	2694	3293	65.9	54.9	84.6	74.2	90.8	84.1	91.9	86.9	96.7	97.1	89.5	76.4	88.7	82.0	87.6	77.0	94.2	90.4
Finland	1886	2154	53.4	45.4	89.2	80.4	94.4	90.6	95.0	91.3	99.0	98.5	98.3	90.8	90.3	84.3	90.1	81.4	95.6	93.1
Denmark	1282	1338	72.9	65.2	93.1	86.5	98.8	95.1	99.0	95.9	98.0	98.1	97.0	92.6	94.0	89.2	93.8	87.5	99.2	96.6
Germany	3743	4034	71.3	50.6	88.2	67.4	95.8	81.0	97.7	85.1	98.5	94.0	96.8	75.0	89.8	70.3	90.4	68.7	98.0	86.2
Netherlands	2990	3476	-	-	87.1	64.7	91.5	75.2	93.2	78.0	99.7	99.2	90.2	75.1	88.4	69.6	92.0	69.2	93.6	79.4
Belgium	1364	1634	72.1	49.9	88.0	65.9	92.2	73.3	93.2	76.7	98.8	98.3	94.9	76.9	89.8	70.6	91.6	71.8	94.2	79.8
Austria	1756	1881	70.7	43.8	94.6	65.3	98.1	73.9	98.4	76.4	99.7	97.9	99.0	68.8	95.2	67.0	95.4	67.9	98.6	77.1
Ireland	1586	1979	53.7	32.3	84.2	55.1	89.7	66.3	90.6	69.1	99.6	99.1	92.6	58.6	85.8	58.8	87.8	60.7	91.0	71.5
France	3067	3557	57.7	42.5	71.2	52.1	90.8	71.3	92.5	75.6	86.2	90.8	79.0	62.5	74.9	59.0	73.4	53.6	93.9	79.0
Italy	3952	4903	57.3	30.0	74.7	40.3	86.7	49.5	87.9	52.2	94.9	97.2	91.2	52.8	77.7	44.9	77.3	49.2	89.4	55.1
Spain	3648	4289	55.6	24.9	78.0	40.7	88.1	53.7	90.0	56.9	99.6	99.6	90.5	51.8	81.7	48.7	83.0	42.1	91.5	61.4
Portugal	2916	3294	70.5	48.6	88.4	61.6	94.0	70.6	95.0	73.3	99.3	98.8	93.9	68.7	89.9	66.2	90.4	66.2	95.3	75.1
Greece	1812	2746	62.3	23.8	81.8	32.7	90.6	43.0	91.4	45.7	99.8	99.3	93.7	43.2	84.9	40.3	83.9	41.3	92.6	50.9

Notes. Sample: aged 16-64, excluding the self-employed, the military and those in full-time education. Sweden is excluded as no wage information is available in any wave; Luxembourg is excluded as no wage information is available after wave 3. Source: PSID and ECHPS.
Sample inclusion rules by column:

1. Employed full year in 1999;
2. Employed at time of survey in 1999;
3. Wage imputed from other waves when missing (-2,+1 for US; -2,+2 for EU);
4. Wage imputed from other waves when missing (-4,+1 for US; -5,+2 for EU);
5. Impute wage<median(wage|gender) if wage is missing;
6. Impute wage<median(wage|gender) if wage is missing and & individual is unemployed;
7. Impute wage<median(wage|gender) if wage is missing & education<upper secondary & experience<10;
   Impute wage>median(wage|gender) if wage is missing & education >=higher ed. & experience>=10;
8. Impute wage<median(wage|gender) if wage is missing & spouse income in bottom quartile;
9. (4) and (6).

Table 5
Adjusted wage gaps, 1999, decompositions at the mean
Under alternative sample inclusion rules

	Employed full year in 1999			Employed in 1999 at time of survey			Wage imputed from other waves when missing		
	raw	chars.	coefs.	raw	chars.	coefs.	raw	chars.	coefs.
USA	0.273	0.099	0.174	0.302	0.118	0.184	0.303	0.119	0.184
UK	0.258	0.089	0.169	0.245	0.092	0.153	0.247	0.094	0.152
Finland	0.140	0.005	0.136	0.161	0.039	0.121	0.174	0.074	0.101
Denmark	0.125	0.044	0.081	0.118	0.034	0.084	0.134	0.039	0.095
Germany	0.197	0.065	0.132	0.217	0.072	0.144	0.218	0.071	0.146
Netherlands	-	-	-	0.202	0.050	0.152	0.213	0.057	0.157
Austria	0.216	0.073	0.143	0.225	0.067	0.158	0.249	0.075	0.175
Ireland	0.146	-0.013	0.159	0.148	0.025	0.124	0.179	0.045	0.134
France	0.128	0.052	0.076	0.108	0.044	0.064	0.155	0.060	0.095
Italy	0.056	-0.058	0.114	0.063	-0.056	0.118	0.082	-0.041	0.124
Spain	0.083	-0.049	0.132	0.124	-0.010	0.134	0.188	0.036	0.151
Portugal	0.072	-0.060	0.132	0.086	-0.051	0.137	0.125	-0.017	0.141
Greece	0.092	-0.007	0.098	0.088	-0.015	0.103	0.160	0.056	0.104

Notes.
1. Characteristics included are: regional or state dummies, education dummies, experience and its square, occupation and industry dummies, public sector.
2. Sample: aged 16-64, excluding the self-employed, the military and those in full-time education. Sweden is excluded as no wage information is available in any wave. Belgium is excluded due to small sample size. Source: PSID and ECHPS.

Table 6
Adjusted wage gaps, 1999, decompositions at 25th percentile
Under alternative sample inclusion rules

	Employed full year in 1999				Employed in 1999 at time of survey				Wage imputed from other waves when missing			
	raw	chars.	coefs.	unobs.	raw	chars.	coefs.	unobs.	raw	chars.	coefs.	unobs.
USA	0.273	0.115	0.162	-0.004	0.324	0.130	0.191	0.003	0.342	0.122	0.215	0.005
UK	0.243	0.095	0.168	-0.020	0.240	0.112	0.145	-0.017	0.246	0.109	0.130	0.007
Finland	0.091	0.035	0.060	-0.004	0.102	0.034	0.103	-0.035	0.118	0.010	0.117	-0.009
Denmark	0.089	-0.016	0.084	0.021	0.074	-0.029	0.115	-0.012	0.103	0.033	0.093	-0.023
Germany	0.197	0.058	0.138	0.001	0.239	0.047	0.155	0.037	0.230	0.051	0.164	0.015
Netherlands	-	-	-	-	0.184	0.054	0.155	-0.025	0.196	0.064	0.149	-0.016
Austria	0.236	0.114	0.136	-0.014	0.267	0.156	0.122	-0.011	0.265	0.172	0.135	-0.042
Ireland	0.152	-0.033	0.152	0.034	0.141	0.055	0.084	0.002	0.160	0.059	0.102	-0.002
France	0.118	0.064	0.059	-0.005	0.095	0.047	0.065	-0.017	0.145	0.000	0.112	0.033
Italy	0.063	-0.032	0.104	-0.010	0.053	-0.039	0.118	-0.027	0.058	-0.034	0.145	-0.053
Spain	0.129	-0.024	0.118	0.035	0.167	0.028	0.160	-0.020	0.222	0.083	0.170	-0.030
Portugal	0.150	0.086	0.108	-0.045	0.146	0.061	0.100	-0.015	0.172	0.100	0.104	-0.032
Greece	0.102	0.084	0.048	-0.031	0.101	0.103	0.045	-0.047	0.151	0.092	0.071	-0.012

Notes.

1. Characteristics included are: regional or state dummies, education dummies, experience and its square, occupation and industry dummies, public sector.
2. Sample: aged 16-64, excluding the self-employed, the military and those in full-time education. Sweden is excluded as no wage information is available in any wave. Belgium is excluded due to small sample size. Source: PSID and ECHPS.

Table 7
Adjusted wage gaps, 1999, decompositions at 50th percentile
Under alternative sample inclusion rules

	Employed full year in 1999				Employed in 1999 at time of survey				Wage imputed from other waves when missing			
	raw	chars.	coefs.	unobs.	raw	chars.	coefs.	unobs.	raw	chars.	coefs.	unobs.
USA	0.275	0.101	0.133	0.042	0.311	0.129	0.156	0.025	0.322	0.136	0.169	0.017
UK	0.229	0.108	0.139	-0.018	0.224	0.092	0.123	0.009	0.233	0.093	0.115	0.025
Finland	0.116	-0.021	0.120	0.016	0.114	0.031	0.094	-0.011	0.127	0.063	0.099	-0.034
Denmark	0.082	-0.014	0.063	0.033	0.081	-0.007	0.060	0.029	0.093	-0.004	0.063	0.034
Germany	0.187	0.071	0.138	-0.022	0.194	0.054	0.142	-0.003	0.198	0.056	0.139	0.002
Netherlands	-	-	-	-	0.157	0.026	0.151	-0.020	0.167	0.029	0.150	-0.011
Austria	0.216	0.061	0.153	0.002	0.218	0.082	0.155	-0.019	0.230	0.100	0.158	-0.028
Ireland	0.188	0.069	0.138	-0.019	0.213	0.046	0.136	0.031	0.237	0.085	0.113	0.039
France	0.096	-0.026	0.083	0.038	0.083	0.019	0.054	0.010	0.131	0.039	0.113	-0.021
Italy	0.061	-0.062	0.101	0.022	0.049	-0.076	0.104	0.021	0.075	-0.045	0.112	0.009
Spain	0.068	-0.016	0.142	-0.058	0.116	0.008	0.155	-0.046	0.185	0.072	0.160	-0.047
Portugal	0.132	0.010	0.137	-0.015	0.161	0.026	0.132	0.003	0.178	0.072	0.132	-0.026
Greece	0.133	0.051	0.052	0.029	0.112	0.052	0.058	0.002	0.192	0.186	0.046	-0.040

Notes.

1. Characteristics included are: regional or state dummies, education dummies, experience and its square, occupation and industry dummies, public sector.
2. Sample: aged 16-64, excluding the self-employed, the military and those in full-time education. Sweden is excluded as no wage information is available in any wave. Belgium is excluded due to small sample size. Source: PSID and ECHPS.

Table 8
Adjusted wage gaps, 1999, decompositions at 75th percentile
Under alternative sample inclusion rules

	Employed full year in 1999				Employed in 1999 at time of survey				Wage imputed from other waves when missing			
	raw	chars.	coefs.	unobs.	raw	chars.	coefs.	unobs.	raw	chars.	coefs.	unobs.
USA	0.283	0.107	0.188	-0.012	0.289	0.113	0.189	-0.013	0.295	0.114	0.191	-0.009
UK	0.268	0.154	0.117	-0.004	0.258	0.129	0.103	0.026	0.256	0.141	0.113	0.002
Finland	0.183	-0.019	0.173	0.029	0.192	-0.006	0.171	0.027	0.219	0.042	0.167	0.010
Denmark	0.169	0.101	0.077	-0.009	0.163	0.107	0.057	0.000	0.156	0.058	0.064	0.035
Germany	0.205	0.061	0.136	0.007	0.215	0.065	0.139	0.011	0.218	0.084	0.141	-0.007
Netherlands	-	-	-	-	0.211	0.058	0.182	-0.029	0.216	0.071	0.171	-0.026
Austria	0.196	0.069	0.132	-0.005	0.209	0.062	0.138	0.009	0.231	0.085	0.133	0.013
Ireland	0.184	-0.015	0.178	0.021	0.187	0.014	0.164	0.009	0.225	0.032	0.172	0.020
France	0.120	0.049	0.069	0.002	0.103	-0.003	0.088	0.018	0.138	0.029	0.114	-0.005
Italy	0.038	-0.079	0.083	0.034	0.044	-0.087	0.081	0.050	0.054	-0.072	0.107	0.019
Spain	0.010	-0.218	0.145	0.084	0.067	-0.102	0.130	0.039	0.161	-0.019	0.168	0.012
Portugal	-0.032	-0.342	0.101	0.208	0.004	-0.270	0.117	0.156	0.062	-0.202	0.149	0.115
Greece	0.082	-0.059	0.099	0.042	0.064	-0.061	0.087	0.037	0.163	0.027	0.105	0.032

Notes.

1. Characteristics included are: regional or state dummies, education dummies, experience and its square, occupation and industry dummies, public sector.
2. Sample: aged 16-64, excluding the self-employed, the military and those in full-time education. Sweden is excluded as no wage information is available in any wave. Belgium is excluded due to small sample size. Source: PSID and ECHPS.

Table 9
Sample sizes for Tables 5-8

	Employed full year		Employed at time of survey		Wage imputed from other waves when missing	
	M	F	M	F	M	F
USA	2630	2220	2808	2872	2860	3027
UK	1677	1628	2120	2131	2295	2430
Finland	513	470	941	922	1026	1098
Denmark	627	615	716	711	775	803
Germany	2232	1597	2669	2037	2971	2531
Netherlands	-	-	2472	1805	2617	2018
Austria	1220	798	1624	1159	1685	1309
Ireland	732	553	1135	860	1216	1041
France	1651	1417	2031	1731	2631	2462
Italy	2099	1391	2719	1824	3160	2279
Spain	1949	1042	2725	1631	3133	2175
Portugal	1971	1540	2443	1904	2617	2224
Greece	1064	636	1391	851	1539	1161

Table 10:
JMP (1991) decomposition of the difference between the gender wage gap in the UK and in each other EU country
Based on base 1999 sample

	1 D_{UK}-D_c	2 Observed characteristics	3 Observed prices	4 Gap effect	5 Unobserved prices	6 Total charact.	7 Wage structure	8 No. obs males	9 No. obs females
Finland	0.092	0.027	0.044	-0.085	0.106	-0.058	0.150	932	900
Denmark	0.127	-0.036	0.057	-0.011	0.116	-0.047	0.173	700	697
Germany	0.031	0.024	0.025	-0.123	0.106	-0.099	0.131	2521	1904
Netherlands	0.050	0.022	-0.096	0.023	0.101	0.045	0.005	2424	1761
Austria	0.012	0.027	-0.023	-0.102	0.109	-0.075	0.086	1541	1103
Ireland	0.064	0.028	-0.015	-0.044	0.095	-0.016	0.080	1203	934
France	0.117	-0.005	0.024	-0.005	0.103	-0.010	0.127	1937	1654
Italy	0.192	0.111	0.035	-0.057	0.104	0.054	0.139	2663	1759
Spain	0.120	0.108	0.000	-0.081	0.092	0.027	0.092	2728	1633
Portugal	0.157	0.131	0.031	-0.097	0.092	0.034	0.123	2509	1951
Greece	0.151	0.110	-0.003	-0.050	0.094	0.060	0.091	1396	852

Notes.

1. Sample: aged 16-64, employed in 1999, excluding the self-employed, the military and those in full-time education. The US is excluded as slight data differences did not allowed for an identical specification of the wage equation to that of other countries; Sweden is excluded as no wage information is available in any wave; Belgium is excluded due to small sample size. Source: ECHPS.
2. The decomposition is based on an identical male wage equation across countries, including education dummies, experience and its square, occupation and industry dummies, and public sector.
3. (6)=(2)+(4); (7)=(3)+(5).

Table 11:
JMP (1991) decomposition of the difference between the gender wage gap in the UK and in each other EU country
Based on 1999 sample & imputed wage observations from other waves

	1 D_{UK}-D_c	2 Observed characteristics	3 Observed prices	4 Gap effect	5 Unobserved prices	6 Total charact.	7 Wage structure	8 No. obs males	9 No. obs females
Finland	0.087	0.009	0.03	-0.060	0.108	-0.051	0.138	1011	1058
Denmark	0.120	0.060	0.017	-0.071	0.114	-0.011	0.131	752	764
Germany	0.021	0.023	0.027	-0.135	0.107	-0.112	0.134	2776	2336
Netherlands	0.037	0.037	-0.088	-0.014	0.102	0.023	0.014	2542	1961
Austria	-0.012	0.010	-0.020	-0.110	0.107	-0.100	0.087	1601	1248
Ireland	0.040	0.017	-0.023	-0.049	0.095	-0.032	0.072	1273	1095
France	0.073	0.027	-0.008	-0.051	0.105	-0.024	0.097	2500	2335
Italy	0.173	0.112	0.029	-0.07	0.103	0.042	0.132	3044	2157
Spain	0.062	0.106	-0.037	-0.102	0.095	0.004	0.058	3111	2141
Portugal	0.126	0.124	0.024	-0.115	0.092	0.009	0.116	2648	2203
Greece	0.085	0.121	-0.070	-0.057	0.091	0.064	0.021	1546	1135

Notes.

1. Sample: aged 16-64, employed in 1999, with wage imputation from other waves if missing, [-4,+1] for EU, [-5,+2], excluding the self-employed, the military and those in full-time education. The US is excluded as slight data differences did not allowed for an identical specification of the wage equation to that of other countries; Sweden is excluded as no wage information is available in any wave; Luxembourg is excluded as no wage information is available after wave 3; Belgium is excluded due to small sample size. Source: ECHPS.
2. The decomposition is based on an identical male wage equation across countries, including education dummies, experience and its square, occupation and industry dummies, and public sector.
3. (6)=(2)+(4); (7)=(3)+(5).

Table A1: Descriptive statistics for the samples used

	US					UK				Finland				Denmark			
	males		females			males		females		males		females		males		females	
	mean	sd	mean	sd		mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
lgwh	2.786	0.668	2.484	0.623	lgwh	3.490	0.501	3.245	0.489	5.575	0.486	5.414	0.441	6.374	0.345	6.255	0.307
edu1	0.140	0.347	0.118	0.322	ed1	0.261	0.439	0.260	0.439	0.194	0.396	0.169	0.375	0.137	0.344	0.143	0.351
edu2	0.588	0.492	0.620	0.485	ed2	0.073	0.260	0.103	0.304	0.526	0.500	0.399	0.490	0.543	0.498	0.549	0.498
edu3	0.272	0.445	0.262	0.440	ed3	0.666	0.472	0.637	0.481	0.279	0.449	0.432	0.496	0.320	0.467	0.308	0.462
exp	21.117	17.990	17.284	16.638	exp	19.047	13.273	20.029	13.707	19.242	13.073	19.315	12.627	25.385	11.385	24.841	11.044
exp2/100	7.695	17.548	5.755	15.740	exp2/100	5.389	6.029	5.889	6.343	5.410	6.097	5.323	5.492	7.739	6.240	7.389	5.717
Occ1	0.199	0.399	0.273	0.445	Occ1	0.188	0.391	0.117	0.321	0.084	0.277	0.049	0.216	0.087	0.281	0.038	0.191
Occ2	0.150	0.357	0.119	0.324	Occ2	0.127	0.333	0.121	0.326	0.169	0.375	0.230	0.421	0.226	0.419	0.174	0.380
Occ3	0.058	0.233	0.046	0.209	Occ3	0.116	0.320	0.170	0.376	0.141	0.349	0.222	0.416	0.177	0.382	0.274	0.446
Occ4	0.060	0.238	0.274	0.446	Occ4	0.110	0.313	0.260	0.439	0.062	0.241	0.126	0.332	0.075	0.264	0.231	0.422
Occ5	0.205	0.404	0.025	0.155	Occ5	0.085	0.279	0.225	0.418	0.063	0.243	0.232	0.422	0.043	0.204	0.183	0.387
Occ6	0.093	0.290	0.064	0.245	Occ6	0.011	0.104	0.003	0.053	0.013	0.112	0.005	0.073	0.020	0.139	0.003	0.053
Occ7	0.075	0.264	0.009	0.097	Occ7	0.177	0.382	0.016	0.127	0.264	0.441	0.031	0.175	0.176	0.381	0.010	0.099
Occ8	0.054	0.226	0.015	0.121	Occ8	0.124	0.329	0.033	0.180	0.142	0.350	0.022	0.146	0.123	0.329	0.028	0.165
Occ9	0.002	0.050	0.000	0.000	Occ9	0.062	0.242	0.054	0.226	0.063	0.243	0.082	0.275	0.073	0.260	0.059	0.236
Occ10	0.012	0.109	0.003	0.053	Ind1	0.015	0.122	0.003	0.057	0.028	0.164	0.017	0.131	0.034	0.180	0.006	0.075
Occ11	0.091	0.287	0.167	0.373	Ind2	0.021	0.144	0.008	0.086	0.016	0.125	0.004	0.066	0.025	0.157	0.006	0.075
Occ12	0.000	0.019	0.005	0.070	Ind3	0.030	0.171	0.015	0.120	0.023	0.151	0.022	0.146	0.036	0.187	0.020	0.139
Ind1	0.026	0.160	0.005	0.070	Ind4	0.008	0.089	0.013	0.112	0.006	0.080	0.012	0.109	0.001	0.037	0.007	0.084
Ind2	0.008	0.088	0.002	0.042	Ind5	0.034	0.182	0.019	0.136	0.081	0.273	0.033	0.178	0.032	0.176	0.010	0.099
Ind3	0.089	0.285	0.011	0.107	Ind6	0.054	0.226	0.021	0.144	0.033	0.179	0.025	0.156	0.022	0.148	0.020	0.139
Ind4	0.245	0.430	0.131	0.338	Ind7	0.061	0.240	0.011	0.106	0.088	0.284	0.010	0.098	0.073	0.260	0.020	0.139
Ind5	0.118	0.323	0.040	0.197	Ind8	0.093	0.291	0.031	0.175	0.055	0.229	0.033	0.178	0.049	0.216	0.021	0.144
Ind6	0.144	0.351	0.143	0.350	Ind9	0.067	0.249	0.010	0.099	0.154	0.361	0.007	0.080	0.101	0.301	0.017	0.129
Ind7	0.045	0.207	0.080	0.271	Ind10	0.141	0.348	0.152	0.359	0.097	0.296	0.123	0.328	0.102	0.303	0.086	0.280
Ind8	0.071	0.256	0.044	0.204	Ind11	0.028	0.165	0.071	0.257	0.019	0.137	0.051	0.220	0.010	0.098	0.007	0.084
Ind9	0.009	0.096	0.040	0.195	Ind12	0.095	0.294	0.044	0.204	0.125	0.331	0.033	0.178	0.082	0.275	0.041	0.198
Ind10	0.011	0.105	0.007	0.083	Ind13	0.056	0.229	0.067	0.249	0.009	0.092	0.029	0.169	0.039	0.194	0.035	0.184
Ind11	0.129	0.335	0.422	0.494	Ind14	0.114	0.318	0.091	0.288	0.103	0.304	0.091	0.288	0.094	0.291	0.055	0.228
Ind12	0.105	0.307	0.075	0.263	Ind15	0.077	0.267	0.078	0.268	0.040	0.197	0.043	0.204	0.085	0.279	0.134	0.340
Public sector	0.194	0.396	0.266	0.442	Ind16	0.038	0.191	0.115	0.318	0.047	0.211	0.111	0.314	0.094	0.291	0.117	0.321
					Ind17	0.036	0.186	0.207	0.405	0.045	0.207	0.291	0.454	0.057	0.233	0.316	0.465
					Ind18	0.031	0.174	0.046	0.210	0.031	0.173	0.067	0.251	0.064	0.245	0.084	0.278
					Public sector	0.169	0.375	0.330	0.470	0.254	0.436	0.459	0.499	0.314	0.465	0.592	0.492
Observations	2808		2872			2120		2131		941		922		716		711	

Table A1 (contd.): Descriptive statistics for the samples used

	Germany				Netherlands				Belgium				Austria			
	males		females		males		females		males		females		males		females	
	mean	sd	mean	sd	Mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
lgwh	4.534	0.540	4.318	0.531	4.891	0.473	4.689	0.465	7.411	0.377	7.363	0.417	6.347	0.492	6.122	0.491
ed1	0.189	0.392	0.194	0.395	0.878	0.328	0.761	0.426	0.254	0.437	0.118	0.323	0.224	0.417	0.284	0.451
ed2	0.568	0.495	0.589	0.492	0.042	0.200	0.080	0.271	0.335	0.473	0.404	0.492	0.708	0.455	0.622	0.485
ed3	0.242	0.429	0.217	0.412	0.081	0.272	0.159	0.366	0.410	0.493	0.478	0.501	0.068	0.251	0.094	0.292
exp	23.245	13.114	22.213	12.501	22.019	11.923	18.098	11.928	14.671	13.316	10.753	11.246	21.199	11.863	19.572	11.657
exp2/100	7.122	7.539	6.496	6.642	6.269	5.867	4.697	5.089	3.915	5.807	2.414	4.741	5.900	5.424	5.188	5.096
Occ1	0.055	0.227	0.018	0.134	0.153	0.360	0.053	0.223	0.064	0.245	0.017	0.129	0.074	0.262	0.025	0.156
Occ2	0.145	0.352	0.115	0.320	0.191	0.393	0.197	0.398	0.162	0.369	0.163	0.370	0.041	0.199	0.066	0.248
Occ3	0.143	0.350	0.316	0.465	0.200	0.400	0.273	0.445	0.168	0.375	0.213	0.411	0.173	0.378	0.174	0.380
Occ4	0.069	0.253	0.197	0.398	0.078	0.268	0.223	0.416	0.069	0.255	0.292	0.456	0.092	0.289	0.276	0.447
Occ5	0.049	0.216	0.183	0.387	0.060	0.238	0.158	0.365	0.092	0.291	0.225	0.419	0.089	0.285	0.277	0.448
Occ6	0.011	0.104	0.010	0.099	0.025	0.156	0.007	0.081	0.000	0.000	0.000	0.000	0.010	0.102	0.009	0.097
Occ7	0.315	0.465	0.050	0.217	0.151	0.358	0.017	0.130	0.185	0.389	0.000	0.000	0.344	0.475	0.044	0.205
Occ8	0.145	0.352	0.043	0.202	0.094	0.292	0.016	0.124	0.162	0.369	0.017	0.129	0.115	0.319	0.023	0.151
Occ9	0.069	0.253	0.069	0.253	0.048	0.214	0.058	0.234	0.098	0.299	0.073	0.261	0.062	0.242	0.105	0.307
Ind1	0.019	0.138	0.015	0.120	0.023	0.149	0.012	0.110	0.000	0.000	0.006	0.075	0.017	0.128	0.008	0.088
Ind2	0.031	0.175	0.006	0.080	0.012	0.110	0.003	0.053	0.029	0.168	0.017	0.129	0.041	0.198	0.006	0.078
Ind3	0.024	0.154	0.016	0.126	0.033	0.178	0.024	0.153	0.035	0.184	0.034	0.181	0.032	0.176	0.019	0.137
Ind4	0.009	0.092	0.018	0.134	0.007	0.085	0.002	0.041	0.029	0.168	0.017	0.129	0.010	0.099	0.029	0.169
Ind5	0.026	0.159	0.022	0.145	0.030	0.169	0.022	0.145	0.029	0.168	0.017	0.129	0.043	0.203	0.018	0.133
Ind6	0.070	0.256	0.042	0.200	0.038	0.192	0.010	0.099	0.075	0.264	0.034	0.181	0.020	0.139	0.015	0.120
Ind7	0.150	0.357	0.048	0.214	0.049	0.215	0.012	0.107	0.069	0.255	0.000	0.000	0.116	0.321	0.031	0.174
Ind8	0.104	0.306	0.045	0.208	0.060	0.238	0.020	0.140	0.035	0.184	0.028	0.166	0.087	0.282	0.027	0.161
Ind9	0.124	0.329	0.022	0.147	0.095	0.293	0.019	0.136	0.104	0.306	0.000	0.000	0.148	0.355	0.027	0.161
Ind10	0.109	0.312	0.184	0.387	0.129	0.335	0.140	0.347	0.075	0.264	0.157	0.365	0.110	0.313	0.198	0.398
Ind11	0.011	0.104	0.022	0.147	0.017	0.128	0.026	0.159	0.035	0.184	0.067	0.251	0.018	0.135	0.070	0.255
Ind12	0.069	0.254	0.034	0.182	0.083	0.276	0.038	0.190	0.133	0.341	0.028	0.166	0.084	0.278	0.028	0.166
Ind13	0.033	0.179	0.054	0.225	0.043	0.204	0.049	0.215	0.035	0.184	0.045	0.208	0.037	0.189	0.049	0.216
Ind14	0.040	0.195	0.051	0.219	0.127	0.334	0.117	0.321	0.110	0.314	0.079	0.270	0.033	0.179	0.041	0.199
Ind15	0.082	0.275	0.100	0.300	0.101	0.302	0.070	0.256	0.023	0.151	0.045	0.208	0.104	0.305	0.079	0.269
Ind16	0.031	0.174	0.081	0.273	0.064	0.244	0.105	0.307	0.075	0.264	0.101	0.302	0.035	0.184	0.096	0.294
Ind17	0.038	0.192	0.202	0.402	0.060	0.237	0.296	0.457	0.046	0.211	0.253	0.436	0.031	0.173	0.189	0.392
Ind18	0.028	0.166	0.038	0.192	0.030	0.169	0.037	0.189	0.064	0.245	0.073	0.261	0.034	0.181	0.071	0.257
Public sector	0.215	0.411	0.374	0.484	0.218	0.413	0.337	0.473	0.173	0.380	0.264	0.442	0.248	0.432	0.312	0.464
Observations	2669		2037		2472		1805		173		178		1624		1159	

Table A1 (contd.): Descriptive statistics for the samples used

	Ireland				France				Italy				Spain			
	males		females		males		females		males		females		males		females	
	mean	sd	mean	sd	Mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
lgwh	3.465	0.584	3.317	0.536	5.653	0.524	5.546	0.509	4.195	0.402	4.133	0.410	8.414	0.504	8.290	0.543
ed1	0.362	0.481	0.263	0.440	0.636	0.481	0.526	0.499	0.455	0.498	0.320	0.466	0.513	0.500	0.368	0.482
ed2	0.424	0.494	0.470	0.499	0.089	0.285	0.126	0.332	0.431	0.495	0.538	0.499	0.201	0.401	0.215	0.411
ed3	0.214	0.410	0.267	0.443	0.275	0.446	0.348	0.477	0.114	0.318	0.143	0.350	0.286	0.452	0.417	0.493
exp	19.413	13.362	16.541	11.973	25.319	17.446	25.463	18.705	19.667	12.773	16.940	11.848	19.817	13.071	15.744	12.095
exp2/100	5.553	6.263	4.168	4.708	9.452	10.450	9.980	11.220	5.499	6.356	4.272	5.237	5.635	6.303	3.941	5.156
Occ1	0.080	0.272	0.044	0.206	0.068	0.253	0.030	0.171	0.028	0.166	0.003	0.057	0.028	0.166	0.010	0.102
Occ2	0.122	0.327	0.173	0.379	0.117	0.321	0.101	0.302	0.063	0.242	0.167	0.373	0.105	0.306	0.205	0.404
Occ3	0.105	0.306	0.107	0.309	0.206	0.404	0.240	0.427	0.116	0.321	0.123	0.328	0.096	0.295	0.113	0.316
Occ4	0.059	0.236	0.226	0.418	0.083	0.276	0.287	0.452	0.195	0.396	0.317	0.466	0.070	0.256	0.175	0.380
Occ5	0.093	0.291	0.278	0.448	0.060	0.238	0.178	0.383	0.091	0.287	0.139	0.346	0.109	0.312	0.223	0.417
Occ6	0.018	0.132	0.002	0.048	0.018	0.134	0.006	0.076	0.021	0.142	0.009	0.096	0.020	0.141	0.005	0.070
Occ7	0.191	0.393	0.013	0.112	0.230	0.421	0.020	0.139	0.241	0.427	0.093	0.291	0.268	0.443	0.055	0.228
Occ8	0.159	0.366	0.070	0.255	0.157	0.364	0.047	0.211	0.143	0.350	0.048	0.213	0.168	0.374	0.035	0.184
Occ9	0.173	0.378	0.087	0.282	0.061	0.239	0.092	0.290	0.103	0.304	0.100	0.301	0.135	0.342	0.178	0.383
Ind1	0.041	0.197	0.012	0.107	0.017	0.130	0.008	0.090	0.037	0.189	0.022	0.148	0.043	0.203	0.019	0.137
Ind2	0.039	0.193	0.002	0.048	0.022	0.146	0.009	0.093	0.029	0.168	0.008	0.090	0.022	0.148	0.002	0.049
Ind3	0.056	0.229	0.026	0.158	0.035	0.184	0.021	0.143	0.027	0.162	0.018	0.133	0.034	0.182	0.029	0.169
Ind4	0.013	0.114	0.014	0.117	0.010	0.101	0.019	0.137	0.024	0.152	0.084	0.278	0.016	0.126	0.052	0.221
Ind5	0.032	0.175	0.012	0.107	0.024	0.152	0.013	0.112	0.028	0.165	0.013	0.114	0.032	0.176	0.008	0.089
Ind6	0.041	0.199	0.031	0.174	0.052	0.222	0.026	0.159	0.040	0.197	0.031	0.173	0.040	0.197	0.025	0.155
Ind7	0.029	0.168	0.012	0.107	0.070	0.256	0.020	0.141	0.096	0.295	0.025	0.155	0.070	0.255	0.010	0.099
Ind8	0.072	0.259	0.059	0.236	0.093	0.290	0.032	0.175	0.072	0.258	0.030	0.170	0.058	0.234	0.017	0.128
Ind9	0.137	0.344	0.007	0.083	0.097	0.295	0.010	0.099	0.099	0.299	0.007	0.081	0.182	0.386	0.011	0.105
Ind10	0.100	0.300	0.140	0.347	0.129	0.335	0.137	0.344	0.084	0.277	0.111	0.315	0.112	0.316	0.143	0.351
Ind11	0.029	0.168	0.090	0.286	0.019	0.136	0.024	0.152	0.016	0.125	0.029	0.168	0.043	0.202	0.069	0.253
Ind12	0.100	0.300	0.036	0.187	0.075	0.264	0.033	0.179	0.079	0.270	0.030	0.171	0.079	0.270	0.030	0.171
Ind13	0.026	0.160	0.049	0.216	0.033	0.180	0.042	0.200	0.040	0.195	0.032	0.177	0.031	0.173	0.025	0.157
Ind14	0.048	0.215	0.073	0.261	0.087	0.281	0.087	0.282	0.038	0.191	0.056	0.230	0.048	0.214	0.093	0.291
Ind15	0.100	0.300	0.055	0.227	0.092	0.289	0.118	0.323	0.122	0.328	0.103	0.304	0.081	0.273	0.080	0.272
Ind16	0.063	0.244	0.123	0.329	0.071	0.257	0.155	0.362	0.050	0.219	0.187	0.390	0.044	0.205	0.139	0.346
Ind17	0.031	0.173	0.169	0.375	0.049	0.215	0.176	0.381	0.063	0.243	0.134	0.341	0.025	0.155	0.137	0.344
Ind18	0.044	0.205	0.092	0.289	0.026	0.159	0.070	0.256	0.055	0.228	0.079	0.270	0.040	0.195	0.110	0.313
Public sector	0.271	0.445	0.317	0.466	0.274	0.446	0.422	0.494	0.306	0.461	0.412	0.492	0.190	0.393	0.289	0.454
Observations	1135		860		2031		1731		2719		1824		2725		1631	

Table A1 (contd.): Descriptive statistics for the samples used

	Portugal				Greece			
	males		females		males		females	
	mean	sd	mean	sd	Mean	sd	mean	sd
lgwh	7.908	0.529	7.822	0.659	8.872	0.512	8.784	0.529
ed1	0.793	0.405	0.654	0.476	0.375	0.484	0.243	0.429
ed2	0.131	0.337	0.172	0.377	0.387	0.487	0.427	0.495
ed3	0.076	0.265	0.174	0.380	0.239	0.426	0.330	0.471
exp	20.241	13.589	17.251	12.836	18.472	11.687	13.854	10.396
exp2/100	5.943	6.565	4.622	5.753	4.777	5.042	2.999	3.817
Occ1	0.019	0.136	0.006	0.076	0.030	0.171	0.012	0.108
Occ2	0.059	0.236	0.117	0.321	0.144	0.351	0.207	0.405
Occ3	0.067	0.250	0.115	0.319	0.070	0.256	0.129	0.336
Occ4	0.087	0.282	0.145	0.352	0.141	0.348	0.264	0.441
Occ5	0.109	0.312	0.182	0.386	0.129	0.335	0.155	0.362
Occ6	0.049	0.216	0.016	0.127	0.013	0.113	0.011	0.102
Occ7	0.300	0.458	0.099	0.298	0.240	0.427	0.073	0.260
Occ8	0.149	0.357	0.069	0.253	0.159	0.366	0.022	0.148
Occ9	0.160	0.367	0.252	0.434	0.074	0.262	0.127	0.333
Ind1	0.063	0.244	0.030	0.172	0.014	0.116	0.009	0.097
Ind2	0.023	0.148	0.002	0.046	0.051	0.220	0.009	0.097
Ind3	0.032	0.175	0.026	0.158	0.032	0.177	0.031	0.172
Ind4	0.032	0.175	0.122	0.328	0.019	0.135	0.068	0.252
Ind5	0.027	0.161	0.010	0.099	0.021	0.143	0.021	0.144
Ind6	0.038	0.190	0.016	0.125	0.037	0.190	0.013	0.113
Ind7	0.048	0.214	0.011	0.102	0.036	0.186	0.008	0.090
Ind8	0.032	0.177	0.020	0.140	0.029	0.167	0.012	0.108
Ind9	0.214	0.410	0.011	0.104	0.139	0.346	0.011	0.102
Ind10	0.145	0.352	0.111	0.315	0.130	0.337	0.146	0.353
Ind11	0.047	0.212	0.074	0.261	0.055	0.229	0.065	0.246
Ind12	0.066	0.248	0.025	0.155	0.098	0.298	0.026	0.159
Ind13	0.025	0.156	0.019	0.138	0.037	0.188	0.063	0.244
Ind14	0.028	0.166	0.046	0.210	0.032	0.177	0.061	0.240
Ind15	0.110	0.313	0.084	0.277	0.127	0.333	0.105	0.306
Ind16	0.034	0.180	0.148	0.355	0.072	0.258	0.186	0.389
Ind17	0.018	0.133	0.128	0.334	0.030	0.171	0.113	0.317
Ind18	0.020	0.139	0.118	0.322	0.040	0.197	0.054	0.226
Public sector	0.189	0.391	0.299	0.458	0.343	0.475	0.412	0.493
Observations	2443		1904		1391		851	

Notes to A1.
The descriptive statistics refer to the 16-64 male and female employed samples in 1999, excluding self-employed, military and full-time students. Source: PSID and ECHPS.
Variables:
Educ1=1 if Less than grade 12 (US); =1 if Less than upper secondary education (EU). Omitted category.
Educ2=1 if Grade 12 completed (US); =1 if Upper secondary education completed (EU)
Educ3=1 if Grade 16 completed (US); =1 if Higher education (EU)
Exp: Actual full-time or part-time experience in years (US); Current age – age started first job (EU)

US	professional, technical and kindred workers	**EU**	
Occ1	managers and administrative	Occ1	Legislators, senior officials and managers
Occ2	sales workers	Occ2	Professionals
Occ3	clerical & kindred workers	Occ3	Technicians and associated professionals
Occ4	craftsmen & kindred workers	Occ4	Clerks
Occ5	operatives except transportation	Occ5	Service workers and shop and market sales workers
Occ6	transport equipment operatives	Occ6	Skilled agricultural and fishery workers
Occ7	laborers except farms	Occ7	Craft and related trades workers
Occ8	farmers and farm managers	Occ8	Plant and machine operators and assemblers
Occ9	farm laborers and farms foremen	Occ9	Elementary occupations
Occ10	service workers (no private hh)	Ind1	Agriculture, hunting and forestry + Fishing
Occ11	private household workers	Ind2	Mining and quarrying + Electricity, gas and water supply
Occ12	Mining	Ind3	Manufacture of food products, beverages and tobacco
Ind1	Construction	Ind4	Manufacture of textiles, clothing and leather products
Ind2	Manufacturing	Ind5	Manufacture off wood and paper products; publishing and printing
Ind3	Transportation, communications & public utilities	Ind6	Manufacture of coke, refined petroleum/chemicals/rubber & plastic/…products etc.
Ind4	wholesale and retail trade	Ind7	Manufacture of metal products, machinery and equipment n.e.c.
Ind5	finance	Ind8	Other manufacturing
Ind6	insurance and real estate	Ind9	Construction
Ind7	business and repair services	Ind10	Wholesale and retail trade; repair of motor vehicles, motorcycles and personal/hh goods
Ind8	personal services	Ind11	Hotels and restaurants
Ind9	entertainment and rec. services	Ind12	Transport, storage and communication
Ind10	professional and related services	Ind13	Financial intermediation
Ind11	public administration	Ind14	Real estate, renting and business activities
Ind12	professional, technical and kindred workers	Ind15	Public administration and defense; compulsory social security
		Ind16	Education
		Ind17	Health and social work
		Ind18	Other community, social and personal service activities; private households with employed persons; extra-territorial organizations and bodies

Table A2:
Proportions of imputed wage observations in total nonemployment

	Male	Female
USA	0.549	0.517
UK	0.478	0.493
Finland	0.534	0.558
Denmark	0.852	0.694
Germany	0.802	0.541
Netherlands	0.477	0.378
Belgium	0.429	0.319
Austria	0.702	0.319
Ireland	0.406	0.312
France	0.740	0.490
Italy	0.523	0.199
Spain	0.545	0.273
Portugal	0.571	0.305
Greece	0.526	0.193

Notes. Figures report the proportion of individuals who were not employed in 1999 but were employed in at least another year in the sample period over the total number of nonemployed individuals in 1999.

Table A3:
Aggregate real wage growth

	Males				Females			
	Coef.	(s.e.)	No. obs.	R^2	Coef.	(s.e.)	No. obs.	R^2
USA	0.021***	0.002	20317	0	0.023***	0.002	22376	0.01
UK	0.025***	0.002	23963	0.01	0.034***	0.001	24907	0.02
Finland	0.014***	0.003	9648	0	0.018***	0.002	9933	0.01
Denmark	0.022***	0.002	10762	0.01	0.018***	0.002	10016	0.01
Germany	0.003*	0.001	35106	0	0.003*	0.001	27904	0
Netherlands	0	0.002	20796	0	0.002	0.002	17563	0
Belgium	0.012***	0.002	9994	0	0.013***	0.002	8569	0
Austria	0.012***	0.002	12225	0	0.010***	0.003	8963	0
Ireland	0.027***	0.002	11861	0.01	0.035***	0.003	9276	0.02
France	0.008***	0.002	20166	0	0.013***	0.002	16927	0
Italy	0.004***	0.001	25341	0	0.008***	0.001	16578	0
Spain	0.013***	0.001	24119	0	0.009***	0.002	14246	0
Portugal	0.030***	0.002	20232	0.01	0.037***	0.002	15280	0.02
Greece	0.021***	0.002	13121	0.01	0.022***	0.002	8110	0.01

Notes. Results from regressions of log gross hourly wages on a linear time trend. Sample: employed males and females aged 16-64, excluding elf-employed, military and full-time students. Source: PSID and ECHPS, 1994-2001.