

ORIGINAL ARTICLE

Open Access



# Measuring the effect of gender segregation on the gender gap in time-related underemployment

Juan Acosta-Ballesteros<sup>1,2</sup> , María del Pilar Osorno-del Rosal<sup>1,2</sup> and Olga María Rodríguez-Rodríguez<sup>1,2\*</sup>

## Abstract

This paper focuses on the impact that gender segregation in the labour market exerts on the underemployment gender gap for young adult workers in Spain. In order to analyse the relative importance of segregation in this gap, we develop a methodology based on two counterfactual simulations that provides a detailed decomposition of the gap into endowments and coefficients effects as well as the interaction of these effects. To the best of our knowledge, we are the first to perform a decomposition using bivariate probit models with sample selection. Using annual samples of the Spanish Labour Force Survey 2006–2016, the results show that working in female-dominated occupations or industries hinders working as many hours as desired, especially for women. Furthermore, we conclude that the gender gap in underemployment is mainly due to the different distribution of male and female workers across occupations and industries. Additionally, the different impact by gender that working in the same gender-typing jobs exerts on the risk of underemployment contributes to widening the gap.

**Keywords:** Gender gap, Occupational and Industry segregation, Time-related underemployment, Counterfactual simulations, Decomposition analysis

**JEL Classification:** J16, J22

## 1 Introduction

Time-related underemployment, which refers to those workers who would like to work more hours than available, is a persistent problem in labour markets and the Spanish one is no exception. Moreover, this problem increased during the Great Recession in many countries (Bell and Blanchflower 2013; Acosta-Ballesteros et al. 2018).

Furthermore, women experience this situation more often than men do (Weststar 2011; Kjeldstad and Nymoen 2012a, b; Vuluku et al. 2013; Acosta-Ballesteros et al. 2018), therefore, we must pay attention to the reasons for this gender gap. Particularly, occupational and

industry segregation may be an important factor, as suggested by Barret and Doiron (2001), since a higher underemployment rate has been linked to female-dominated occupations and industries (Kjeldstad and Nymoen 2012a, Kjeldstad and Nymoen, 2012b, and Kameråde and Richardson, 2018). Moreover, as Spain has experienced higher levels of segregation than other European countries in a persistent way (Iglesias-Fernández et al. 2012), analysing the effect of segregation on the underemployment gender gap in the Spanish labour market emerges as an interesting research issue.

To the best of our knowledge, only Vuluku et al. (2013) have tried to explain the underemployment gender gap, but they did not include any occupational and industry segregation indicators in their study. Thus, to overcome this shortcoming in the literature, the main objective of this article is to carry out an in-depth analysis of the underemployment gender gap. Specifically, we intend

\*Correspondence: orguez@ull.es

<sup>1</sup> Departamento de Economía, Contabilidad y Finanzas, Universidad de La Laguna (ULL), La Laguna (Tenerife), Spain

Full list of author information is available at the end of the article

to test whether segregation plays an important role in explaining it, as well as to quantify how much is due to men and women working in different industries and occupations, and how much is due to men and women facing different underemployment risks when they work in the same gender-typing jobs.

To do this, first, we quantify the effect of occupational and industry segregation on workers' underemployment risk using a detailed measure of gender segregation. Estimating this impact through bivariate models with selection enables us to handle the potential sample selection bias due to estimating the probability of underemployment just for employed people. Second, as we are not aware of a methodology that allows decomposing a gap using this kind of model, we develop one that is inspired by the Fairlie technique (1999, 2005, and 2017). It is based on two counterfactual simulations that provide a detailed decomposition of the gap into effects due to workers having different characteristics, effects due to these characteristics having different returns, and the interaction of both these effects.

We focus on young workers because this collective is especially affected by underemployment. Data from the Spanish Labour Force Survey (LFS) indicate that in 2017 the underemployment rate for workers under 35 was 14.7%, while the figure was 8.5% for workers older than 34 years old. Additionally, by looking at people right at the beginning of their careers we can avoid many of the cumulative advantages/disadvantages that people may have experienced throughout their careers. Thus, focusing on young workers allows us to have a current view of underemployment patterns and of the gender gap in it, avoiding possible gender differences from the past.

This paper is organised as follows. It begins with the conceptual framework on time-related underemployment and puts forward our working hypotheses. The next section describes the methodological approach used. Then, the data and variables used in the econometric model are presented. This is followed by the results, while a discussion of these findings is provided at the end.

## 2 Conceptual framework and hypotheses

According to neoclassical theory, individuals can choose their working hours freely from a continuous time distribution; these hours are chosen by maximising a utility function subject to a particular budget constraint. Nevertheless, employers and trade unions' decisions, the degree of labour mobility and economic conditions determine the actual hours offered to employees (Simic 2002). Therefore, workers' preferred and actual hours may not coincide, so some individuals will work either more (overemployed) or less (underemployed) than they want. Thus, time-related underemployment means that

some employed people would like to work more hours than available.

The demographic and job factors that determine time-related underemployment have been previously analysed in the literature (Hakim 1997; Weststar 2011; Prause and Dooley 2011; McKee-Ryan and Harvey 2011; Kjeldstad and Nymoen 2012a, 2012b; Wilkins 2006; Acosta-Ballesteros et al. 2018). Particularly, significant differences in underemployment have been found across occupations and industries (Kjeldstad and Nymoen 2012a, 2012b; Valletta et al. 2016). However, very few studies have linked these differences to occupational and industry gender segregation in the labour market. Thus, Kjeldstad and Nymoen (2012a) and Kjeldstad and Nymoen (2012b) find a higher underemployment risk in those occupations and sectors that are traditionally female-dominated, with a stronger effect for men, although they do not include specific variables for segregation in their econometric model. Kameråde and Richardson (2018) consider segregation measures in their analysis, and they also find a higher likelihood of underemployment in female-dominated occupations; however, this effect is not so clear across industries. Additionally, Dueñas-Fernández et al. (2016), who do not focus specifically on this issue, analyse involuntary part-time work in Spain and find that segregation, especially occupational, is strongly related to part-time work (particularly for women).

There are several reasons that explain higher underemployment risks in female-dominated occupations and industries. In this sense, as Kameråde and Richardson (2018) point out, women are mainly employed in labour intensive jobs where employers can change the number of hours their employees work to adapt to fluctuations in demand. Therefore, part-time or short-schedule jobs are more likely to be found in female-dominated occupations, which may lead to underemployment. Moreover, female-dominated occupations usually require low qualifications. In addition, female workers tend to cluster in industries that offer comparatively low payment for the same level of qualification, such as in education, health and social work activities (Boll et al. 2016). This pattern also translates into a frequent desire to work more hours. Conversely, male-dominated occupations are typically characterised by better-paid jobs and are usually related to more stable, full-time contracts (Hegewisch et al. 2010). Furthermore, male workers are overrepresented in industries that offer high rewards for the same level of qualification (particularly manufacturing). Therefore, male-dominated occupations and industries often lead to low underemployment rates.

Despite these arguments suggesting the important role that occupational and industry segregation plays in the likelihood of underemployment, accurate estimates of

its impact have not been achieved in the aforementioned research. In this article, we overcome this shortcoming using a more suitable estimation strategy. Specifically, we propose and test the following hypothesis:

*Hypothesis 1: working in female-dominated occupations and industries implies a higher probability of time-related underemployment than being employed in male-dominated ones, both for men and women.*

As women face a higher risk of underemployment than men do, there is a gender gap regarding this handicap and occupational and industry segregation may have an important impact on it. Furthermore, this effect may be partially due to the uneven distribution of men and women across different jobs, as suggested by Barrett and Doiron (2001). Additionally, differences in the returns that working in female or male-dominated jobs imply should also be considered. These authors, as a simple exercise, give women the average male distribution across occupations and industries and conclude that the main reason that explains women being involuntary part-timers more often than men is simply being employed in different industries and occupations.

Interestingly, previous research highlights the fact that men may benefit from their minority status in female-dominated jobs in several ways (Simpson 2004). In this sense, as reviewed in Lupton (2006), men progress more quickly than women do to senior positions avoiding the problem of the “glass ceiling” inherent in vertical segregation. Additionally, men may be channelled into certain specialties in occupations that are regarded as more appropriate to their gender. As a third advantage, men are paid more than women are in female-dominated occupations (Torre 2018). By contrast, women may face negative outcomes in male-dominated jobs (Simpson 1997, 2000). Thus, for example, as Martin and Barnard (2013) find, formal and covert organisational practices, which maintain gender discrimination and bias, are the main challenges that women face. These arguments may also apply regarding underemployment, so female workers may face a higher risk of underemployment than men both in female and male-dominated jobs.

Nevertheless, to the best of our knowledge, only Vuluku et al. (2013) have tried to identify the reasons behind the underemployment gender gap, though they do not include any measure of gender segregation in their analysis and use univariate models, which can lead to biased estimations. We fill this gap in the literature using a new methodology that allows us to propose and test the following hypotheses:

*Hypothesis 2a: The gender gap in underemployment*

*is mainly due to the different distribution of male and female workers across occupations and industries.*

*Hypothesis 2b: The different impact that working in certain occupations and industries exerts on the risk of male and female underemployment contributes to widening the gender gap.*

### 3 Methodology

As a first step, to analyse the effect of occupational and industry segregation on time-related underemployment, we estimate two bivariate probit selection models (Greene 2012), one for men and another for women. These models enable us to handle the potential sample selection bias due to estimating the probability of underemployment just for employed people, as Acosta-Baltes et al. (2018) have already shown.

Let us define  $y_1^*$  and  $y_2^*$  as the latent variables reflecting the likelihood of being underemployed and employed, respectively. Thus, the model can be specified as follows:

$$y_{i1}^* = x_{i1}\gamma_1 + \varepsilon_{i1}, y_{i1} = 1 \quad \text{if } y_{i1}^* > 0, 0 \text{ otherwise} \quad (1)$$

$$y_{i2}^* = x_{i2}\gamma_2 + \varepsilon_{i2}, y_{i2} = 1 \quad \text{if } y_{i2}^* > 0, 0 \text{ otherwise} \quad (2)$$

with  $(y_{i1}, x_{i1})$  observed only when  $y_{i2} = 1$ .

In these equations,  $y_{i1}$  indicates if worker  $i$  is underemployed and  $y_{i2}$  if the individual is employed; row vector  $x_{i1}$  contains the variables explaining underemployment;  $x_{i2}$  reflects the variables determining employment. As usual, the independent variables that have a qualitative nature are included in the model as dummy variables or as groups of them. Finally,  $\varepsilon_{i1}$  and  $\varepsilon_{i2}$  are the error terms, which follow a bivariate normal distribution with mean zero, variance equal to 1 and covariance  $\rho$ .

To test if working in female-dominated occupations and industries implies a higher probability of underemployment than working in male-dominated ones (*Hypothesis 1*), we analyse the estimated marginal effects of occupational and industry segregation on the probability of underemployment. Since the model is bivariate with selection, these partial effects (like those regarding the rest of variables) are obtained using the conditional probability of underemployment given employment. In addition, the marginal effects on the probability of employment are computed using the selection equation.<sup>1</sup>

To simplify notation, we redefine the variables and coefficients in Eqs. (1) and (2) as follows. The variables

<sup>1</sup> The interdependency caused by groups of dummies has been taken into account in order to calculate the marginal effects.

considered in both equations for individual  $i$  are gathered in  $x_i$ , which is a row vector including vectors  $x_{i1}$  and  $x_{i2}$ . Additionally, vector  $\beta_1$  contains the estimated values for  $\gamma_1(\widehat{\gamma}_1)$  and takes value zero for those variables in  $x_{i2}$  which are not included in  $x_{i1}$ . In a similar way,  $\beta_2$  includes components equal to zero for those variables considered in Eq. (1) but not in Eq. (2). Thus,  $x_i\beta_1 \equiv x_{i1}\widehat{\gamma}_1$  and  $x_i\beta_2 \equiv x_{i2}\widehat{\gamma}_2$ .

According to this notation, the estimated probability of being underemployed conditioned to being employed for individual  $i$  is:

$$F_i(x_i\beta_1^j, x_i\beta_2^j, \rho^j) = \frac{BVN(x_i\beta_1^j, x_i\beta_2^j, \rho^j)}{\Phi(x_i\beta_2^j)} \quad (3)$$

where  $\Phi$  is the cumulative standard normal distribution function and BVN is the joint cumulative distribution of the bivariate normal. Superscript  $j$  refers to men ( $M$ ) or women ( $W$ ).

As stated above, our main objective is to identify the most relevant factors explaining the gender gap in underemployment and, more specifically, to test if gender segregation accounts for an important portion of it. To achieve this goal and test *Hypotheses 2a* and *2b*, a detailed decomposition of the gap is required.

The traditional Oaxaca-Blinder two-fold decomposition (Blinder, 1973 and Oaxaca, 1973) of the gap into endowments (portion of the gap due to group differences in observable characteristics) and coefficients effects (the “unexplained” portion of the gap) cannot be applied because our model is not linear. Previous research (Even and Macpherson, 1990; Doiron and Riddell, 1994; Fairlie 1999, 2005, 2017; Yun, 2004, 2008; Powers et al., 2011; and Bazen et al., 2017) has decomposed the gap in probit and logit models, with the Fairlie and Yun techniques being the two most widely applied. However, as we estimate a nonlinear model with two equations, we develop a new procedure to decompose the gap, which extends the Fairlie technique to this kind of model. We have chosen the Fairlie approach as our starting point because it uses a non-linear function to obtain the gap decomposition, while in the Yun procedure, the curvature of the corresponding function is not considered.

According to the Fairlie technique, the contribution of each observable variable to the explained portion of the gap is equal to the change in the average predicted probability from replacing (for instance) the female distribution with the male distribution of that variable (keeping constant the rest). The procedure he proposed is matching one-to-one individuals in the female and male subsamples and switching the distributions of variables sequentially from a woman to a man. Nevertheless, the order of switching is potentially important because in

non-linear models, the independent contribution of one variable to the gap depends on the value of the other variables, which may imply a path dependence problem. Moreover, Fairlie methodology does not identify the coefficients effect corresponding to a specific variable,<sup>2</sup> which is required to test *Hypothesis 2b*.

The aggregate decomposition in our methodology, which is a direct extension of Fairlie’s, is defined by Eqs. (4) to (6), where  $E_W$  reflects the endowments effect using as weights women’s coefficients, and  $C_M$  quantifies the coefficients effect using as weights men’s characteristics:

$$Gap = C_M + E_W \quad (4)$$

$$C_M = \frac{\sum_{\forall Men} F_i(x_i^M\beta_1^M, x_i^M\beta_2^M, \rho^M)}{N^M} - \frac{\sum_{\forall Men} F_i(x_i^M\beta_1^W, x_i^M\beta_2^W, \rho^W)}{N^M} \quad (5)$$

$$E_W = \frac{\sum_{\forall Men} F_i(x_i^M\beta_1^W, x_i^M\beta_2^W, \rho^W)}{N^M} - \frac{\sum_{\forall Women} F_i(x_i^W\beta_1^W, x_i^W\beta_2^W, \rho^W)}{N^W} \quad (6)$$

Summations in (5) and (6) are across the subsample of the employed, as we decompose differences in the average predicted probabilities of being underemployed conditioned to being employed. Thus,  $N^W$  and  $N^M$  indicate the sample size for employed women and men, respectively.

An alternative decomposition (Eq. 7) with each component evaluated using as weights the other gender coefficients or endowments is also possible. However, we do not define and explain it here because it is symmetric to this one.

$$Gap = E_M + C_W \quad (7)$$

To obtain a detailed decomposition, we develop a methodology based on two counterfactual simulations that identify the contribution of each variable to both  $E_W$  and  $C_M$ . These simulations can be used together to approximate the total impact of a specific variable on the underemployment gender gap.

The first one provides a detailed decomposition of the endowments effect and has been designed for discrete variables<sup>3</sup> (as most of the variables in the labour market

<sup>2</sup> Fairlie does not focus on the “unexplained” portion of the gap because of the difficulty in interpreting results. As we explain below, this shortcoming can be overcome using the Kim (2013) methodology.

<sup>3</sup> For a continuous variable, we propose equalling the distribution of frequencies by gender as described, but considering the variable is discrete with a very large number of categories.

literature). It is inspired by Fairlie, who pointed out that a potential solution to the path dependence problem “is to estimate each contribution by switching the variable of interest first” (Fairlie 2005, page 313), as our method does. Specifically, we calculate the contribution of a single variable  $k$  as the change in women’s average conditional probability of underemployment resulting from switching women from the categories where they are over-represented to those where they are under-represented. This procedure is carried out until women’s relative frequencies across the categories of  $k$  are equal to men’s ones. The selection of women who are switched is random, so the procedure is repeated 50 times to ensure consistency, and then the results are averaged.<sup>4</sup> As the changes described affect 10% of the observations or less for most variables, the change in the probability of underemployment is due to a relatively small change in the data.

Specifically, the contribution of a single variable  $k$ , denoted as  $\theta_E^W(k)$ , can be computed as described in Eq. (8):

$$\theta_E^W(k) = \frac{\sum_{\forall \text{Women}} F_i(x_i^{W \rightarrow M}(k) \beta_1^W, x_i^{W \rightarrow M}(k) \beta_2^W, \rho^W)}{N^W} - \frac{\sum_{\forall \text{Women}} F_i(x_i^W \beta_1^W, x_i^W \beta_2^W, \rho^W)}{N^W} \quad (8)$$

where  $x_i^{W \rightarrow M}(k)$  contains the same information as  $x_i^W$  but variable  $k$  has been modified as described<sup>5</sup> and  $k=1, 2, \dots, n$ , where  $n$  refers to the number of categorical variables included in  $x_i$ .

The sum of the individual contributions of all the variables does not exactly equal the endowments effect. Thus, the summing up property, which the method proposed by Fairlie has, does not satisfy. So we can write:

$$E_W = \sum_{k=1}^n \theta_E^W(k) + D_E^W \quad (9)$$

An approximation error ( $D_E^W$ ) emerges because the endowments effect ( $E_W$ ) in the aggregate decomposition is computed by switching all the variables simultaneously. Conversely, in our simulation, we switch only one variable at a time. As the conditional probability of underemployment is not linear, both results are slightly different. Although when both expressions,  $E_W$  and  $\sum_{k=1}^n \theta_E^W(k)$ , are linearised they coincide, a disparity emerges from the

differences in Taylor expansion remainders (see Appendix 1 in Additional file 1).

To calculate the detailed decomposition of the coefficients effect, we propose a second counterfactual simulation following a similar procedure to that used in the first one. Oaxaca and Ransom (1999) show that this decomposition is destined to suffer from an identification problem, since the detailed coefficients effect attributed to dummy or categorical variables is not invariant to the choice of reference groups. Gardeazabal and Ugidos (2004) and Yun (2005) propose methods to solve this problem. Despite being widely used, these approaches show some limitations.<sup>6</sup> Thus, we use the grand-mean method that Kim (2013) proposes. This method appears to be a good option for analyses regarding labour market outcomes because it accurately estimates the extent to which each variable contributes to the group differences. Additionally, it gives a meaning to the intercept term and to the coefficient component of each dummy variable.

Specifically, we calculate the coefficient effect related to a specific variable  $k, \theta_\beta^M(k), k=1, \dots, n$ , as the change

in men’s average conditional probability of underemployment if the parameter of a specific characteristic were that of women. Additionally, it is necessary to include the change corresponding to parameter  $\rho, \theta_\rho^M$ . These effects are described in Eqs. (10) and (11), where  $\beta_1$  and  $\beta_2$  are the estimated transformed coefficients of the model according to Kim (2013) method.<sup>7</sup> In  $\beta_1^{\sim M \rightarrow W}(k)$  and  $\beta_2^{\sim M \rightarrow W}(k)$ , men’s coefficients corresponding to the variable  $k$  have been replaced by women’s ones.

$$\theta_\beta^M(k) = \frac{\sum_{\forall \text{Men}} F_i(x_i^M \beta_1^{\sim M \rightarrow W}(k), x_i^M \beta_2^{\sim M \rightarrow W}(k), \rho^M)}{N^M} - \frac{\sum_{\forall \text{Men}} F_i(x_i^M \beta_1^{\sim M}, x_i^M \beta_2^{\sim M}, \rho^M)}{N^M} \quad (10)$$

<sup>6</sup> According to Fortin et al. (2011) and Kim (2013), these normalizations have several limitations: they may leave the estimation and decomposition without a simple meaningful interpretation; they will likely be sample specific and make comparisons across studies impossible; and they are sensitive to the number of categories and to the grouping method.

<sup>7</sup> The estimated coefficients are transformed by subtracting from each of them the grand-mean weighted sum of the coefficients of each variable  $\beta_1^{\sim M}(k)$  or  $\beta_2^{\sim M}(k)$ ,  $k=1, \dots, n$ . It is also necessary to sum  $\sum_{k=1}^n \beta_1^{\sim M}(k)$  or  $\sum_{k=1}^n \beta_2^{\sim M}(k)$  to the intercepts in order to transform them.

<sup>4</sup> This number of repetitions was selected after an analysis of sensitivity. We decided to choose 50 because the average difference in the results found with respect to using 200 was around  $10^{-5}$  and the standard errors could be computed in a reasonable time.

<sup>5</sup> Note that each discrete variable  $k$  is included as a set of dummies in  $x_i$ .

$$\theta_{\rho}^M = \frac{\sum_{\forall Men} F_i \left( x_i^M \tilde{\beta}_1^M, x_i^M \tilde{\beta}_2^M, \rho^W \right)}{N^M} - \frac{\sum_{\forall Men} F_i \left( x_i^M \tilde{\beta}_1^M, x_i^M \tilde{\beta}_2^M, \rho^M \right)}{N^M} \quad (11)$$

Again, summing the individual contributions of the variables does not exactly equal the coefficients effect,  $C_M$ . An approximation error ( $D_C^M$ ) emerges for the same reasons already explained (see Appendix 2 in Additional file 1). Thus, we can write<sup>8</sup>:

$$C_M = - \left\{ \left( \sum_{k=1}^n \theta_{\beta}^M(k) \right) + \theta_{\rho}^M \right\} + D_C^M \quad (12)$$

Since our detailed decomposition of the gap is the sum of both expressions ( $E_W$  and  $C_M$ ), the approximation errors imply that the sum of individual contributions of all the variables does not equal the gap. To assess the magnitude of this disparity, in the Results section, we display the approximation errors of our decomposition.

Despite our decomposition of the gap not being exact, it provides technical advantages compared to Fairlie decomposition procedure, as well as being applicable to a bivariate probit model. Thus, its economic interpretation is straightforward, and it avoids the path dependence problem, since it always uses the same starting point, real women (or men) in the sample, and only one characteristic is modified. Moreover, our approach does not require a one-to-one matching of individuals, since we replicate the distribution of each specific variable and the number of women who have a specific characteristic changed is just those strictly necessary, so we keep almost real individuals. Conversely, in Fairlie decomposition technique, each woman is randomly matched with a man in the sample, and she takes his characteristics sequentially until she becomes that man. As the sequential change of characteristics is made, it is likely that the remaining combination of characteristics will be unreal. In addition, our methodology offers a simulation that allows us to approximate a detailed decomposition of the coefficients effect.

Even though we could test our hypotheses using a two-fold decomposition, it is increasingly common in the literature to use a three-fold one (Daymont and Andrisani, 1984), which has the advantage that endowments and coefficients effects are computed from the same starting

point, so they are more easily interpreted. This three-fold decomposition when the starting point is women<sup>9</sup> can be easily obtained from Eqs. (4) or (7) and can be expressed as:

$$Gap = E_W + C_W + (C_M - C_W) \quad (13)$$

$$Gap = E_W + C_W + (E_M - E_W) \quad (14)$$

The new term,  $C_M - C_W = E_M - E_W$ , can be interpreted as an interaction component that indicates the portion of the gap that occurs when both endowments and coefficients change simultaneously. Alternatively, it is the portion of the gap that remains after controlling for the endowments and coefficients effects. This interaction component is more difficult to interpret than the first two and is often disregarded. However, we believe, as Etezady et al. (2021), that neglecting it provides a substantially incomplete picture of the total influences of endowments and coefficients to the gap. Thus, our analysis is based on Eq. (14).

Some final comments regarding our methodology need to be pointed out. First, to obtain the standard errors for the results of both counterfactual simulations, which are necessary to test if the corresponding changes in the probability of underemployment are statistically significant, Krinsky and Robb's (1986) method has been applied,<sup>10</sup> as Dowd et al. (2014) explain.

Second, the survey structure of our data has been taken into account in the methodology. Thus, the bivariate probit selection models have been estimated considering sample weights and cluster-robust standard errors. Additionally, the sample weights have been considered in both counterfactual simulations by replicating each observation according to its weight.<sup>11</sup>

Third, our methodology is displayed for bivariate probit models with sample selection, but it can also be easily applied to single equation models like the probit or logit ones. This fact allows us to carry out some robustness analyses. Thus, we specify univariate probit models to explain underemployment and we obtain the three-fold decomposition of the gap. These results are compared to

<sup>9</sup> Equations 13 and 14 can also be proposed for men.

<sup>10</sup> In studies based on survey data not only the outcome variable but also the predictors are subject to sampling variation (Jann 2008). It implies that the standard errors may be underestimated, especially those regarding the endowment component. However, the results of the models shown in Additional file 2 seem to indicate that this is not the case, since the standard errors estimated using our methodology are very similar to those obtained using the Oaxaca-Blinder or Yun methodologies.

<sup>11</sup> This process is required in the first simulation in order to switch the value of a specific variable from those categories where women (men) are over-represented. In the second simulation, weighting each observation according to its raising factor is enough.

<sup>8</sup> Note that the minus sign is required in Eq. (12) because in the simulation the average man is the starting point, while when computing  $C_M$ , the average man is the final point.

those obtained using Fairlie and Yun methodologies for a probit model and to those achieved using Oaxaca-Blinder technique for a linear model. Table S1 in Additional file 2 shows these results, which are similar to those obtained with our methodology.

#### 4 Data and variables

In this article, we use the definition of time-related underemployment directly provided by the Spanish Statistical Office. Specifically, the criteria applied in the Spanish LFS to classify workers as underemployed (in line with the International Labor Organization Bureau of Statistics recommendations) are: they would like to work more hours, they are available to do so, and they work less than the usual weekly hours of full-timers in their industry. Thus, underemployment is a more accurate indicator of labour underutilization than involuntary part-time employment. It reflects non-desired workdays for all types of workers, capturing the preference of both part-timers and full-timers to have longer workdays.

The data used come from the 2006–2016 annual samples of the Spanish LFS.<sup>12</sup> Therefore, our database is a pool of cross-sectional annual observations, since each individual is included only once in the annual sample. Our sample contains young people aged 16 to 34 who were active. The few individuals with inconsistent answers or who do not provide the necessary information for the analysis have been removed. The final sample includes 70,445 women: 73.6% are employed, and among them, 16.5% are underemployed. The corresponding figures for the male subsample are 80,962, 75.1% and 11.8%, respectively.

The independent variables included in the econometric analysis<sup>13</sup> (displayed in Table 1) reflect the main factors previously found to determine underemployment. In order to classify occupations and industries as gender-dominated or integrated, we follow the relative concept of Anker (1998). Thus, the dividing line between gender-dominated and integrated occupations (or industries) is established in relation to the average percentage of female workers in the labour force as a whole (44% over the period analysed). Specifically, we consider female-dominated occupations or industries are those having more than 1.25 times the mean percentage female, while male-dominated ones are those having less than 0.75 times the mean percentage female. If the percentage of women is between both limits, the occupation or industry is labelled as gender-integrated. Applying this

criterion, we obtain a band similar to the one in Hakim (1998), where gender-integrated occupations are characterised by a proportion of women ten percentage points around the percentage of women in total employment.

Our gender segregation measures have been computed using the three-digit codes from both occupations (according to the National Classification of Occupations, 1994 and 2011) and industries (National Classification of Economic Activities, 1993 and 2009). However, when the number of people working in a certain occupation or industry is less than 50, segregation has been defined according to two-digit or one-digit codes.<sup>14</sup> Additionally, given the methodological change in both classifications, it has been necessary to calculate the value of each segregation variable for two different sub-periods. As both gender segregation measures are correlated, we have solved the collinearity problem by defining an interaction variable with nine categories that integrates both occupational and industry segregation.

As education plays an important role in the risk of underemployment (Acosta-Ballesteros et al. 2018), we define 43 educational categories using the information provided by the LFS on education level and field of study, and according to the National Classification of Education (2000 and 2014). Ten specializations for vocational training and university degrees are distinguished.<sup>15</sup> Moreover, whether workers took longer than usual in completing their studies is also considered.

The remaining explanatory variables include nationality, having children under 16, and some additional regressors reflecting household composition; whether the individual is enrolled in formal studies is also taken into account. We also consider professional status (self-employed or employed in the public or private sector with a fixed-term or permanent contract), the size of the firm, having a recent job (tenure up to 12 months and depending on the worker's age), the unemployment rate by gender in the Autonomous Regions,<sup>16</sup> as well as a dummy variable that takes value one if the observation corresponds to the period after the labour reform of 2012.<sup>17</sup>

<sup>14</sup> This fact only occurs in a few occupations (industries) that account for 0.15% (0.89%) of workers in our sample.

<sup>15</sup> Before the Bologna Process, the Spanish education system distinguished short-cycle (three years) and long-cycle (more than three years, usually five) university degrees. The new degrees under the European Higher Education Area are included as short-cycle programmes.

<sup>16</sup> This is the only continuous variable in the model. As it is already defined by gender, it is not necessary to develop the procedure explained above. Thus, in the first simulation, the change in this characteristic has been carried out by simply attributing each woman the unemployment rate she would face if she were a man (and vice-versa).

<sup>17</sup> This reform, among other measures, allows firms to reduce the working hours of their employees more easily than before, and may partially explain the relatively high underemployment rate observed since 2012 in Spain.

<sup>12</sup> The accuracy of the results in this paper using these data is our sole responsibility.

<sup>13</sup> The frequencies of the independent variables are provided in Table S2 in Additional file 2.

**Table 1** Average marginal effects: underemployment and employment by gender

Variable	Women		Men	
	Under	Employ	Under	Employ
Educational attainment (omitted: primary education or less)				
Compulsory secondary	− 0.006	0.077***	0.003	0.082***
Non-compulsory secondary	− 0.024***	0.141***	− 0.013	0.142***
Lower vocational training				
Education	− 0.026	0.116***	0.049	0.181***
Arts and humanities	0.012	0.170***	0.066	0.116***
Social sciences	− 0.092**	0.342***	− 0.109***	0.149
Business, administration, law	− 0.014	0.134***	− 0.009	0.133***
Sciences	− 0.012	0.194***	0.036	0.029
ICT	− 0.052	0.124**	0.060*	0.135***
Technology	− 0.026	0.096***	0.002	0.138***
Agriculture	0.007	0.121**	− 0.021	0.163***
Health	− 0.036***	0.183***	0.005	0.127***
Social services	− 0.031***	0.145***	− 0.006	0.122***
Higher vocational training				
Education	− 0.019	0.171***	0.073*	0.190***
Arts and humanities	− 0.016	0.136***	− 0.011	0.138***
Social sciences	0.025	0.153***	0.020	0.179***
Business, administration, law	− 0.034***	0.171***	− 0.024*	0.169***
Sciences	− 0.054**	0.174***	− 0.045	0.160***
ICT	− 0.028	0.187***	− 0.042***	0.206***
Technology	− 0.005	0.174***	− 0.016*	0.179***
Agriculture	− 0.005	0.030	0.020	0.156***
Health	− 0.038***	0.181***	− 0.003	0.164***
Social services	− 0.036**	0.153***	− 0.013	0.155***
Short-cycle university				
Education	− 0.028**	0.224***	0.021	0.185***
Arts and humanities	0.062*	0.125***	− 0.070***	0.117***
Social sciences	0.012	0.132***	0.055	0.147***
Business, administration, law	− 0.059***	0.169***	− 0.042***	0.189***
Sciences	− 0.046	0.146***	− 0.022	0.170***
ICT	− 0.060**	0.161***	− 0.060***	0.214***
Technology	0.007	0.152***	− 0.054***	0.174***
Agriculture	− 0.020	0.160***	− 0.058**	0.114**
Health	− 0.040***	0.230***	− 0.004	0.200***
Social services	− 0.067***	0.182***	− 0.033	0.204***
Long-cycle university				
Education	0.000	0.203***	0.008	0.167***
Arts and humanities	0.015	0.127***	− 0.001	0.112***
Social sciences	− 0.017	0.130***	− 0.008	0.125***
Business, administration, law	− 0.074***	0.192***	− 0.069***	0.183***
Sciences	− 0.030*	0.180***	− 0.070***	0.184***
ICT	− 0.099***	0.270***	− 0.098***	0.246***
Technology	− 0.012	0.149***	− 0.057***	0.201***
Agriculture	− 0.048*	0.169***	− 0.093***	0.147***
Health	− 0.075***	0.272***	− 0.078***	0.267***
Social services	− 0.027	0.210***	− 0.051***	0.161***
Student	− 0.011*	− 0.058***	− 0.001	− 0.072***



**Table 1** (continued)

Variable	Women		Men	
	Under	Employ	Under	Employ
Time to complete studies (omitted: within appropriate time)				
Less than three years late	0.013***	− 0.044***	0.009	− 0.028***
Three or more years late	0.047***	− 0.074***	0.010	− 0.052***
Age (omitted: 16–19)				
20–24	0.019	0.158***	0.010	0.084***
25–29	0.008	0.232***	0.009	0.135***
30–34	0.008	0.269***	0.003	0.153***
Live with parents	0.018***	− 0.099***	− 0.004	− 0.092***
Live in couple	− 0.006	− 0.010	− 0.028***	0.048***
Children under 16	0.003	− 0.086***	0.017**	− 0.029***
Non-Spanish	0.050***	− 0.057***	0.071***	− 0.076***
Occupational segregation (omitted: male-dominated)				
Integrated	0.007		− 0.001	
Female-dominated	0.046***		0.028***	
Industry segregation (omitted: male-dominated)				
Integrated	0.023***		0.026***	
Female-dominated	0.045***		0.016**	
Tenure 12 months or less	0.039***		0.041***	
Professional status (omitted: self-employed)				
Public sector employee with permanent contract	− 0.000		0.001	
Public sector employee with fixed-term contract	0.072***		0.067***	
Private sector employee with permanent contract	0.040***		0.029***	
Private sector employee with fixed-term contract	0.124***		0.092***	
Firm size (omitted: up to 10 workers)				
More than 10 workers	− 0.040***		− 0.021***	
Unknown	− 0.042***		− 0.041***	
Unemployment rate <sup>a</sup>	0.002***		0.003***	
NUTS Region (omitted: Northwest)				
Northeast	− 0.003**	0.043***	− 0.002***	0.019***
Madrid	− 0.002**	0.032***	− 0.001	0.010
Centre	0.002**	− 0.028***	− 0.000	0.002
East	− 0.002**	0.026***	0.000	− 0.001
South	0.003**	− 0.056***	0.006***	− 0.059***
Canary Islands	0.003**	− 0.045***	0.007***	− 0.064***
After 2012 labour reform	0.041***		0.026***	

The year when the survey was carried out is controlled for, but not reported

Under: underemployed. Employ: employed

<sup>a</sup> Unemployment rate by gender and year in the Autonomous Region of residence

\*  $p < .10$ . \*\*  $p < .05$ . \*\*\*  $p < .01$

Finally, the employment equation includes variables as educational attainment, taking longer than usual to graduate, age, household composition, enrolled in school, the area of residence (which reflects the general conditions of local demand for work) and the year when the worker was interviewed.

## 5 Results

### 5.1 Marginal effects

In this section, we present the results of the bivariate probit selection models for both subsamples, men and women. The estimations indicate that the rho coefficients are 0.207 ( $p$ -value = 0.01964) and 0.361 ( $p$ -value = 0.0000),

respectively, justifying the use of these models to handle the sample selection bias due to estimating the underemployment probability just for employed people.

The marginal effects of the explanatory variables have been computed from the estimated coefficients and are displayed in Table 1. The ones corresponding to the employment equation (Columns 2 and 4) follow the same direction as in previous studies.

The results obtained for underemployment are shown in Columns 1 and 3. Focusing on the role of gender segregation,<sup>18</sup> we observe that workers in female-dominated occupations are those with the highest probability of underemployment, with a larger impact for women. Specifically, we find that young women working in female-dominated occupations are 4.6 percentage points more likely to be underemployed than those in male-dominated occupations; the corresponding increase in the male subsample is 2.8 points. When segregation is defined in relation to activity sectors, our results indicate that women in female-dominated industries also show the highest likelihood of underemployment (an increase of 4.5 percentage points). Men in female-dominated activities are also more often underemployed than in male-dominated ones. However, men in gender-balanced industries are the most prone to suffer this handicap. Altogether, these results support *Hypothesis 1*, confirming a higher underemployment risk associated with working in female-dominated jobs than in male-dominated ones.

The marginal effects of the remaining explanatory variables are not discussed due to space constraints. We only want to briefly point out some of the results regarding educational attainment, due to the relevance of this variable in almost any labour market outcome. The figures in Table 1 suggest that having a long-cycle university degree in almost any field reduces underemployment. Overall, business, administration and law, and information and communication technologies (ICTs) seem to be the best specializations at most education levels. Additionally, it is worth noting that education means greater differences in the risk of being underemployed for women than for men. Some additional comments are included in the sensitivity section.

## 5.2 Analysis of the underemployment gender gap

According to the results of the bivariate probit models, women's estimated conditional probability of underemployment is 0.163; the corresponding figure for men is 0.124. Therefore, young female workers in Spain are 1.31 times more likely of being underemployed than their

male peers. Thus, the underemployment gender gap ( $-0.039$ ) is not negligible.

As stated above, the main objective of this article is to carry out an in-depth analysis of this gap and to determine the impact of gender segregation on it. To do this, we have carried out a detailed three-fold decomposition of the gap obtained from Eq. (14) and the results are shown in Table 2.

According to the last row in Table 2, the overall endowments effect,  $E_W$ , is very important and seems to explain the underemployment gender gap (130.8%), while the coefficient effects,  $C_W$ , is not significant. Moreover, the interaction term, which is positive (0.0126), reduces the gap ( $-32.7\%$ ).

Regarding the detailed decomposition of the gap, the sum of the individual contributions of all the variables to each component of the gap shows that the errors terms from our counterfactual simulations are small and not statistically significant. These results validate our approach.

The results in columns 1–3 allow us to conclude that the different distribution of men and women across jobs is crucial to explain the gender gap, as stated in *Hypothesis 2a*. Thus, if women were distributed across occupations and industries as men are (maintaining female coefficients), the gap would be reduced to the greatest extent ( $-0.0453$ ). This result suggests that gender segregation leads to a kind of discrimination against women. Indeed, it cannot be argued that women work more frequently in certain (female-dominated) jobs because they prefer shorter work-schedules. Conversely, their larger risk of underemployment, which means they would like to work more hours than available more often than their male peers, is mainly due to the kind of jobs they work in.

Focusing on the coefficients effect (Columns 4–6 in Table 2), the contribution of the intercept term of the underemployment equation (Intercept 1) quantifies the extent to which women are, on average, treated differently to men. This contribution can be interpreted as the average extent of discrimination (Kim 2013). Our results indicate that the contribution of Intercept 1 to the gap is small and not significant, so women and men would face the same average risk of underemployment. However, the coefficient effect of gender segregation indicates an important deviation from the mean discriminatory level. Specifically, the gender gap in underemployment would reduce ( $-0.0146$ ) if women had men's returns in the same gender-typing jobs (maintaining female characteristics). Therefore, the different impact on underemployment that working in certain occupations and industries exerts on the risk of experiencing it contributes to widening the gender gap, supporting our *Hypothesis 2b*. In fact, this is the most important contribution to the coefficient

<sup>18</sup> The marginal effects corresponding to each gender segregation measure have been obtained from the estimated parameters of the interaction variable capturing both occupational and industry segregation.

**Table 2** Three-fold decomposition of underemployment gender gap

Variable	Endowments			Coefficients			Interaction			Total
	Contribution	SE	%Gap	Contribution	SE	%Gap	Contribution	SE	%Gap	%Gap
Educational attainment	0.0073**	0.00314	− 18.9	0.0053**	0.00218	− 13.8	− 0.0061***	0.00364	15.7	− 17.0
Student	0.0003*	0.00018	− 0.9	0.0001	0.00007	− 0.2	− 0.0003	0.00024	0.9	− 0.2
Time to complete studies	− 0.0001***	0.00004	0.3	− 0.0003**	0.00012	0.7	0.0001	0.00006	− 0.2	0.8
Age	− 0.0021***	0.00004	5.4	− 0.0001*	0.00005	0.2	− 0.0052*	0.00029	13.4	19.0
Live with parents	0.0016***	0.00054	− 4.2	0.0019**	0.00078	− 4.8	− 0.0018*	0.00072	4.8	− 4.2
Live in couple	0.0006	0.00063	− 1.7	− 0.0006*	0.00033	1.5	0.0019*	0.00087	− 4.8	− 5.0
Children under 16	− 0.0003	0.00047	0.7	0.0006	0.00049	− 1.6	− 0.0011	0.00072	2.8	1.9
Non— Spanish	− 0.0009***	0.00012	2.4	0.0009	0.00057	− 2.3	− 0.0001	0.00015	0.2	0.3
Gender segregation	− 0.0453***	0.00355	117.0	− 0.0146***	0.0032	37.8	0.0217***	0.00473	− 56.2	98.6
Professional status	− 0.0028***	0.00038	7.3	− 0.0041***	0.00091	10.7	0.0005**	0.00058	− 1.2	16.8
Recent job	− 0.0008***	0.00022	2.1	− 0.0003	0.0008	0.7	0.0000	0.00026	0.1	2.9
Firm size	− 0.0030***	0.00032	7.7	− 0.0016***	0.00056	4.2	0.0011*	0.00042	− 2.8	9.1
Unemployment rate	− 0.0059***	0.00077	15.1	0.0057	0.01078	− 14.7	− 0.0002**	0.0011	0.6	1.0
NUTS Region	0.0000	0.00001	0.0	0.0000	0.00003	0.1	0.0000	0.00002	− 0.1	0.0
After 2012 reform	− 0.0009***	0.0001	2.2	− 0.0013***	0.00047	3.2	0.0005***	0.00012	− 1.2	4.2
Year	− 0.0003**	0.00012	0.8	0.0005**	0.00022	− 1.2	− 0.0002	0.00016	0.5	0.1
Intercept 1				0.0016	0.00894	− 4.1				− 4.1
Intercept 2				− 0.0025**	0.00102	6.4				6.4
$\theta_p^M$				0.0114***	0.00936	− 29.3				− 29.3
Approximation error	0.0017	0.00441	− 4.4	− 0.0033	0.0046	8.5	0.00199	0.0054	− 5.1	− 1.0
Aggregate	− 0.0506***	0.0037	130.8	− 0.0007	0.00454	1.9	0.0126*	0.0051	− 32.7	100.0

Standard errors (SE) have been computed using Krinsky and Robb's method (1,000 draws). %Gap: Change as a percentage of the total underemployment gender gap

effect affecting the underemployment gender gap, reinforcing our previous finding regarding the discriminating effect of segregation against women.

The interaction term captures the effect of changing endowments and coefficients simultaneously. The positive sign of this portion suggests that the risk of underemployment in female-dominated jobs, where women are indeed highly represented, is significantly larger for women than for men. Therefore, this interaction term reflects that, once the returns have been changed from women to men, the additional contribution of distributing women across jobs as men is only  $-0.0236$  (which is 0.0217 smaller than the contribution of segregation to the endowments effects).

Regarding the total impact of each individual variable to the underemployment gender gap, figures in the last column in Table 2 show gender segregation is the most important one explaining it (98.6%). Moreover, gender differences in age also widen the gap, explaining 19% of it. A similar conclusion is obtained for professional status (16.8%). Conversely, educational attainment reduces the underemployment gap ( $-17\%$ ).

### 5.3 Sensitivity analysis

In this subsection, we carry out different sensitivity analyses to check the robustness of our results. Table 3 shows the estimated gender gap in underemployment and the specific contribution of occupational and industry segregation to it, according to Eq. (14).

In Model I, we re-estimate our baseline model using an alternative definition of gender segregation. Specifically, we have re-defined integrated occupations and industries as those where the female percentage is between 0.5 and 1.5 times the average female share of employment in the labour force. The following two models consider gender segregation only in occupations (Model II) or industries (Model III) to analyse their separate effects. Thus, we can check if our results change when no threshold is established to classify occupations and industries as gender-dominated or integrated. Additionally, the baseline model is re-estimated splitting the original sample into two subsamples: workers with tertiary education (Model IV) and those without (Model V). Therefore, taking into account the marginal effects already analysed, we can test our hypotheses for more homogeneous groups of education. Model VI is the same as the benchmark model but includes inactive people together with unemployed ones

**Table 3** Sensitivity analysis: Contribution of occupational and industry segregation to the underemployment gender gap

	Gap	Endowments			Coefficients			Interaction			Total
		Contribution	SE	%Gap	Contribution	SE	%Gap	Contribution	SE	%Gap	
Baseline model	-0.0387	-0.0453	0.00355	117.0	-0.0146	0.0032	37.8	0.0217	0.00473	-56.2	98.6
Model I	-0.0387	-0.0439	0.00422	113.4	-0.0147	0.00368	38.1	0.0215	0.00539	-55.5	96.0
Model II	-0.0387	-0.0386	0.00313	99.9	-0.0118	0.0029	30.5	0.0184	0.00426	-47.5	82.9
Model III	-0.0387	-0.0295	0.00239	76.1	-0.0102	0.00222	26.4	0.0166	0.00313	-43.0	59.5
Model IV	-0.0369	-0.0361	0.00322	98.1	-0.0007	0.00368	1.9	0.0083	0.00530	-22.5	77.5
Model V	-0.0536	-0.0524	0.00608	97.9	-0.0254	0.01318	47.4	0.0351	0.00750	-65.5	79.8
Model VI	-0.0387	-0.0451	0.00355	116.6	-0.0143	0.00419	37.0	0.0215	0.00473	-55.5	98.1
Model VII	-0.0742	-0.0615	0.00237	82.8	-0.0124	0.00359	16.7	0.0301	0.00332	-40.5	59.0
Model VIII	-0.0387	-0.0429	0.00371	111.0	-0.0163	0.00368	42.1	0.0170	0.05017	-44.0	109.1
Model IX	-0.0384	-0.0466	0.00364	121.2	-0.0154	0.00329	40.0	0.0232	0.00482	-60.4	100.8
Model X	-0.0387	-0.0453	0.00355	117.0	-0.0179	0.00417	46.2	0.0217	0.00473	-56.2	107.0

Standard errors (SE) have been computed using Krinsky and Robb's method (1,000 draws). %Gap: Change as a percentage of the total underemployment gender gap. All the values for the contribution of segregation are significant at 99% except the coefficient and interaction components in Model IV, which are not significant

when estimating the employment equation. Model VII includes the same independent variables as the baseline model but only involuntary part-time workers are considered underemployed, as is often the case in the literature.

The last three models include some methodological changes. In Model VIII, we adjust the variable gender segregation last instead of first to check if the order of switching affects the results. As Heckman-type selection models are mostly identified by assumptions about error distributions, in Model IX we use Inverse Probability Weighting (IPW) based on a probit model as an alternative technique to address sample selection problems (see Seaman and White 2013 for a review).<sup>19</sup> Finally, Model X is estimated using the methodology proposed in Gardeazábal and Ugidos (2004) to address the identification problem that arises for categorical variables (instead of Kim's approach).

In general terms, the results in the first column of Table 3 are quite similar to those from the main analysis and few differences can be found. In particular, the largest estimated underemployment gender gap is found for workers without tertiary education (Model V) and when only involuntary part-timers are classified as underemployed (Model VII).<sup>20</sup>

When we carry out the detailed three-fold decomposition, our main conclusions remain unchanged.

<sup>19</sup> To facilitate comparisons with the baseline model, the reweighting approach has been applied only when estimating the coefficients of the probit that explains underemployment, but not when the decomposition of the gap has been computed. Therefore, the results obtained with IPW are conditioned to being employed.

<sup>20</sup> This result is not surprising, since involuntary part-time work in Spain is mainly a female problem. In fact, the average percentage of involuntary part-timers among male workers in the last 15 years is just 3.5%, while this figure is 11.6% for women.

As expected, in Models II and III the three effects are smaller than those obtained in the baseline model. It is noteworthy that the portion of the gap explained by segregation is larger when we consider only occupational segregation. Some differences are also found if only involuntary part-time workers are considered underemployed (Model VII). This decision implies estimating our models using an alternative endogenous variable, so the results obtained are likely to change significantly. However, the different distribution of men and women across occupations and industries is what still drives the gap.

The most interesting results are those obtained for workers with different levels of education (Models IV and V). It is worth noting that the estimated underemployment gap is larger for workers without tertiary education, as might be expected. Gender segregation, however, is the factor that mainly explains the gap regardless of workers' educational level, explaining almost the same percentage of it in both subsamples. Moreover, the different distribution of men and women across jobs widens the gap to a larger extent for less educated workers than for more educated ones. Additionally, only in the sample of workers without tertiary education, the different returns associated with women and men widen the gap. Thus, we can conclude that tertiary education reduces the gap not only because women and men are more similarly distributed across jobs (in comparison with less educated workers), but also because working in the same gender-typing jobs leads to a similar risk of underemployment for male and female.

Regarding the last three models, there are some small differences with the baseline in the three components, especially in the coefficients effect (that seems to increase). These differences translate into a larger proportion of the gap explained by segregation in the three

models. However, our main conclusions still hold, so our procedure is robust to the methodological changes they include.

Overall, the estimated underemployment gender gap is mainly explained by the different distribution of male and female workers across occupations and industries in every specification of the model, which clearly supports *Hypothesis 2a*. Moreover, *Hypothesis 2b* is also confirmed because, in most cases, the different impact that working in certain gender-typing jobs exerts on the risk of male and female underemployment contributes to widening the gender gap. Hence, the results in this subsection give robustness to our findings.

## 6 Conclusions and discussion

This article provides evidence of the crucial impact of occupational and industry segregation on the time-related underemployment gender gap for people aged 16 to 34 using the annual samples of the Spanish LFS from 2006–2016. It is worth noting that despite the Spanish labour market being deeply gender segregated, the effect of this feature on the gap has not been addressed before in the literature. Furthermore, to the best of our knowledge, we are the first to perform a decomposition using bivariate probit models with sample selection. To do this, we have developed a methodology based on two counterfactual simulations that provides a three-fold detailed decomposition of the underemployment gender gap into endowments and coefficients effects as well as the interaction of these effects.

Our methodology, inspired by the Fairlie technique (1999, 2005, and 2017), has several advantages. Thus, it allows identifying the coefficients effect corresponding to a specific variable (while Fairlie technique does not). In fact, this effect can be easily interpreted using Kim's (2013) method to solve the identification problem. Moreover, we keep almost real individuals for two reasons. First, our approach does not require a one-to-one matching of individuals, since we replicate the distribution of each specific variable, and the number of women who have a specific characteristic adjusted is just those strictly necessary. Second, the proposed procedure always uses the same starting point, and we estimate each contribution by switching the variable of interest first (as suggested by Fairlie 2005). Thus, only one characteristic is modified at a time. Although we are aware that the order of switching the variables is potentially important, this decision provides a potential solution to the path dependence problem already pointed out by Fairlie.

The methodology proposed has some limitations. It does not show the summing up property of the Fairlie method, however, the estimated approximation errors are small and not significant. Moreover, using the

Krinsky-Robb approach could lead to underestimating the standard errors. Despite our results indicating this is not the case, nonparametric bootstrap could be considered as an alternative.

The procedure has been tested in several ways and the main conclusions still hold. Firstly, the results obtained applying our procedure to a univariate probit model are very similar to those obtained through other decomposition techniques like Fairlie, Yun (2004) and Oaxaca-Blinder (Oaxaca 1973 and Blinder 1973). Secondly, the main results of our baseline model are robust to methodology changes, definition of some variables and to the sample used.

Our results demonstrate that working in female-dominated occupations and industries implies a higher probability of time-related underemployment than in male-dominated ones, confirming our first hypothesis. Moreover, we find that the disadvantage in terms of underemployment that implies working in a female-dominated occupation or industry is greater for women than for men.

Furthermore, according to our results, the estimated underemployment gender gap for young workers in Spain is 3.9 percentage points. As underemployment has a negative impact on income, welfare dependency and life satisfaction (Wilkins 2007), the higher underemployment risk faced by women implies negative consequences related to experience, earnings, and possibly promotions (Weststar, 2011). Therefore, designing effective policies leading to more gender equality will only be possible if the factors behind the gender gap in underemployment are clearly identified. To the best of our knowledge, the reasons for this difference have been little investigated. As an exception, Barrett and Doiron (2001) affirm that the main reason that explains women being involuntary part-timers more often than men is the fact they are employed in different industries and occupations. Nevertheless, Vuluku et al. (2013), who are the only ones who have decomposed the underemployment gender gap, did not include occupational or industry segregation as an explanatory factor, while we do. This fact could be one of the reasons why they find that only 5.4% of the gap is explained by female-male differences in characteristics, while 94.6% is unexplained, while our results are the opposite.

The results obtained from our simulations lead us to conclude that the fact that men and women work in different industries and occupations is what widens the underemployment gap to the greatest extent. This effect is even increased due to women facing a different risk of underemployment than men when they work in the same gender-typing jobs. So, *Hypotheses 2a* and *2b* are supported and we can state that the gender gap in

underemployment would be largely reduced if men and women were more evenly distributed across occupations and industries. Thus, segregation (especially occupational) is not only a source of gender differences in terms of wages and job quality (Stier and Yaish 2014), but also a key factor explaining the underemployment gender gap. Moreover, as the World Economic Forum (2017) states, a crucial factor for further progress in reducing the overall global gender gap is the closing of occupational gender gaps. Therefore, policy measures should be designed and implemented to fight against segregation in the labour market in order to achieve gender equality in Spain.

In this respect, education is an important factor to be highlighted. Thus, we can consider two different ways in which education can influence underemployment. Firstly, education has a direct impact on underemployment since the marginal effects show that some educational attainments contribute to reduce the risk of underemployment. Moreover, the results from our simulations using the baseline model, as well as those obtained after splitting the sample into workers with tertiary education and without it, allow us to affirm that education reduces the underemployment gender gap. Specifically, our results suggest a higher educational attainment helps women to escape from this disadvantage in the labour market.

Second, education may affect segregation in the marketplace due to the link between educational presorting and occupational and industry segregation (Borghans and Groot 1999; Shauman 2006; Smyth and Steinmetz 2008). The statistical evidence on the strength of the link between segregation in education and in employment is mixed<sup>21</sup> (Bettio et al. 2009). Nevertheless, appropriate remedies that address the barriers women experience to enter male dominated jobs should include changes in the education and training of women and girls, such as introducing gender aware career counselling/guidance (National Foundation for Australian Women 2017). Particularly, young women should be encouraged to enrol in previously male-dominated education programmes in order to gain access to a wider range of jobs. Additionally, encouraging young men to join female-dominated specialties could translate into a more mixed gender educational profiles. In fact, as Bettio et al. (2009) point out, since women are outperforming men in levels of education attained—up to the first stage of tertiary education—choice of field is the primary channel through which education can influence de-segregation in the labour market in the future.

Especially desirable would be more women specialising in ICTs since our results show this field seems to be

linked to a lower risk of experiencing underemployment. This result is in line with the literature attributing ICTs the potential capacity to reduce gender inequalities, since they improve the occupational and professional position of women (Castaño et al. 1999), particularly in the Spanish labour market (Iglesias-Fernández et al. 2010a, 2010b). Specifically, ICTs reduce both the need for manual labour and physical effort in favour of knowledge, teamwork and communication skills (WWW-ITC 2004). This fact, in turn, promotes changes in the sectoral distribution and educational requirement of jobs and the demand for occupations (Iglesias-Fernández et al. 2012), leading to more opportunities for women. However, the ratio of female/male graduates in ICTs is only 0.14 in Spain (World Economic Forum 2017).

Although educational segregation by gender plays a significant role in shaping gender segregation within the labour market, as Smyth and Steinmetz (2008) point out, women and men who choose similar fields do not have exactly the same occupational outcomes. Thus, educational policies should be complemented with other instruments. For instance, reinforcing policies that encourage employers to hire female workers in male intensive occupations and industries could also help reduce both segregation and the gender gap in underemployment. Additionally, the economic conditions of female occupations should be improved to raise both men's and women's interest in female-dominated occupations. In order to achieve an egalitarian distribution of men and women across jobs, eradicating the disincentives to work in female-dominated occupations is necessary (Torre 2018). Finally, any other measures devoted to fighting against gender stereotypes and discrimination would be welcome to reduce inequality in the Spanish labour market.

#### Abbreviation

LFS: Labour Force Survey.

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12651-021-00305-0>.

**Additional file 1: Appendix 1.** Approximation error in the endowments effect. **Appendix 2.** Approximation error in the coefficients effect.

**Additional file 2: Table S1.** Comparison of methods for the three-fold decomposition of the underemployment gender gap. **Table S2.** Frequencies (%) of the independent variables by gender: not underemployed (NU), underemployed (U), or unemployed (Unem)

#### Acknowledgements

We thank three anonymous referees for their insightful comments and suggestions, which have helped to improve this article considerably.

#### Authors' contributions

All authors wrote, read and approved the final manuscript.

<sup>21</sup> Analysing the causality relation between educational and occupational segregation is beyond the scope of this study.

**Funding**

Authors declare that they have received no specific funding to conduct this research.

**Availability of data and materials**

The data that support the findings of this study are available from Instituto Nacional de Estadística (INE). Purchase terms do not allow authors sharing these data.

**Code availability**

The results have been obtained using STATA.

**Declarations****Ethics approval and consent to participate**

The authors of this study declare this manuscript does not infringe the copyright, moral rights, or other intellectual property rights of any other person. The authors of this study testify that this paper is original and has not incurred in any type of plagiarism. The authors of this study declare that the article is original, has not already been published in a journal, and is not currently under consideration by another journal. The authors of this study agree to the terms of the SpringerOpen Copyright and License Agreement.

**Consent for publication**

Not applicable.

**Competing interests**

Authors declare that no competing interests exist.

**Author details**

<sup>1</sup>Departamento de Economía, Contabilidad y Finanzas, Universidad de La Laguna (ULL), La Laguna (Tenerife), Spain. <sup>2</sup>Facultad de Economía, Empresa y Turismo, Universidad de La Laguna (ULL), Campus de Guajara, Apartado 456, Postal Code 38200, La Laguna (Tenerife), Spain.

Received: 27 July 2020 Accepted: 11 November 2021

Published online: 28 November 2021

**References**

- Acosta-Ballesteros, J., Osorno-del Rosal, M.P., Rodríguez-Rodríguez, O.M.: Underemployment and employment among young workers and the business cycle in Spain: the importance of education level and specialisation. *J. Educ. Work* **31**(1), 28–46 (2018)
- Anker, R.: Gender and jobs: sex segregation of occupations in the world. International Labour Office, Geneva (1998)
- Barrett, G.F., Doiron, D.J.: Working part time: by choice or by constraint. *Can. J. Econ.* **34**(4), 1042–1065 (2001)
- Bazen, S., Joutard, X., Magdalou, B.: An Oaxaca decomposition for nonlinear models. *J. Econ. Soc. Meas.* **42**(2), 101–121 (2017)
- Bell, D., Blanchflower, D.: Underemployment in the UK Revisited. *Natl. Inst. Econ. Rev.* **224**(1), F8–F22 (2013)
- Bettio, F., Verashchagina, A., Mairhuber, I., Kanjuo-Mrčela, A.: Gender segregation in the labour market: Root causes, implications and policy responses in the EU. Publications Office of the European Union, Luxembourg (2009)
- Blinder, A.S.: Wage discrimination: reduced form and structural estimates. *J. Hum. Resour.* **8**(4), 436–455 (1973)
- Boll, C., Leppin, J., Rossen, A., Wolf, A.: Magnitude and impact factors of the gender pay gap in EU countries. Report prepared for and financed by the European Commission–Directorate-General for Justice, European Union (Hrsg.), Hamburg (2016) <https://op.europa.eu/en/publication-detail/-/publication/aa27ac5b-fc83-11e5-b713-01aa75ed71a1/language-en>
- Borghans, L., Groot, L.: Educational presorting and occupational segregation. *Labour Econ.* **6**, 375–395 (1999)
- Castaño, C., Iglesias, C., Mañas, E., Sánchez-Herrero, M.: Diferencia o discriminación: La situación de las mujeres en el mercado de trabajo y el impacto de las tecnologías de la información. Consejo Económico y Social, Madrid (1999)
- Daymont, T.N., Andrisani, P.J.: Job preferences, college major, and the gender gap in earnings. *J. Hum. Resour.* **19**, 408–428 (1984)
- Doiron, D.J., Riddell, W.C.: The impact of unionization on male-female earnings differences in Canada. *J. Hum. Resour.* **29**(2), 504–534 (1994)
- Dowd, B.E., Greene, W.H., Norton, E.C.: Computation of standard errors. *Health Serv. Res.* **49**(2), 731–750 (2014)
- Dueñas-Fernández, D., Iglesias-Fernández, C., Llorente-Heras, R.: La involuntariedad en el empleo a tiempo parcial y la Gran Recesión: un análisis de género en España. *EKONOMIAZ. Revista Vasca De Economía* **90**(02), 284–317 (2016)
- Etezady, A., Shaw, F.A., Mokhtarian, P.L., Circella, G.: What drives the gap? Applying the Blinder-Oaxaca decomposition method to examine generational differences in transportation-related attitudes. *Transportation* **48**, 857–883 (2021)
- Even, W.E., Macpherson, D.A.: Plant size and the decline of unionism. *Econ. Lett.* **32**(4), 393–398 (1990)
- Fairlie, R.W.: The absence of the African-American owned business: an analysis of the dynamics of self-employment. *J. Labor Econ.* **17**(1), 80–108 (1999)
- Fairlie, R.W.: An extension of the Blinder-Oaxaca decomposition technique to logit and probit models. *J. Econ. Soc. Meas.* **30**(4), 305–316 (2005)
- Fairlie, R.W.: Addressing path dependence and incorporating sample weights in the nonlinear Blinder-Oaxaca decomposition technique for logit, probit and other nonlinear models. Stanford Institute for Economic Policy Research, Working Paper No. 17–013, University of California (2017)
- Fortin, N., Lemieux, T., Firpo, S.: Decomposition methods in economics. In: Ashenfelter, O., Card, D. (eds.) *Handbook of labor economics*, Vol 4a, *Handbooks in Economics*, pp. 1–102. North Holland, Great Britain (2011)
- National Foundation for Australian Women Gender segregation in the workplace and its impact on women's equality. Report submitted to Senate Finance and Public Administration References Committee. [https://www.aph.gov.au/Parliamentary\\_Business/Committees/Senate/Finance\\_and\\_Public\\_Administration/Gendersegregation/Report](https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Finance_and_Public_Administration/Gendersegregation/Report) (2017)
- Gardeazábal, J., Ugidos, A.: More on identification in detailed wage decompositions. *Rev. Econ. Stat.* **86**(4), 1034–1036 (2004)
- Greene, W.H.: *Econometric analysis*, 7th edn. Pearson, Upper Saddle River (2012)
- Hakim, C.: A sociological perspective on part-time work. In: Blossfeld, H.P., Hakim, C. (eds.) *Between equalization and marginalization: women working part-time in Europe and the United States of America*, pp. 22–70. Oxford University Press, Oxford (1997)
- Hakim, C.: Social Change and Innovation in the Labour Market: Evidence from the Census SARs on Occupational Segregation and Labour Mobility, Part-time Work and Student Jobs. Homework and Self-employment. Oxford University Press, Oxford (1998)
- Hegewisch, A., Liepmann, H., Hayes, J., Hartmann, H.: Separate and not equal? Gender segregation in the labor market and the gender wage gap. Institute for Women's Policy Research, Briefing Paper No. 377, Washington (2010)
- Iglesias-Fernández, C., Llorente-Heras, R., Dueñas-Fernández, D.: Diferencias de Género en el Empleo TIC. *Cuad. Econ.* **33**(92), 105–138 (2010b)
- Iglesias-Fernández, C., Llorente-Heras, R., Dueñas-Fernández, D.: Profundizando en la segregación laboral. Sectores, ocupación y TIC en España. Instituto Universitario de Análisis Económico y Social. Serie de Documentos de Trabajo 07/2012, Universidad de Alcalá (2012)
- Iglesias-Fernández, C., Llorente-Heras, R., Dueñas-Fernández, D.: Mujer y Empleo TIC en España. In C. Castaño (Dir.), *Género y TIC. Presencia, Posición y Políticas*, pp. 83–106. UOC Ediciones, Barcelona (2010a)
- Jann, B.: The Blinder-Oaxaca decomposition for linear regression models. *Stata J.* **8**(4), 453–479 (2008)
- Kameräde, D., Richardson, H.: Gender segregation, underemployment and subjective well-being in the UK labour market. *Hum. Relat.* **71**(2), 285–309 (2018)
- Kim, C.: Detailed wage decompositions: revisiting the identification problem. *Sociol. Methodol.* **43**, 346–363 (2013)
- Kjeldstad, R., Nymoen, E.H.: Part-time work and gender: worker versus job explanations. *Int. Labour Rev.* **151**(1–2), 85–107 (2012a)
- Kjeldstad, R., Nymoen, E.H.: Underemployment in a gender-segregated labour market. *Econ. Ind. Democracy* **33**(2), 207–224 (2012b)

- Krinsky, I., Robb, A.L.: On approximating the statistical properties of elasticities. *Rev. Econ. Stat.* **68**(4), 715–719 (1986)
- Lupton, B.: Explaining men's entry into female-concentrated occupations: issues of masculinity and social class. *Gender Work Organ.* **13**(2), 103–128 (2006)
- Martin, P., Barnard, A.: The experience of women in male-dominated occupations: A constructivist grounded theory inquiry. *Journal of Industrial Psychology* **39**(2), 1–12 (2013)
- McKee-Ryan, F.M., Harvey, J.: I have a job, but...: a review of underemployment. *J. Manage.* **37**(4), 962–996 (2011)
- Oaxaca, R.: Male-female wage differentials in urban labor markets. *Int. Econ. Rev.* **14**(3), 693–709 (1973)
- Oaxaca, R., Ransom, M.: Identification in detailed wage decompositions. *Rev. Econ Stat* **81**(1), 154–157 (1999)
- Powers, D.A., Yoshioka, H., Yun, M.S.: mvdcmp: multivariate decomposition for nonlinear response models. *Stata J.* **11**(4), 556–576 (2011)
- Prause, J., Dooley, D.: Youth Underemployment. In: Feldman, D., Maynard, D. (eds.) *Underemployment: psychological, economic, and social challenge*, pp. 59–80. Springer, New York (2011)
- Seaman, S., White, I.R.: Review of inverse probability weighting for dealing with missing data. *Stat. Methods Med. Res.* **22**(3), 278–295 (2013)
- Shauman, K.: Occupational sex segregation and earning of occupations: what causes the link among college-educated workers? *Soc. Sci. Res.* **35**, 577–619 (2006)
- Simic, M.: Underemployment and overemployment in the UK. *Labour Market Trends* **110**(8), 399–414 (2002)
- Simpson, R.: Have times changed? Career barriers and the token woman manager. *Brit. J. Manage.* **8**(s1), 121–130 (1997)
- Simpson, R.: Gender mix and organizational fit: how gender imbalance at different levels of the organization impacts on women managers. *Women Manag. Rev.* **15**(1), 5–20 (2000)
- Simpson, R.: Masculinity at work: the experiences of men in female dominated occupations. *Work Employ. Soc.* **18**(2), 349–368 (2004)
- Smyth, E., Steinmetz, S.: Field of study and gender segregation in European labour markets. *Int. J. Comp. Sociol.* **49**(4–5), 257–281 (2008)
- Stier, H., Yaish, M.: Occupational segregation and gender inequality in job quality: a multi-level approach. *Work Employ. Soc.* **28**(2), 225–246 (2014)
- Torre, M.: Stopgappers? The occupational trajectories of men in female-dominated occupations. *Work Occup.* **45**(3), 283–312 (2018)
- Valletta, R.G., Bengali, L., Van der List, C.: Cyclical and Market Determinants of Involuntary Part-time Employment. IZA Discussion Paper No.9738, Bonn (2016)
- Vuluku, G., Wambugu, A., Moyi, E.: Unemployment and underemployment in Kenya: a gender gap analysis. *Economics* **2**(2), 7–16 (2013)
- Weststar, J.: A review of women's experiences of three dimensions of underemployment. In: Feldman, D., Maynard, D. (eds.) *Underemployment: psychological, economic, and social challenge*, pp. 105–125. Springer, New York (2011)
- Wilkins, R.: Personal and job characteristics associated with underemployment. *Aust. J. Labour Econ.* **9**(4), 371–393 (2006)
- Wilkins, R.: The consequences of underemployment for the underemployed. *J. Ind. Relat.* **49**(2), 247–275 (2007)
- World Economic Forum: The Global Gender Gap Report. [http://www3.weforum.org/docs/WEF\\_GGGR\\_2017.pdf](http://www3.weforum.org/docs/WEF_GGGR_2017.pdf) (2017)
- WWW-ITC: Conceptual framework and state of the art. Widening women's work in information and communication technology. <http://www.ftu-namur.org/fichiers/D12-print.pdf> (2004)
- Yun, M.S.: Decomposing differences in the first moment. *Econ. Lett.* **82**(2), 275–280 (2004)
- Yun, M.S.: A simple solution to the identification problem in detailed wage decompositions. *Econ. Inq.* **43**, 766–772 (2005)
- Yun, M.S.: Identification problem and detailed Oaxaca decomposition: a general solution and inference. *J. Econ. Soc. Meas.* **33**(1), 27–38 (2008)

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---