ECONOMICS
& Sociology

Colom Andrés, M. C., & Molés Machí, M. C. (2023). Gender gap decomposition in employment rate of young people. *Economics and Sociology*, *16*(1), 71-84. doi:10.14254/2071-789X.2023/16-1/5

GENDER GAP DECOMPOSITION IN EMPLOYMENT RATE OF YOUNG PEOPLE

M. Consuelo Colom Andrés
University of Valencia,
Valencia, Spain
E-mail: consuelo.colom@uv.es
ORCID 0000-0002-1923-7273

M. Cruz Molés Machí
University of Valencia,
Valencia, Spain
E-mail: cruz.moles@uv.es
ORCID 0000-0003-3828-2237

Received: May, 2022
1st Revision: December, 2022
Accepted: January, 2023

DOI: 10.14254/2071-789X.2023/16-1/5

ABSTRACT. Differences in behavior between women and men have long been observed in the labor market. Occupational segregation, lack of equal opportunities and lower wages are still linked to the lower opportunity cost that leaving the labor market implies for women. In this paper we analyze, using decomposition techniques, the gap in the employment rate between young women and men. These techniques allow us to separate the weight of observed characteristics from the weight of preferences or unobserved factors. Our results reveal that both types of factors are relevant in the gender gap in the employment rate. Preferences or unobserved factors, such as the different perception that both genders have of family and involvement of women in housework, which leads to greater female labor abandonment, are likely to be behind the gender gap in the case of already having job.

JEL Classification*:* J16, J21, C25

Keywords: gender gap, employment rate, multilevel logit model, decomposition techniques

Introduction

In most developed countries, the gender gap in many segments of the labor market has been narrowing in recent decades, but there are still differences in important aspects, such as wages or the employment rate, where women continue to be at a disadvantage.

Women's preferences for jobs that allow them to balance working hours and family responsibilities may help to explain their greater representation in occupations with worse working conditions, lower wages, less responsibility and difficult access to managerial positions.

In the labor employment rate (defined as the ratio of young employed people to the total number of young people) there has been a rapprochement between women and men, although the rate of women is below that of men in a large part of welfare states. Women tend to leave the labor market or to reduce their working day with the arrival of children, and only in a few cases, it is men who give up their working life.

Incorporation into the labor market depends on the human capital and on the family and personal characteristics associated with the individual. Educational level, age or gender are determining factors in the insertion of young people into the labor market (Grigoli et al., 2018). The current economic situation also has a significant impact on the labor market. In times of economic recession, the labor market becomes precarious and young people are negatively affected by falling employment (Alawad et al, 2020; Bayrak and Tatli, 2018; Bell and Blanchflower, 2011).

The existence of an unobservable cultural factor is possibly responsible for the lower participation of women in the labor market and for gender differences (Antecol, 2000; Banerjee, 2019; Fitzenberger et al., 2004). Occupational segregation, stereotypes, lack of real equality of opportunities and lower wages continue to be linked to the lower opportunity cost that leaving the labor market implies for women.

In Spain, as in other neighboring countries, the situation in the labor market is heterogeneous between regions. These regional disparities may be caused by differences in the economic situation and structure (different levels of industrialization that condition the labor supply, both in terms of quantity and type of occupation), differences in the real estate market (e.g., the housing prices), as well as social and cultural differences (regions with a larger rural population generally have more conservative features).

The aim of this paper is to study the labor market insertion of young people in Spain after the economic crisis caused by the bursting of the real estate bubble. We use data from 2018, a moment in time in which an economic recovery is already visible and access to a job is easier. This analysis will upgrade the knowledge of the labor situation of young people in a period of economic expansion, which will help to carry out policies better aimed at improving the employment of young people.

In order to carry out the study, firstly, we estimate separately for women and men a multilevel logit model that allows us to collect the existing regional heterogeneity in the Spanish labor market. With results of the estimation of the model we will learn what is the influence of both the sociodemographic and economic characteristics of the young person and the characteristics of the region on labor insertion. Estimating the same model for women and men will allow assessing whether the determinants of the decision to work, and their weight, are the same for both groups of young people.

Secondly, we analyze the gender gap in the employment rate of Spanish young people. In order to do it, we use decomposition techniques that allow us to separate the effect of the observed factors from the effect of the unobserved factors or preferences on the overall gap.

Decomposition techniques were pioneered by Blinder (1973) and Oaxaca (1973), who analyzed the wage gap between women and men for an average individual. Fairlie (2005) performed an extension of this technique for a dichotomous outcome variable. Decomposition methods examining differences in the entire distribution of the outcome variable (not only in average) have been developed, as DiNardo et al. (1996). Machado and Mata (2005) proposed a technique using quantile regressions, which essentially performs the Blinder-Oaxaca decomposition across the entire distribution of the outcome variable, and applied it to changes in wage distributions.

In this paper, firstly, we perform the decomposition of the gender gap in the employment rate for an average individual, following the extension proposed by Fairlie (2005). Next, since it is possible that the gap, and the causes that generate it, vary throughout the entire distribution, we examine the differences in the different percentiles of the distribution of the outcome variable. For that, we adapted the technique proposed by Machado and Mata (2005) to a dichotomous variable.

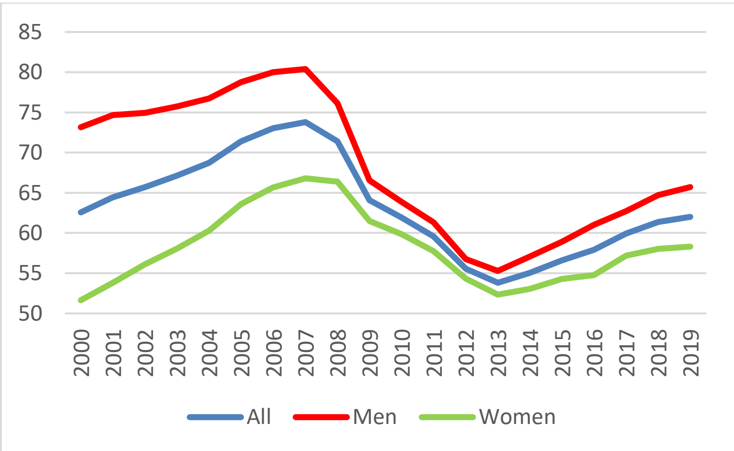
Knowing which factors (observed or not observed) have greater weight in the gender gap in the employment rate of Spanish young people provides a better understanding of their labor market situation and their process of transition to adulthood. This knowledge can help to implement labor policies more efficiently which mitigate labor discrimination against women, mainly in the period of forming a family.

The following section contains the evolution of the employment rate in Spain throughout the 21th century. Section 2 shows the model and the decomposition technique used in analysis. Section 3 presents the description of data and variables. The corresponding analysis of the results of the estimation of the model and of the application of the decomposition technique can be found in section 4. And the main conclusions are in the last section.

1. Labor employment rate of young people in Spain

In this section, we analyze the evolution of the employment rate of young people in the period 2000-2019. *Graph 1*, based on the data provided by the Active Population Survey (EPA) carried out by the National Institute of Statistics (INE), shows the employment rate of young people between 20 and 34[1] years old, disaggregated by gender. These years cover an economic growth period (until 2007), the crisis caused by the bursting of the real estate bubble (between 2008 and 2013) and the subsequent economic recovery (from 2014).

The highest value of the employment rate (73.8%) was reached in 2007. The bursting of the real estate bubble had a very negative impact on the labor market of young Spanish people and represented a pronounced and continuous loss of employment of this group. In consequence, the employment rate experimented a drastic drop, reaching 53.8% in 2013. Since 2014, a recovery in youth employment can be seen, whose rate stood at 61% in 2018 and at 62% in 2019, although the levels of 2007 were not reached.



Graph 1. Employment rate. Young people between 20 and 34 years old
Source: *EPA*

However, as of 2020, due to the crisis caused by Covid-19, a change in trend will predictably be seen again (Berry and McDaniel, 2020). Also, the events that are currently caused by the Ukraine war will produce changes in the world economic situation that will affect the labor market.

[1] The EPA presents results for five-year age groups, starting with 16-year-olds. The five-year group of young people between 16 and 19 years old has not been considered to obtain the employment rate in *Graph 1*, since a large part of these young people are still in training and are out of the labor market. However, the rate has been recalculated considering this five-year group and it has been seen that the profile of the graph does not change, only the magnitude is modified.

Focusing on the employment rate by gender, we see that in the expansionary stage before the crisis there is a big difference between women and men. In the year 2000 there is a difference of more than 21 points. In 2007, the gap has been reduced practically by half, since, for women, a faster growth rate is observed, possibly due to the increase in the participation of women in the labor market, the increase in the opportunity cost of not participating or changes in their preferences.

In years of economic crisis, the difference by gender in employment continues to decrease and, in 2012, it was barely 2.4 percentage points. Generally, the level of employment of women tends to remain more stable than that of men in periods of recession, since men tend to have occupations that are more dependent on the economic cycle, such as jobs related to construction. However, in 2014, with the economic recovery, there was an increase in the employment rate for both women and men, but at different paces, which once again led to an increase in the gender gap, reaching a difference of 6.7 points in 2018 and 7.4 points in 2019.

2. Empirical strategy

2.1. *Multilevel logit model*

In order to analyze the labor insertion of young Spanish people, we consider a dichotomous variable that takes the value 1 if the young person has a job and the value zero otherwise. We use a multilevel binomial logit model, which allows us to consider the existence of variability in the labor market between regions.

The multilevel binomial logit model consists of the following two equations

$$log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \gamma_{0j} + \boldsymbol{\gamma}_1 \boldsymbol{z}_{ij} \quad (1)$$

$$\gamma_{0j} = \lambda_{00} + \boldsymbol{\lambda}_{01}\boldsymbol{v}_j + \mu_{0j} \quad (2)$$

where the subscript *i* represents the individual and *j* the region in which the individual resides; p_{ij} is the response probability, $\boldsymbol{z}_{ij}$ the vector of characteristics of the individual; $\boldsymbol{\gamma}_1$ the vector of fixed parameters of the first level and γ_{0j} the response of the second level, being $\boldsymbol{v}_j$ the specific characteristics of the region, λ_{00} the constant term, $\boldsymbol{\lambda}_{01}$ the fixed parameter vector and μ_{0j} the error term associated to the regions (random effect), for which a Normal distribution is admitted with mean 0 and variance τ_{00}.

Substituting (2) in (1) we obtain the following expression:

$$log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \lambda_{00} + \boldsymbol{\lambda}_{01}\boldsymbol{v}_j + \boldsymbol{\gamma}_1 \boldsymbol{z}_{ij} + \mu_{0j} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} \quad (3)$$

where $\boldsymbol{x_{ij}} = (\boldsymbol{v}_j, \boldsymbol{z}_{ij})$ is the vector of observed variables and $\boldsymbol{\beta} = (\lambda_{00}, \boldsymbol{\lambda}_{01}, \boldsymbol{\gamma}_1, \mu_{0j})$ is the vector of model parameters.

To see the adequacy of the data to a multilevel model, the intra-class correlation coefficient is defined, *ICC*, which gives the percentage of total variance that is explained by differences between groups.

The total variance can be decomposed as the sum of the within-group variance (or first-level variance) and the between-group variance (or second-level variance); thus, the *ICC* in the multilevel logit model is given as:

$$ICC = \frac{variance\ between\ groups}{variance\ between\ grups + variance\ \text{within}\ grups} = \frac{\tau_{00}}{\tau_{00} + \frac{\pi^2}{3}} \quad (4)$$

A high value for the *ICC* coefficient will confirm the existence of high variability between the groups. Thus, a non-negligible value of the *ICC* coefficient of a null model, without explanatory variables at either of the two levels, will indicate that the differences in employment rate between the regions are significant, confirming the suitability of using a multilevel model.

2.2. *Decomposition methodology*

In a multilevel binomial logit model, the decomposition of the overall gap between two groups *A* and *B* in the mean value, following the extension proposed by Fairlie (2005) is given by:

$$\bar{P}_A(\boldsymbol{x}_A'\widehat{\boldsymbol{\beta}}_A) - \bar{P}_B(\boldsymbol{x}_B'\widehat{\boldsymbol{\beta}}_B) = [\bar{P}_A(\boldsymbol{x}_A'\widehat{\boldsymbol{\beta}}_A) - \bar{P}_{AB}(\boldsymbol{x}_B'\widehat{\boldsymbol{\beta}}_A)] + [\bar{P}_{AB}(\boldsymbol{x}_B'\widehat{\boldsymbol{\beta}}_A) - \bar{P}_B(\boldsymbol{x}_B'\widehat{\boldsymbol{\beta}}_B)] = \Delta_X + \Delta_\varepsilon \quad (5)$$

where $\bar{P}_A$ and $\bar{P}_B$ are the average probabilities of the multilevel binomial logit model calculated in groups *A* (women) and *B* (men), $\widehat{\boldsymbol{\beta}}_A$ and $\widehat{\boldsymbol{\beta}}_B$ the coefficients of the model estimated with the sample of group *A* and *B*, respectively, and $\boldsymbol{x}_{iA}$ and $\boldsymbol{x}_{iB}$ are the vectors of observable characteristics of individual *i* belonging to group *A* and *B*. The counterfactual probability $\bar{P}_{AB}$ represents the average probability of women calculated with the characteristics of men.

According to (5), the difference in the response variable (overall gap) in the mean value between groups *A* and *B* is decomposed into an *explained part or endowment effect*, Δ_X, attributed to the existence of differences between the observable characteristics of the two groups, and in an *unexplained part or return effect*, Δ_ε , which is associated with differences between unobservable characteristics and preferences. The unexplained part symbolizes the pure effect of discrimination in terms of the influence of unobserved predictors, such as the preferences of individuals in both groups.

In order to establish comparisons in the percentiles of the distribution of a discrete variable, we have adapted the procedure proposed by Machado and Mata (2005). The decomposition of the difference of the outcome variable for each percentile will be brought out as

$$\xi_A(\alpha) - \xi_B(\alpha) = [\xi_A(\alpha) - \xi_{AB}(\alpha)] + [\xi_{AB}(\alpha) - \xi_B(\alpha)] \quad (6)$$

being $\xi_A(\alpha)$ and $\xi_B(\alpha)$ the α percentile of groups *A* and *B*, respectively, and $\xi_{AB}(\alpha)$ the α percentile of the counterfactual distribution of group *A* (α percentile considering group *A* with the characteristics of group *B*).

Thus, for each percentile, the first term on the right side of (6) includes the *explained part or endowment effect*, and the second term is the corresponding *unexplained part or return effect*.

3. Data and variables

Empirical information is obtained from the Family Budget Survey (EPF) for the year 2018 conducted by INE. This survey offers annual information about income and expenditure of consumption and on the living conditions of households, as well as their members.

Individuals between the ages of 18 and 35 who do not co-reside with their parents have been selected for the study. Young people who live in the parental home have not been considered as a large part of them are still in their formative stage and are not part of the active population. After removing invalid observations, we have a final sample of 3292 young people living independently (1875 women and 1417 men).

The explanatory variables of the first level include sociodemographic and economic characteristics of the young adult: age, educational level, whether the young adult does not have Spanish nationality, whether the young adult lives with a partner, whether they have children under 15 years and the income of other house members.

Age and level of education, both introduced in the model with three dummy variables, are presumably key characteristics in labor insertion and it is expected that the employment possibilities will increase as age and educational level. The nationality and the family characteristics (*Partner* and *Children*) seem to be good predictors of employment decisions of the young adult and may affect women and men differently in their incorporation into the labor market. And the income of other members of the household, predictably, will discourage the young person from entering the labor market.

The variables of the second level collect the social and economic situation of the region of residence of the young person. The housing price indicates the situation of the real estate market. The social structure of the regions is collected with the proportion of rural population (resident population in municipalities with less than 5000 inhabitants) and the proportion of immigrant population in the region.

Table 1 presents the definition of all these variables. *Table 2* shows a statistical summary of the socio-demographic and economic characteristics (first level variables) according to gender. The descriptive statistics of the regional variables are in *Table 3*.

Table 1. Definition of variables

	Definition
Dependent variable	
Employment	1 = With employment
Individual Variables	
Age18	1 = Between 18 and 23 years
Age24	1 = Between 24 and 29 years
Age30	1 = Between 30 and 35 years
Primary [a]	1 = Primary school education
Secondary	1 = Secondary school education
University	1 = University education
Partner	1 = Living with a partner
Children	1 = Having children under the age of 15
Foreign	1 = Non-Spanish nationality
Non-Individual Income [b]	Income of other household members (euros)
Regional variables	
Price [b, c]	Price of purchase of housing per square meter
Rural [d]	Rural population ratio (%)
Immigration [d]	Immigrant population ratio (%)

Notes: [a] *Reference variable;* [b] *in logarithms*
Sources: [c] *Ministry of Development;* [d] *INE*

Table 2 shows differences between genders in some of the variables that reflect the characteristics of the young person. Regarding the academic formation, we see that, in the first two levels, the percentage of men is higher than that of women; while in university studies it is women who present a higher value, with a difference of 10 percentage points. The *Foreign* variable shows a difference of almost 5 points in favor of women and in the variables that reflect the family situation (*Partner* and *Children*) the difference is around 8 and 12 points, respectively.

Table 2. Descriptive statistics of individual variables

	Women		***Men***	
Variables	***Mean***	***Std. Dev.***	***Mean***	***Std. Dev.***
Age	30.78	3.69	31.00	3.82
Non-Individual Income	1247.88	911.66	929.32	990.50
Categorical Variables	**%**		**%**	
Primary	30.45		36.56	
Secondary	35.31		39.17	
University	34.24		24.28	
Foreign	23.52		18.98	
Partner	82.99		75.30	
Children	56.37		43.68	

Sources: *own calculation*

From *Table 3* we can see that the regions with the highest prices are Madrid and the País Vasco and the ones with the lowest values are Extremadura and Castilla-La Mancha. These latter regions are also among those with the highest percentages of rural population. And the highest percentage of the immigrant population is in the Baleares, almost 20%, while in Asturias, Extremadura and Galicia it is just below 4%.

Table 3. Descriptive statistics of regional variables

	House prices[a]	**% Rural population**[b]	**% Immigrant population**[b]
Andalucía	1261.5	10.64	7.61
Aragón	1186.9	23.98	10.80
Asturias	1274.2	6.81	3.93
Balears (Illes)	2214.2	5.53	19.94
Canarias	1458.2	4.11	14.34
Cantabria	1450.2	21.28	5.37
CastillaLeón	1036.1	33.69	5.56
Castilla-La Mancha	864.9	31.68	8.47
Cataluña	1948.6	10.14	12.05
Comunidad Valenciana	1181.0	9.24	13.09
Extremadura	857.9	37.02	2.97
Galicia	1197.7	15.39	3.53
Madrid	2481.5	2.20	11.14
Murcia	958.8	0.93	13.45
Navarra	1367.0	32.03	8.47
País Vasco	2381.2	10.20	5.31
Rioja	1110.9	23.00	10.82

Sources: [a] *Ministry of Development;* [b] *INE*

4. Results

Before estimating the model with all the variables, the *ICC* was calculated in a null model (without explanatory variables). This coefficient gives us the percentage of total variance that is attributed to the differences between the regions without considering the effect of the other explanatory variables.

The value obtained for the *ICC* with this null model is 1.27% for the subgroup of women and 2.53% for the subgroup of men, in both cases non-negligible. This indicates that, in the employment of young people, there is a certain heterogeneity between the Spanish regions and that it is appropriate to use a multilevel model.

4.1. Multilevel logit model for labor insertion

In order to establish comparisons in non-linear models between groups whose sample information is different (as in our case) and in which the standard tests of equality of coefficients are not adequate, there are various solutions in the econometric literature. In this paper we follow the solution proposed by Long and Mustillo (2018) and use the marginal effects to compare the weight of the explanatory variables in the insertion in the labor market between the independent young women and men groups.

Table 4 shows the marginal effects of the estimation of the multilevel logit model of labor insertion for both groups of young people. The estimated coefficients of the model are in *Table A1* of the Appendix.

Table 4. Marginal Effects of the Employment model

Fixed Effects	**Women**	**Men**
Age24	0.1624^{***}	0.1674^{***}
Age30	0.2516^{***}	0.2098^{***}
Secondary	0.1591^{***}	0.0534^{***}
University	0.2396^{***}	0.0782^{***}
Foreign	-0.1407^{***}	-0.0355
Partner	0.0787^{**}	0.0791^{***}
Children	-0.1421^{***}	-0.0303
Non-Individual Income	-0.0186^{***}	-0.0074^{**}
Price	0.1022^{***}	0.0861^{**}
Rural	-0.0055	0.0035^{***}
Immigration	0.0041	0.0053^{*}

Note: $^{*}\ p < 10\%$; $^{**}\ p < 5\%$; $^{***}\ p < 1\%$
Sources: *own calculation*

Results show certain differences regarding the significance of the coefficients associated with the characteristics of the individual (first level variables) between the two groups of young people. *Foreign* and *Children* are two variables that do not affect the probability of being employed in the group of men; while these variables have an important weight in the group of women and their effect is negative. Not having Spanish nationality or having children under 15 years of age discourages women from having a job.

The difference between both groups of young people in the influence of *Foreign* may be associated with the type of labor occupation that each performs. Foreign women, on many occasions, carry out jobs related to caring for the elderly or children or domestic service without regularizing their employment situation; while this is not appreciated in the collective of men. On the other hand, children lead, in many cases, women to leave the labor market or reduce the working day, something that rarely happens in the case of men, as has already been pointed out in previous investigations.

The findings indicate that both age and education are determinant in both groups of young people and, in general, the effect of these variables is greater in the group of women. As expected, as the age or academic level of the young person increases, so does the tendency to have a job.

The income of other members of the household also is determining for both women and men, and its weight is greater in the subgroup of women. Its effect is negative, discouraging labor insertion. If other income is received at household, young adult is less likely to have a job.

Regarding the variables of the second level, we see that the situation of the regional real estate market, represented by the housing purchase price, is decisive in the labor insertion in the two groups; although its influence is greater for women. If the housing price is high, the tendency to have a job grows. The variables that collect the regional population structure (*Rural* and *Immigration*) are only determinants in the subgroup of men; while they do not have effect for women. The probability that a young man has a job is higher in regions with a higher proportion of rural population and immigrant population.

The results of the estimation of the multilevel logit model (*Table 4*) show us some variations in the effect of the explanatory variables between women and men. And also, in *Table 2* (descriptive statistics), several differences by gender have been observed in the mean values of the explanatory variables included in model. This indicates that the gender difference in having a job is possibly caused both by differences in unobserved factors (such as stereotyped gender beliefs that attribute specific characteristics to women and men) and by differences in the characteristics observed.

In the next section we decompose the employment rate gender gap. This will allow us to separate the effect of observed characteristics from the effect of unobserved factors and preferences.

4.2. *Gender gap in the labor insertion*

Firstly, following the extension proposed by Fairlie (2005) for a dichotomous outcome variable, we perform the decomposition of the gender gap in the employment rate for an average individual according to (5). *Table 5* presents the overall gap and its decomposition into the explained part and the unexplained part, considering women as group *A* and men as group *B*.

Table 5. Decomposition of employment rate gender gap at the mean value

Overall gap (Women-Men)	**-0.0782757**	**100%**
Endowment Effect Δ_X	0.0205603	-26.27%
Return Effect Δ_ε	-0.0988360	126.27%

Sources: *own calculation*

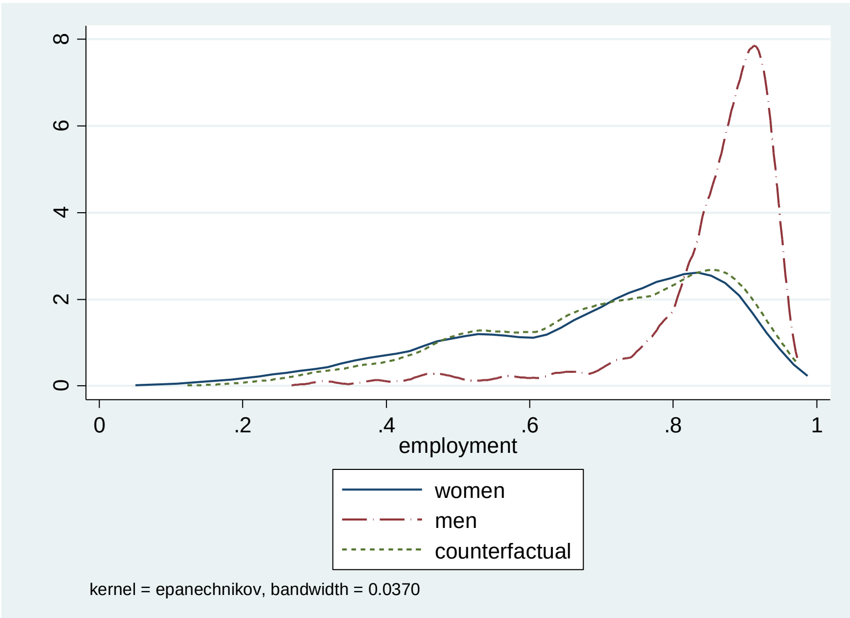
It can be seen that the gender gap in the average value of the employment rate is more than 7 points and with a negative sign. This shows that, on average, the employment rate of women is lower than that of men.

Generally, in periods of economic recovery, such as in 2018 (moment of time analyzed), an increase in the employment rate is observed, although this increase presents different pace in women and men, propitiating an increase in the employment gender gap with respect to a period of economic recession.

From *Table 5*, we see that the value of *endowment* effect has the opposite sign to the overall gap and the *return* effect has the same sign, which indicates that, at the mean value, if women had the same characteristics as men, the gender gap in the employment rate would be even more accentuated.

Regarding the gap decomposition, we see that unobserved factors, tastes or preferences (*return* effect, Δ_ε) explain more of the gender gap in the average employment rate than the observed characteristics (*endowment* effect, Δ_X). Unobserved factors such as occupational segregation, stereotypes, the lack of real equality of opportunities, lower wages or the perception of the family that women have which, on many occasions, leads women to leave the labor market, seem to be responsible for the difference between women and men.

Next, we analyze the gender gap throughout the entire distribution, since the gender difference varies throughout the distribution of the outcome variable, and also the causes that generate it can vary. We calculate the probability density function of the labor insertion for women, that for men and the corresponding counterfactual, from the model estimates (Appendix *Table A1*). The counterfactual density function allows us to know how the probability density function of women would be if they had the same observed characteristics as men. *Graph 2* presents the profile of these functions.



Graph 2. Probability densities of employment
Sources: *own calculation*

We see a considerable probability mass in the upper tail for men, indicating that many of the men present a high probability of having a job. Women have a flatter density function profile, with a non-negligible probability mass in the middle and upper part of the distribution.

Focusing on the counterfactual density function, we see that this function is very close to the probability density function of women throughout the entire distribution. This indicates that if women had the same observed characteristics as men, the probability that a woman would have a job would change little. Thus, the observed characteristics do not seem to be the main responsible for the gender gap in labor insertion.

In order to obtain the differences in the different percentiles of the probability distribution of labor insertion, the cumulative distribution function (CDF) for women, that for men, and the corresponding counterfactual are calculated. *Table 6* presents the tabulated values of the cumulative probability for the main percentiles and *Table 7* shows the values of the overall gender gap, its decomposition and the percentage explained by the observed characteristics (*endowment* effect) and the unobserved factors (*return* effect).

Table 6. Cumulative employment probability at main percentiles

Percentile	**Group A: Women**	**Group B: Men**	**AB: Counterfactual**
10	0.40	0.75	0.46
20	0.51	0.82	0.54
30	0.60	0.85	0.63
40	0.67	0.86	0.69
50	0.72	0.88	0.73
60	0.77	0.90	0.79
70	0.82	0.91	0.83
80	0.85	0.92	0.87
90	0.88	0.93	0.90

Sources: *own calculation*

In *Table 6* we see notable differences in the employment probability by gender, and that men present higher probability values than women in all percentiles of the distribution. At the 20th percentile, men already have a probability of 82%, but that probability is not reached until the 70th percentile in the case of women.

From *Table 7* can be seen that the overall gap is negative in all the percentiles of the distribution, and of a different magnitude depending on the percentile considered. At the low percentiles, at the 10th and 20th, there is a gender difference of more than 30 percentage points. Little by little this difference is reduced, until the 90th percentile where it is only 5 points.

Table 7. Decomposition of the employment gender gap at main percentiles

Percentile	**Overall gap (A-B)**	**Endowment: (A-AB)**	**Return: (AB-B)**	**% Endowment**	**% Return**
10	-0.35	-0.06	-0.29	17.2%	82.8%
20	-0.31	-0.03	-0.28	9.7%	90.3%
30	-0.25	-0.03	-0.22	12%	88%
40	-0.19	-0.02	-0.17	10.5%	89.5%
50	-0.16	-0.01	-0.15	6.3%	93.7%
60	-0.13	-0.02	-0.11	15.4%	84.6%
70	-0.09	-0.01	-0.08	11.1%	88.9%
80	-0.07	-0.02	-0.05	28.6%	71.4%
90	-0.05	-0.03	-0.02	60%	40%

Sources: *own calculation*

When analyzing the decomposition of the overall gap, it has been obtained that it is the residual component (*return* effect) that explains the gender differences in most of the percentiles, as was the case when we previously performed the decomposition at the mean value of the distribution. The observed characteristics only have a prominent weight in explaining the overall gap at the last percentiles. At the 80th percentile this *endowment* effect represents almost 29% and at the 90th percentile it is already 60% (a weight greater than that of the residual part).

Results of gender gap decomposition highlight that the labor discrimination suffered by women is associated, mainly, with the fact of being a woman, not with the possible differences in the characteristics observed in each group of young people. Occupational segregation is surely behind this, together with the differences in the perception of the family that women and men have, which lead to greater female labor abandonment at certain moments of life.

Conclusion

This paper confirms that in the employment rate of Spanish young people there are clear differences by gender, not only in magnitude, but also in what are the determinants in each group of young adults.

Results of the multilevel binomial logit model have shown that the characteristics of the individual have greater weight in the group of women than in the group of men. Regarding the characteristics of the region of residence (second level variables), the housing price impact in both groups, while the proportion of rural population and the proportion of immigrant population only shows an effect in the group of men.

The analysis of the gender gap in the employment rate revealed that, in Spain, young women have a considerably less employment rate than men. In the first percentiles of the distribution, differences of more than 30 points have been found, however these differences are gradually being reduced.

Gender gap decomposition indicated that the observed characteristics are not the main responsible for the differences, but rather it is the residual component that explains most of the gap. The observed characteristics have greater weight in the last percentiles of the distribution.

Our results highlight that the gender difference in the employment rate is associated, mainly, with the occupational segregation of women (women look for occupations that, although advantageous in terms of schedules or labor flexibility, in general, offer lower salaries), to the differences in the perception of the family that women and men have, which leads to labor female abandonment at certain moments of life.

The analysis of the employment rate gender gap along the percentiles of the entire distribution can be useful for the improvement of labor policies in the group of young people. The decomposition throughout the entire distribution makes it possible to identify where the greatest gender differences are in the employment rate. This identification will allow the design of governmental lines of action aimed to mitigate the labor discrimination against women in the family formation period, and to help those more disadvantaged or in a more precarious situation young people collectives.

References

Alawad, A., Kreihan, F., & Selim, M. (2020). Determinants of Youth Unemployment: Evidence from Jordan. *International Journal of Economics and Business Administration*, *4*(VIII), 152-165. doi:10.35808/ijeba/576

Antecol, H. (2000). An examination of cross-country differences in the gender gap in labor force participation rates. *Labour Economics*, *7*, 409-426. doi.org/10.1016/S0927-5371(00)00007-5

Banerjee, M. (2019). Gender Equality and Labour Force Participation: Mind the Gap. *Antyajaa: Indian Journal of Women and Social Change*, *4*(1), 113-123. doi:10.1177/2455632719831827

Bayrak, R. & Tatli, H. (2018). The Determinants of Youth Unemployment: A Panel Data Analysis of OECD Countries. *The European Journal of Comparative Economics, 15*(2), 231-248. doi.org/10.25428/1824-2979/201802-231-248

Bell, N.F. & Blanchflower, D.G. (2011). Young people and the Great Recession. *Oxford Review of Economic Policy*, 25(2), 241-267. doi: 10.1093/oxrep/grr011

Berry, C. & McDaniel, S. (2020). Post-crisis precarity: Understanding attitudes to work and industrial relations among young people in the UK. *Economic and Industrial Democracy*, 43(1), 322-343. doi:10.1177/0143831X19894380

Blinder, A. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, *8*(4), 436-455. doi:10.2307/144855

DiNardo, J., Fortin, N. M. & Lemieux, T. (1996). Labor Market Institutions and the Distribution of Wages. 1973-1992: A Semiparametric Approach. *Econometrica*, *64*, 1001-1044. doi:10.2307/2171954

Fairlie, R.W. (2005). An extension of the Blinder-Oaxaca decomposition technique to logit and probit models. *Journal of Economic and Social Measurement*, *30*(4), 305-16. doi:10.3233/JEM-2005-0259

Fitzenberger, B.; Schnabel, R. & Wunderlich, G. (2004). The gender gap in labor market participation and employment: A cohort analysis for West Germany. *Journal of Population Economics, 17*, 83-116. doi:10.1007/s00148-003-0141-6

Grigoli, F. Koczan, Z. & Topalova, P. (2018). Drivers of Labor Force Participation in Advanced Economies: Macro and Micro Evidence. *International Monetary Fund, WP/18/150.* doi:10.5089/9781484361528.001

INE (2018). *Household budget survey (EPF)* [Data file]. https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176806&menu=resultados&idp=1254735976608#!tabs-1254736195147

INE (2000-2019). *Active Population Survey (EPA)* [Data file]. https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176918&menu=ultiDatos&idp=1254735976595

Long, J.S., & Mustillo, S.A. (2018). Using predictions and marginal effects to compare groups in regression models for binary outcomes. *Sociological Methods and Research*. doi:10.1177/0049124118799374

Machado, J.A. & Mata, J. (2005). Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression. *Journal of Applied Econometrics, 20*, 445-465. doi:10.1002/jae.788

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, *14*(3), 693-709. doi:10.2307/2525981

Appendix

Table A1. Multilevel logit model of employment

Fixed effects	***Women***	***Men***
γ_{00}	-4.3997***	-6.3582**
Age24	0.8934***	1.4489***
Age30	1.3839***	1.8163***
Secondary	0.8753***	0.4627***
University	1.3179***	0.6775***
Foreign	-0.7738***	-0.3070
Partner	0.4328**	0.6847***
Children	-0.7814	-0.2623
Non-Individual Income	-0.1023***	-0.0640**
Price	0.5623***	0.7457**
Rural	-0.0030	0.0310***
Immigration	0.0226	0.0459**
Random effects		
τ_{00}	0.0001	0.0348

Notes: * $p < 10\%$; ** $p < 5\%$; *** $p < 1\%$
Sources: *own calculation*