GRAPE

Group for Research in APplied Economics

GRAPE Working Paper # 41

The changing nature of gender selection into employment over the Great Recession

Juan J. Dolado, Cecilia Garcìa-Penalosa, Linas Tarasonis

FAME | GRAPE, 2020

Foundation of Admirers and Mavens of Economics
Group for Research in Applied Economics

Redistributive allocation mechanisms

Juan J. Dolado
Universidad Carlos III de Madrid

Cecilia Garcìa-Penalosa
Aix-Marseille University, EHESS, CNRS, Central Marseille & AMSE

Linas Tarasonis
Vilnius University & Bank of Lithuania

Abstract
The Great Recession has strongly influenced employment patterns across skill and gender groups in EU countries. We analyze how these changes in workforce composition might distort comparisons of conventional measures of gender wage gaps via non-random selection of workers into EU labour markets. We document that male selection (traditionally disregarded) has become positive during the recession, particularly in Southern Europe. As for female selection (traditionally positive), our findings are twofold. Following an increase in the LFP of less-skilled women, due to an added-worker effect, these biases declined in some countries where new female entrants were able to find jobs, whereas they went up in other countries which suffered large female employment losses. Finally, we document that most of these changes in selection patterns were reversed during the subsequent recovery phase, confirming their cyclical nature.

Keywords:
sample selection, gender wage gaps, gender employment gaps

JEL Classification
J31

Corresponding author
Linas Tarasonis, linas.tarasonis@evaf.vu.lt.

Acknowledgements
We are grateful to G. Jolivet, C. Schluter, G. Spanos, H. Turon and seminar participants for their useful suggestions. Financial support from the ADEMU project (EC-H2020, Grant no. 6649396), the Spanish Ministerio de Econom´ıa y Competitividad, the French National Research Agency (ANR-17-EURE-0020 and ANR-18-CE41-0003-01) and the Research Council of Lithuania (Lietuvos Mokslo Taryba [S-LL-19-3]) is gratefully acknowledged. The views expressed are those of the authors and do not necessarily represent those of the Bank of Lithuania. All errors are our own.

Published by: FAME | GRAPE
ISSN: 2544-2473

Foundation of Admirers and Mavens of Economics
ull. Mazowiecka 11/14
00-052 Warszaw
Poland

W | grape.org.pl
E | grape@grape.org.pl
TT | GRAPE_ORG
FB | GRAPE.ORG
PH | +48 799 012 202

1 Introduction

There has been an extensive debate both in the academic literature and the media about the effects of the Great Recession on household income inequality. Yet, its impact on gender wage inequality remains far less explored.[1] This is somewhat surprising since industries which differ markedly in their relative use of male and female labour have experienced quite unequal fluctuations in employment and labour-force participation, both of which could affect male and female wages unequally through their effects on the workforce composition. In particular, these changes have been very relevant in some European Union (EU) member states, where the recession was longer and more severe than in the US and other high-income countries.[2] Consequently, the EU provides an interesting laboratory to analyze how gender wage gaps react to differences in the way men and women self-select into labour markets when faced with large shifts in labour demand and labour supply, like those taking place during the Great Recession.[3]

To account for non-random selection by gender over the business cycle it is important to distinguish between the *raw* gender gap (RG), i.e. the difference between observed wages of male and female employees, and the *potential* gender gap (PG), i.e. the difference that would be observed if all men and women of working age were employed. In effect, when comparing wages across two population groups, non-random selection into employment implies that RGs could be above or below PGs, depending on the sign of the selection biases. The literature typically assumes no selection whatsoever for the majority group (white, natives, men, etc.), while both positive or negative selection is considered for the minority. As a result, a large body of research on gender gaps has stressed that accounting for selection is paramount to obtain less distorted measures of the gender gaps. Hence, the need to pay growing attention to PGs rather than just report RGs.[4]

When looking at the EU, our focus on selection is further dictated by the available evidence highlighting their key role in explaining differences in RGs in the recent past, prior to the crisis. For example, relying on imputed wage distributions for the male an female working age populations, Olivetti and Petrongolo (2008) have documented that PGs in Southern Mediterranean countries (*Southern EU*, hereafter) were

[1] See, for example, Jenkins et al. (2012)

[2] This is so since the Great Recession in most of the EU not only covers the global financial crisis in 2008-09, but also the subsequent sovereign debt crisis in the Euro area from late 2009 to mid 2012.

[3] More precisely, the gender wage gap is defined in the sequel as the difference between male and female hourly wages in log points.

[4] See, *inter alia*, Heckman (1979), Johnson et al. (2000), Neal (2004), Mulligan and Rubinstein (2008), Olivetti and Petrongolo (2008), and Arellano and Bonhomme (2017).

considerably larger than RGs from the mid 1990s to the early 2000s, whereas these differences were fairly small in other EU countries (*Rest of EU*, henceforth) and the US. Olivetti and Petrongolo (2008) rationalize this finding by convincingly arguing that, while male labour-force participation (LFP) rates are high everywhere, the historically lower female LFP rates in Southern EU countries are often related to positive selection among participating women, as those who work often have relatively high-wage characteristics. By contrast, selection issues become irrelevant in the Rest of EU and the US since most of these countries exhibit rather high female LFP rates. Accordingly, lower RGs in Southern EU are mainly explained by positive selection in their female workforces which increases the average wage among female employees, relative to other countries were female selection is not a relevant issue. The main finding of this influential paper is that, once selection-bias corrections are implemented, the previous ranking gets reversed, leading to higher PGs in Southern EU than in the Rest of the EU.

In view of these considerations, the aim of this paper is to explore whether the above regularities on gender non-random sorting into the EU labour markets have changed as a result of the Great Recession, as well as to check to what extent the subsequent recovery phase has led to a reversal of those changes. To address this issue, use is made of the EU-SILC longitudinal dataset on wages, which is available for several EU member states covering periods before and after the global financial crisis. To estimate how selection biases behave over the relevant subsamples, we estimate changes in PGs from a wide range of imputations for non-observed wages, and then proceed to compare them with the observed changes in RGs, whose evolution over this period is taken as given.[5] We argue that, as a result of changes in workforce composition during the slump, male selection has become more important than previously thought, whereas female selection may have become stronger or weaker, depending on the economic forces at play. We refer to this phenomenon as "the *changing nature*" of selection by gender during the Great Recession.[6]

Our main insight for the emergence of positive male selection is that, following massive job destruction in sectors intensive in low-skilled male workers (e.g. in the construction and manufacturing sectors in some EU economies), the distribution of observed male wages has become a censored version of the imputed distribution.

[5] A number of recent reports, most notably OECD (2014), have documented that RGs have continued to narrow in most EU countries during the Great Recession. Potential explanations of this fact could be women´s over-representation in the public sector (where gender gaps are generally lower) and the widespread use of early retirement policies (mainly affecting elderly male employees with long professional careers and high wages).

[6] To the best of our knowledge, Arellano and Bonhomme (2017) is the only paper that documents positive male selection into the labour market. Their focus is on the UK prior to the Great Recession.

This would lead to a higher average wage of male employees, implying that RGs would be larger than PGs, rather than the opposite. As regards female selection, two contrasting effects are at play. First, it is likely that the existence of a so-called "added-worker" effect during the crisis – whereby less-skilled women who were previously inactive enter the labour market to help restore household income levels as male breadwinners become jobless– has increased female LFP at the bottom of the wage distribution, therefore reducing female positive selection biases. In line with previous findings by Bentolila and Ichino (2008), Bredtmann et al. (2018, Table 2) have recently shown that this effect is particularly strong in Southern Mediterranean countries, probably due to their less generous welfare states.[7]. If new female entrants from the bottom of the skill distribution succeed in finding jobs during the slump, male and female selection would change in different directions (male up, female down), so that the difference between RGs and PGs would become larger. However, even under an increasing female LFP, if labour demand for both male and female less-skilled workers experienced large adverse shifts during the downturn, it could well be the case that both male and female selection may have become more positive, so that the sign of the difference between RG and PG would be ambiguous. We argue that this rise in female selection characterizes well the experience of some Southern EU countries with high shares of temporary contracts (dual labour markets), since women are over-represented in fixed-term jobs (e.g. in the services sector) which were massively destroyed during the slump due to having much lower termination costs than open-ended contracts.

In sum, while unskilled men's employment has been subject to a large negative labour demand shift, women's employment patterns have been subject to both supply and demand shifts, and depending on which dominates, female selection may have moved in line or in opposite direction to male selection. Moreover, insofar as these phenomena are driven by a cyclical collapse in labour demand, one should observe a reversal of the changing patterns in selection once the recovery started, an issue on which we also provide evidence.

Two empirical strategies are used to construct PG in EU countries. Following Olivetti and Petrongolo (2008), we first apply the sample-selection correction methodology advocated by Johnson et al. (2000) and Neal (2004). This approach imputes missing wages for non-employed workers relative to the median (rather than the ac-

[7]Bredtmann et al. (2018) – using the same database (EU-SILC; see Section 3) and a similar sample period as ours – find evidence of a high responsiveness of women's labor supply to their husband's loss of employment. Given that this evidence is based on the same panel dataset we use here and for a similar sample period (2004-13), in the sequel we take the "added-worker" effect as a given stylised fact for this set of countries.

tual level of missing wages). An advantage of this approach is that it avoids arguable exclusion restrictions often invoked in the standard econometric (Heckit) approach to extrapolate the wage distribution below the reservation wage.[8] However, a potential drawback of this procedure is that the reliability of its results hinges strongly on the plausibility of assumptions underlying the imputation rules. Therefore, to check how robust our findings are under a more conventional control-function approach, we also report results based on Arellano and Bonhomme's (2017) estimation procedure of quantile wage regressions by gender subject to selectivity corrections. Notice that, besides being suitable for median regression, the main reason for using a quantile approach is that our rationalization of changes in the gender wage gap relies on the different behaviour of male and female workers with different skills, namely, those at the bottom and other parts of the wage distribution.

Our empirical findings broadly support the mechanisms outlined above. First, we document that the traditional assumption of no male selection prior to the crisis may not be a valid during the Great Recession. Strong evidence of positive male selection is found for several EU countries, particularly in Southern EU. Second, we show that patterns of female selection are mixed. On the one hand, we document that a significant rise of less-skilled female LFP in some EU countries has reduced female selection relative to what was found before the slump. On the other hand, in those countries where the rise in female LFP has not translated into new jobs and female unemployment rates have also surged (particularly in dual labour markets), female selection has become stronger than before the crisis.

Related literature

This paper contributes to a vast literature on gender outcomes in developed (and developing) countries; cf. Blau et al. (2013) and Goldin (2014) for comprehensive overviews. While most of this research analyzes the determinants of secular trends in gender wage gaps (typically using RGs), our paper complements this approach by focusing on their behaviour at particularly relevant business cycle phases, like those taking place during the Great Recession and the subsequent recovery.

There is some previous research on this topic that is worth highlighting. For example, the issue of how male hourly real wages change over the US business cycle has been addressed in a well-known paper by Keane et al. (1988) which uses the stan-

[8]For example, this might be the case regarding number of children or being married (as proxies for household chores). Such variables are often assumed as only affecting labour-market participation via reservation wages. However, one could argue that they might as well affect effort at market-place work, and therefore productivity and wages.

dard Heckman (1979)'s techniques to correct for non-random selection.[9] We differ from this forerunner in several respects. First, we focus on gender wage gaps instead of exclusively on male wages. Second, our evidence refers to a cross-country comparison of gender wage gaps in EU countries which have been subject to much less research than the US (see e.g. Blau et al., 2013). Third, we identify new channels on how the Great Recession in particular and business cycles in general affect selection by gender. Lastly, while most the papers on this topic apply a conventional Heckit approach, our results rely on the two alternative econometric techniques mentioned earlier, which are less problematic in correcting for selection biases.

The rest of the paper is organized as follows. Section 2 provides some theoretical underpinning of the main mechanisms at play and derives testable implications in terms of signs of changes in selection biases and LFP/ employment rates by gender. Section 3 describes the EU-SILC longitudinal dataset used throughout the paper. Section 4 explains our two empirical approaches (imputation rules around the median and quantile selection models) to compute the potential wage distributions and correct for selectivity biases. Section 5 presents the empirical results yielded by both econometric procedures. Section 6 interprets the main empirical findings of the paper in the light of the hypotheses outlined in Section 2. Finally, Section 7 concludes. An Appendix provides further details on the model (parts A and B) and on the construction of hourly wages (part C), while an Online Appendix gathers additional results on alternative imputation procedures and further descriptive statistics for the 13 European countries included in our sample.

2 A Simple Theoretical Framework

2.1 The basic model

To provide some simple theoretical underpinning for the main mechanisms at play, we start by reviewing the basic effects of selection on the measurement of gender wage gaps. Following Mulligan and Rubinstein (2008),we consider a conventional *mincerian* equation for the determination of the (logged) hourly *potential wage*:

$$w_{it} = \mu_t^w + g_i \gamma_t + \varepsilon_{it} \tag{1}$$

where w_{it} denotes individual i's potential hourly wage in year t, g_i is a gender indicator variable (males have $g = 0$, females have $g = 1$), μ_t^w represents (an index of) the

[9]See also Bowlus (1995) and Gayle and Golan (2012) for further examples in the gender gap literature accounting for the dynamics of employment selection over the business cycle.

determinants of wages that are common to *all* workers, while γ_t captures those determinants of female wages common to all women but not applicable to men (including discriminatory practices by employers). Finally, ε_{it} is an error term normalized to have a unit variance (for both males and females) such that $m(\varepsilon_{it}|\, \mu_t^w, g_i) = 0$, where $m(\cdot)$ denotes the (conditional) *median* function.[10]

If potential wages were available for all individuals in the working age population, then the *potential* median gender wage gap at year t, PG_t, would be defined as:

$$PG_t \equiv m(w_{it}|g_i = 0) - m(w_{it}|g_i = 1) = -\gamma_t, \tag{2}$$

where one would expect that $PG_t > 0$ (i.e. $\gamma_t < 0$) on historical grounds (see Olivetti and Petrongolo, 2016).

However, to the extent that selection into employment is not a random outcome of the male and female populations, the observed (raw) gender gap in median wages, RG_t , in a sample restricted to *employed* individuals will differ from the PG_t, namely:[11]

$$\begin{aligned} RG_t &\equiv m(w_{it}|g_i = 0, L_{it} = 1) - m(w_{it}|g_i = 1, L_{it} = 1) \\ &= -\gamma_t + m(\varepsilon_{it}|g_i = 0, L_{it} = 1) - m(\varepsilon_{it}|g_i = 1, L_{it} = 1) \\ &= PG_t + \underbrace{b_t^m - b_t^f}_{\text{selection bias differential}}, \end{aligned} \tag{3}$$

where L_{it} is an indicator for whether individual i is employed in year t, and $b_t^m = m(\varepsilon_{it}|g_i = 0, L_{it} = 1)$ and $b_t^f = m(\varepsilon_{it}|g_i = 1, L_{it} = 1)$ are the (median) selection biases of males and females, respectively. These two terms differ from zero to the extent that non-employed males and females have different potential wages than the employed ones. As discussed above, Olivetti and Petrongolo (2008) argue that: (i) the inequality $b_t^m < b_t^f$ holds in Southern EU countries prior to the Great Recession, so that $RG_t < PG_t$; and (ii) $b_t^m \simeq b_t^f$ held in Rest of EU countries and the US, implying that $RG_t \simeq PG_t$.

Using (3), the change (Δ) in the observed RGs over time becomes:

$$\Delta RG_t = \Delta PG_t + \Delta b_t^m - \Delta b_t^f. \tag{4}$$

Equation (4) has three terms. The first one ($\Delta PG_t = -\Delta\gamma_t$) is the change in the gender-specific component of wages, which may exist due to changes in gender wage

[10]Consistent with the empirical section, our focus in this section is on median rather than mean gender gaps. This choice is without loss of generality since the results can be rewritten in terms of mean gaps and selection biases. As is well known, in this case the latter become functions of the inverse Mill's ratio, as in Mulligan and Rubinstein (2008).

[11]The discussion below reproduces the well-known arguments on selection biases in the seminal work by Gronau (1974) and Heckman (1979), albeit based on gaps in median wages rather than on average wages, as these authors consider.

discrimination, relative market valuation of skills, or relative human capital accumulation when considering *all* men and women. The second and third terms in (4) capture in turn the changes in the selection biases of males and females, respectively, which constitute the main focus of this paper.[12]

Traditionally, this setup has been used to predict which females are employed using a potential market wage equation determining w_{it}, as in (1), plus an additional equation determining the reservation wage, r_{it},such that individuals would accept a job if $w_{it} > r_{it}$. We extend this conventional framework by adding an extra equation determining productivity, x_{it}, to capture labour-demand constraints that could affect both men and women. This leads to the following three-equation model (where equation (1) is repeated below in (5) for convenience):

$$w_{it} = \mu_t^w + g_i \gamma_t + \varepsilon_{it} \tag{5}$$

$$x_{it} = \mu_t^x + u_{it} \tag{6}$$

$$r_{it} = g_i(\mu_t^r + v_{it}), \tag{7}$$

such that μ_t^x in (6) represents (an index of) the determinants of the average productivity of a worker, μ_t^r in (7) captures the determinants of female reservation wage (notice that the male reservation wage is normalized to zero in (7), since $g_i = 0$ for men), u_{it} is a productivity shock, and v_{it} is a reservation-wage shock. The normalization $r_{mt} = 0$ is used as a shortcut to capture the fact that male LFP rates are very high everywhere. Furthermore, since the shock in the wage equation (5) should mainly reflect unexpected productivity changes, it is assumed for simplicity that,

$$u_{it} = (1 + \rho)\varepsilon_{it},$$

with $\rho > 0$. Therefore, a productivity shock of size $(1 + \rho)\varepsilon_{it}$ only shifts the wage by the lower amount ε_{it}, reflecting some wage rigidity.[13] This assumption allows us to capture the fact that some individuals sorting themselves into the labour market during a recession may not be able to find jobs when wages are partially rigid, as it has been the case in several EU countries. Finally, whereas ε_{it} has a continuous

[12]Note that, had we allowed for changes in the variance in the error term ε_{it}, an additional term would appear in (4), namely $(b_t^m - b_t^f)\Delta\sigma_t^\varepsilon$, where σ_t^ε is its time-varying standard deviation . This term captures changes in the dispersion of wages which has been shown to play an important role in explaining female selection in the US (see Mulligan and Rubinstein, 2008). Yet, these changes are ignored in the sequel. The reason is that, as shown in Figure A1 in the Online Appendix where wage dispersion is measured by the logarithm of the ratio between wages at the 90th and 10th percentiles, no major trends seem to be present over 2004-2012, with the possible exceptions of Greece and Portugal.

[13]This is particularly the case in most European countries, where unions play a more important role in wage setting than in the US. Our model implies symmetry in wage response to positive and negative productivity shocks, although it could be easily generalized to allow for asymmetric responses.

support, to simplify matters we constrain the female reservation wage shock to only take two values: a high one, $\overline{v}$, with probability $p \in (0,1)$ and a low one, $\underline{v}$, with probability $1-p$. This simplified two-mass distribution suffices to capture the lower LFP rate of less-skilled women by assuming that $\overline{v} > \underline{v}$.

Accordingly, individual i works at time t if her/his reservation wage is higher than her/his potential market wage (labour supply condition), i.e. $w_{it} > r_{it}$, and if her/his productivity is greater than the wage, leaving a positive surplus for the firm (labour demand condition), i.e. $x_{it} - w_{it} > 0$. As a result, there are labour supply (LS) and labour demand (LD) threshold values of the productivity shock ε_{it}, determining whether the worker participates and the firm creates/ destroys jobs. In the sequel these cut-off values will be respectively labelled $a_t^{LS}(g_i)$ and $a_t^{LD}(g_i)$, and their derivation can be found in Appendix A. Since the worker's decision to participate and the firm's decision to create a job implies that ε_{it} should exceed a given cut-off value, notice that the LD and LS conditions will be the binding ones whenever $a_t^{LS}(g_i) < a_t^{LD}(g_i)$ and $a_t^{LD}(g_i) < a_t^{LS}(g_i)$, respectively.

The main implications of this simple model can be summarised as follows. First, the LD constraint $a_t^{LD}(g_i = 0)$ is the only binding one for men, due to the assumption that they always participate ($r_m = 0$). Second, as regards women, the LD constraint $a_t^{LD}(g_i = 1)$ binds (i.e. $a_t^{LD} > a_t^{LS}$) whenever: (i) their potential wage $(\mu_t^w + \gamma_t)$ is larger than the reservation wage (μ_t^r) but it is below their expected productivity (μ_t^x), implying they would like to participate but firms do not create new female jobs and would even terminate existing ones ; and (ii) wages are more rigid, i.e. ρ is large. Conversely, whenever female productivity is high, their reservation wage is low and wages are more flexible, the LS constraint becomes the binding one ($a_t^{LD} < a_t^{LS}$) . For example, in more traditional societies (such as those in Southern EU), where the average female reservation wage is high due to cultural and social norms, and the surplus is small due to lower productivity in these countries, the LS condition becomes the binding one. On the contrary, in more modern societies (such as in the Rest of the EU), where the average female reservation wage is low and the surplus is high, the LD condition turns out to be the binding constraint. Moreover, the LS constraint is also more likely to bind for lower-educated women in all countries given that they are often more heavily involved in household chores than higher-educated women.

Finally, in Appendix B, we derive comparative statics of male and female observed median wages with respect to changes in μ_t^x and μ_t^r. The former captures changes in productivity due to business cycle fluctuations, whereas the latter captures changes in (female) outside-option values due to, for example, added- worker effects. The main findings here are as follows:

(i) male and female median wages go down as μ_t^x falls (e.g. in a recession); this leads to growing positive selection for both genders as low-productivity (low-wage) workers are the ones more likely to lose their jobs during a downturn (i.e. $\Delta b_t^m > 0$ and $\Delta b_t^f > 0$ in expression (4) above), and

(ii) female median wages go down as μ_t^r falls. This is because less-skilled (married) women, who were not participating before the recession, are the ones who start searching for jobs during the slump as their reservation wages fall due to the large job losses suffered by their less-skilled partners (i.e. $\Delta b_t^f < 0$ in (4)).

Summing up, the main implication of the previous analysis is that, while the male median wage is bound to increase in a downturn, the female median wage may increase or decrease, depending upon which of the two opposite forces (LD and LS constraints) dominates as a result of the recession. The opposite effects would prevail during expansions.

2.2 Gender-gap scenarios over the Great Recession

The implications of the previous analysis result in a range of hypotheses about gender gaps that can emerge (individually or jointly), depending on how employment and LFP rates change by gender. The Great Recession has had two key effects for our purposes. On the one hand, there was a large shedding of unskilled low-paid jobs; this increase in job destruction has not only affected *male* labour-intensive industries but also *female* workers in some countries as well. On the other, as documented by Bredtmann et al. (2018), the slump led to a rise in less-skilled female LFP (particularly in Southern EU labour markets), as a response to a decline in the employment rate of less-skilled men. When the LS constraint binds, then the added-worker effect implies that new less-skilled female entrants in the labour market will succeed in finding jobs; by contrast, when LD is the binding constraint, the rise in less-skilled female LFP would not translate into new jobs, and even some of those who were already working may become dismissed, resulting in higher female unemployment rates. Denoting employment rates at time t by E_t^{ij}, where $i = f, m$ denotes gender and $j = u, s$ whether the individual is unskilled or skilled, we can then outline the main testable implications of our analysis as follows:

- **Hypothesis I**: *Gender differences in job destruction rates among less-skilled workers.*

– **Hypothesis I$_m$**: If the recession has mainly hit low-paid jobs in *male* labour-intensive industries, this implies that $\Delta E_t^{mu} < 0$, while $\Delta E_t^f = \Delta E_t^{ms} \approx 0$. As a result, male selection becomes positive during the slump ($\Delta b_t^m > 0$) while female bias does not change ($\Delta b_t^f = 0$). From equation (4), this implies that $\Delta RG_t > \Delta PG_t$.

– **Hypothesis I$_f$** : If the recession has mainly hit low-paid jobs in *female* labour-intensive industries, then it holds that $\Delta E_t^{fu} < 0$, while $\Delta E_t^m = \Delta E_t^{fs} = 0$. As a result, female selection becomes even more positive ($\Delta b_t^f > 0$) during the slump, while male selection does not change ($\Delta b_t^m = 0$). Thus, from (4), $\Delta RG_t < \Delta PG_t$.

- **Hypothesis II**: *Added-worker effect and creation/destruction of female less-skilled jobs.*

  – **Hypothesis II$_{fe}$.** When less-skilled female LFP increases and LS is the binding constraint for this type of women (as in the added-worker effect), they will enjoy job gains, i.e. $\Delta E_t^{fu} > 0$. Thus, female selection becomes *less* positive ($\Delta b_t^f < 0$) during the slump. Moreover, if Hypothesis I$_m$ also holds ($\Delta E_t^{mu} < 0$), male selection (previously absent) becomes positive ($\Delta b_t^m > 0$). Hence, from (4), $\Delta RG_t \gg \Delta PG_t$.

  – **Hypothesis II$_{fu}$.** When less-skilled female LFP increases and LD is the binding constraint for thus type of women, they will experience job losses, i.e. $\Delta E_t^{fu} < 0$. Thus, female selection becomes even *more* positive ($\Delta b_t^f > 0$) during the slump. Moreover, if Hypothesis I$_m$ also holds ($\Delta E_t^{mu} < 0$), male selection remains positive ($\Delta b_t^m > 0$), and therefore ΔRG_t could be larger or smaller than ΔPG_t, depending on the relative sizes of the positive changes in selection.

Notice that, while Hypothesis I can be seen as an individual hypothesis regarding whether job destruction affects mostly either men (subscript m) or women (f), Hypotheses II + I$_m$ is a joint hypothesis that combines male job destruction in both instances with either female employment gains (fe) or higher female unemployment (fu) in response to an increase in female less-skilled LFP rates. Two key conclusions arise from this analysis. First, if the adverse employment shock during the Great Recession translated into large job losses among less-skilled men, positive male selection appears as a distinct possibility that should be taken into account when computing PGs. Second, the relative pattern of RGs and PGs during the crisis is highly contextualised, depending on both the differential labour demand responses for men and women and their (endogenous) labour supply decisions.

3 Data

In order to compute both RGs and PGs, we use the European Statistics on Income and Living Conditions (EU-SILC) data set.[14] This is an unbalanced household-based panel survey which has replaced the European Community Household Panel Survey (ECHPS) as the standard data source for many gender wage gap studies in Europe, including the aforementioned Olivetti and Petrongolo (2008). It collects comparable multidimensional annual micro-data on a few thousand households per country, starting in 2004. Our core sample focuses on the Great Recession and covers the period 2007-2012, where 2007 captures the pre-crisis situation. However, data for a longer period (2012-2016) will be used to check how our main theoretical implications change once the recovery phase started.

The countries in our sample are classified in two groups: (i) "Southern EU": Greece, Italy, Portugal and Spain, and (ii) "Rest of EU": Austria, Belgium, Denmark, Finland, France, Ireland, The Netherlands, UK, and Norway. Within the latter, in some instances we distinguish among three blocks: *Continental* EU (Austria, Belgium, France, and The Netherlands), *Nordic* (Denmark, Finland and Norway), and *Anglosaxon* (Ireland and the UK).[15]

We restrict our sample to individuals aged 25-54 as of the survey date, and we use self-defined labour market status to exclude those in self-employment, full-time education, and military service.[16] To derive hourly wages, we follow a similar methodology to Engel and Schaffner (2012). A detailed account of this procedure is provided in the Appendix C.

The educational attainment categories (no college and college) correspond to ISCED 0-4 and 5-7, respectively. Descriptive statistics are reported in the Online Appendix A. Finally, throughout the empirical analysis, observations are weighted using population weights when available.[17]

[14]Existing literature using EU-SILC data for international comparisons of gender gaps includes Christofides et al. (2013), who use OLS and quantile regressions to document the differences in the gender gap across the wage distribution in a number of countries.

[15]It is noteworthy that Germany is not included in our sample due to lack of longitudinal information in EU-SILC on several key variables affecting wages. Moreover, though Norway is only an associated member of the EU, for simplicity we will refer to it and the remaining full member estates as EU countries,

[16]One of the shortcomings of the EU-SILC data is that income information is only available for the income reference period while labour market status and additional variables are recorded at the moment of the interview during the survey year, which for most countries does not cover the same period. In fact, the income reference period corresponds to the previous calendar year for all countries except the UK (where the income reference period is the current year) and Ireland (where the income reference period is the 12 months preceding the interview).

[17]Specifically, we use personal base weights, PB050. For Denmark, Finland, Sweden and The Nether-

Figure 1: Labour market attachment by gender, 2007-2012.

Before proceeding to the results, it is convenient to consider gender differences in the LFP and employment responses to the downturn. As shown in Figure 1a–where changes in female LFP rates (in pp., vertical axis) during the crisis are plotted against changes in male LFP rates (in pp., horizontal axis)–, most EU countries exhibit a much larger rise in female LFP than men's since 2007 (i.e., at the beginning of the recession), with Finland and Ireland being the exceptions. Yet, as stressed earlier, higher LFP by women may not necessarily translate into female employment gains during the recession. According to Figure 1b– where changes in female employment rates (in pp.,vertical axis) are displayed against the corresponding changes in male employment rates (in pp., horizontal axis)–, both turn out to be negative in almost half of the countries under consideration.[18] As can be seen, Greece, Ireland, Portugal and Spain exhibit much larger drops in male, as compared to female, employment rates (points above the 45^o line), capturing large job destruction in their male-intensive industries. However, even within Southern EU countries, there are interesting diverging patterns. For example employment changes in Italy are more muted than in the other three members of this block. By contrast, the Rest of EU countries exhibit much fewer male and female job losses (with the exception of Denmark and Ireland, which also experienced the bursting of housing bubbles).

When LFP and employment changes are analyzed by workers' educational attain-

lands, income data is only available for selected respondents. We use personal base weights for selected respondents, PB080, for these countries. Personal weights are not available for Norway and Ireland.

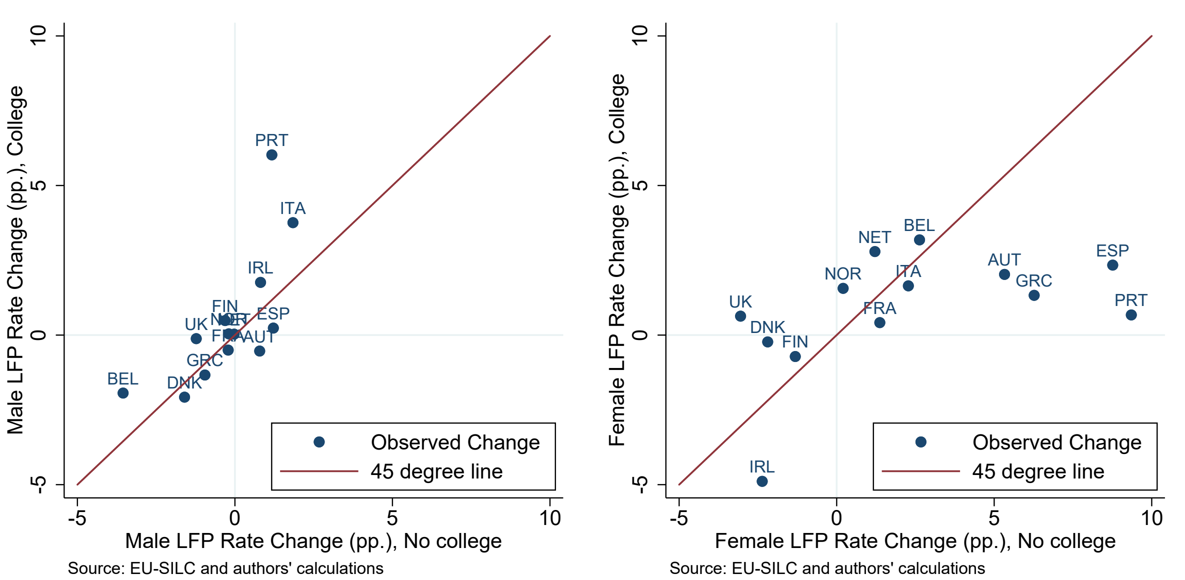
[18]Employment rates are defined as the ratios between employment and the labour force.

ment (for males in Figure 2a and 3a and for females in Figure 2b and 3b), it becomes clear that the fall in employment among less-educated (no-college) male workers has been much more pronounced. This has been particularly the case not only in Ireland and Spain, as a result of the collapse of their real estate sectors, but also in Greece, following the sovereign debt crisis this country suffered. Likewise, regarding participation, it can be seen that most of the gains in LFP in Southern EU countries are due to married females with lower educational attainments, in line with the added-worker hypothesis outlined above. Overall, we take this preliminary evidence as providing considerable support to the mechanism underlying Hypothesis II in Section 2.2.

Figure 2: Changes in LFP by gender and education, 2007-2012.

(a) Males

(b) Females

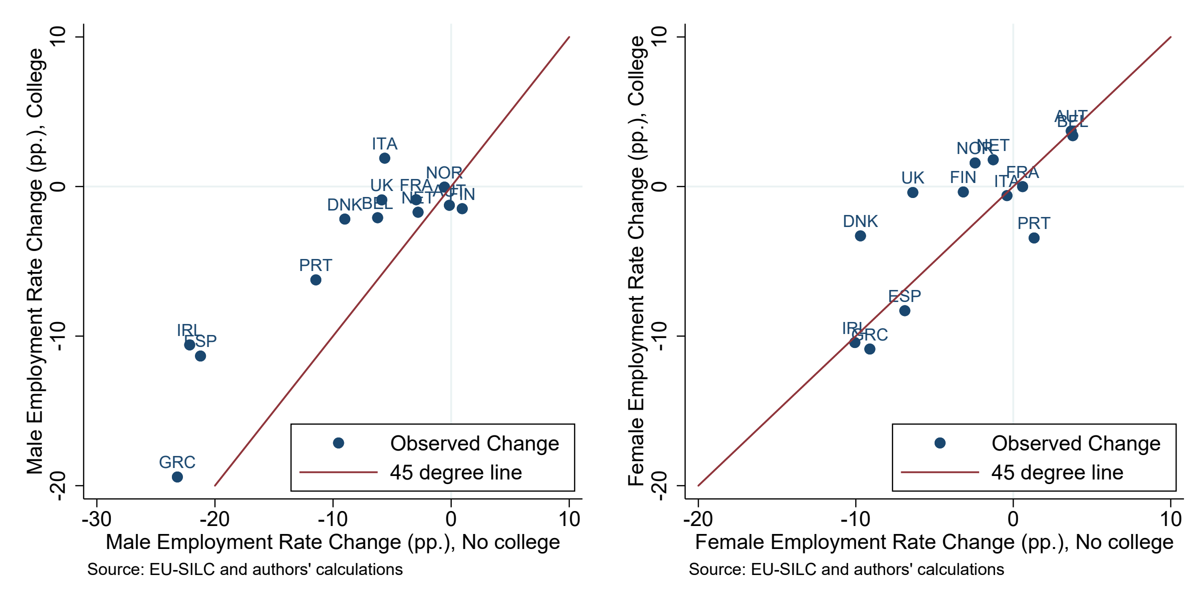


4 Econometric methods

In this section we describe the two econometric procedures used to test the main hypotheses discussed above on how changes in selection biases by gender have translated into changes in RGs and PGs during the downturn and the subsequent recovery. Both procedures provide corrections for the selection biases which arise in the estimation of standard wage *mincerian* regressions based on reported employees' wages, as in (1), when those who are employed exhibit different potential wage distributions than the non-employed ones.

Figure 3: Changes in employment rates by gender and education, 2007-2012.

(a) Males (b) Females



4.1 Imputation around the median

As discussed in Olivetti and Petrongolo (2008), the imputation around the median estimator uses a transformed dependent variable which equals w_{it} for those who are employed at time t, $L_{it} = 1$, and some arbitrary (low or high) imputed value, $\underline{w}_t$ and $\overline{w}_t$ respectively, for those in the non-employment, $L_{it} = 0$.[19] The main insight behind this procedure is that, contrary to the mean, the observed median of the distribution of observed and imputed wages yields an unbiased estimator of the true median of potential wages insofar as the missing observations are imputed on the correct side of the median.[20]

A small number of observable characteristics, X_i, is used to make assumptions about the position of the imputed wage with respect to the median of the gender-specific wage distribution. We define a threshold value for X_i below which non-employed workers would earn wages below the gender-specific median, and another threshold value above which individuals would earn above-median wages.

Specifically, our core specification relies on standard human capital theory, and therefore uses both observed *educational* attainment and labour market *experience* ("Imputation on EE") to predict the position of the missing wages. The imputed

[19] As noted earlier, this approach is closely related to Johnson et al. (2000) and Neal (2004).

[20] To simply illustrate this property, suppose that the true realization of the wage for five individuals (ranked in increasing order) is $\{1, 3, 5, 6, 10\}$ and that the first and last observations (i.e. 1 and 10) happen to be missing. If imputations for these missing values are equal to 2 and 29, the new estimated median will remain unbiased (=5) whereas the mean will be severely biased (changing from 5 to 8).

dependent variable is set to equal a low value, $\underline{w}_t$, if an individual has low education and limited labour market experience, and a high value, $\overline{w}_t$, when an individual is highly educated and has extensive labour market experience.[21] In addition, to take into account non-employed individuals with low (high) education and long (limited) experience, we follow Olivetti and Petrongolo (2008) in fitting a probit model for the probability that the wage of employed individual lies above the gender specific median, based on education, experience (and its square), and the interaction of both variables. In this way, predicted probabilities for the non-employed are obtained. An imputed sample using all individuals in the sample is then constructed using these predicted probabilities as sample weights.

Since these imputation methods for missing wages follow an educated guess, we provide two procedures to assess their goodness of fit. Following Olivetti and Petrongolo (2008), the first procedure (Goodness Method 1) makes use of wage information for non-employed individuals from other waves in the panel in which individuals report having received a wage. In this way, it is possible to check whether the relative position as regard the median of imputed wages using information of the aforementioned demographics corresponds to the actual one when the wage is actually observed. We propose a second method (Goodness Method 2) which considers all employed workers and computes the fraction of those with wage observations on the correct side of the median as predicted by the imputation rule.

Finally, as an alternative imputation method which does not rely on using somewhat arbitrary assumptions based on observable characteristics, as above, we follow Olivetti and Petrongolo (2008) in exploiting the panel nature of the data. In particular, for all those not employed in year t, we recover their wages from the nearest wave, t'. The identifying assumption is that the wage position with respect to the median when an individual is not employed can be proxied by the observed wage in the nearest wave. While this procedure, labelled "Imputation on Wages from Other Waves" ("WOW") relies exclusively on wages, and therefore has the advantage of incorporating selection on time-invariant unobservables, it has the disadvantage of not providing any wage information on individuals who never worked during the sample period. Thus, this method will be relatively conservative in assessing the effects of positive selection in the countries with a relatively low labour market attachment of

[21]This methodology implies a trade-off between the likelihood of imputing an individual's wage correctly (which increases with the number of covariates) and the share of observations for which we cannot ascertain the position relative to the mean (which also increases with the number of covariates). Following Olivetti and Petrongolo (2008) we only use two explanatory variables, which provide a reasonable compromise. We performed robustness tests with a larger number of covariates as discussed in Table A4 in the Appendix.

females. Moreover, since the panel dimension of our data set is relatively short, this procedure yields less satisfactory results in terms of goodness of fit.[22] Consequently, we relegate its results to the Online Appendix.

4.2 Quantile selection models

As acknowledged above, estimation of selection biases using imputations of missing values around the median wage may be problematic in a context of short panels (like ours) and a large fraction of people who never worked throughout the panel. Hence, it seems convenient to compare the results yielded by the imputation rules with those stemming from a more conventional control-function approach which takes advantage of the longitudinal structure of the data.

Recalling that the key ingredients of our theoretical argument are that male job destruction and changes in female LFP and employment have mostly affected less-skilled workers (i.e., those in the lower part of the wage distribution), it seems natural to implement selection corrections in a *quantile* regression framework. If our interpretation is correct, the insight behind this approach is that we should observe more positive selection biases at the lower quantiles of the observed male wage distribution than at the other quantiles. By the same token, selection bias should be more positive in the female wage distribution if the adverse shifts in LD dominate the favourable shifts in LS (due to the added-worker effect) or, conversely, less positive when LS acts as the binding constraint. To do so, we apply the methodology recently developed by Arellano and Bonhomme (2017; AB hereafter).

In AB's (2017) quantile model, sample selection is modeled via a bivariate cumulative distribution function, or *copula*, of the errors in the wage and the selection equations. In particular, the following selection model is considered for the latent (potential) wage of each individual of gender g ($g = m, f$), labeled as w^{*g}, and their decision to accept a job:

$$
\begin{aligned}
w^{*g} &= X^{g'} \beta^{g}(U), && (8)\\
D^{g} &= \mathbf{1}\{V \leq p(Z^{g})\}, && (9)\\
w^{g} &= w^{*g} \text{ if } D^{g} = 1, && (10)
\end{aligned}
$$

where $\beta^{g}(U)$ in (8) is increasing in a random variable uniformly distributed on the unit interval, U, independent of the set of covariates determining wages, X^{g}, such

[22] The longitudinal component of EU-SILC allows to follow each household for four years, with the exception of France, where each household is followed for eight consecutive years.

that $Q(\tau, X^g) = X^{g'}\beta^g(\tau)$ is the τ-th conditional quantile of w^{*g} given X^g. Moreover, (9) represents the selection equation where $\mathbf{1}\{\cdot\}$ is an indicator function, while $Z^g = (X^g, B^g)$, such that B^g are those extra covariates which appear in the participation equation but not in the wage equation; finally V is the rank of the error term in this equation, which is also uniformly distributed on the support $(0, 1)$. Assuming that (U, V) is jointly statistically independent of Z^g given X^g, denoting the c.d.f. of (U, V) as $C(u, v)$, and finally defining $p(Z^g) = \Pr(D^g = 1| Z^g) > 0$, the presence of dependence between U and V is the source of the sample selection bias. In particular, this dependence is captured by $G(\tau, p; \rho^g) = C(\tau, p; \rho^g)/p$ which is the the conditional copula of U given V, defined on $(0, 1) \times (0, 1)$. In this respect, notice that a negative copula means positive selection since individuals with higher wages (higher U) tend to participate more (lower V) and, conversely, a positive copula implies negative selection.

Then, AB (2017) show that

$$\beta^g(\tau) = \arg\min_{b(\tau)} E\left[\left(D^g(G_{\tau Z^g}(w^g - X^{g'}b^g(\tau))^+ + (1 - G_{\tau Z^g})(w^g - X^{g'}b^g(\tau))^-\right)\right],$$

where $a^+ = \max(a, 0)$, $a^- = \max(-a, 0)$, and $G_{\tau Z^g} = G(\tau, F^{-1}(z^{g\prime}\gamma^g); \rho^g)$ denotes the rank of $X^{g'}\beta^g(\tau)$ in the selected sample $D^g = 1$, conditional on $Z^g = z^g$. Since the above optimization problem is a linear program, given γ^g and ρ^g, the parameters $\beta^g(\tau)$ can be estimated in a τ-by-τ fashion by solving linear programs, just like with the conventional check function in standard quantile regressions (see Koenker and Bassett Jr (1978)). The only difference is that, in quantile regressions, τ replaces $G_{\tau Z^g}$; in other words, correcting for selection in quantile regressions implies that one needs to rotate the check function depending on Z^g. AB (2017) suggest two previous steps in order to compute $\beta^g(\tau)$: estimation of the propensity score $p(Z^g)$ in (9) (e.g., via a probit model) and estimation by means of a grid-search GMM of the degree of selection (i.e., the copula parameter ρ^g) using a Frank copula, though they also cover more general cases.

5 Empirical Results

In this section we present the main results from the two econometric approaches discussed above: (i) imputations around the median, and (ii) selection bias corrections in quantile regressions. For brevity, in (i) we focus exclusively on the evidence drawn from imputation on EE, which yields the best goodness-of- fit results (see below). The corresponding results for the imputation rule based on wages from other waves can be found in the Online Appendix.

5.1 Imputation around the median wage

Table 1 presents results for our EE imputation method. Recall that two education categories are being considered: those individuals with upper secondary education or less are considered to be "less-educated", while those with some tertiary education are defined as "high-educated". Similarly, we define as "low (high) experienced individuals" those with less than (at least) 15 years of work experience.

Table 1: Median Wage Gaps under Imputation on Education and Experience 2007-2012

	Levels in 2007						Changes over 2007-2012					
	Raw Wage Gap	Potential Wage Gap	Selection Bias M	Selection Bias F	Employment Rate M	Employment Rate F	Raw Wage Gap	Potential Wage Gap	Selection Bias M	Selection Bias F	Employment Rate M	Employment Rate F
Greece	.182	.445	.025	.288	.887	.540	-.089	-.067	.059***	.081**	-.220	-.081
Italy	.035	.277	.034	.276	.863	.559	.051	.024	.010***	-.017	-.045	.006
Portugal	.172	.223	.036	.087	.875	.707	-.059	-.105	.024**	-.021*	-.106	.010
Spain	.131	.248	.017	.134	.889	.638	-.020	.002	.066***	.088***	-.179	-.064
Southern	**.130**	**.298**	**.028**	**.196**	**.879**	**.611**	**-.030**	**-.037**	**.040**	**.033**	**-.138**	**-.032**
Austria	.189	.300	.012	.124	.893	.695	.015	-.007	-.007	-.029	-.002	.036
Belgium	.074	.142	.022	.090	.897	.732	-.019	-.060	.003	-.038***	-.046	.036
France	.114	.161	.008	.055	.917	.808	.005	-.019	.010***	-.014*	-.020	.012
Netherlands	.158	.199	.004	.044	.963	.823	-.048	-.038	-.001	.009	-.023	.007
Continental	**.133**	**.201**	**.011**	**.079**	**.917**	**.765**	**-.012**	**-.031**	**.001**	**-.018**	**-.023**	**.023**
Ireland	.170	.303	.020	.153	.862	.674	-.039	-.069	.003	-.026**	-.151	-.070
UK	.247	.301	.011	.065	.934	.804	-.064	-.045	.009**	.028*	-.032	-.024
Anglosaxon	**.208**	**.302**	**.015**	**.109**	**.898**	**.739**	**-.052**	**-.057**	**.006**	**.001**	**-.091**	**-.047**
Denmark	.116	.126	-.002	.009	.966	.905	-.036	-.048	-.012***	-.023***	-.066	-.065
Finland	.203	.221	.016	.035	.862	.831	-.049	-.086	.015	-.022**	.005	-.011
Norway	.154	.161	.006	.013	.969	.931	.020	.003	-.006***	-.023***	-.002	-.002
Nordic	**.158**	**.170**	**.007**	**.019**	**.932**	**.889**	**-.022**	**-.044**	**-.001**	**-.023**	**-.021**	**-.026**

Source: EU-SILC and authors' calculations. Note: Selection bias = an increase in observed wage due to selection. Wage imputation rule: Impute wage $<$ median when non-employed and education $\leq$ upper secondary and experience $<$ 15 years; impute wage $>$ median when non-employed and education $\geq$ higher education and experience $\geq$ 15 years. All raw and potential wage gaps are significant at the 1% level. *, **, *** denotes statistical significance at 10, 5 and 1 percent levels.

Table 1 presents the results for the four Southern EU and the nine Rest of EU countries split into the three blocks as defined above (Anglosaxon, Continental EU and Nordic). In the left panel we report the RGs and PGs in levels (log. points), as well as the selection biases and employment rates by gender in 2007 (at the onset of the Great Recession).[23] Selection biases are measured as percentage point (pp.)

[23] In the Online Appendix (see Table A2 in section A), we present evidence on how female LFP rates have increased in the four Southern EU economies and in a few Rest of EU countries, and that this rise has been much higher among less-educated women everywhere.

changes in the median wage once missing wages are imputed. The right panel in turn shows the corresponding changes of these variables between 2007 and 2012 (during the Great Recession) with asterisks denoting statistical significance of changes in selection biases.[24] To help interpret our findings, it is useful to recall from equation (4) that changes in PG equal changes in RG plus changes in the female bias minus changes in the male bias, i.e. $\Delta PG_t = \Delta RG_t + \Delta b_t^m - \Delta b_t^f$.

In agreement with the findings of Olivetti and Petrongolo (2008), the left panel of Table 1 shows that, before the slump, Southern EU countries exhibited on average a much lower RGs (13 pp.) than PGs (30 pp.), as well as higher gender employment gaps in favour of men than the Rest of EU countries. With regard to RGs, it can be seen that only the Continental EU countries exhibit a similar gap to the one in Southern EU countries while, in relation to PG, only the Anglosaxon countries fare similarly. As a result, the most salient features of this evidence can be summarised as follows: (i) the difference $PG - RG$ is much higher (17 pp.) in Southern EU than in the Rest of the EU (5 pp. on average), (ii) the female selection bias is also much higer in Southern EU (19.6 pp.), broadly explaining the difference of 17 pp. between PGs and RGs, as it is also the case for male selection biases, which are almost three times larger in the South than elsewhere (2.8 pp. against 1.1 pp.), in agreement with the lower aggregate employment rates in this block of countries.

The right panel of Table 1 indicates that, in line with the evidence in OECD (2014), RGs declined in most countries over the Great Recession, with Italy being the noticeable exception. However, our findings indicate that the slump also involved considerable changes in selection which triggered an even larger drop in PGs for a majority of countries. Hence, in relative terms, this means that changes in RGs overestimate changes in PGs over the downturn, which contrasts with the results reported by Olivetti and Petrongolo (2008) before the recession, when PGs exceeded RGs in several EU countries. Two findings explain this new pattern. First, the male selection bias became more positive, notably in Southern EU where it rose by almost 6.6 pp.[25] The largest increases in male selection seem to have taken place in those countries where the decline in male employment was largest. Second, the evolution of female selection varies across countries. We can observe two different patterns, with two thirds of the countries exhibiting a reduction in female selection and the remaining one third ex-

[24]To test for the null of no selection changes between 2007 and 2012, we run a gender-specific median quantile regression of both latent and raw wages on a constant, a dummy for latent wages, a dummy for 2012 and an interaction of the two. The standard errors are bootstrapped and clustered by year, and population weights are used in the regression. The t-ratio on the interacted coefficient tests for the null of no changes in selection biases. The same procedure is applied in Table 2 below to test the null hypothesis of no change between 2012 and 2016.

[25]There are a few exceptions, notably Denmark and Norway, where male selection fell.

Table 2: Median Wage Gaps under Imputation on Education and Experience 2012-2016

	Changes over 2012-2016					
	Raw Wage Gap	Potential Wage Gap	Selection Bias M	Selection Bias F	Employment Rate M	Employment Rate F
Greece	-.030	-.130	-.026	-.126***	.013	.021
Italy	-.028	.063	.000	.091**	-.003	-.011
Portugal	.000	.027	-.050***	-.023**	.060	.039
Spain	-.047	-.078	-.040***	-.072***	.050	.075
Southern	**-.026**	**-.029**	**-.029**	**-.032**	**.030**	**.031**
Austria	-.004	-.026	-.005	-.027	-.005	.019
Belgium	.017	.028	-.012*	-.001	.018	.022
France	-.006	.007	-.003	.010*	-.012	.002
Netherlands	.073	.051	.001	-.021**	-.005	-.008
Continental	**.020**	**.015**	**-.005**	**-.010**	**-.001**	**.009**
Ireland	-.054	-.117	.021***	-.042***	-.711	-.604
UK	-.005	-.024	-.008*	-.027	.040	.002
Anglosaxon	**-.030**	**-.071**	**.007**	**-.034**	**-.335**	**-.301**
Denmark	.026	.017	.029***	.019**	.009	.071
Finland	-.021	-.015	-.011*	-.004	-.029	-.009
Norway	-.013	-.013	.002	.002*	-.004	.004
Nordic	**-.003**	**-.004**	**.007**	**.006**	**-.008**	**.022**

Source: EU-SILC and authors' calculations. Note: Selection bias = an increase in observed wage due to selection. Wage imputation rule: Impute wage $<$ median when non-employed and education $\leq$ upper secondary and experience $<$ 15 years; impute wage $>$ median when non-employed and education $\geq$ higher education and experience $\geq$ 15 years. *, **, *** denotes statistical significance at 10, 5 and 1 percent levels.

periencing an increase. These differences are clearly illustrated within the Southern EU block. Female selection biases experienced substantial reductions in Portugal (-2.1 pp.) and to a lesser extent in Italy (-1.7 pp. but not statistically significant), the only two countries in the South where female employment rates fared relatively well during the crisis. On the other hand, female employment rates plummeted in Greece and Spain (by -8.1 pp. and -6.4 pp., respectively) leading to growing (more positive) female selection biases (changes above 8 pp.). Another country where female selection bias has increased markedly is the UK (2.8 pp.), due to its drop in employment being largely driven by the dismissals of young, and hence below-median-wage workers.[26]

[26] Male employment changes in the UK over the recession have been characterised by both a decline in youth male and female employment, that tended to increase positive selection among men and women, and job destruction in the male-dominated and high-paid financial sector; see Bell and

These differences indicate the importance of the supply and demand forces discussed above for female selection. Consequently, the evolution of the PG depends on which of these dominates, as we can see in the contrasting findings between Portugal, where the PG fell by 10 pp., and Spain, where it remained stable. Further details on the differences between these two countries will be discussed further below.

Table 2 presents the changes in the variables reported in Table 1 during the recovery period (2012-2016). It should be noticed that, due to the sovereign debt crisis, recovery was delayed by one or two years in some of the Southern EU countries (see the case of Portugal below). As can be observed, RGs and PGs decrease in most countries, with the exception of the Continental EU block and Denmark where they go up. The most salient finding, however, is that the increase in male selection during the slump is partially reversed during the subsequent recovery, particularly in Southern EU. This is explained by a higher demand for less-skilled male labour once growth took off.[27] Likewise, female selection biases that had grown in these countries during the crisis, now decline, being fuelled by higher demand for less-skilled female labour during the upturn. A notable exception is Portugal, where positive female selection declined both during the crisis and the recovery. This indicates that the higher demand for less-skilled women during the crisis remained strong when it was over (see discussion further below), whereas in other countries a parallel rise in the demand for high-skilled women took place. Notwithstanding this exception, we take the reversed signs of selection biases from the downturn to the upturn as supportive evidence of their business-cycle nature. Male selection is countercyclical, increasing in the downturn and falling in the upturn, while female selection also follows the business cycle, although its sign depends on the interplay between labour supply and demand

To provide a graphical illustration of how the LS and LD constraints operate, we focus on the experiences of Portugal and Spain, the two neighbouring Iberian countries badly hit by the recession. The left panels in Figures 4 and 5 display the estimated selection biases by gender in each country from 2007 to 2016. For comparison, the right panels present employment rates by gender. As can be seen, male selection biases (dashed lines) surged in both countries during the Great Recession (i.e. the LD constraint binds for men). Yet, striking differences appear as regards female selection biases: while the positive female selection declines in Portugal (the LS constraint binds for women), it goes up in Spain (the LD constraint binds for women). These contrasting patterns are related to the fact that, while only male employment

Blanchflower (2010). The joint effect of these two forces is a negligeable change in male selection and an increase in female selection.

[27] Note that there is no reversal in France, where the lack of change in male employment is accompanied by a small and insignificant change in male selection.

Figure 4: Selection bias and employment rates by gender, Portugal, 2007-2012.

(a) Selection Bias (b) Employment Rate

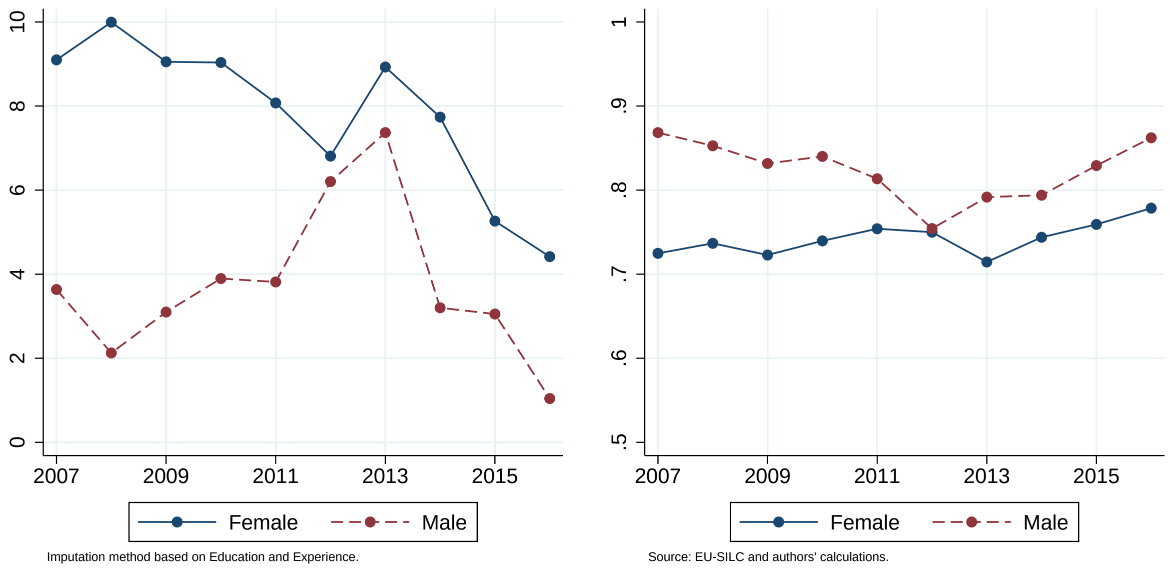
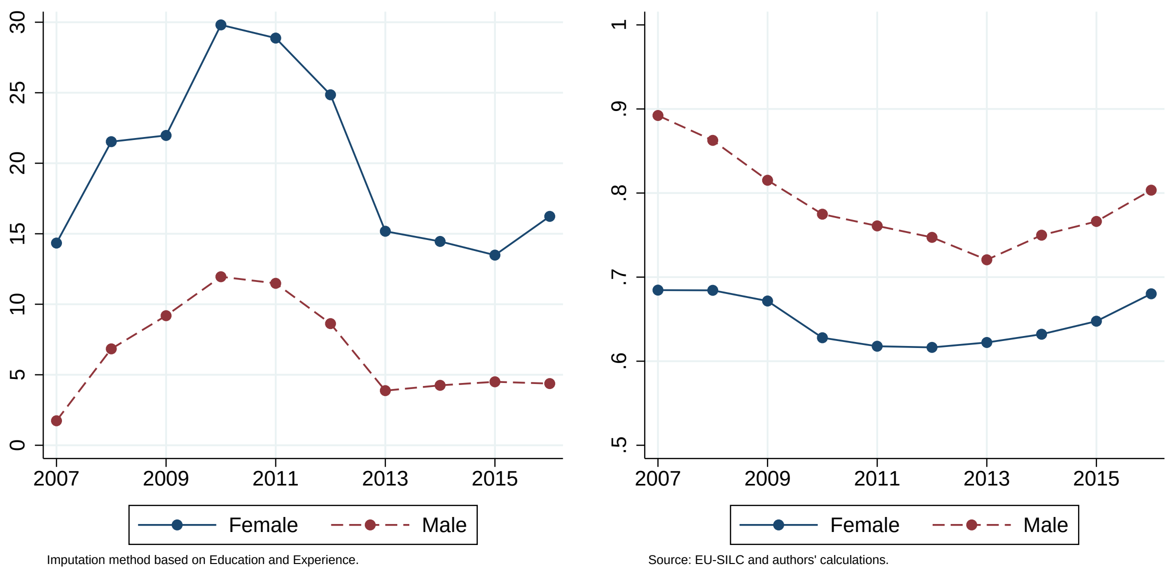


Figure 5: Selection bias and employment rates by gender, Spain, 2007-2012.

(a) Selection Bias (b) Employment Rate



fell in Portugal, job destruction hit both female and male jobs in Spain. The worse performance of the Spanish labour market during the downturn can be attributed to two factors. First, before the crisis, Spain had less wage flexibility and more generous unemployment benefits than Portugal (see Bover et al. (2000)). Second, Spain had a much higher rate of female temporary workers (above 30% of employees), most of which were massively destroyed once the crisis hit. The much higher rate of fixed-term contracts in Spain than in Portugal had to do both with its much larger weight of employment in the construction and ancillary sectors (reaching 15% of total employment in 2007) and a higher gap between firing costs (including red-tape costs) for workers under permanent (open-ended) and temporary contracts, which inhibited direct hiring of workers under permanent contracts and *temp-to-perm* contract conversions in this country(see Dolado, 2016). As already pointed out above, once the recovery started these selection patterns changed. Male selection biases declined in both countries as a result of the recovery of unskilled male employment (the LD constraint was less binding for men), particularly in the hospitality and retail sectors. Female selection biases went down drastically in Spain (signalling that the LD constraint for women became weaker as well), while they follow a non–monotonic pattern in Portugal: first up and then down (as during the downturn). The initial hike in female selection in Portugal was due to greater hiring of more educated women at the beginning of the upturn, which later on was more than offset by a much higher demand for less-skilled women as a result of a boom in tourism resulting from political instability in competing destination countries located in Northern Africa (which also occurred in Spain).

Table 3: Rate and Goodness of Imputation on Education and Experience

	2007						2012						2016					
	Imputation Rate		Goodness Method 1		Goodness Method 2		Imputation Rate		Goodness Method 1		Goodness Method 2		Imputation Rate		Goodness Method 1		Goodness Method 2	
	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F
Greece	.42	.69	.88	.88	.84	.85	.37	.57	.78	.78	.80	.79	.38	.54	.85	.70	.80	.83
Italy	.53	.73	.82	.72	.70	.69	.52	.71	.82	.76	.73	.73	.62	.72	.78	.77	.71	.74
Portugal	.38	.55	.59	.61	.71	.76	.44	.43	.63	.53	.65	.77	.34	.46	.73	.51	.81	.86
Spain	.39	.63	.66	.70	.75	.79	.54	.65	.72	.65	.73	.77	.39	.61	.46	.73	.76	.75
Southern	**.43**	**.65**	**.74**	**.73**	**.75**	**.77**	**.47**	**.59**	**.74**	**.68**	**.73**	**.76**	**.43**	**.58**	**.70**	**.68**	**.77**	**.79**
Austria	.32	.51	.85	.76	.76	.81	.32	.48	.77	.70	.83	.80	.42	.55	.73	.84	.81	.81
Belgium	.42	.56	.86	.84	.80	.80	.50	.60	.84	.77	.78	.82	.54	.65	.81	.85	.83	.83
France	.42	.58	.83	.77	.80	.79	.46	.61	.76	.72	.81	.80	.52	.61	.75	.74	.80	.78
Netherlands	.34	.58	.77	.83	.81	.75	.46	.58	.80	.76	.81	.79	.42	.52	.49	.62	.87	.82
Continental	**.38**	**.56**	**.83**	**.80**	**.79**	**.79**	**.43**	**.57**	**.79**	**.74**	**.81**	**.80**	**.47**	**.58**	**.70**	**.76**	**.83**	**.81**
Ireland	.37	.53	.85	.80	.83	.81	.39	.45	.70	.68	.73	.76	.46	.46	.57	.65	.73	.75
UK	.40	.51	.54	.70	.75	.74	.42	.55	.89	.75	.75	.71	.59	.55	.81	.64	.75	.74
Anglosaxon	**.39**	**.52**	**.69**	**.75**	**.79**	**.77**	**.41**	**.50**	**.80**	**.71**	**.74**	**.73**	**.52**	**.51**	**.69**	**.64**	**.74**	**.74**
Denmark	.23	.46	.55	.82	.67	.76	.37	.28	.09	.64	.69	.68	.41	.41	.78	.85	.78	.80
Finland	.60	.41	.80	.73	.75	.78	.53	.45	.67	.60	.76	.75	.50	.45	.84	.53	.74	.72
Norway	.37	.38	.66	.70	.75	.79	.35	.41	.72	.65	.73	.77	.39	.45	.46	.73	.76	.75
Nordic	**.40**	**.42**	**.67**	**.75**	**.73**	**.78**	**.42**	**.38**	**.49**	**.63**	**.72**	**.73**	**.43**	**.44**	**.70**	**.70**	**.76**	**.76**

Source: EU-SILC and authors' calculations. Note: Wage imputation rule: Impute wage $<$ median when non-employed and education $\leq$ upper secondary and experience $<$ 15 years; impute wage $>$ median when non-employed and education $\geq$ higher education and experience $\geq$ 15 years. Imputation Rate = proportion of imputed wage observations in total non-employment. Goodness Method 1 = proportion of imputed wage observations on the same side of the median as wage observations from other waves in the panel. Goodness Method 2 = proportion of employed workers on the same side of the median as predicted by the imputation rule.

Lastly, a brief comment is due on the reliability of the results obtained by using the imputation on EE rule. Table 3 reports results on our two above-mentioned measures of goodness of fit, computed for men and women separately, for the years 2007, 2012 and 2016. We report both the imputation rates for each year and the share of imputations that place the individual on the correct side of the median. As can be inspected, both measures indicate a satisfactory fit for about 75% of the individuals of either gender in our sample. Furthermore, there is no indication that we do a better job in imputing female than male missing wages.[28]

5.2 Quantile regressions

Using the AB's (2017) method described above, we estimate wage quantile regressions separately for male and female wages, allowing for sample selection using EU-SILC unbalanced panel data for 2007-2012. The dependent variable is the log-hourly wage, covariates X^g contain experience and its square, marital status, the two education indicators mentioned earlier, a set of dummies for region of residence (NUTS) in each country, and year effects. As for B^g (determinants of participation that do not affect wages directly), we take the number of children in 6 age brackets and their interaction with marital status, non-labour income and a dummy variable of whether the corresponding spouse lost his/her job in the previous year interacted with marital status (added-worker effect or AWE in short). Notice that, if the latter effect holds, we would expect a positive effect of this variable on the probability of participating in the labour market. Unfortunately, as discussed earlier(in footnote 16), the AWE indicator is not available for Nordic countries and The Netherlands, since information on labor market experience in both countries is restricted to a single member of the household and not for both spouses. Thus, these countries are omitted in this sub-section.

Table 4 presents evidence for the nine remaining EU countries where the information requirements to run these quantile regressions is available. For brevity, the reported results correspond to the male and female selection biases for three relevant quantiles at the bottom, centre and upper part of the wage distribution: $\tau = 0.2$, 0.5, and 0.8.

As with the previous approach, the increase in male selection appears relevant in most countries, being stronger at $\tau = 0.2$ than for the other higher quantiles, in line with the much higher job destruction rate for the less-skilled workers than for other higher-skilled groups. The exception to this rule is the Anglosaxon block, where the rise in male selection is stronger at $\tau = 0.5$, and 0.8, possibly due to the dismissals of many young, and hence relatively lower-paid, workers in high-pay sectors such as the

[28] In order to check the robustness of our imputation method, the Online Appendix B reports estimates based on a probit model. The results are qualitatively similar to our findings in Table 1.

Table 4: Quantile Regression Estimates Corrected for Selection

Quantile	Changes in Selection Bias over 2007-2012					
	20		50		80	
	M	F	M	F	M	F
Greece	.178	.151	.068	.093	.088	.063
Italy	.009	-.004	.004	-.001	-.003	.001
Portugal	.031	-.021	.026	.005	.033	-.005
Spain	.113	.086	.082	.058	.050	.039
Southern	**.083**	**.053**	**.045**	**.039**	**.042**	**.025**
Austria	.018	.011	.000	.016	-.003	.035
Belgium	.007	-.034	.002	-.018	-.014	-.044
France	-.011	-.002	.003	-.007	.001	-.009
Continental	**.005**	**-.008**	**.002**	**-.003**	**-.005**	**-.006**
Ireland	.001	.048	.047	-.006	.035	-.038
UK	.036	.026	.032	.027	.037	-.002
Anglosaxon	**.019**	**.037**	**.040**	**.010**	**.036**	**-.020**

Source: EU-SILC and authors' calculations. Covariates in the Participation eqn. are described in the main text. Matlab code at: `https://drive.google.com/file/d/0B13ohL0_ULTDaDE2N0d1ZnEzZ1U/view`

banking and financial industries. Moreover, in agreement with the evidence reported in Table 1, this rise in male selection is much stronger in Southern EU countries (except Italy) than in the other countries where the decline in male employment rates was much less intense. Second, in contrast to the strong rise in Greece and Spain and to a lesser extent in the Anglosaxon block, female selection goes down in Portugal (particularly at $\tau = 0.2$) and in Belgium, supporting the the findings on this issue presented in Table 1.

Table 5 reports the estimated copulas and correlations between the error terms in the wage and participation equations, denoted as corr (U, V). As can be seen, all copulas and correlations are negative over the Great Recession period and, in most instances, copulas turn out to be statistically significant. As discussed earlier, negative copulas imply positive selection which takes places both among men and women. Notice that female selection remains positive by the end of the recession, even in countries where it experienced a sizeable reduction (like in Italy and Portugal), the reason being that it was initially (in 2007) very high and positive.

Finally, two additional empirical findings are worth discussing, though they are

Table 5: Quantile Regression Estimates Corrected for Selection

	Copula		corr(U,V)	
	M	F	M	F
Greece	-4.78***	-3.13***	-0.63	-0.46
Italy	-0.12*	-0.70**	-0.02	-0.12
Portugal	-0.91***	-1.42***	-0.15	-0.23
Spain	-2.19***	-0.86***	-0.34	-0.14
Southern	**-2.00**	**-1.53**	**-0.28**	**-0.24**
Austria	-1.37***	-1.37***	-0.22	-0.22
Belgium	-0.06	-0.30**	-0.01	-0.05
France	-0.12*	-0.36**	-0.02	-0.06
Continental	**-0.52**	**-0.68**	**-0.08**	**-0.11**
Ireland	-0.06	-0.42**	-0.01	-0.07
UK	-0.30**	-0.06	-0.05	-0.01
Anglosaxon	**-0.18**	**-0.24**	**-0.03**	**-0.04**

Source: EU-SILC and authors' calculations. Covariates in the Participation eqn. are described in the main text. *, **, *** denotes statistical significance at 10, 5 and 1 percent levels. Replication codes at: `https://drive.google.com/open?id=0B13ohL0_ULTDMVhBN0s1OXhldWc`.

not reported for the sake of brevity. First, we have checked how selection patterns have changed over time by estimating copulas using cross-section quantile regressions with selection corrections for three specific years: 2007, 2012 and 2016. In general, we find that the male copulas are more negative in 2012 than in 2007, while they are less negative in 2016 than in 2012. This agrees with our earlier evidence regarding an increase of male selection during the recession period and a reduction over the recovery period. As for female selection, the results vary in line with the evidence reported in Table 1. In countries, like Greece, Spain and the UK where female employment went severely down over the downturn, female copulas are more negative in 2012 than in 2007, while the opposite happens for countries like Italy, Portugal, Ireland and those in the Nordic block. Second, we find that the estimated coefficient on AWE in the participation probit equations for men is often negative and statistically insignificant in most countries. By contrast, the corresponding coefficient for women is positive and highly significant, particularly in Southern countries and Ireland, meaning that male job losses triggered higher female LFP. In line with the evidence presented by Bredtmann et al. (2018), this is seemingly consistent with the conjectured added-worker effect for less-educated married women. Overall, we take these results as being fairly in agreement with the previous evidence based on median imputation methods.

6 Interpreting the findings

In view of the previous empirical evidence drawn from the two chosen selection-correction methods, we complete our analysis by providing an overview of how these results fit in the theoretical scenarios laid out in Section 2.2 about the main potential drivers of gender wage gaps in the EU during the Great Recession. Relying on the results in Tables 1 and 4, and Figures 2 and 3, we summarize our interpretation of the evidence in Table 6.

The first conclusion to be drawn is that neither the male (Hypothesis $\mathbf{I_m}$) nor the female version (Hypothesis $\mathbf{I_f}$) of Hypothesis I (i.e. destruction of less-skilled jobs) hold *per se* for any of the countries in our sample. This is because our evidence points to sizeable changes in both male and female selection in parallel, perhaps with the exception of Norway. Hence, from this finding one can infer that the estimated selection biases and the observed employment changes in EU countries should be rationalized through a combination of the individual hypotheses listed in Section 2.2.

Within Southern EU, the patterns for Italy and Portugal conform neatly to the implications of the combined Hypotheses $\mathbf{I_m}$+II_{fe} (i.e. added-worker effect plus large losses in unskilled male employment losses with no major changes in female employment rates), which jointly lead to a substantial reduction (resp. increase) in female (resp. male) selection, so that $\Delta RG > \Delta PG$. By contrast, the patterns in Greece and Spain seem to be better rationalized by the combined Hypotheses $\mathbf{I_m}$+II_{fu} (i.e. added worker effect plus a collapse in both male and female unskilled employment rates), leading to a simultaneous rise in the selection biases for both genders. Since our evidence points out to a larger increase in the female bias than in the male bias, this would imply that $\Delta PG > \Delta RG$ in these two countries.

Among the Rest of EU countries, where employment losses have been much more muted than in Southern EU – except in Denmark and Ireland–, we find two distinct patterns. On the one hand, several countries in the Continental EU block provide nice examples of Hypothesis II_{fe}, i.e. added worker effect plus unskilled female employment gains, leading to a reduction in female selection. Likewise, the substantial drop in male unskilled employment and in the female selection biases in Finland seem best explained by the combined Hypotheses $\mathbf{I_m}$+$\mathbf{II_{fe}}$ (added worker effect with

Table 6: Summary of Findings over the Great Recession

	Consistent Hypotheses			
	I_m	I_f	II_{fe}	II_{fu}
Southern				
Greece	✓			✓
Italy	✓		✓	
Portugal	✓		✓	
Spain	✓			✓
Continental				
Austria			✓	
Belgium			✓	
France	✓		✓	
Netherlands	✓			
Anglosaxon				
Ireland	✓		✓	
United Kingdom	✓			✓
Nordic				
Denmark	✓		✓	
Finland	✓		✓	
Norway			✓	

Notes: Hypothesis I_m (I_f): higher job destruction rate among low-skilled *male* (*female*) workers. Hypothesis II_{fe} : added-worker effect with female employment *gains*. Hypothesis II_{fu} : added-worker effect with female employment *losses*.

female employment gains and male employment losses). Lastly, the findings for the Anglosaxon block are more ambiguous. While the Irish pattern is akin to the one for Italy and Portugal, and so rationalized by more positive male and less positive female selection (the combined Hypotheses $\mathrm{I}_{\mathbf{m}}$+$\mathrm{II}_{\mathbf{fe}}$), the rationalization for the UK experience seems to fit better with a milder version of the more positive selection for both genders (combined Hypotheses $\mathrm{I}_{\mathbf{m}}+\mathrm{II}_{\mathbf{fu}}$) that were previously applied to Greece and Spain, albeit at a much lower scale.

Overall, the existence of positive male selection emerges as a uniform and robust finding in most countries, despite being much more pronounced in Southern EU than in Rest of EU. In relation to positive female section, depending on whether LD or LS shifts dominate, we find instances where it has gone up and others where it goes down. Consequently, rationalization of the these contrasting patterns in female selection call for a combination of factors. Among the Southern EU countries most badly hit by the crisis, it seems that in those economies where female LFP was higher to start with (e.g. in Portugal, whose female LFP rate was close to those prevailing in Continental EU before the crisis), or where crisis has been milder (e.g. Italy or some of the Continental EU countries), the female selection bias has declined. Conversely, countries where female LFP was initially lower and had more dualized labour markets (Greece and Spain), have witnessed a further rise in the female selection bias. Nonetheless, the case of the UK stands out as an exception to this rule, since the female selection bias has increased despite having high female LFP in 2007. Yet, as argued earlier, a potential explanation of this finding is the fact that unemployment in the UK hit young women particularly hard.

7 Conclusions

This paper analyzes how the conventional patterns of female and workers' self-selection into EU labour markets have exhibited relevant changes as a result of the large shifts in labour demand and supply brought about by the Great Recession. Based on a large body of empirical evidence, it has been traditionally assumed that male selection biases were negligible before the crisis, due to high male LFP rates everywhere. By contrast, due to their lower LFP rates (particularly in southern Europe), working women were favourably selected since not many unskilled women worked. Our main hypothesis here is that, if the large job losses experienced during the crisis have mainly affected unskilled male-dominated sectors, then male selection may have become significantly positive over that period. In addition, if non-participating unskilled women had increased their LFP rates due to an added-worker effect during the recession , then female selection should have become less positive than prior to the crisis. However, the overall impact the downturn on the change in the female bias could be a priori ambiguous, since the rise in female labour supply could have been more than offset by adverse labour demand shifts during the recession, in which case the female selection bias could become even more positive than it was before.

Using an imputation technique for the wages of non-participating individuals in EU-SILC datasets for a large group of EU countries, as well as quantile wage regressions corrected for selection biases, our findings yield strong support for the hypothesis of a rise in male selection during the recession. This has been especially the case in the Southern EU economies, and to a lesser extent in France and the UK, where there have been considerable male employment losses in response to the decline of low-productivity industries during the slump. With regard to female selection, our results are mixed . We find that, in line with the added-worker effect, female selection has become less positive (particularly in the Continental EU and Nordic blocks, and in Italy and Portugal), while in other instances (most notably Greece and Spain, but also the UK) it has become even more positive because widespread job destruction has also meant substantial reductions in female employment rates, either for less-educated or less-experienced women.

Our results highlight the importance of correcting for male selection in computing gender wage gaps. For example, according to the EE imputation rule for missing wages, the potential gender gap (PG) in Spain barely changes (0.2 pp.) over the Great Recession once male selection is taken into account. However, had we ignored male selection and only corrected for female selection, as it is traditionally done, the estimated PG would have *increased* by 6.8 pp. Hence, future research about measuring gender gaps might require corrections for the two gender groups.

Given the cyclical nature of changes in selection into the workforce, we also provide evidence about whether changes in PG have reversed over the subsequent recovery period (2012-16). We find that the positive male selection bias during the downturn goes down in most countries during the recovery, as the less-skilled workers who were laid off during the slump regained jobs when employment growth picked up. By the same token, in those countries where the female selection bias has increased (decreased) during the crisis, we find that this bias goes down (up) during the recovery, pointing to a to a favourable labour demand shift for less-skilled (skilled) women, especially in those countries where either group of women faced big employment losses during the slump. Overall, the decrease in the female selection bias is likely to be a long- lasting phenomenon since female LFP is likely to rise in the future at both tails of the skills distribution, in line with the job polarization hypothesis documented by Autor and Dorn (2013) for the US and Goos et al. (2009) for some EU countries.

References

Arellano, Manuel and Stéphane Bonhomme, "Quantile selection models with an application to understanding changes in wage inequality," *Econometrica*, 2017, *85* (1), 1–28.

Autor, David and David Dorn, "The growth of low-skill service jobs and the polarization of the US labor market," *The American Economic Review*, 2013, *103* (5), 1553–1597.

Bell, David NF and David G Blanchflower, "UK unemployment in the great recession," *National institute economic review*, 2010, *214* (1), R3–R25.

Bentolila, Samuel and Andrea Ichino, "Unemployment and consumption near and far away from the Mediterranean," *Journal of Population Economics*, 2008, *21* (2), 255–280.

Blau, Francine D, Marianne A Ferber, and Anne E Winkler, *The economics of women, men, and work*, Pearson; 7 edition, 2013.

Bover, Olympia, Pilar Garcia-Perea, and Pedro Portugal, "Labour market outliers: Lessons from Portugal and Spain," *Economic Policy*, 2000, *15*, 380–428.

Bowlus, Audra J, "Matching workers and jobs: Cyclical fluctuations in match quality," *Journal of Labor Economics*, 1995, pp. 335–350.

Bredtmann, Julia, Sebastian Otten, and Christian Rulff, "Husband's unemployment and wife's labor supply: the added worker effect across Europe," *ILR Review*, 2018, *71* (5), 1201–1231.

Christofides, Louis N, Alexandros Polycarpou, and Konstantinos Vrachimis, "Gender wage gaps, 'sticky floors' and 'glass ceilings' in Europe," *Labour Economics*, 2013, *21*, 86–102.

Dolado, Juan J., "EU Dual Labour Markets: Consequences and Potential Reforms," in R. Blundell et al., ed., *Economics without Borders. Economic Research for European Policy Challenges*, Cambridge: Cambridge University Press, 2016.

Engel, Melissa and Sandra Schaffner, "How to Use the EU-SILC Panel to Analyse Monthly and Hourly Wages," *Ruhr Economic Paper*, 2012, (390).

Gayle, George-Levi and Limor Golan, "Estimating a Dynamic Adverse-Selection Model: Labour-Force Experience and the Changing Gender Earnings Gap 1968–1997," *The Review of Economic Studies*, 2012, *79* (1), 227–267.

Goldin, Claudia, "A grand gender convergence: Its last chapter," *The American Economic Review*, 2014, *104* (4), 1091–1119.

Goos, Maarten, Alan Manning, and Anna Salomons, "Job polarization in Europe," *The American Economic Review*, 2009, pp. 58–63.

Gronau, Reuben, "Wage Comparisons–A Selectivity Bias," *The Journal of Political Economy*, 1974, pp. 1119–1143.

Heckman, James J, "Sample Selection Bias as a Specification Error," *Econometrica*, 1979, *47* (1), 153–161.

Jenkins, S. P., A. Brandolini, J. Micklewright, and B. (Eds.) Nolan, *The great recession and the distribution of household income*, Oxford University Press, 2012.

Johnson, Norman L, Samuel Kotz, and Narayanaswamy Balakrishnan, *Continuous univariate distributions, vol. 1-2*, New York: John Wiley & Sons, 1994.

Johnson, William, Yuichi Kitamura, and Derek Neal, "Evaluating a simple method for estimating black-white gaps in median wages," *American Economic Review*, 2000, pp. 339–343.

Keane, Michael, Robert Moffitt, and David Runkle, "Real wages over the business cycle: Estimating the impact of heterogeneity with micro data," *The Journal of Political Economy*, 1988, pp. 1232–1266.

Koenker, Roger and Gilbert Bassett Jr, "Regression quantiles," *Econometrica*, 1978, pp. 33–50.

Mulligan, Casey B and Yona Rubinstein, "Selection, investment, and women's relative wages over time," *The Quarterly Journal of Economics*, 2008, *123* (3), 1061–1110.

Neal, Derek, "The Measured Black-White Wage Gap among Women Is Too Small," *Journal of Political Economy*, 2004, *112* (S1), S1–S28.

OECD, "Employment Outlook 2014," *Paris: Organisation for Economic Co-operation and Development*, 2014.

Olivetti, Claudia and Barbara Petrongolo, "Unequal pay or unequal employment? A cross-country analysis of gender gaps," *Journal of Labor Economics*, 2008, *26* (4), 621–654.

_ **and** _ , "The Evolution of the Gender Gap in Industrialized Countries," *Annual Review of Economics*, 2016, *8*, 405–434.

Appendix

A Derivation of LD and LS constraints

Given the wage, productivity and outside value equations in system (5) to (7) in the main text, we derive here the values of the relevant thresholds of the productivity thresholds determining LS and LD decisions.

* Labour supply (LS) cut-off values

As for the LS thresholds, men participate when $w_{it} > r_{it}$. Since the male reservation wage has been normalized to zero, equations (5) and (7) with $g_i = 0$ imply that their productivity shock ε_{it} has to exceed the LS cut-off value, $a_t^{LS}(g_i = 0)$, given by:

$$a_t^{LS}(g_i = 0) = -\mu_t^w, \tag{A1}$$

where, for simplicity, it is assumed that the inequality $\varepsilon_{it} > a_t^{LS}(g_i = 0)$ always holds. This implies that that men would always wish to participate in the labour market and, as a result, that their LS constraint does not bind.

As regards women, their labour supply (LS) condition, $w_{it} > r_{it}$ is satisfied iff ε_{it} exceeds the following two LS thresholds, depending on the value of the reservation wage shock, v_{it}:

$$a_t^{LS}(g_i = 1, v_{it} = \overline{v}) \equiv \overline{a}_t = \mu_t^r + \overline{v} - \mu_t^w - \gamma_t, \tag{A2a}$$

$$a_t^{LS}(g_i = 1, v_{it} = \underline{v}) \equiv \underline{a}_t = \mu_t^r + \underline{v} - \mu_t^w - \gamma_t. \tag{A2b}$$

* Labour demand (LD) cut-off values.

With regard to the LD condition to create/maintain a job, $w_{it} < x_{it}$, it holds iff ε_{it} exceeds the following LD threshold:

$$a_t^{LD}(g_i) \equiv \frac{\mu_t^w + g_i\gamma_t - \mu_t^x}{\rho}. \tag{A3}$$

for $g_i = 1, 0$.

The conditions above yield gender-specific lower bounds for ε_{it} implying that *only one* of the two constraints above binds.

Notice that the previous assumption on zero male reservation wage implies that the LD threshold $a_t^{LD}(g_i = 0)$ is the only one that binds for them. By contrast, both LD and LS constraints may be binding for female workers. For example, concerning women with a high reservation-wage shock, the LD constraint would be binding iff: $a_t^{LS}(g_i = 1, v_{it} = \overline{v}) < a_t^{LD}(g_i = 1)$ or equivalently:

$$\frac{\mu_t^x - (\mu_t^w + \gamma_t)}{\overline{a}_t} < \rho, \tag{A4}$$

whereas for women with low reservation wage shock, the corresponding LD condition becomes:[29]

$$\frac{\mu_t^x - (\mu_t^w + \gamma_t)}{\underline{a}_t} < \rho. \tag{A5}$$

Intuitively, equations (A4) and (A5) hold when: (i) the potential female wage is high relative to productivity, i.e. when the numerator $\mu_t^x - (\mu_t^w + \gamma_t)$ in (A5) is small; (ii) the reservation wage is low relative to potential wage, i.e., when the denominators in (A5) $\underline{a}_t$ and $\overline{a}_t$ are high;and (iii) the surplus is high, i.e., when ρ is much larger than zero. By contrast, when $\mu_t^x - (\mu_t^w + \gamma_t)$ is high, $\underline{a}_t$ and $\overline{a}_t$ are low and ρ is close to unity, it is likely that $a_t^{LD} < a_t^{LS}$, so that the LS constraint would be the binding one.

B Comparative statics

* Male Participation

In order to examine male LFP, for illustrative purposes we make use of the following result concerning the median of a (standardized) normal distribution which is truncated from below (see Johnson et al., 1994). Assuming $\varepsilon_{it} \sim \mathcal{N}[0,1]$ and denoting the c.d.f. of the standardized normal distribution by $\Phi(\cdot)$, then the median, $\underline{m}(a)$, of the truncated from below distribution of ε_{it}, such that $a < \varepsilon_{it}$, is given by:

$$\underline{m}(a) = \Phi^{-1}\left[\frac{1}{2}(1 + \Phi(a))\right].$$

Using this result, the observed male wage, for which the LD constraint binds, $a_t^{LS}(g=0) < a_t^{LD}(g=0)$, has the following closed-form solution:

$$\begin{aligned} w_t^m &\equiv m(w_{it}|g_i = 0, L_{it} = 1) = m(w_{it}|g_i = 0, a_t^{LD}(g=0) < \varepsilon_{it}) \\ &= \mu_t^w + \underline{m}(a_t^{LD}(g=0)). \end{aligned}$$

then, given the properties of $\Phi(\cdot)$, it holds that the $\underline{m}$ $(\cdot)$ term is a non-negative increasing function of $a_t^{LD}(g=0)$ which measures the strength of the male selection bias, namely, $b_t^m = m(\varepsilon_{it}|g_i = 0, L_{it} = 1) = \underline{m}(a_t^{LD}(g=0))$.

Next, the response of w_t^m with respect to a change in μ_t^x is given by:

$$\frac{dw_t^m}{d\mu_t^x} = \frac{\partial \underline{m}}{\partial a_t^{LD}(g=0)} \times \frac{\partial a_t^{LD}(g=0)}{\partial \mu_t^x} < 0, \tag{B1}$$

since $a_t^{LD}(g=0)$ is decreasing in μ_t^x. Hence, if we identify the Great Recession as a drop in expected productivity, $\Delta\mu_t^x < 0$, then the median of the observed male wage

[29]Note that, since $\underline{a}_t < \overline{a}_t$, the LD condition is more likely to be the binding one for women with a high reservation-wage shock than for women with a low reservation-wage shock.

distribution increases, due to a stronger positive selection of males into employment, $\Delta b_t^m > 0$.[30] In other words, less-skilled male workers with lower wages will not show up in the observed wage distribution because they become unemployed, leading to a rise in the median wage for employed men.

* Female Participation

Under our assumption on the female reservation-wage shocks v_{it}, it is easy to check that the median of female wages , $\underline{m}(a(v))$, of the truncated-from-below distribution of ε_{it}, such that $a(v) < \varepsilon_{it}$, is given by:

$$\underline{m}(a(v)) = \Phi^{-1}\left[\frac{1}{2}(1 + p\Phi(\bar{a}) + (1-p)\Phi(\underline{a}))\right].$$

Mutatis mutandis, the female wage among the employed workers is given by:

$$\begin{aligned} w_t^f &\equiv m(w_{it}|g_i = 1, L_{it} = 1) = m(w_{it}|g_i = 1, a_t^f(v) < \varepsilon_{it}) \\ &= \mu_t^w + \gamma_t + \underline{m}(a_t^f(v)) \end{aligned}$$

$$a_t^f(v) \equiv \begin{cases} a_t^{LS}(g=1;v) & : a_t^{LS}(g=1;v) > a_t^{LD}(g=1) \\ a_t^{LD}(g=1) & : a_t^{LS}(g=1;v) < a_t^{LD}(g=1) \end{cases}$$

Thus, the observed female wage will depend on which of the LS and LD constraints is binding. Again, the strength of the selection bias for females is measured by the $\underline{m}(\cdot)$ term, that is, $b_t^f = m(\varepsilon_{it}|g_i = f, L_{it} = 1) = \underline{m}(a_t^f(v))$. If the binding constraint is LD, i.e., $a_t^{LS}(g=1;v) < a_t^{LD}(g=1)$, then a reduction in labour productivity ($d\mu_t^x < 0$) during the Great Recession will have the same impact on observed female wages as the one discussed before for male wages, namely:

$$\frac{dw_t^f}{d\mu_t^x} = \frac{\partial \underline{m}(a_t^f(v))}{\partial a_t^{LD}(g=1)} \times \frac{\partial a_t^{LD}(g=1)}{\partial \mu_t^x} < 0. \tag{B2}$$

That is, observed female median wages will increase due to an even stronger positive selection of women into employment when productivity goes down, since those at the bottom of the wage distribution are the ones losing their jobs.

However, if the LS constraint is the binding one, $a_t^{LS}(g=1;v) > a_t^{LD}(g=1)$, then:

$$\frac{dw_t^f}{d\mu_t^r} = \frac{\partial \underline{m}(a_t^f(v))}{\partial a_t^{LS}(g=1;v)} \times \frac{\partial a_t^{LS}(g=1;v)}{\partial \mu_t^r} > 0. \tag{B3}$$

Hence, insofar as the downturn has generated an added-worker effect among previous female non-participants in the less-skilled segment of the labour market,

[30] Note that the converse argument could be used to model the effects of a rise in early retirement. Because older male workers have longer experience and this typically leads to higher wages, early retirement would imply stronger negative selection, $\Delta b_t^m < 0$.

this would translate into a reduction in the reservation wage, $\Delta\mu_t^r < 0$. This results in a reduction of the observed female wage due to a less positive selection, $\Delta b_t^f < 0$, since less-skilled women entering the labour market would get jobs.

C Deriving Hourly Wages

The main challenge in deriving hourly wages is to combine annual income (PY010) and monthly economic status information (PL210A-PL210L up to 2009 and PL211A-PL211L onwards) for the previous calendar year with the number of hours usually worked per week (PL060) at the date of the interview.

To do this we combine the longitudinal files from the period 2005-2017 and use the imputed annual hours of work

$$hours_{annual} = months_{annual} \times 4.345 \times hours_{week}$$

to compute hourly wages. The following sequential set of rules are used to impute missing annual hours of work during the previous calendar year:

1. For those workers who have only one employment spell (with no changes in full-time/part-time status), we use the number of months of this spell and the number of hours from the previous survey.

2. For those workers who have only one employment spell (with no changes in full-time/part-time status), we use the number of months of this spell and the number of hours declared at the date of the interview if the person hasn't changed job since last year (PL160).

In the case of United Kingdom, we only use the number of hours at the date of the interview since the income reference period coincides with the year of the interview.

3. For those workers who have only one employment spell (with no changes in full-time/part-time status), we use the number of months of this spell and approximate the number of hours by the year- gender- full-time/part-time status- specific mean.

4. For those workers who have multiple employment spells, we use the number of months of each spell and the number of hours for each spell approximated by the year- gender- full-time/part-time status- specific mean.

On-line appendix (not for publication)

A Descriptive statistics

This Appendix provides tables of descriptive statistics of the original micro data (Table A1), as well as indicators of broad labour market patterns, such as LFP (Table A2) and employment (Table A3) rates by education, wage inequality by gender (Figure A1), median wage gaps under imputation using a Probit model (Table A4) and wages from other waves (Table A5), as well as rates of imputation (Table A6).

B Additional results: Imputations

In order to check the robustness of our imputation method, Table A4 reports estimates based on a probit model. The imputation technique proceeds in two steps. First, we estimate a probit model for the probability of earning a wage below the gender-specific median, controlling for education dummies, experience and its square. The estimated probabilities, $\hat{P}_i$, are then used as sampling weights to impute the wages of the non-employed individuals. Specifically, each non-employed individual appears twice in the imputed sample: with a wage above the median and a weight $\hat{P}_i$, and with a wage below the median and a weight $1 - \hat{P}_i$. To account for a bias in the reference median wage in the first step, we enlarge our base sample with wage observations from other waves. The conclusions from this probabilistic model are fairly robust to a more general specification that includes marital status, the number of children, and the position of spouse income in their gender-specific distribution.

Our alternative imputation method attributes to non-employed individuals (who are observed as having been employed in other waves of the panel) their wage in the nearest year for which it is available. The results are reported in Table A5.

Unfortunately, since the panel dimension of our data is not long, we only have a limited number of available observations to impute. As can be seen from Table **??**, we impute around one third of observations and, particularly, few women in Southern Europe. These figures are much lower than those in Olivetti and Petrongolo (2008), who have a much longer panel. Low imputation rates imply that a lower gap is found between the Southern EU and Rest of EU countries concerning female selection in 2007. Changes in selection biases since the onset of the Great Recession smaller than those obtained under the other imputation methods. This is particularly the case for female selection, except in Greece. This smaller variation is not surprising since, e.g. in Spain the imputation rates for 2007 and 2012 are 23% and 30%, while they were

66% and 73% with Imputation EE. Yet, as with the other imputation methods, we still document a sizeable increase in male selection in peripheral countries, making this finding rather robust.

Table A1: Descriptive Statistics of Samples Used

	2007				2012				2016			
	Males		Females		Males		Females		Males		Females	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Greece												
Employed	.85	.35	.54	.50	.60	.49	.42	.49	.68	.46	.48	.50
Unemployed	.11	.31	.11	.31	.35	.48	.28	.45	.28	.45	.29	.45
Inactive	.04	.19	.35	.48	.05	.22	.30	.46	.03	.18	.23	.42
Annual Earnings	21.63	15.73	16.10	10.44	17.90	13.07	14.29	9.22	16.40	10.02	13.82	8.38
Annual Hours	2073	506	1770	575	1943	564	1716	616	2048	620	1823	626
Log(hourly wage)	2.29	.57	2.17	.56	2.07	.52	1.98	.52	1.98	.46	1.93	.46
Age	38.80	8.42	38.92	8.43	38.88	8.48	39.38	8.41	39.62	8.13	40.00	8.25
Educ1	.27	.44	.28	.45	.20	.40	.20	.40	.17	.38	.16	.37
Educ2	.41	.49	.35	.48	.41	.49	.38	.48	.40	.49	.34	.47
Educ3	.26	.44	.30	.46	.30	.46	.33	.47	.34	.47	.38	.49
Experience	16.76	9.60	10.19	9.08	16.48	10.12	11.47	10.08	15.62	9.34	10.81	9.06
Temporary	.21	.41	.23	.42	.13	.34	.17	.38	.19	.40	.23	.42
Nb of Observations	1799		2320		1840		2369		4742		6053	
Italy												
Employed	.85	.36	.56	.50	.81	.39	.56	.50	.82	.38	.55	.50
Unemployed	.09	.29	.10	.30	.16	.36	.12	.32	.12	.32	.09	.29
Inactive	.06	.24	.35	.48	.04	.19	.32	.47	.06	.24	.36	.48
Annual Earnings	19.05	8.90	14.45	7.35	19.24	10.45	14.70	8.01	19.91	13.96	14.89	8.15
Annual Hours	2089	436	1716	521	2019	447	1718	501	2020	398	1748	503
Log(hourly wage)	2.28	.42	2.24	.46	2.21	.49	2.12	.52	2.19	.55	2.10	.57
Age	39.68	8.21	40.08	8.05	40.44	8.21	41.19	8.01	40.34	8.45	41.35	8.35
Educ1	.44	.50	.40	.49	.38	.48	.35	.48	.38	.49	.33	.47
Educ2	.39	.49	.38	.49	.45	.50	.45	.50	.43	.49	.44	.50
Educ3	.13	.34	.16	.37	.15	.35	.17	.38	.18	.39	.21	.41
Experience	16.82	9.58	11.54	9.18	17.32	9.34	12.86	9.38	15.35	9.53	11.43	9.47
Temporary	.10	.30	.14	.35	.10	.30	.13	.34	.16	.36	.17	.38
Nb of Observations	7848		9534		4349		5317		4741		5635	
Portugal												
Employed	.84	.37	.71	.45	.71	.46	.71	.46	.85	.35	.78	.42
Unemployed	.10	.30	.10	.30	.26	.44	.19	.39	.12	.33	.13	.34
Inactive	.06	.24	.19	.39	.03	.18	.10	.30	.02	.15	.09	.29
Annual Earnings	10.91	7.10	8.81	6.10	10.62	6.74	9.08	5.35	12.16	8.04	9.75	5.60
Annual Hours	2092	431	1863	505	2035	561	1912	560	2160	471	1984	473
Log(hourly wage)	1.61	.51	1.49	.58	1.56	.47	1.49	.47	1.59	.50	1.48	.47
Age	38.45	8.73	39.61	8.57	37.40	8.65	38.89	8.54	40.67	8.22	41.32	8.08
Educ1	.72	.45	.66	.47	.58	.49	.46	.50	.52	.50	.45	.50
Educ2	.16	.37	.15	.36	.25	.43	.28	.45	.26	.44	.24	.43
Educ3	.11	.32	.18	.38	.16	.37	.25	.43	.20	.40	.30	.46
Experience	19.65	10.67	17.22	10.70	17.36	10.79	17.01	10.05	21.35	10.22	19.72	10.03
Temporary	.17	.38	.21	.41	.17	.38	.18	.38	.14	.35	.17	.37
Nb of Observations	1880		2250		1282		1531		3272		3868	
Spain												
Employed	.88	.32	.67	.47	.72	.45	.59	.49	.81	.40	.67	.47
Unemployed	.09	.28	.11	.31	.27	.44	.26	.44	.18	.38	.21	.41
Inactive	.03	.17	.22	.41	.01	.11	.15	.35	.02	.13	.12	.32
Annual Earnings	17.35	9.39	13.00	8.52	16.54	11.07	13.14	9.62	17.51	12.52	14.18	11.45
Annual Hours	2099	483	1760	587	1928	570	1659	645	2028	491	1737	574
Log(hourly wage)	2.11	.51	1.98	.58	2.10	.60	1.99	.63	1.99	.87	1.96	.75
Age	38.17	8.27	38.69	8.27	39.72	8.10	40.12	8.06	40.60	8.01	40.72	8.01
Educ1	.41	.49	.39	.49	.42	.49	.36	.48	.37	.48	.31	.46
Educ2	.24	.43	.25	.43	.24	.42	.23	.42	.24	.43	.24	.43
Educ3	.34	.47	.36	.48	.34	.47	.41	.49	.38	.49	.45	.50
Experience	17.81	9.73	12.93	9.18	15.75	10.95	11.56	10.16	18.30	9.24	14.77	8.91
Temporary	.25	.43	.29	.45	.20	.40	.24	.43	.25	.43	.28	.45
Nb of Observations	7461		8873		5660		6664		3311		3854	

Source: EU-SILC and authors' calculations. Note: The descriptive statistics refer to the base samples, aged 25-54, excluding the self-employed, those in the military, and those in full-time education. Description of variables: Employed, unemployed, and inactive are self-defined. Educ1=1 if less than upper secondary education. Educ2=1 if upper secondary education completed. Educ3=1 if higher education. Married=1 if living in a couple.

Table A1: Descriptive Statistics of Samples Used (Continued)

	2007				2012				2016			
	Males		Females		Males		Females		Males		Females	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Austria												
Employed	.88	.32	.71	.45	.88	.32	.72	.45	.89	.32	.75	.43
Unemployed	.07	.25	.05	.23	.08	.27	.06	.24	.07	.26	.05	.23
Inactive	.05	.22	.23	.42	.04	.20	.22	.41	.04	.19	.19	.40
Annual Earnings	36.28	21.59	22.97	36.20	43.29	31.89	24.64	17.84	46.01	33.48	27.71	19.59
Annual Hours	2121	430	1627	624	2108	491	1605	598	2094	464	1584	626
Log(hourly wage)	2.87	.51	2.65	.55	2.91	.64	2.69	.58	2.93	.60	2.72	.61
Age	40.41	8.17	40.25	8.24	40.74	8.70	40.91	8.44	40.82	8.69	40.98	8.63
Educ1	.09	.29	.16	.37	.10	.30	.17	.38	.08	.28	.14	.35
Educ2	.59	.49	.50	.50	.56	.50	.48	.50	.55	.50	.48	.50
Educ3	.21	.41	.19	.39	.22	.42	.18	.39	.36	.48	.34	.47
Experience	21.27	9.27	16.61	9.57	22.10	9.79	17.68	9.62	21.99	10.02	18.12	9.91
Temporary	.04	.19	.06	.24	.05	.21	.06	.24	.06	.23	.07	.26
Nb of Observations	2329		2647		1522		1769		1472		1683	
Belgium												
Employed	.87	.34	.74	.44	.83	.37	.77	.42	.86	.35	.78	.41
Unemployed	.07	.25	.08	.27	.10	.30	.08	.28	.07	.26	.06	.25
Inactive	.07	.25	.18	.39	.07	.25	.14	.35	.07	.26	.15	.36
Annual Earnings	35.46	18.82	25.38	13.26	40.28	22.03	30.77	17.25	42.13	21.18	32.99	18.65
Annual Hours	2048	510	1650	555	2019	479	1648	546	2009	486	1691	539
Log(hourly wage)	2.89	.42	2.79	.45	2.94	.39	2.88	.41	2.94	.41	2.87	.39
Age	39.89	8.47	39.97	8.61	40.01	8.50	39.95	8.74	39.78	8.65	39.96	8.64
Educ1	.25	.43	.23	.42	.20	.40	.18	.38	.18	.38	.16	.36
Educ2	.36	.48	.33	.47	.34	.47	.31	.46	.38	.49	.30	.46
Educ3	.37	.48	.42	.49	.41	.49	.47	.50	.42	.49	.53	.50
Experience	18.31	9.93	15.06	10.01	16.47	9.87	14.32	9.97	16.04	9.91	14.00	9.81
Temporary	.05	.23	.11	.31	.07	.26	.10	.30	.08	.27	.10	.30
Nb of Observations	2458		2802		1517		1715		1413		1562	
France												
Employed	.92	.28	.82	.39	.88	.33	.81	.39	.89	.32	.82	.39
Unemployed	.06	.24	.07	.25	.10	.30	.08	.27	.10	.29	.08	.27
Inactive	.02	.14	.11	.32	.02	.14	.11	.31	.02	.14	.11	.31
Annual Earnings	24.40	16.81	16.64	10.53	25.88	16.98	18.49	11.47	28.46	20.43	20.38	13.12
Annual Hours	2070	516	1684	579	2032	531	1728	576	2036	522	1766	574
Log(hourly wage)	2.46	.51	2.30	.60	2.48	.50	2.31	.58	2.51	.52	2.36	.57
Age	40.26	8.20	40.50	8.31	40.13	8.33	40.48	8.39	41.08	8.23	41.15	8.21
Educ1	.19	.39	.22	.42	.14	.34	.15	.36	.16	.37	.14	.35
Educ2	.49	.50	.43	.50	.51	.50	.44	.50	.45	.50	.42	.49
Educ3	.32	.47	.35	.48	.36	.48	.41	.49	.39	.49	.44	.50
Experience	19.06	9.93	16.00	9.89	18.81	9.66	16.18	9.64	19.28	9.39	16.79	9.54
Temporary	.10	.29	.16	.36	.11	.32	.13	.34	.11	.31	.15	.35
Nb of Observations	4121		4624		3810		4205		2972		3353	
Netherlands												
Employed	.93	.25	.80	.40	.90	.30	.80	.40	.92	.28	.84	.36
Unemployed	.02	.13	.04	.19	.08	.28	.07	.25	.06	.23	.06	.24
Inactive	.05	.22	.16	.37	.02	.12	.13	.33	.03	.16	.10	.29
Annual Earnings	44.00	33.61	24.12	14.97	47.30	26.73	28.71	18.84	50.87	38.11	31.05	21.18
Annual Hours	1949	367	1358	477	1934	392	1388	466	1960	365	1440	463
Log(hourly wage)	3.13	.48	2.89	.58	3.13	.51	2.96	.53	3.17	.46	2.95	.50
Age	40.32	8.41	39.96	8.28	40.54	8.56	40.61	8.36	41.22	8.43	40.79	8.81
Educ1	.18	.38	.20	.40	.15	.36	.17	.37	.11	.32	.13	.33
Educ2	.37	.48	.42	.49	.38	.49	.42	.49	.38	.49	.40	.49
Educ3	.42	.49	.33	.47	.44	.50	.40	.49	.51	.50	.47	.50
Experience	17.81	9.79	14.09	8.78	17.74	9.44	14.93	9.02	19.11	8.99	17.08	9.34
Temporary	.12	.33	.14	.35	.12	.32	.14	.35	.13	.34	.14	.35
Nb of Observations	2315		2712		1700		1896		1238		1610	

Source: EU-SILC and authors' calculations. Note: The descriptive statistics refer to the base samples, aged 25-54, excluding the self-employed, those in the military, and those in full-time education. Description of variables: Employed, unemployed, and inactive are self-defined. Educ1=1 if less than upper secondary education. Educ2=1 if upper secondary education completed. Educ3=1 if higher education. Married=1 if living in a couple.

Table A1: Descriptive Statistics of Samples Used (Continued)

	2007				2012				2016			
	Males		Females		Males		Females		Males		Females	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Ireland												
Employed	.85	.36	.67	.47	.71	.45	.59	.49	.79	.41	.66	.47
Unemployed	.11	.32	.03	.17	.27	.44	.09	.28	.18	.39	.07	.25
Inactive	.04	.19	.30	.46	.03	.16	.32	.47	.03	.17	.27	.44
Annual Earnings	44.67	35.95	27.34	21.69	46.43	107.75	31.42	32.00	46.13	36.23	32.23	26.27
Annual Hours	2015	543	1467	633	1896	609	1507	636	1931	586	1530	609
Log(hourly wage)	2.98	.56	2.81	.62	2.99	.61	2.87	.62	3.01	.59	2.90	.62
Age	41.00	8.33	41.26	8.28	39.80	8.12	39.47	8.12	40.48	7.82	40.44	7.69
Educ1	.34	.47	.30	.46	.23	.42	.18	.38	.21	.40	.12	.33
Educ2	.23	.42	.25	.43	.23	.42	.23	.42	.23	.42	.21	.41
Educ3	.35	.48	.35	.48	.49	.50	.48	.50	.51	.50	.54	.50
Experience	20.66	9.70	15.83	9.01	18.31	9.40	14.36	8.97	18.52	9.39	15.54	9.02
Temporary	.04	.21	.08	.28	.08	.27	.07	.26	.09	.29	.08	.26
Nb of Observations	1326		1820		1392		1827		1130		1518	
United Kingdom												
Employed	.94	.23	.81	.40	.91	.29	.78	.41	.94	.24	.79	.41
Unemployed	.03	.17	.02	.12	.06	.23	.04	.19	.03	.17	.03	.17
Inactive	.03	.16	.18	.38	.04	.19	.18	.38	.03	.17	.18	.39
Annual Earnings	47.77	35.88	28.00	21.33	42.47	43.03	26.52	23.70	41.94	34.18	27.43	23.47
Annual Hours	2267	509	1694	663	2237	560	1712	672	2202	520	1700	644
Log(hourly wage)	3.10	.55	2.85	.60	2.79	.58	2.61	.54	2.78	.59	2.62	.59
Age	40.09	8.01	40.05	8.14	39.95	8.28	40.00	8.29	39.73	8.64	39.43	8.78
Educ1	.09	.29	.11	.31	.12	.32	.10	.30	.17	.38	.14	.35
Educ2	.54	.50	.56	.50	.42	.49	.42	.49	.34	.47	.33	.47
Educ3	.32	.47	.31	.46	.46	.50	.48	.50	.49	.50	.53	.50
Experience	19.58	9.63	15.84	9.01	19.09	9.77	17.10	9.90	18.83	9.55	15.60	9.44
Temporary	.03	.17	.04	.19	.03	.17	.03	.18	.03	.18	.04	.21
Nb of Observations	2825		3748		3654		4436		1680		2089	
Denmark												
Employed	.98	.12	.94	.23	.90	.29	.88	.33	.85	.35	.91	.29
Unemployed	.01	.08	.01	.12	.06	.24	.11	.31	.10	.31	.06	.23
Inactive	.01	.09	.04	.21	.04	.19	.02	.13	.04	.20	.03	.18
Annual Earnings	47.98	26.57	36.72	15.99	55.33	34.01	42.41	18.27	60.20	41.48	47.12	20.52
Annual Hours	2064	409	1829	362	2026	451	1793	375	1981	483	1823	419
Log(hourly wage)	3.11	.69	3.01	.61	3.27	.41	3.14	.40	3.35	.43	3.23	.38
Age	40.07	8.17	39.98	8.10	40.22	8.22	40.57	8.23	40.24	8.65	40.86	8.39
Educ1	.19	.39	.16	.37	.10	.30	.10	.30	.12	.33	.08	.27
Educ2	.47	.50	.41	.49	.48	.50	.42	.49	.42	.49	.36	.48
Educ3	.34	.47	.43	.49	.43	.49	.48	.50	.46	.50	.55	.50
Experience	18.44	9.50	15.89	9.67	18.89	10.55	18.02	10.76	18.70	10.77	18.64	10.39
Temporary	.00	.00	.00	.00	.08	.27	.07	.25	.06	.24	.08	.27
Nb of Observations	1503		1762		559		633		1106		1215	
Finland												
Employed	.90	.30	.86	.34	.87	.34	.82	.39	.85	.36	.85	.36
Unemployed	.09	.29	.04	.20	.12	.32	.07	.25	.14	.34	.08	.27
Inactive	.01	.11	.09	.29	.02	.13	.12	.32	.01	.11	.07	.26
Annual Earnings	36.19	22.83	25.69	14.12	41.89	25.12	31.52	17.60	43.16	23.72	32.91	20.38
Annual Hours	1985	500	1813	485	1982	441	1815	459	1958	439	1765	479
Log(hourly wage)	2.96	.49	2.75	.45	2.97	.50	2.86	.41	3.03	.44	2.90	.44
Age	39.66	8.63	40.00	8.65	39.82	8.57	40.08	8.64	39.61	8.53	39.69	8.70
Educ1	.12	.33	.11	.31	.12	.33	.05	.21	.08	.27	.03	.17
Educ2	.49	.50	.39	.49	.44	.50	.36	.48	.49	.50	.33	.47
Educ3	.38	.49	.50	.50	.42	.49	.58	.49	.42	.49	.63	.48
Experience	16.59	9.84	15.90	10.19	18.30	11.13	17.23	11.12	16.54	9.22	15.27	9.31
Temporary	.11	.31	.19	.39	.10	.30	.14	.34	.07	.25	.14	.35
Nb of Observations	1128		1254		2055		2293		1139		1204	

Source: EU-SILC and authors' calculations. Note: The descriptive statistics refer to the base samples, aged 25-54, excluding the self-employed, those in the military, and those in full-time education. Description of variables: Employed, unemployed, and inactive are self-defined. Educ1=1 if less than upper secondary education. Educ2=1 if upper secondary education completed. Educ3=1 if higher education. Married=1 if living in a couple.

Table A1: Descriptive Statistics of Samples Used (Continued)

	2007				2012				2016			
	Males		Females		Males		Females		Males		Females	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Norway												
Employed	.97	.16	.91	.28	.97	.17	.93	.26	.96	.20	.92	.27
Unemployed	.01	.10	.02	.12	.02	.13	.02	.14	.03	.16	.02	.15
Inactive	.02	.13	.07	.26	.01	.12	.05	.23	.01	.12	.06	.23
Annual Earnings	58.85	111.64	35.52	19.93	78.03	46.09	51.91	28.25	66.83	36.55	47.47	27.32
Annual Hours	2107	451	1764	511	2110	357	1789	446	2075	381	1842	450
Log(hourly wage)	3.26	.71	3.03	.69	3.54	.60	3.31	.60	3.31	.55	3.09	.63
Age	39.59	8.14	39.79	8.23	41.27	8.28	41.03	7.91	40.81	8.57	41.02	8.62
Educ1	.17	.37	.14	.35	.12	.32	.11	.32	.14	.35	.10	.30
Educ2	.43	.50	.35	.48	.42	.49	.30	.46	.35	.48	.29	.45
Educ3	.37	.48	.48	.50	.41	.49	.57	.49	.50	.50	.60	.49
Experience	18.02	9.65	15.98	9.25	19.91	8.97	17.43	8.70	18.78	9.06	17.77	9.10
Temporary	.05	.21	.10	.30	.04	.21	.08	.28	.04	.20	.08	.27
Nb of Observations	1379		1222		1496		1610		1689		1798	

Source: EU-SILC and authors' calculations. Note: The descriptive statistics refer to the base samples, aged 25-54, excluding the self-employed, those in the military, and those in full-time education. Description of variables: Employed, unemployed, and inactive are self-defined. Educ1=1 if less than upper secondary education. Educ2=1 if upper secondary education completed. Educ3=1 if higher education. Married=1 if living in a couple.

Table A2: LFP Rates by Education

	LFP Rate in 2007				Changes over 2007-2012				Changes over 2012-2016			
	College		No college		College		No college		College		No college	
	M	F	M	F	M	F	M	F	M	F	M	F
Greece	.988	.872	.958	.566	-.013	.013	-.010	.063	-.012	.011	.011	.036
Italy	.939	.877	.947	.594	.038	.016	.018	.023	-.004	-.016	.008	.017
Portugal	.911	.920	.967	.787	.060	.007	.012	.094	.015	.057	-.006	.004
Spain	.985	.879	.967	.660	.002	.023	.012	.088	.006	.045	.004	.085
Southern	**.956**	**.887**	**.960**	**.652**	**.022**	**.015**	**.008**	**.067**	**.001**	**.024**	**.004**	**.036**
Austria	.978	.805	.947	.743	-.005	.020	.008	.053	.012	.015	.006	.004
Belgium	.973	.871	.968	.778	-.019	.032	-.035	.026	.010	.037	.001	-.029
France	.990	.945	.973	.858	-.005	.004	-.002	.014	-.021	-.010	-.038	-.097
Netherlands	.986	.914	.977	.795	.000	.028	.000	.012	-.001	-.102	-.016	-.007
Continental	**.982**	**.884**	**.966**	**.793**	**-.007**	**.021**	**-.008**	**.026**	**.000**	**-.015**	**-.012**	**-.032**
Ireland	.967	.862	.972	.606	.018	-.049	.008	-.024	-.984	-.813	-.980	-.583
United Kingdom	.978	.864	.969	.798	-.001	.006	-.012	-.031	.013	-.003	.014	-.013
Anglosaxon	**.972**	**.863**	**.970**	**.702**	**.008**	**-.021**	**-.002**	**-.027**	**-.486**	**-.408**	**-.483**	**-.298**
Denmark	.997	.950	.989	.953	-.021	-.002	-.016	-.022	.018	.039	.004	.023
Finland	.984	.899	.991	.901	.005	-.007	-.003	-.013	.006	.020	.001	-.005
Norway	.995	.952	.992	.945	.000	.016	-.002	.002	.003	.014	-.002	.002
Nordic	**.992**	**.933**	**.991**	**.933**	**-.005**	**.002**	**-.007**	**-.011**	**.009**	**.024**	**.001**	**.006**

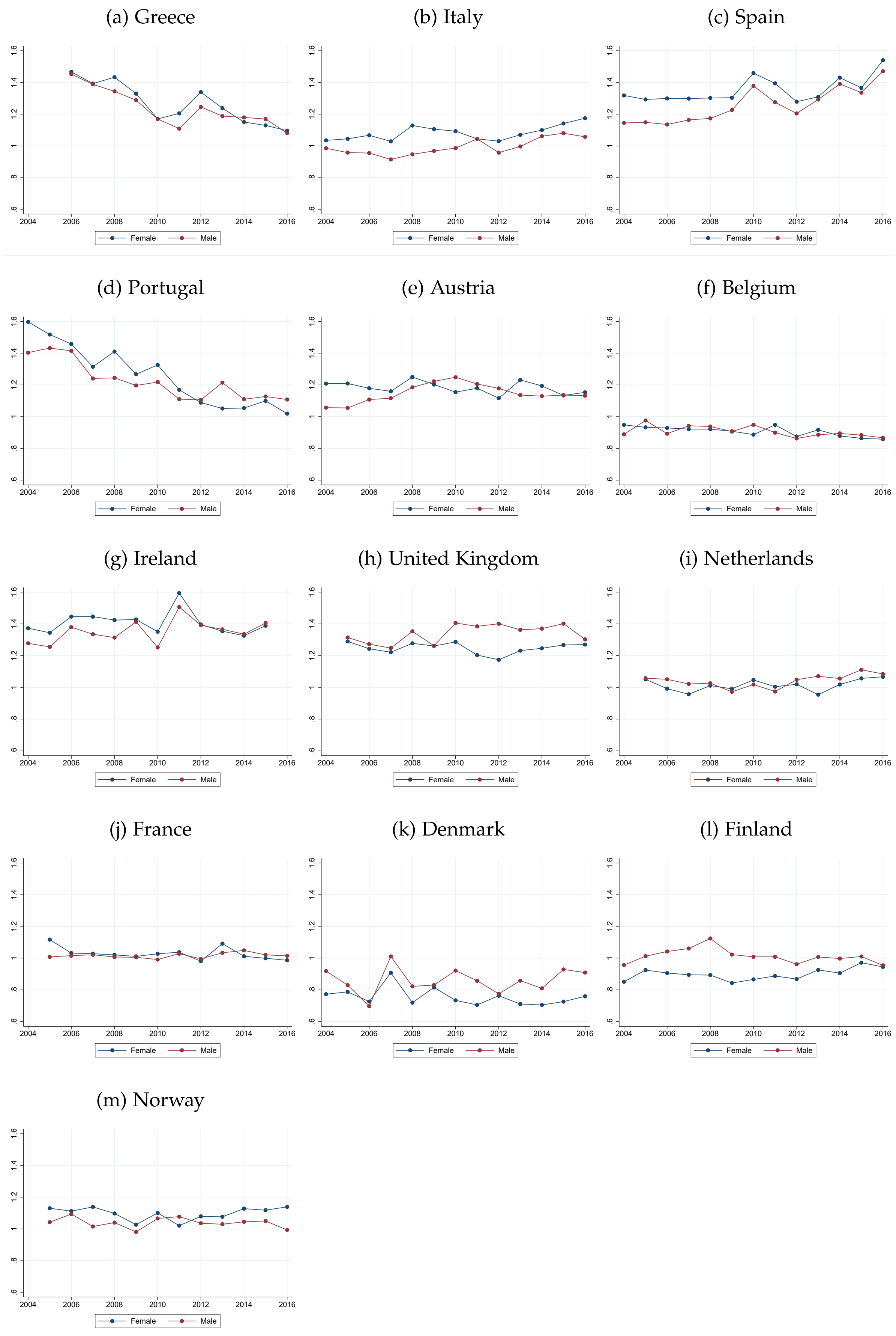
Source: EU-SILC and authors' calculations.

Table A3: Employment Rates by Education

	Employment Rate in 2007				Changes over 2007-2012				Changes over 2012-2016			
	College		No college		College		No college		College		No college	
	M	F	M	F	M	F	M	F	M	F	M	F
Greece	.917	.787	.878	.444	-.194	-.109	-.232	-.091	.054	-.032	-.014	.027
Italy	.857	.787	.864	.521	.019	-.006	-.056	-.004	-.020	-.039	-.002	-.013
Portugal	.854	.839	.879	.674	-.062	-.034	-.115	.013	.095	.074	.050	.018
Spain	.932	.811	.867	.542	-.113	-.083	-.212	-.069	.069	.069	.032	.063
Southern	**.890**	**.806**	**.872**	**.545**	**-.088**	**-.058**	**-.154**	**-.038**	**.049**	**.018**	**.016**	**.023**
Austria	.951	.767	.878	.678	-.013	.037	-.001	.037	-.008	.002	-.013	.008
Belgium	.913	.820	.884	.651	-.021	.034	-.062	.038	.022	.044	.012	-.021
France	.948	.892	.902	.765	-.009	.000	-.030	.006	-.025	.006	-.039	-.103
Netherlands	.976	.906	.954	.780	-.017	.018	-.028	-.013	-.029	-.117	-.063	-.045
Continental	**.947**	**.846**	**.905**	**.718**	**-.015**	**.022**	**-.030**	**.017**	**-.010**	**-.016**	**-.026**	**-.040**
Ireland	.919	.837	.830	.583	-.106	-.104	-.221	-.101	-.813	-.733	-.608	-.482
United Kingdom	.950	.849	.926	.780	-.009	-.004	-.059	-.064	.024	.002	.054	-.010
Anglosaxon	**.934**	**.843**	**.878**	**.682**	**-.057**	**-.054**	**-.140**	**-.082**	**-.395**	**-.366**	**-.277**	**-.246**
Denmark	.942	.912	.977	.900	-.022	-.033	-.090	-.097	-.009	.058	.020	.075
Finland	.945	.861	.814	.800	-.015	-.004	.009	-.032	-.038	-.008	-.023	-.020
Norway	.978	.934	.963	.928	.000	.016	-.006	-.024	.002	.008	-.011	-.006
Nordic	**.955**	**.902**	**.918**	**.876**	**-.012**	**-.007**	**-.029**	**-.051**	**-.015**	**.019**	**-.005**	**.016**

Source: EU-SILC and authors' calculations.

Figure A1: Cross-country wage inequality, 2004-2016.



Notes.— Wage inequality is measured by logarithm of the ratio between wages at 90th and 10th percentiles. Source: EU-SILC and authors' calculations.

Table A4: Median Wage Gaps under Imputation on Education and Experience - Probabilistic Model

	Levels in 2007				Changes over 2007-2012				Changes over 2012-2016			
	Raw Wage Gap	True Wage Gap	Selection Bias M	Selection Bias F	Raw Wage Gap	True Wage Gap	Selection Bias M	Selection Bias F	Raw Wage Gap	True Wage Gap	Selection Bias M	Selection Bias F
Greece	.182	.411	.018	.247	-.089	-.093	.061	.058	-.030	-.065	-.017	-.052
Italy	.035	.184	.015	.164	.051	.037	.015	.001	-.028	-.028	-.006	-.007
Portugal	.172	.219	.012	.059	-.059	-.090	.020	-.010	.000	.019	-.020	-.001
Spain	.131	.201	.013	.083	-.020	-.011	.029	.039	-.047	-.042	.003	.008
Southern	**.130**	**.254**	**.014**	**.138**	**-.030**	**-.039**	**.031**	**.022**	**-.026**	**-.029**	**-.010**	**-.013**
Austria	.189	.269	.012	.093	.015	-.012	-.007	-.034	-.004	-.013	.008	.000
Belgium	.074	.142	.022	.091	-.019	-.049	.003	-.028	.017	.021	-.010	-.005
France	.114	.143	.006	.035	.005	-.013	.012	-.007	-.006	.010	.000	.015
Netherlands	.158	.179	.000	.021	-.048	-.028	.005	.025	.073	.043	.018	-.012
Continental	**.133**	**.183**	**.010**	**.060**	**-.012**	**-.026**	**.003**	**-.011**	**.020**	**.015**	**.004**	**.000**
Ireland	.170	.272	.043	.145	-.039	-.031	.040	.048	-.054	-.078	.003	-.020
UK	.247	.260	.013	.027	-.064	-.045	.013	.032	-.005	.011	-.014	.002
Anglosaxon	**.208**	**.266**	**.028**	**.086**	**-.052**	**-.038**	**.026**	**.040**	**-.030**	**-.033**	**-.006**	**-.009**
Denmark	.116	.122	-.003	.003	-.036	-.042	.003	-.003	.026	.016	.023	.012
Finland	.203	.209	.016	.022	-.049	-.070	.010	-.011	-.021	-.010	-.014	-.002
Norway	.154	.157	.004	.006	.020	.017	.000	-.003	-.013	-.013	.002	.002
Nordic	**.158**	**.162**	**.006**	**.010**	**-.022**	**-.032**	**.004**	**-.005**	**-.003**	**-.002**	**.003**	**.004**

Source: EU-SILC and authors' calculations. Note: Selection bias = an increase in observed wage due to selection. Wage imputation rule: Impute wage $<(>)$ median with probability $\hat{P}_i$ (respectively, $1-\hat{P}_i$) if non-employed, where $\hat{P}_i$ is the predicted probability of earning a wage below the gender-specific median, as estimated from a probit model including education dummies, experience and its square on an enlarged base sample with wage observations from other waves.

Table A5: Median Wage Gaps under Imputation Based on Wages from Other Waves

	Levels in 2007				Changes over 2007-2012				Changes over 2012-2016			
	Raw Wage Gap	Potential Wage Gap	Selection Bias M	Selection Bias F	Raw Wage Gap	Potential Wage Gap	Selection Bias M	Selection Bias F	Raw Wage Gap	Potential Wage Gap	Selection Bias M	Selection Bias F
Greece	.182	.194	.016	.027	-.089	-.115	.046	.021	-.030	-.017	-.037	-.024
Italy	.035	.045	.010	.021	.051	.045	.016	.011	-.028	-.015	-.012	.001
Portugal	.172	.177	.006	.011	-.059	-.069	.011	.001	.000	.012	-.006	.005
Spain	.131	.142	.012	.023	-.020	-.038	.032	.014	-.047	-.008	-.029	.010
Mean	**.130**	**.140**	**.011**	**.021**	**-.030**	**-.044**	**.026**	**.012**	**-.026**	**-.007**	**-.021**	**-.002**
Austria	.189	.190	.016	.017	.015	.032	.000	.017	-.004	-.002	.001	.004
Belgium	.074	.075	.006	.007	-.019	-.021	.011	.009	.017	.009	-.009	-.017
France	.114	.127	.006	.020	.005	-.009	.007	-.007	-.006	-.003	-.004	-.002
Netherlands	.158	.155	.012	.009	-.048	-.036	-.002	.010	.073	.062	.000	-.011
Continental	**.133**	**.137**	**.010**	**.013**	**-.012**	**-.008**	**.004**	**.007**	**.020**	**.017**	**-.003**	**-.006**
Ireland	.170	.190	.014	.035	-.039	-.061	.015	-.008	-.054	-.032	-.028	-.006
UK	.247	.250	.010	.013	-.064	-.075	.004	-.007	-.005	.002	-.008	-.002
Anglosaxon	**.208**	**.220**	**.012**	**.024**	**-.052**	**-.068**	**.009**	**-.007**	**-.030**	**-.015**	**-.018**	**-.004**
Denmark	.116	.119	.003	.006	-.036	-.045	.005	-.004	.026	.034	-.006	.001
Finland	.203	.184	.026	.008	-.049	-.044	-.007	-.001	-.021	-.013	-.008	.000
Norway	.154	.159	.009	.013	.020	.014	.000	-.005	-.013	-.006	-.004	.002
Nordic	**.158**	**.154**	**.013**	**.009**	**-.022**	**-.025**	**-.001**	**-.004**	**-.003**	**.005**	**-.006**	**.001**

Source: EU-SILC and authors' calculations. Note: Selection bias = an increase in observed wage due to selection. Wage imputation rule: Impute wage from other waves when non-employed.

Table A6: Rate of Imputation Based on Wages from Other Waves

	2007		2012		2016	
	M	F	M	F	M	F
Greece	.37	.15	.38	.22	.22	.15
Italy	.36	.18	.46	.23	.33	.17
Spain	.53	.36	.54	.38	.35	.32
Portugal	.75	.78	.69	.74	.50	.60
Southern	**.50**	**.37**	**.52**	**.39**	**.35**	**.31**
Austria	.40	.42	.52	.47	.43	.39
Belgium	.33	.29	.44	.39	.41	.26
France	.62	.55	.63	.59	.51	.53
Netherlands	.65	.41	.42	.27	.40	.23
Continental	**.50**	**.42**	**.50**	**.43**	**.44**	**.35**
Ireland	.28	.20	.19	.11	.19	.16
UK	.37	.34	.33	.28	.44	.27
Anglosaxon	**.33**	**.27**	**.26**	**.20**	**.31**	**.22**
Denmark	.38	.51	.65	.58	.16	.33
Finland	.70	.63	.57	.61	.40	.55
Norway	.60	.70	.47	.50	.36	.37
Nordic	**.56**	**.61**	**.56**	**.56**	**.31**	**.42**

Source: EU-SILC and authors' calculations. Note: Wage imputation rule: Impute wage from other waves when non-employed. Imputation Rate = proportion of imputed wage observations in total nonemployment.