ISSN: 2281-1346

UNIVERSITÀ DI PAVIA
Department of Economics and Management

DEM Working Paper Series

Gender pay gap: a route from the North to the South of Italy

Maria Elena Filippin
(University of Pavia)

 # 176 (12-19)

Via San Felice, 5
I-27100 Pavia

economiaweb.unipv.it

Gender pay gap: a route from the North to the South of Italy*

Maria Elena Filippin†

December 10, 2019

Abstract

This paper analyzes the gender pay gap across different regions in Italy, using the Oaxaca-Blinder decomposition method. We expect regional heterogeneity, both in terms of the gender pay gap and in its determinants. Our results show that, on a regional basis, the retribution gap widely varies, as its percentages of the explained and unexplained parts. Workers' observable characteristics, related to both labor and personal features, that justify the explained part at a national level are confirmed by the regional data. Furthermore, data on the activity rate show that both at a national and regional level, female participation to the labor market, although it has been improving in recent years, is still profoundly lower than the male one. Therefore, we implement the Heckman correction, which reveals that women's model coefficients are overestimated, at a national level and in half of the Italian regions. This result suggests that, although female participation in the labor market is lower than the male one, the fewer women participating in the labor market, on average, have higher productivity than men.

JEL classification: J16, J21, J31, O15, R1
Keywords: Gender Pay Gap, Oaxaca-Blinder, Italian regions

*The author thanks Carolina Castagnetti for the ongoing support during the research project, and Annalisa Ferrando for her valuable suggestions. The usual disclaimer applies.

†e-mail: mariaelena.filippin01@universitadipavia.it

1 Introduction

The literature on the gender pay gap (GPG) is vast and wide-ranging, and it focuses mainly on the dynamics over time. Several studies have shown a decline in gender earning disparities: this trend had been dramatic in the United States between 1970 and 1990 (Mandel and Semyonov, 2006), even though there has been a slowdown of the trend towards the new millennium (Blau and Kahn 2006; Blau et al., 2006). Castagnetti, Rosti and Töpfer (2018) have analyzed the effect of public-contest recruitment on the gender pay gap in Italy between 2005 and 2014, and they have found that the GPG vanishes as the public-contest promotes a more merit-based and gender-fair wage setting. However, women continue to earn considerably less than men; about 20% in the United States of America and 15% in Europe. Despite the empirical finding that the difference in pay between men and women has decreased in the last decades, a consistent part of the GPG remains unexplained, and this part has not declined over time but has been roughly stable over the past 30 years (Blau and Kahn, 2017). However, although the literature on the causes and consequences of the gender pay gap has grown and become substantial, there is one aspect that has been neglected and this is about the geographical aspect of the GPG. In this paper, we focus on the inter-regional GPG in Italy, for which the analysis are still scarce.

Since when the Kingdom of Italy was proclaimed, in March 1861, there had been an exceptionally pronounced inter-regional gap between the northern and southern regions of the country. In spite of Italy's remarkable industrialization during the twentieth century, dualism remains even now its greatest unresolved problem (Toniolo, 2016). Beyond the varied social-economic situation of Italy, the woman's role differs considerably from the North to the South of the country. More specifically, there are two elements which underline the different female occupational situation in the southern regions. First, a long domestic tradition for which women are in charge of all the housework and family care. Second, the lack or reduced presence of support facilities to women's work, such as public nurseries and full-time schools, which makes the decision to enter the labor market even less attractive. We focus on the female participation to the labor market later in our work, and we analyze how this impact on our study of the GPG. In this way, our work gives an important contribution to the economic analysis of inter-regional wage differentials in Italy.

To our knowledge, there are no works that has attempted to study the size and reasons of the inter-regional GPG in Italy. Bozzano (2014) has explored and evaluated the gender gap by means of the Italian Gender Gap Index, an indicator developed from the Global Gender Gap Index by the Word Economic Forum and adapted to the regional context in Italy. However, she only marginally focuses on the earning gaps between men and women. Our work contributes substantially to the topic for several reasons. First, we study the gender pay gap by means of the well-established model developed by Oaxaca and Blider. Second, we use a rich dataset provided by ISFOL-PLUS, which covers several data waves from 2005 up to 2016. In her work, Bozzano has encountered many problems in collecting data. First, for the low data availability, her study is conducted for the year 2008 only. Secondly, she used low quality collection of disaggregated data at both gender and regional level.

Therefore, despite the valuable previous literature, our work constitutes the first attempt

to question about the causes and consequences of the wage gap across Italian regions.

The paper is organized as follows. Section 2 presents the econometric approach, while Section 3 describes the data and the preliminary analysis. Section 4 reports and discusses the results of the GPG decomposition and their explanatory characteristics. Section 5 extends the analysis by accounting for sample selection bias. Finally, Section 6 concludes.

2 Econometric model

There are several alternative techniques for decomposing the gender pay gap via the use of regression equations. We decide to implement the procedure proposed by Oaxaca (1973) and Blinder (1973). This twofold decomposition method distinguishes between two distinctive portions in earning differentials: a portion that is explained by gender differences in work-related observed characteristics, collected in the set of the x_{ki}; and the unexplained portion of the gap that cannot be accounted for by mean differences in wage determinants.

The Oaxaca-Blinder decomposition is implemented in two stages: the regression analysis and the decomposition analysis. The regression analysis is carried out separately for the earnings structure of men (M) and the earnings structure of women (F), as detailed in the following equations:

$$\ln w_i^M = \beta_0^M + \sum_{k=1}^{K} x_{ki}^M \beta_k^M + u_i^M \,, \tag{1}$$

$$\ln w_i^F = \beta_0^F + \sum_{k=1}^{K} x_{ki}^F \beta_k^F + u_i^F \,, \tag{2}$$

where $\ln w_i$ represents the natural logarithm of hourly wages of individual, x_{ki} are set of variables covering the observed characteristics that may impact on log hourly wage of individual, β_0 is a constant and β_k are the parameters for the corresponding variables covering the observed characteristics, u_i is the disturbance term for observation i.

The Oaxaca-Blinder decomposition uses the following regression property for the means of the log hourly earnings of men and women:

$$\overline{\ln w^G} = \hat{\beta}_0^G + \sum_{k=1}^{K} \overline{x_k^G} \hat{\beta}_k^G \,, \tag{3}$$

being G the gender (male or female).

In the second step, the difference between the means of log hourly wages of men and women is decomposed into two parts: the explained one and the unexplained one. Within the decomposition approach, it must be decided which earnings structure constitutes the non-discriminatory benchmark against which to decompose the difference between the means of hourly wages of men and women. In accordance with the definition of the unadjusted gender pay gap, it is usually assumed that the male earnings structure constitutes this benchmark. Thus, the male constant and coefficients are treated as non-discriminatory wage structure.

By following this standard approach, a counter-factual equation is constructed by replacing male constant and coefficients in women's equation:

$$\overline{\ln w^{F}*} = \hat{\beta_0^M} + \sum_{k=1}^{K} \overline{x_k^F}\hat{\beta_k^M} \,. \tag{4}$$

This equation can be interpreted as the log hourly wage that a female worker would have earned if she had been paid on the same basis as an equivalent male worker. By subtracting the counter-factual to the actual mean of log hourly wages of women, it is possible to calculate the unexplained part.

The explained part can be expressed as the difference between the actual mean of the log hourly wages of men and the counter-factual mean of the log hourly wages of women.

The final decomposition equation for the wage differential is as follows:

$$\overline{\ln w^{M}} - \overline{\ln w^{F}} = (\hat{\beta_0^M} - \hat{\beta_0^F}) + \sum_{k=1}^{K} \overline{x_k^M}\hat{\beta_k^F} + \sum_{k=1}^{K} \hat{\beta_k^M}(\overline{x_k^M} - \overline{x_k^F}) \,, \tag{5}$$

where the difference between the log hourly wages for male and female samples, evaluated at the mean, is the gender pay gap. The last term is the difference in endowments in terms of characteristics at the rate of return of men. Since different endowments have different effects on wages, their difference represents the explained GPG. The first two terms in equation (5) constitute the unexplained GPG because the two coefficients β^M and β^F differ by gender whereas, as the same endowments should have the same effect on wages for both men and women, they should be equal.

In what follows, the gender pay gap analysis will be conducted in the subsequent way: first, the overall explained and unexplained parts, expressed in log unit percentages of the wages differential, will be commented both at national as well as at regional level; then, the contribution of each characteristic will be evaluated for the explained part of the GPG.

3 Data and preliminary analysis

The empirical analysis is based on micro-data collected by the Italian Institute for the Development of Vocational Training for Workers (ISFOL) in the Participation, Labor, Unemployment Survey (PLUS). The data were collected in the context of a joint project with the Italian Ministry of Labor and Social Policy, started in 2005, which seeks to create a data set for the study of wage inequality by gender. The data set provides a large number of sample cases where all the variables are comparable over time, and deliver broad information on the participants' work profiles and territorial backgrounds. The data waves in the data set cover years 2005, 2006, 2008, 2010, 2011, 2014, and 2016, where each year includes panel interviews with participants from the previous sample. We consider the panel dimension in our analysis, taking into account all available years.

Our analysis focuses on full-time employees aged between 18 and 65 years. We restrict the sample by considering only salaried employees and excluding students, trainees, and pensioners. Similarly, self-employed workers are not considered in the study. Those restrictions

are aimed at forming a homogeneous sample of employed individuals. Part-time workers are excluded from the sample as they probably earn less than the average hourly wage as they have more significant dispersion in the pay than full-time workers. Moreover, the number of part-time workers differs significantly between man and women, with women more likely to be employed part-time (Chzhen and Mumford, 2011). The analysis is also constrained to earnings from the main job only which yields the highest part of incomes.

We use the log of the monthly net wages as the dependent variable. In previous studies, the net hourly wages are used instead, but men and women tend to work different hours, with men more likely to be in full-time jobs and women more likely to be employed part-time. Therefore, taking into account the monthly or yearly earning, the gender pay gap enlarges as a consequence of the lower female working hours. The gap increases even more accounting for the low labor force participation of women that, despite having the same productive characteristics of men who work, stay out of the labor market.

We select a group of about twenty independent variables. The variables are ordered based on two clusters of qualities: personal characteristics, including physical (sex, age), domestic (marital status, family size, presence of one or more children), educational (qualification, years of schooling, honors degree), personal wealth (ownership of the house) and geographical (Italian citizenship, district extent, living region); and labor characteristics which include the firm type (public or private), the years of experience, the contractual form (temporary versus permanent job), the profession, the sector of employment and the presence of overtime work. The variable referred to working condition is used to distinguish between who is inside the labor forces and who is not. Finally, year dummies are included in the analysis. Table 6 in Appendix A describes the variables we use for our analysis.

Table 1 presents the results of a preliminary OLS regression of the various explicative variables on the log of net wages. Most of the estimated coefficients have the expected signs. First, being a woman significantly reduces earnings. Second, other variables which have a similar negative impact on wages are mainly related to labor characteristics, such as the one related to people employed in the private sector, or with a temporary contract. Furthermore, most of the professions show negative and significant coefficients. Employees in unskilled works, including housemaids and babysitters, are those whose wages are most negatively affected, followed by retailers, artisans, and farmers. Wages of semi-skilled workers, such as steam-operators, are negatively influenced too, as well as earning of office workers. On the contrary, both senior positions of managers and entrepreneurs and scientific professions show positive and significant coefficients. Those results are in line with empirical findings that highly specialized works allow, on average, higher salaries than jobs which do not require advanced skills. Among the sectors of employment, only be employed in financial activities has positive significant impact on wages, while work in family services has a negative effect, as well as work in the education sector. Those two sectors traditionally employed mainly female workers, a shred of evidence that could support the hypothesis that women earn on average less than men.

However, while wages seem to be most negatively affected by features at the labor market level, some personal characteristics of individuals have a positive impact on earnings. This is the case for the variables referred to age and education. Wages are positively affected if

individuals live in large cities. Nevertheless, while the variable related to years of working experience has positive and significant coefficient, its value squared shows the opposite trend. This fact can be due to the shape of an average worker's experience function, which we fairly assume to be a monotonic parabolic function in the observed interval, but the positive difference in experience weights more as little is the presence on the labor market. In other words, after a while, the impact of the years of experience seems to become less critical. Empirical evidence confirms this fact, as salary stabilizes ones reached the advanced career positions.

Table 1: OLS Estimates of Log Monthly Wages for Italy - Panel
The negative coefficient of the gender dummy indicates that being a woman significantly reduces earnings. Other variables which have a negative impact on wages are mainly related to labor characteristics, such as private sector employment, temporary contract, unskilled professions and artisan business, family service and educational sectors of employment. Some variables indicating personal characteristics of individuals have a positive impact on wages. This is the case for age, education and living in large cities.

Variables	lnwage	Std.Error
female	-0.244***	(0.003)
age	0.006***	(0.000)
schooling	0.024***	(0.001)
honors	0.089***	(0.007)
exper	0.011***	(0.001)
exper2	-0.000***	(0.000)
city_size2	-0.002	(0.003)
city_size3	0.015***	(0.004)
private	-0.098***	(0.004)
temporary	-0.117***	(0.005)
prof_senp	0.143***	(0.018)
prof_scip	0.060***	(0.017)
prof_tech	-0.018	(0.017)
prof_offi	-0.080***	(0.017)
prof_comm	-0.181***	(0.017)
prof_skil	-0.120***	(0.017)
prof_sems	-0.089***	(0.018)
prof_unsk	-0.271***	(0.018)
sect_agri	-0.029	(0.037)
sect_indu	0.025	(0.036)
sect_ener	0.014	(0.037)
sect_cons	0.005	(0.037)
sect_buss	-0.028	(0.036)
sect_tour	-0.049	(0.036)

Variables	lnwage	Std.Error
sect_tran	0.023	(0.036)
sect_comm	-0.031	(0.036)
sect_fina	0.120***	(0.036)
sect_serv	-0.058	(0.036)
sect_publ	-0.050	(0.036)
sect_educ	-0.084**	(0.036)
sect_heal	-0.009	(0.036)
sect_scie	-0.016	(0.036)
sect_fami	-0.104***	(0.036)
overtime	0.064***	(0.003)
_cons	6.622***	(0.040)
Observations	71261	
R-Squared	0.332	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2 shows the results of the regional regressions with focus on the impact of gender on net wages. The coefficients for women are negative and significant in all the regions. Moreover, the coefficients are quite homogeneous too, as well as the fitness of the different models.

We replicate the OLS analysis at regional level.[1] Overall, the national evidence can be confirmed at the regional level. The variable indicating people employed in the private sector and the variable for temporary contracts are negative and significant in all the regions. On the contrary, the variables regarding age and education show positive and significant coefficients in all the regions. Nevertheless, the dummy indicating people graduated with honors is significant at the national level and in 3/5 of the Italian regions, with most of the exceptions located in the North of Italy. Table 3 reports the single regional coefficients for the dummy variable referred to graduation with honors. At the national level, living in large cities has a positive impact on wages. This tendency is verified in large regions such as Lombardia, Lazio, Puglia, and Sicilia, while Molise and Campania present the opposite trend. Here, the variable which takes into account living in a medium city size has positive and significant coefficient.

The most interesting findings are referred to the professions and sectors of employment. Here, the regional analysis show a lot of heterogeneity. Work in a senior position has positive and significant impact on wages especially in the North of the country, whereas the negative significant effect of the low-skilled professions is accentuated in the Centre and in the southern part of Italy. Moreover, while at the national level the significance of the sectors of employment is restricted to a few areas, the situation is different looking when at single regions. In Toscana, besides financial activities, there are several employment sectors which show positive and significant coefficients, such as health, energy and construction, tourism

[1]Tables can be consulted upon request to the author.

and transport. Moving to the South, the significance of the areas of employment improves. In Basilicata, energy and financial activities are positive significant too, as well as business and communication sectors. On the contrary, in Puglia most of the areas present negative significant values. In Sicilia and Sardegna the situation is reversed: the former region has positive and significant values for all the sectors of employment, but for education and family services, which are not significant; the latter's coefficients are all negative significant, except for transport and financial activities that are not significant.

Table 2: OLS Estimates of Log Monthly Wages by region, *Female* - Panel
The coefficients for the gender dummy are negative and significant in all the regions. The coefficients and the fitness of the different models are homogeneous across the Italian regions.

	Piemonte + VdA	Liguria	Lombardia	Trentino-Alto Adige	Veneto	Friuli-Venezia Giulia
female	-0.245***	-0.300***	-0.253***	-0.266***	-0.296***	-0.252***
	(0.012)	(0.015)	(0.009)	(0.016)	(0.011)	(0.016)
Observations	4871	3026	9890	2903	5558	2799
R-squared	0.337	0.399	0.290	0.288	0.321	0.340

	Emilia-Romagna	Toscana	Marche	Umbria	Lazio	Molise	Abruzzo	Campania	Puglia
female	-0.241***	-0.245***	-0.242***	-0.272***	-0.237***	-0.260***	-0.230***	-0.233***	-0.233***
	(0.012)	(0.012)	(0.017)	(0.018)	(0.013)	(0.019)	(0.018)	(0.013)	(0.014)
Observations	5140	4382	2794	2462	4514	1868	2347	4348	3550
R-squared	0.300	0.288	0.321	0.340	0.300	0.373	0.351	0.349	0.343

	Basilicata	Calabria	Sicilia	Sardegna
female	-0.206***	-0.206***	-0.221***	-0.229***
	(0.019)	(0.018)	(0.015)	(0.016)
Observations	2063	2132	3735	2879
R-squared	0.344	0.415	0.345	0.374

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: OLS Estimates of Log Monthly Wages by region, *Honors* - Panel
The coefficient of the dummy variable for people graduated with honors is significant at the national level and in 3/5 of the Italian regions, with most of the exceptions located in the North of Italy.

	Piemonte + VdA	Liguria	Lombardia	Trentino-Alto Adige	Veneto	Friuli-Venezia Giulia
honors	0.101***	0.142***	0.097***	-0.004	0.023	-0.008
	(0.031)	(0.029)	(0.024)	(0.047)	(0.032)	(0.037)
Observations	4871	3026	9890	2903	5558	2799
R-squared	0.337	0.339	0.290	0.288	0.321	0.340

	Emilia-Romagna	Toscana	Marche	Umbria	Lazio	Molise	Abruzzo	Campania	Puglia
honors	0.027	0.070***	0.052	0.027	0.149***	0.056	0.227***	0.119***	0.131***
	(0.025)	(0.024)	(0.037)	(0.044)	(0.026)	(0.040)	(0.042)	(0.027)	(0.028)
Observations	5140	4382	2794	2462	4514	1868	2347	4348	3550
R-squared	0.300	0.373	0.351	0.349	0.343	0.428	0.369	0.393	0.403

	Basilicata	Calabria	Sicilia	Sardegna
honors	-0.026	0.120***	0.145***	0.108***
	(0.054)	(0.045)	(0.035)	(0.033)
Observations	2063	2132	3735	2879
R-squared	0.344	0.415	0.345	0.374

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4 Model and Results

The ISFOL PLUS data are used in the decomposition and estimation of the gender pay gap. Table 4 reports the results of the Oaxaca-Blinder method applied to Italy as a whole. The difference between the means of male and female wages, expressed in logarithmic scale, amounts to 26.30%, among which 0.76% is due to, on average, better characteristics of men compared to women in the labor market, and a gap of 98.86% remains unexplained. The table reports also the sub-components of the explained section. At the national level this is mostly driven by characteristics of the labor market, such as temporary contracts and overtime work, whereas age is the only explanatory component in personal features of individuals. Moreover, both variables related to work experience and its square show coefficients' signs in line with our previous findings from the OLS regressions.

Table 4: Oaxaca-Blinder decomposition for Italy, Explained Part Components - Panel
The difference between the means of male and female wages amounts to 26.30%. The explained part is mostly driven by labour characteristics, such as temporary contract, overtime work, work experience, be employed in unskilled professions, working in education, health, and family services sectors. Variables with negative coefficients partially cancelled out the positive factors which drive the explained part of the gap. This is the case for education, living in large cities and private sector employment.

Variables	lnwage	Std.Error
Overall		
group_1	7.230***	(0.002)
group_2	6.967***	(0.002)
difference	0.263***	(0.003)
explained	0.002	(0.003)
unexplained	0.260***	(0.004)
Explained		
female	0.000	(.)
age	0.012***	(0.001)
schooling	-0.020***	(0.001)
honors	-0.002***	(0.000)
exper	0.019***	(0.003)
exper2	-0.005**	(0.003)
city_size2	-0.000	(0.000)
city_size3	-0.001***	(0.000)
private	-0.014***	(0.001)
temporary	0.004***	(0.000)
prof_senp	0.002*	(0.001)
prof_scip	-0.003	(0.003)

Variables	lnwage	Std.Error
prof_tech	0.000	(0.002)
prof_offi	0.005	(0.004)
prof_comm	0.008***	(0.002)
prof_skil	-0.018***	(0.006)
prof_sems	-0.009**	(0.004)
prof_unsk	0.002***	(0.001)
sect_agri	-0.001	(0.000)
sect_indu	-0.003	(0.005)
sect_ener	-0.003*	(0.002)
sect_cons	-0.005**	(0.003)
sect_buss	0.004*	(0.002)
sect_tour	-0.001**	(0.001)
sect_tran	-0.004	(0.003)
sect_comm	-0.001	(0.000)
sect_fina	0.000	(0.000)
sect_serv	0.002**	(0.001)
sect_publ	-0.004**	(0.002)
sect_educ	0.026***	(0.009)
sect_heal	0.007*	(0.004)
sect_scie	-0.001*	(0.001)
sect_fami	0.002***	(0.001)
overtime	0.004***	(0.000)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Focusing on the characteristics within the professions and sector of employment, we find some interesting insights. Working in unskilled professions contributes positively to explain wage differentials, however be employed in senior positions shows a positive significant coefficient too. This fact can be interpreted as a signal of gender segregation: male workers work, on average, in better paid sectors compared to women.[2] Also, sectors such as business, services for firms, education, health, and family services have a positive effect on the explained part of the gap. In particular, the latter three sectors are mainly dominated by the female component of the labor market. However, from the analysis, it seems that men working in those areas of employment have, on average, better characteristics than women working in the same areas.

The positive factors are partially canceled out by the negative impact of some other variables. This means that women are expected to earn more than men due to possessing on average higher levels of given characteristic than men. For example, a negative gap in educational variables means that women are expected to earn more than men as they have, on average, higher levels of education. The same occurs if the individual lives in a large city

[2]We refer to the NACE classification at section level (17 economic areas).

size, or if he/she works for a private firm. Several sectors of employment thus show negative significant coefficients: energy, construction, tourism, public administration, and scientific professions.

We compute the Italian unexplained GPG to be equal to 98.86%, expressed in log units, of the overall gap. In other words, women with the same average characteristics of men, are expected to earn 98.86% log points less than men due to higher financial returns to men than women. This unexplained portion of wages differential can be interpreted as the presence of discrimination against women. However, the unexplained part can also be due to unobserved characteristics in the labor market, such as the individual productivity. We will investigate this issue in the next session.

Figure 1 shows the explained part and the unexplained part by Italian region, respectively, arranged from smallest to largest GPG%, where the percentages are expressed in log units. Among the Italian regions, the explained GPG varies from -12% in Toscana to 11.79% in Basilicata, whereas the unexplained GPG varies from 88.21% in Basilicata to 112.20% in Lazio. The overall explained gap is negative in ten regions: Toscana, Lazio, Puglia, Trentino-Alto Adige, Piemonte, Val d'Aosta, Sicilia, Lombardia, Liguria, and Veneto.

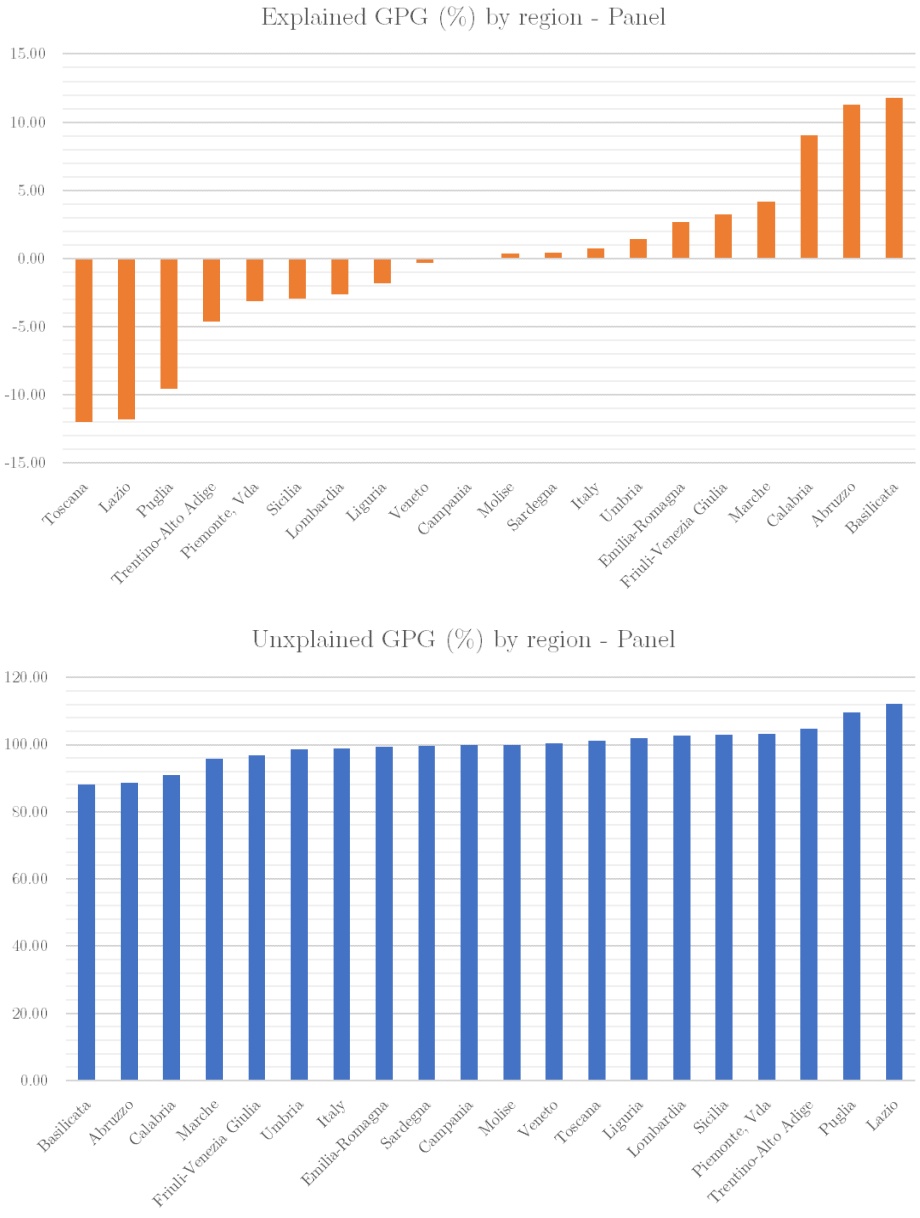
At regional level, the overall GPG is mostly driven by the same characteristics found at the national level. Overall, all those characteristics have a homogeneous effect across Italian regions. This means that gaps for those characteristics are either positive or negative in all of the countries.

The explained GPG is positively mostly driven by three factors: age, temporary contracts, and overtime job. The first two variables show positive and significant coefficient in seventeen regions, while the third does it in fourteen regions.[3] Education and working in a private firm drive the explained part of the wage differential negatively. The variable for year of schooling is negative and significant in all the regions, while the dummy for people graduated with honors shows the same behavior in just half of the Italian countries. These results mean that, on average, women have a higher level of education than men in most regional labor markets. Finally, the variable for employees in the private sector explains the explained GPG negatively everywhere but in Calabria, where the variable is not significant.

A more mixed picture can be observed for professions and sectors of employment. Working in senior positions justify the explained GPG positively at Italian level: this is the case for Piemonte and Val d'Aosta, Friuli-Venezia Giulia, Toscana, and Sicilia. With a couple of exceptions, Bozzano (2014) has found similar results in her analysis of the Italian Gender Gap within the Economic Participation and Opportunity Dimension (ECO). In fact, she found that, in managerial positions, equality between men and women had been reached only in Friuli-Venezia Giulia, Umbria, Abruzzo and Basilicata, while Lazio, Piemonte and Lombardia show low levels of equality. We also found that commerce and unskilled works have positive and significant coefficients for Friuli-Venezia Giulia and Emilia-Romagna. Skilled and semi-skilled works do the same in Piemonte and Val d'Aosta as well as in Toscana. This means that, in all these professions, men have on average better characteristics than women in each regional labor market. Be employed in agriculture and industry sectors positively justified

[3]Tables can be consulted upon request to the author.

Figure 1: Explained and Unexplained GPG (%) by region - Panel
The explained part can be expressed as the difference between the actual mean of the men' wages and the counter-factual mean of the women' wages. By subtracting the counter-factual to the actual mean of wages of women, it is possible to calculate the unexplained part. Among the Italian regions, the explained GPG varies from -12% in Toscana to 11.79% in Basilicata, whereas the unexplained GPG varies from 88.21% in Basilicata to 112.20% in Lazio.

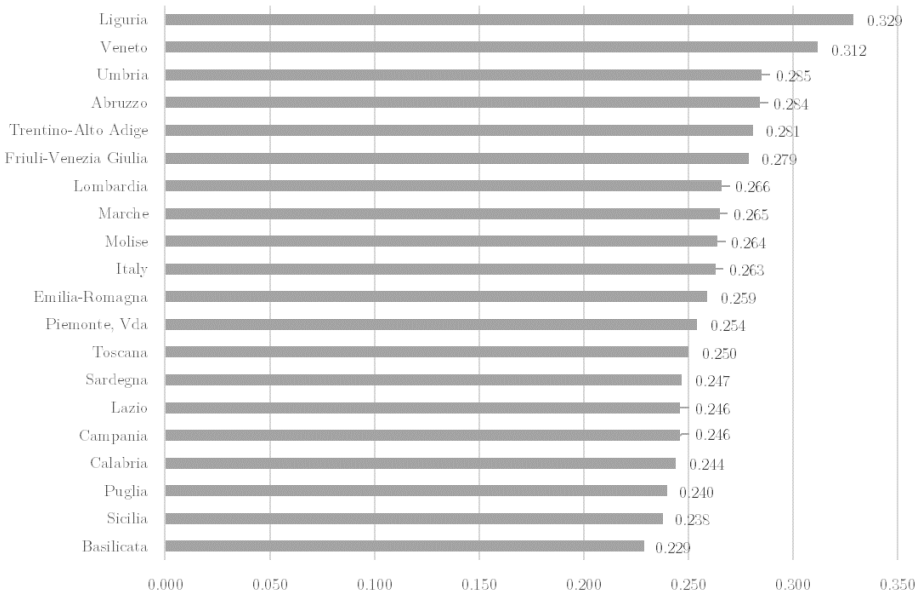


Data source: ISFOL PLUS Questionnaire

the explained part of wage differentials in Umbria, while the reverse occurs in Sardegna. Those two regions are the only ones where areas of employment have the most significant coefficients, whereas in the remaining regions there are few significant evidence.

Figure 2 reports the overall GPG by region expressed in logarithmic scale. Among the Italian regions, the gap goes from 22.9% log units in Basilicata to 32.9% log units in Liguria. The top five regions, where the GPG is lower, are all located in the South of Italy, while the worst five performers are Trentino-Alto Adige, Abruzzo, Umbria, Veneto and Liguria, in increasing order. The latter result contrast with what Bozzano (2014) says in her study. In fact, she found that the overall gender gap is mostly closed in Piemonte, while southern regions, such as Puglia, Basilicata and Molise lags behind. However, she also found that larger gender gaps emerge in the economic sphere. Therefore, the reasons behind the larger gender gap in the southern countries must be searched outside the economic area. For this reason, in the following session, we analyze the labor force participation rate and we implement the Heckman correction.

Figure 2: Gender Pay Gap by region in log units - Panel
Among the Italian regions, the gap goes from 22.9% log units in Basilicata to 32.9% log units in Liguria. The top five regions, where the GPG is lower, are all located in the South of Italy, while the worst five performers are Trentino-Alto Adige, Abruzzo, Umbria, Veneto and Liguria, in increasing order.

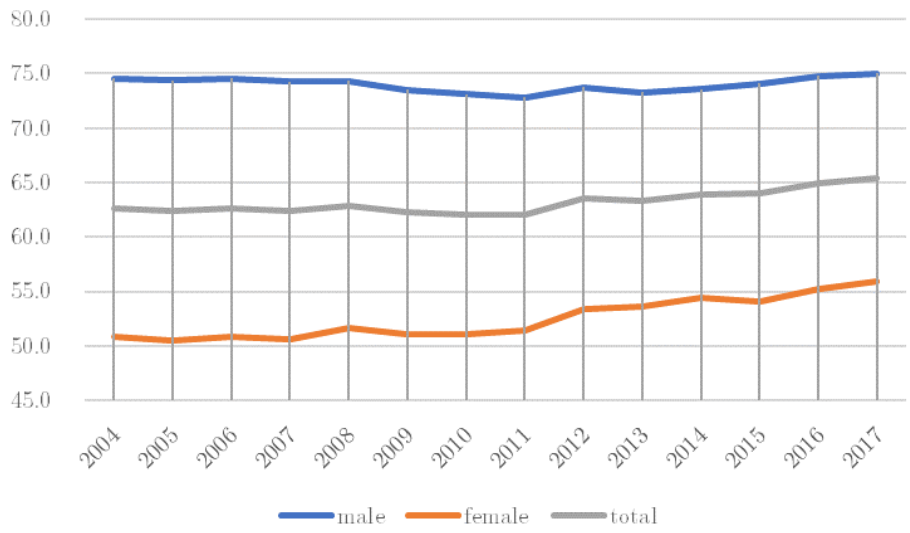


Data source: ISFOL PLUS Questionnaire

5 Heckman correction for the sample selection bias

The data on the Italian Economic Activity Rate (EAR), provided by the Italian National Institute of Statistics (ISTAT), shows a considerable gap in the labor force participation rate between male and female. As Figure 3 shows, the total EAR between 2004 and 2007 has remained stable at around 62%, while the male EAR corresponded to 74% and the female one just to 62%. In 2008, the total EAR increased slightly, as for male and female, but the next year all the trends of the labor force participation rate have seen a negative inflection, maybe because of the financial crisis. This negative trend has lasted until 2012, when the total EAR has jumped up to 63%. During that year, the female EAR has experienced a similar increase, reaching 53%, while the male EAR has increased fewer. Between 2012 and 2017, the EAR trend has been positive both at the aggregate level as well as by gender. In 2017, the total labor force participation rate was 65%, the male one was 75%, and the female one was almost equal to 56%.

Figure 3: Economic Activity Rate (%) Trend, 2004-2017
The data shows a considerable gap in the EAR between male and female, which has remained steady regardless the positive or negative trend of the overall indicator.

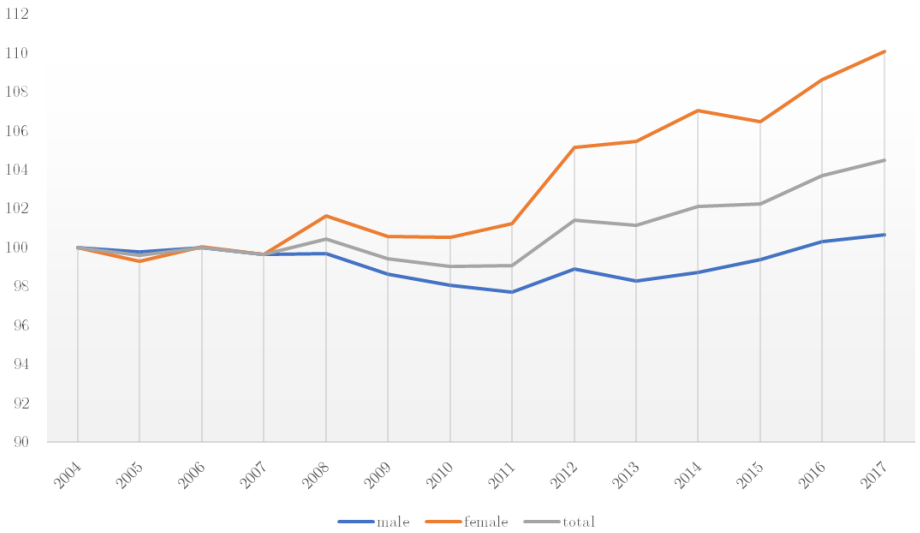


Data source: ISTAT, Dataset Code: DCCV_TAXATVT1

During the past years, we have seen a slow convergence between the male and female economic activity rate. To better underline the different trends, is useful to look at the indexing number for EAR in Figure 4, where 2004 = 100 is the indexed year. Up to 2007, the two EARs were almost identical, but after that year the labor force participation rate of women has jumped, while the male one has slowed down. The negative tendency, which has lately interested female EAR too, has continued until 2012 when both indicators have increased. In particular, the female EAR has scored +5% with respect to the base year. This

positive trend lasts still today, with a slight inflection of women labor force participation rate in 2015.

Figure 4: Economic Activity Rate, 2004 = 100
Starting from 2007, the female EAR has increase, with a slightly negative inflection. The positive trend has lately interested the male EAR too.



Data source: ISTAT, Dataset Code: DCCV_TAXATVT1

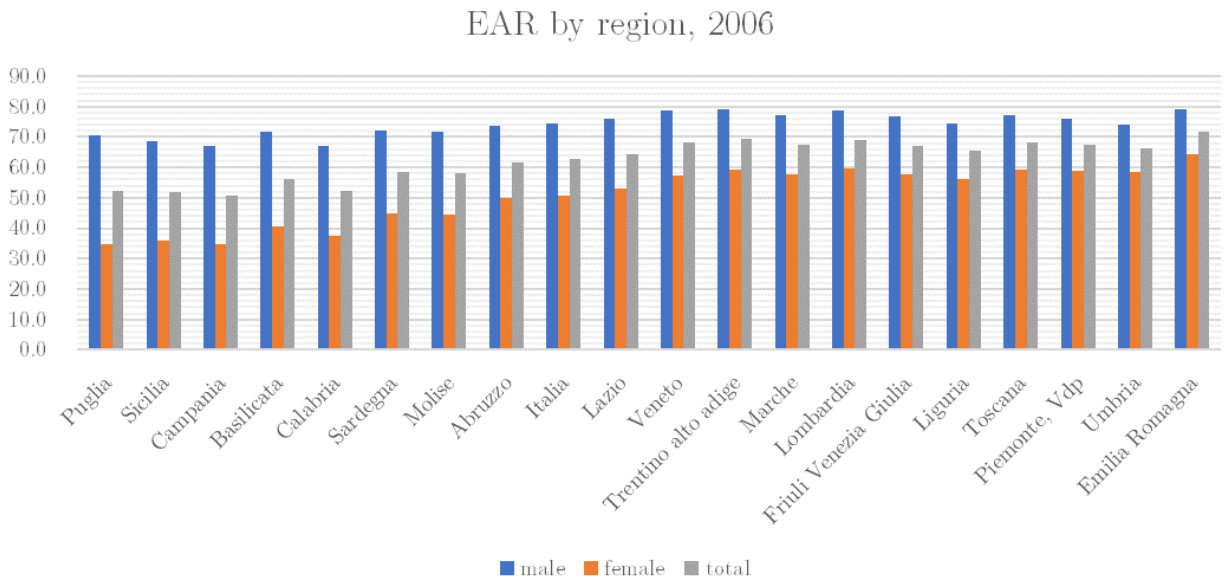
Eventually, the female participation in the labor market in 2017 was +10% than in 2004, while the male one was just +4%. However, despite those improvements, the percentage of women involved in the labor forces is still meager compared to the one of men.

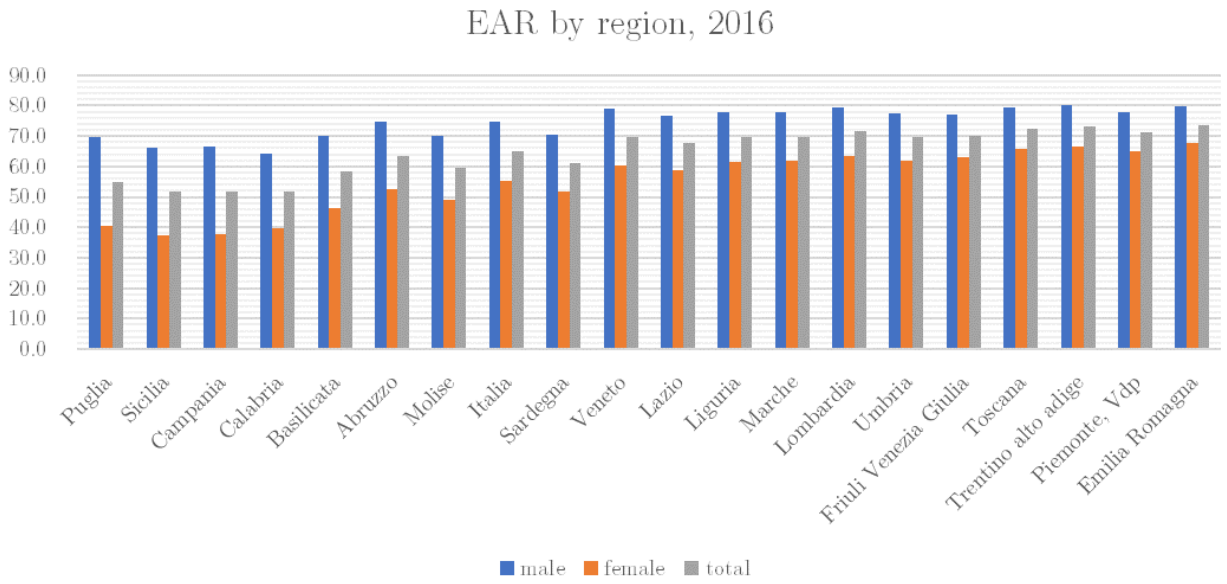
We find an analogous situation considering the economic activity rate by region. Figure 5 shows the EAR divided by region, for the year 2006 and 2016 respectively. In 2006, the minimum gap between male and female EAR was 15 percentage points in Emilia-Romagna, while the maximum was 36 percentage points in Puglia. However, most of the regions presented an EAR gap among gender well above 20 pp, with Puglia, Sicilia, Campania and Basilicata worst performers with a gap higher than 30 pp. In 2016 the situation has improved, with the minimum EAR gap equal to 12 pp in Emilia-Romagna and the maximum gap equal to 30 pp in Puglia. The general improvements in the labor force participation may be due to the progressive adoption of the European Directive on equal treatment between women and men in employment and occupation (Directive 2006/54/EC).[4]

The different labor force participation among genders may give rise to a sample selection bias. This bias arises when non-observed factors, that influence the likelihood of participation

[4]The Commission launched infringement proceedings against Italy in March 2007 for incorrectly transposing Directive 2006/54/EC. Following the Commission's intervention, Italy adopted a new law amending its existing legislation in June 2008. This brought national law into line with the Directive on the issues raised by the Commission.

Figure 5: Economic Activity Rate by region, 2006 & 2016
In 2006, most of the regions present an EAR gap among gender well above 20 percentage points, with Puglia, Sicilia, Campania and Basilicata worst performers. In 2016 the situation has slighly improved across the Italian regions.





Data source: ISTAT, Dataset Code: DCCV_TAXATVT1

of individuals in the labor market, are correlated with non-observed characteristics affecting wages. Under such circumstances, the assumptions to ensure the consistency of the estimated coefficients of wage equations are not met, and this results in wrong conclusions about the degree of wage discrimination against women.

We implement a two-step estimation method to correct for sample selection bias, as proposed by Heckman (1979). In the first stage, we define the participation equation, where individual decision to participate or not to the labor market depends on a set of personal characteristics. We include in the set of covariates the age of the respective individuals, if he/she is married or not and if he/she has one or more children, the family size and educational features, such as the years of schooling and if the respective individual has graduated with honors. Through the estimation, we construct the variable λ and we include it as an additional regressor in the second stage, where the wage equation is estimated. In the second stage, we calculate the Inverse Mills Ratio and we include it in the basic wage equation. The Inverse Mills Ratio of a distribution measures the probability density function over the cumulative distribution function. When the coefficient of the Inverse Mills Ratio is positive, it is said that "positive selection" occurs, with "negative selection" otherwise. Positive selection means that, without the correction, the model estimates are upward-biased, while negative selection results when model estimates are downward-biased. Finally, we obtain consistent regression estimates by estimating the wage equation corrected for selectivity bias.

The interpretation of the coefficient of the Inverse Mills Ratio depends on the authors' choice to model selection (e.g., Heckman) or to model non-selection (e.g., Maddala, 1983). We follow the non-selection model interpretation of the coefficient of lamda, therefore a positive coefficient means that the models estimates are downward-biased and a negative coefficient means that the model estimates are upward-biased. If follows that the meanings of positive selection and of negative selection are reversed.

Table 5 shows the coefficients of lambda, distinguishing between men and women (Model1 and Model2, respectively) for each region and Italy as a whole. At the national level, the male Inverse Mill Ratio coefficient is positive significant. This means that the model estimates for men are downward-biased. The opposite occurs for women: their Inverse Mill Ratio coefficient is negative significant. Consequently, female model estimates are upward-biased. An explanation is that, even if the female labor force participation is lower with respect to the male one, the few women on the labor market have, on average, higher productivity than men.

We find a quite varied situation along with the Italian territory: in Abruzzo, Puglia, Sardegna, and Basilicata male lambda indicates that there is positive selection, so the model estimates for men are downward-biased. On the contrary, the negative significant Inverse Mill Ratio indicates that there is negative selection in Umbria with respect to men. Nevertheless, in more than half of the Italian regions, there is negative selection for women. This is the case for Piemonte and Val d'Aosta, Lombardia, Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia, Emilia-Romagna, Toscana, Marche, Umbria, Basilicata, and Sicilia. In those regions, the lower female economic activity rate affects the wages of women, and in particular the model estimates for women are upward-biased.

Table 5: Oaxaca-Blinder Decomposition by region with Heckman correction, *Lambda* (Inverse Mills Ratio) - Panel Model1 is referred to men, while Model2 is referred to Women. At the national level, the male Inverse Mill Ratio coefficient is positive significant. This means that the model estimates for men are downward-biased. The opposite occurs for women: their Inverse Mill Ratio coefficient is negative significant. Consequently, female model estimates are upward-biased. This is also the case in more than half of the Italian regions. This means that, in these regions, the lower female economic activity rate affects the wages of women, and in particular the model estimates for women are upward-biased.

	Piemonte + VdA	Liguria	Lombardia	Trentino-Alto Adige	Veneto	Friuli-Venezia Giulia
lambda - Model1	0.013	0.094	0.067	0.087	0.045	0.112
	(0.080)	(0.115)	(0.054)	(0.148)	(0.087)	(0.093)
lambda - Model2	-0.106*	-0.048	-0.132***	-0.237***	-0.262***	-0.175***
	(0.061)	(0.059)	(0.036)	(0.065)	(0.050)	(0.064)

	Emilia-Romagna	Toscana	Marche	Umbria	Lazio	Molise	Abruzzo	Campania	Puglia
lambda - Model1	0.004	0.056	0.020	-0.441**	0.026	0.064	0.023**	-0.008	0.207**
	(0.117)	(0.070)	(0.114)	(0.190)	(0.087)	(0.186)	(0.110)	(0.055)	(0.840)
lambda - Model2	-0.203***	-0.199***	-0.204***	-0.233***	-0.079	0.050	-0.069	0.067	-0.042
	(0.067)	(0.055)	(0.072)	(0.075)	(0.055)	(0.080)	(0.071)	(0.045)	(0.046)

	Basilicata	Calabria	Sicilia	Sardegna	Italy
lambda - Model1	0.338***	0.038	0.076	0.224**	0.086***
	(0.112)	(0.078)	(0.071)	(0.087)	(0.020)
lambda - Model2	-0.137*	0.085	-0.090*	0.077	-0.124***
	(0.070)	(0.060)	(0.054)	(0.068)	(0.013)

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

6 Conclusion

This paper addresses the wage differentials among Italian regions. From a preliminary analysis based on simple OLS regressions on a set of personal and labor characteristics, we find that, both at the national level and at the regional level, being a woman significantly reduces earnings. Wages are negatively affected also by some work-related characteristics, such as being employed in the private sector or having a temporary contract. Moreover, middle-low skill works lead to lower wages, especially in the Centre and in the South of Italy. Finally, working in family services and education reduce salaries. However, we also find that some characteristics positively affect wages, such as age or living in large cities. The latter feature is particularly evident in large regions, while small regions show the opposite trend. Moreover, investing in education is a win-win choice: the variable measuring years of schooling is positive in all regions and in Italy as a whole, whereas graduating with honors positively affect wages in 3/5 of Italian regions, with most of the exceptions located in the North. However, in the same geographic area, be employed in senior positions or scientific professions positively affects salaries.

The Oaxaca-Blinder decomposition results shows that in Italy the GPG, expressed in logarithmic scale, is equal to 26.3%, among which 0.76% is explained and 98.86% remains unexplained. At the regional level, the GPG goes from 22.9% in Basilicata up to 32.9% in Toscana, while the explained part goes from -12% in Toscana up to 11.79% in Basilicata and the unexplained part goes from 88.21% in Basilicata up to 112.20% in Lazio. The characteristics which drive the explained part of the wages differential at the national level are confirmed for each region. Men who are elder, have temporary contracts and do overtime job possess, on average, better characteristics than women in the labor market. Moreover, the results regarding the sectors of employment suggest the presence of gender segregation: men work in better paid NACE sectors compared to women. Nevertheless, our findings show that women, on average, have better educational level than men in the labor market. In particular, this is the case in half of the regions for women graduated with honors.

Looking at our results from the Oaxaca-Blinder decomposition, we check for the participation in the labor market, and we find that the female economic activity rate, even if it has followed a positive tendency towards the male one, remains significantly low. This fact can imply a sample selection problem which, in turn, can lead to non-consistent estimated coefficients of the wage equations. For this reason, the Heckman correction has been implemented as a further refinement of our analysis. We find that there is negative selection at the national level and in half of the Italian regions. This means that women model estimates are upward-biased.

All our findings confirm the extant literature and enrich the analysis of inter-regional wage differentials in Italy. Even if we cannot be sure that the unexplained part of the gender pay gap implies gender discrimination with respect to women, wages differential exists and varies across all the Italian regions. Multiple initiatives to lower the GPG have been fostered at the Italian level, all implementing European Directives aimed at promoting gender equality. Within this field, the European policies are many and wide-ranging. In the European Commission's Strategy on Gender Equality 2010-2015, the reconciliation of work,

family and private life (known as "work-life balance principle") was recognized as a key objective. To pursue this objective, reconciliation policies have been introduced, such as family and parental leave schemes, care arrangements for children as well as the development of a working environment structure and organization. To tackle down the GPG, the European Commission has adopted the Action Plan 2018-2019, which includes a broad and coherent set of activities, mutually reinforcing each other. Besides, the Action Plan is accompanied by an Evaluating Report of the Pay Transparency Recommendation, which the Commission adopted in 2014. Pay transparency is a key lever in bringing gender pay differentials within companies to light. Within this field, Italy has become a reference case on the topic of gender quotas worldwide. In fact, since the approval of the 'Golfo–Mosca' law (2011), which imposes a balanced representation for men and women in listed and state-owned Italian companies, the female representation in the boards of directors has increased from 7% (2011) to the actual 33% (Profeta et al., 2014). Moreover, gender quotas in boards may influence the overall selection process, leading to a better quality of board members and, in turn, of corporate governance. However, much more should be done across the spectrum of jobs to improve gender equality.

To conclude, our results underline the relevance of the issue of gender equality for Italy. Most of the current research has investigated the topic at a macro level, while our analysis has been done by means of granular data at regional level. Our study highlights the need to go deeply into the question, and that, despite the several initiatives undertaken, there is still room for improving gender equality in Italy. Our findings also leave space for further refinements of the results at provincial level within the large regions, an additional analysis that could help explaining better the reasons behind the gender wage differentials.

A Definition of Variables

Table 6: Definition of Variables

Variable Name	Definition
	Dependent Variables
montlywage	Monthly net wages
lnwage	The natural log of monthly net wages
lfp	One if the respective individual is inside labour forces, zero otherwise
	Independent Variables
ydate	Years, including 2005, 2006, 2008, 2010, 2011, 2014 and 2016
	Personal characteristics
female	One if the respective individual is a woman, zero otherwise
age	Age of the respective individual (in years)
married	One if the respective individual is married, zero otherwise
fam_size	Number of relatives in the respective individual's family
child	One if the respective individual as at least one child, zero otherwise
qualification	Qualification of the respective individual
schooling	Number of years of schooling completed
honors	One if the maximum degree mark was attained, i.e. *110 e lode*, in case of graduation from university, zero otherwise
homeowner	One if the respective individual owns a house, zero otherwise
it	One if the respective individual holds the Italian citizenship, zero otherwise
city_inhab	Number of city inhabitants
city_size1 - city_size3	City dimension dummies based on the number of inhabitants, for small, medium and large, respectively
region	Region where the respective individual lives
area	Geographical area where the respective individual lives
	Labour characteristics
private	One if the respective individual works in a private firm, zero otherwise
exper	Number of years of work experience
exper2	*exper* squared
temporary	One if the respective individual has temporary contract, zero otherwise

Variable Name	Definition
profession	Profession of the respective individual
prof_senp - prof_milf	Dummies by profession in senior positions, scientific jobs, technical jobs, office work, commerce, skilled work, semi-skill work, unskilled work, military forces, respectively
sector	Sector of employment
sect_agri - sect_inte	Sectorial dummies for employment in agriculture, industry, energy, construction, commerce, tourism, transport, communication, financial activities, services for firms, public administration, education, health, scientific activities, family services, international, respectively
overtime	One if the respective individual does overtime work, zero otherwise
condition	One if the respective individual is occupied, zero otherwise

References

[1] *Adjusted gender pay gap*. Tech. rep. Eurostat, 2018. URL: https://circabc.europa.eu.

[2] F. D. Blau and L. M. Kahn. "The U.S. Gender Pay Gap in the 1990s: Slowing Convergence". In: *Industrial and Labor Relations Review* 60.1 (2006), pp. 45–66. URL: http://www.jstor.org/stable/25067574.

[3] F. D. Blau and L. M. Kahn. "Changes in the Labor Supply Behavior of Married Women: 1980-2000". In: *Journal of Labor Economics* 25.3 (2007), pp. 393–438. URL: http://www.jstor.org/stable/10.1086/513416.

[4] F. D. Blau and L. M. Kahn. "The Gender Wage Gap: Extent, Trends, and Explanations". In: *Journal of Economic Literature* 55 (2017), pp. 789–865. URL: http://www.nber.org/papers/w21913.

[5] F. D. Blau, Ferber M., and A. Winkles. *The Economics of Women, Men and Work*. Pearson, 2006.

[6] A. S. Blinder. "Wage Discrimination: Reduced Form and Structural Estimates". In: *The Journal of Human Resources* 8.4 (1973), pp. 436–455. URL: http://www.jstor.org/stable/144855.

[7] Monica Bozzano. "Assessing Gender Inequality among Italian Regions: The Italian Gender Gap Index". In: *Rivista di Politica Economica* 103.1 (2014), pp. 255–300. URL: https://ideas.repec.org/a/rpo/ripoec/y2014i1p255-300.html.

[8] C. Castagnetti, L. Rosti, and M. Töpfer. *Discriminate Me – if You Can! The Disappearance of the Gender Pay Gap among Public-Contest Selected Employees*. Tech. rep. 158. University of Pavia, Department of Economics and Management, 2018. URL: https://ideas.repec.org/p/pav/demwpp/demwp0158.html.

[9] Y. Chzhen and K. Mumford. "Gender gaps across the earnings distribution for full-time employees in Britain: Allowing for sample selection". In: *Labour Economics* 18.6 (2011), pp. 837–844. URL: http://www.sciencedirect.com/science/article/pii/S0927537111000595.

[10] *Commission steps up effort to tackle down the gender pay gap*. Tech. rep. European Commission, 2017. URL: http://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=607452.

[11] *Gender Equality Glossary*. Tech. rep. Gender Equality Commission of the Council of Europe, 2015. URL: https://edoc.coe.int/en/gender-equality/6947-gender-equality-glossary.html.

[12] J. Heckman. "Sample Selection Bias as a Specification Error". In: *Econometrica* 47.1 (1979), pp. 153–161. URL: http://www.jstor.org/stable/1912352.

[13] B. Jann. "The Blinder-Oaxaca decomposition for linear regression models". In: *Stata Journal* 8.4 (2008), 453–479(27). URL: http://www.stata-journal.com/article.html?article=st0151.

[14] G. S. Maddala. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, 1983.

[15] H. Mandel and M. Semyonov. "A Welfare State Paradox: State Interventions and Women's Employment Opportunities in 22 Countries". In: *American Journal of Sociology* 111.6 (2006), pp. 1910–1949. URL: https://doi.org/10.1086/499912.

[16] D. Neumark. "Employers' Discriminatory Behavior and the Estimation of Wage Discrimination". In: *The Journal of Human Resources* 23.3 (1988), pp. 279–295. URL: http://www.jstor.org/stable/145830.

[17] R. Oaxaca. "Male-Female Wage Differentials in Urban Labor Markets". In: *International Economic Review* 14.3 (1973), pp. 693–709. URL: http://www.jstor.org/stable/2525981.

[18] P. Profeta et al. *Women Directors: The Italian Way and Beyond*. Palgrave Macmillan UK, 2014. URL: http://gen.lib.rus.ec/book/index.php?md5=1B07CFB%20385AF3B9231D08E054D06746.

[19] *Reconciliation of work, family and private life in the European Union: Policy review*. Tech. rep. European Institute for Gender Equality, 2015. URL: http://eige.europa.eu/rdc/eige-publications/reconciliation-work-family-and-private-life-european-union-policy-review.

[20] *Tackling the gender pay gap in the European Union*. Tech. rep. European Commission, 2013. URL: http://eige.europa.eu/rdc/thesaurus/terms/1196.

[21] G. Toniolo. *An Economic History of Liberal Italy (Routledge Revivals): 1850-1918*. Routledge Revivals. Routledge, 2016. URL: https://books.google.lv/books?id=LxLejwEACAAJ.