Italian Economic Journal
https://doi.org/10.1007/s40797-025-00316-7

RESEARCH PAPER

The Gender Wage Gap: Evidence from Organisations in Northern Italy

Chiara Tasselli[1] · **Tindara Addabbo**[1]

Received: 29 January 2024 / Accepted: 5 February 2025

Abstract

The gender wage gap in Italy is persistent and, considering the low and selected female participation in the labour market, appears to be much higher than the unadjusted gender wage gap shown by Eurostat data. The data referring to the individual, using statistical sources that collect wages together with individual and family characteristics, allow the wage differential to be corrected for the non-random selection of women in the labour market. However, these data cannot delve deeper into the analysis of the determinants at the corporate level. With the intent to identify the determinants at the firm, or company, level, we conducted an analysis based on company microdata from organisations located in Northern Italy that voluntarily underwent a gender equality analysis and certification. Our estimates are based on individual-level data, while incorporating select corporate-level variables to account for firm-specific characteristics related to cultural and organisational policies and practices. Econometric analyses confirm the presence of a gender wage gap in favour of men, together with the positive impact of age, seniority, and education on wages. Turning to the firm's fixed effects: formalised policies on part-time work and time flexibility can help reduce the gender pay gap but when a high percentage of people work part-time, this can widen the pay gap. In addition, professional development policies in a firm play a significant role in narrowing the gap.

Keywords Gender equality · Wage differentials · Labour market discrimination · Firm-level determinants · Oaxaca decomposition

JEL Classification J31 · J71 · J81 · J16

✉ Chiara Tasselli
chiara.tasselli@unimore.it

Tindara Addabbo
tindara.addabbo@unimore.it

[1] Department of Economics Marco Biagi, University of Modena and Reggio Emilia, Modena, Italy

Published online: 08 May 2025

1 Introduction

Extensive research has been done on the gender gap in earnings and its negative impact on the status of women in their economic and social lives. The gender pay gap has been related to interconnected demand and supply factors, such as a different allocation of care responsibilities and working time within the household, discontinuous career paths, vertical segregation (sticky floor and glass ceiling), contract types, etc. Discontinuous working profiles are related to the unequal distribution of care work and responsibilities found in heterosexual couples. Women still bear a higher load of care work and are more likely to interrupt their careers or reduce their commitments in the presence of children.

Lower wages have also been associated with other pre-labour market factors, such as the horizontal segregation in education, which sees women clustered in low-paying fields. The relevance of this dynamic, in addition to a woman's lower, future, professional prospects, lies in the fact that the choices made by both sexes at a young age are not completely free but are, instead, gender biased.

The scope of this paper is to understand to what extent the gender pay gap is related to a company's specific characteristics by using data from administrative sources on a sample of heterogeneous organisations that submitted themselves to analysis to reach a certification on gender equality. Due to the nature of this corporate self-selection, it is reasonable to assume that our sample leads to an underestimation of the actual gender wage gap compared to the population of organisations. However, notwithstanding the sub-sample self-selection, a non-negligible gender pay gap can be observed.

The data source chosen allows us to contribute to the existing literature in three ways. First, it allows us to analyse the effects of a set of corporate policies and the extent to which they formalise the gender pay gap in Northern Italy, leading to policy suggestions at the corporate level for that geographical area. Second, it allows us to tackle the issue of attracting and keeping women in the labour force in a country, Italy, characterised by low female employment and a high rate of women leaving employment after childbirth. The Blinder-Oaxaca decomposition allows us to investigate whether this difference is mainly due to gender differences in a worker's actual characteristics or whether, instead, it may be due to the differential effects of characteristics or to unobserved factors that were not considered in the model. This can also lead to disentangling heterogeneity in a company's remuneration policies, which results in both direct discrimination and other employer-specific mechanisms driving the gender wage gap as returns in workers' characteristics. Third, this data source can help us understand the role that organisational policies and practices play in creating and perpetuating gender inequality in the workplace, beyond just the gender pay gap. By examining the presence and trajectory of organisational policies related to issues such as work-life balance, promotion, and career opportunities, as well as policies to prevent discrimination, we can gain insights into the factors that contribute to gender disparities in employment and develop evidence-based recommendations for promoting gender equality in the workplace.

2 Literature Review

The gender wage gap has been found to persist over time (Zizza 2013; Mussida and Picchio 2014; Blau and Kahn 2017) notwithstanding increases in women's education and in their participation rates in the labour market.

Moreover, the gender wage gap is higher when corrected for the non-random selection into employment (as found by, amongst others, Olivetti and Petrongolo 2008; Addabbo 2018; McKay and Mussida 2018). This is related to a woman's lower participation in labour markets and differences in the characteristics of employed and not employed women.

Wage inequalities are also linked to horizontal and vertical segregation (Blau and Kahn 2017). Centra and Cutillo (2009) address the problem of endogeneity related to the segregation of women into female jobs/sectors by the estimation of a bivariate selection model that controls for the selection into employment and the selection in female occupations on Isfol wage differential data for Italy. They define female occupations (in terms of employment sector and profession) as those where the percentage of female workers is higher than 59%. By correcting for the female employment selection, the part of the Oaxaca-Blinder decomposition related to the coefficients of the variables introduced in the model is still higher than the one due to differences in the characteristics, but is lower than the case when there is no control for the double selection.

Busch and Holst (2011) found that working in a more 'feminised' workplace has a negative effect on women's wages, including for women in senior positions. They provide models with fixed effects on wage differentials by considering the probability of access to senior positions on the German Socio-Economic Panel Study (GSOEP) data.

Busch (2020), by using GSOEP data (1992–2015), finds evidence of an increase in the wage gap between women working in male-type occupations and women working in other occupations linked to the rise in wages of women who work in male-type occupations and a decline in discrimination in traditionally male occupations in the last two decades in Germany.

Addison et al. (2018) use longitudinal Current Population Survey data to analyse the impact of gender composition on earnings and find that its impact is reduced when they control for observed heterogeneity and even more so when they control for unobserved heterogeneity and detect, when using synthetic panels of ageing cohorts, larger wage penalties for younger cohorts in predominantly female-type occupations.

In analysing the unexplained part of the gender wage gap, Goldin (2014) highlights the role of different, non-monetary characteristics in the workplace:

> "As women have increased their productivity enhancing characteristics and as they "look" more like men, the human capital part of the wage difference has been squeezed out. What remains is largely how firms reward individuals who differ in their desire for various amenities. These amenities are various aspects of workplace flexibility." (Goldin 2014, p. 1094).

Goldin (2014) then demonstrates that the unexplained differences by occupation in earnings are related to the value that organisations place on hours and job continuity, and that the occupations with higher elasticities in annual earnings, with respect to weekly hours, have greater negative log earnings gender gaps. In these occupations, those who need to work fewer hours or those who require time flexibility (more likely to be female workers), will experience a high wage penalty. In short, the rewards for longer work hours are higher and time flexibility is penalised.

Bertrand (2018) applies this framework to her study of the persistent glass ceiling phenomenon on U.S. data, whilst also taking into account other possible explanations (gender differences in fields of education, psychological attributes, such as risk aversion, and attitudes toward negotiation and competition). As she explains:

> "Many of the higher-paying jobs in the economy involve long hours and inflexible schedules. Also, those financially more rewarding careers require continuous labour force attachment in order to stay on the 'fast track', which makes it difficult to combine those careers with job interruptions. Because women remain the dominant providers of childcare as well as of other forms of non-market work, these various job features might be particularly detrimental to them." (Bertrand 2018, p.215).

The residual gender pay gap can then be correlated to the difference in demand for flexibility by gender in higher paying jobs.

Destefanis et al. (2024) apply Goldin's framework (already applied with reference to the U.S. labour market) to the Italian labour market, taking into account the differences between the two labour markets and detecting cases where the framework is well suited to explaining a higher residual gender gap in wages (as in the higher parts of the wage distribution characterised by a lower weight of union coverage and unionisation). Destefanis et al. (2024) use the Structure of Earnings Survey (collecting both employee earnings and individual characteristics and employer data) for Italy, matching it with the Italian sample survey on professions and the Occupational Information Network survey and focusing on a within-occupation gender gap. Their estimation provides results which are consistent with Goldin's (2014) hypothesis, not only for higher segments of the Italian labour market but also for lower ones, and confirms that higher elasticity occupations are characterised by wider gender pay gaps.

Casarico and Lattanzio (2024) use a wide employer-employee dataset on private sector companies and employees in Italy from 1995 to 2015 to estimate the impact of the firms' heterogeneity on the gender wage gap. The results of their work show that about 34% of the mean gender wage gap can be explained by differences in the company's remuneration policy, the majority of it being due to the between-firm component. Moreover, Casarico and Lattanzio (2024) are able to disentangle the within or between firms impact on the gender pay gap at different levels of the pay distribution. They find that the within-firm components play a major role at the top of the pay distribution with an increasing impact over time.

Employer-employee datasets have, more recently, also been used to better account for the impact on the gender pay gap of corporate heterogeneity in terms of company-specific pay policies and premiums. By using the Portuguese Ministry of Employment census of private sector employee data matched with longitudinal data on the hourly wages of Portuguese workers and firms' data, Card et al. (2016) found that sorting

(2/3) and bargaining power (1/3) channels explain approximately 20% of the gender wage gap. Women are more likely to work in companies which pay small premiums.

By estimating wage regressions of high dimensional fixed effects and applying Gelbach's decomposition method to the longitudinally-matched employer-employee job title dataset from Portugal, Cardoso et al. (2016) show that one fifth of the conditional gender wage gap for workers with the same general labour market experience and the same seniority in the company can be attributed to their being allocated to companies of different quality; another one fifth can be allocated to jobs of different quality. They also show that it is a woman's lower access to higher paying companies, rather than worker allocation to jobs, that explains the observed glass ceiling effect.

Morchio and Moser (2024) use employer-employee data from Brazil and find a large gender pay gap due to women working for lower paying employers. These employers offer better characteristics other than wages, with compensating differentials explaining half of the gender pay gap.

Attention has also been paid to the impact of parenthood in the interaction between workplaces and job characteristics, related or not to pay on different career paths by gender (Hotz et al. 2018). Hotz et al. (2018) use a Swedish large employer-employee longitudinal dataset showing that the differences in the attributes of workplaces and jobs in terms of the management of the workplace, skills and gender composition of the workforce and other non-wage characteristics play a crucial role in explaining the gender gaps in early careers. The reduction in the motherhood penalty is found to be determined more by the positive impact of family-friendly policies in making it possible for mothers to work more hours than on improving wages with respect to non-mothers. However, by moving towards more family-friendly workplaces, mothers are found to lose out in the skill content of their jobs, with a negative impact on the wage gap over their life.

The persistence of wage inequalities, including those based on gender and occupation, points to deeper structural problems not only in labour markets. Addressing these problems requires comprehensive policies that consider both individual and organisational factors, also in order to promote a better intra-family balance.

3 Data and Descriptive Statistics

We used primary data to analyse the gender wage gap. The data came from administrative sources and twenty-five legal entities in Northern Italy. The sample concerned 34,634 employees, collected from late 2019 to the beginning of 2024. The companies are heterogeneous in size, industry (NACE), and location and provided their data by submitting themselves to a gender equality analysis and certification. The voluntary nature of participation and a reduced representation of all the types of firms located in the area can lead to non-random selection of the observed sample. A more in-depth discussion of the sample characteristics and the nature of self-selection can be found in Appendix A. It is important to note that this sample does not represent Northern Italian firms in terms of size or commitment to gender equality. Although the weighting procedure employed, described in Appendix A, cannot solve the issues related to self-selection, the data can be considered, thanks to the weighting procedure, representative of the working population of the area.

As the data collection was designed to assess gender equality within a company, we can also control corporate policies and ratings regarding other gender equality aspects not exclusively related to wages. Moreover, this administrative source allows us to analyse the entire population within and between companies.

In the data cleaning procedure, to normalise the database, we excluded workers whose contracted hours were less than 50%[1] of the hours specified in a standard full-time contract and whose hourly wage was below 7 euros.

In addition, by following the common trend in the literature (Blau and Kahn 2017), which deals with similar datasets, we also excluded workers over 65 years old, workers with a temporary agency contract and all employees whose working hours or whose income could not be obtained (self-employed workers, collaborators).

Our main dependent variable is the log of the average hourly earnings (*lnYhour*). Annual earnings (*Income*) were calculated by adding the variable income items to the gross salary.

We controlled for worker characteristics, such as:

(i) Gender (*female*).
(ii) Worker *age* and Company *seniority.*
(iii) Whether the employee has a *permanent* contract.
(iv) Position (*Blue collar, Employee, Middle manager, Manager*).
(v) Employee working pattern: whether they are a "smart worker" (*Smart_worker*) and whether they work part-time (*partTime_worker*) (defined, in our context, as those who work up to 60% of the weekly hours specified in a standard work contract).
(vi) The level of feminisation of the occupation[2] in which the worker is placed (*High_fem*).

Moreover, we control for the observable corporate fixed effects (at the legal entity level) as:

(i) the age in terms of years of the firm (*FirmAge*).
(ii) the logarithmic form of the company revenue (*Ln_revenue*).

Finally, we also added as corporate fixed effects other variables related to the corporate composition at the top and formalised work-life balance policies:

(i) Duncan and Duncan (1955)[3] segregation index calculated by corporate organisational areas and levels/grades in the profession (*Duncan*)[4].
(ii) percentage of female managers (*%F_Manager*).

[1] The threshold focuses on the part-time percentage rather than absolute values (e.g., 20 h per week) because there are different collective bargaining agreements (CCNL) and a full-time contract does not always correspond to 40 h.

[2] Occupations were identified based on the intersection of a worker's classification and the economic activity sector in the INPS 2022 open access data, as done by Centra and Cutillo (2009).

[3] $1/2 \cdot \Sigma \left| \frac{F_i}{F} - \frac{M_i}{M} \right|$; Where F_i and M_i are the percentages of the employed women and men in the i-th major occupation group relative to their respective totals.

[4] To avoid collinearity but to exploit all the possible information, we enter the interaction between the Duncan index computed on horizontal segregation and the vertical one.

Table 1 Inferential statistics: Individual characteristics and corporate allocation by gender

	(1) Female		(2) Male		(3) t-test ($\overline{x_f} - \overline{x_m}$)	
	Mean	Linearized sd	Mean	Linearized sd	diff.	t
Income	31,476.87	306.61	39,116.52	471.49	− 7639.65***	(− 13.10)
Yhour	16.59	0.15	19.09	0.23	− 2.5***	(− 8.95)
Age	43.50	0.21	42.05	0.24	1.45***	(4.43)
Seniority	9.36	0.16	10.33	0.16	− 0.97***	(− 3.90)
Permanent	0.90	0.01	0.85	0.01	0.05***	(4.48)
Blue collar	0.39	0.01	0.47	0.01	− 0.08***	(− 7.22)
White collar	0.55	0.01	0.42	0.01	0.13***	(10.23)
Middle man	0.05	0.00	0.07	0.00	− 0.02***	(− 4.50)
Manager	0.01	0.00	0.03	0.00	− 0.02***	(− 5.01)
Smart_worker	0.36	0.01	0.27	0.01	0.09***	(− 8.00)
Part-T_work	0.15	0.01	0.02	0.00	0.13***	(31.96)
N	16,338		18,296		34,634	

For the analysis, IPF weights were applied to make the sample representative of the labour force in the geographical area of Northern Italy

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

(iii) percentage of female middle managers (*%F_Middle_Man.*).

(iv) the company level of inclusive cultural practices (*Cultural_Practices*): a continuous variable created from the fuzzy aggregation of the following indicators: the existence of a whistleblowing system, the existence of a corporate gender equity coordination/body; the existence of climate/satisfaction/well-being surveys from a gender perspective; the existence of a gender balance report; the existence of a budget for initiatives supporting a culture of gender equality.

(v) a continuous variable that measures the flexible working hours in terms of entry/exit times and distribution of hours (excluding time banking) in the company: *Flex_hours*.

(vi) a continuous variable that measures the average percentage of overtime in the company:*Average_corporate_overtime.*

(vii) percentage of part-time workers (*%_PT_worker*).

(viii) percentage of smart workers (*%_Smart_worker*).

(ix) whether there is a formalised regulation on part-time working *(RegPartTime*).

(x) whether there is a formalised regulation on smart working *(RegSmartW*).

(xi) whether there is a formalised regulation on hourly flexibility *(RegFlexibility*).

(xii) whether there are policies on professional development *(Profes_Dev*).

As corporate fixed effects, these latter ones are constant variables for workers within a company but they change between companies.

Table 6 in the appendix—shows the inferential statistics.[5] Table 1 reproduces the mean of the covariates by gender showing significant differences (and expected signs) for most variables. Men earn an average of 7639.65 euros per year more than women:

[5] Descriptive statistics of our sample are in Table 5 in the appendix.

this is partly due to a greater number of hours worked per week (as many as 15% of women work part-time, while only 2% of men have a part-time job), but a positive difference remains even when we control for the number of hours (in fact, men earn 2.5 euros more per hour worked). Even though women in the sample have a slightly higher percentage of permanent contracts they are positioned at the lower end of the grades, while management positions are male-dominated. Moreover, although women are older, they have less seniority: a probable consequence of a discontinuous career due to motherhood and delayed entry into the workforce caused by their higher level of education.

The descriptive and inferential statistics regarding the firms' determinants can be found in the Tables 8, 9, 10 in Appendix C, along with corresponding comments in Appendix A.

Table 2 reports the bivariate correlation analysis between the gender pay gap (GPG) in its various specifications and firm determinants. Some of the expected findings align

Table 2 Correlation between the unadjusted GPG and firm determinants

	(1) Unadjusted GPG (yearly earnings)	(2) Unadjusted GPG corrected by working position (yearly earnings)	(3) Unadjusted GPG (hourly income)	(4) Unadjusted GPG corrected by working position (hourly income)
%F_company[a]	0.34***	0.22***	0.19***	0.02***
%F_Manager	0.01*	0.06***	0.08***	0.15***
%F_Middle_man	– 0.27***	– 0.15***	– 0.33***	– 0.22***
FirmAge	– 0.19***	– 0.08***	– 0.02***	0.13***
Ln_revenue	0.18***	– 0.03***	0.10***	– 0.14***
Cultural_Practices	0.44***	0.16***	0.42***	0.11***
Flex_hours	0.09***	– 0.12***	0.09***	– 0.12***
Av.corporate_overt	– 0.10***	0.09***	– 0.29***	– 0.13***
%_PT_worker	0.13***	0.25***	– 0.10***	– 0.04***
%_Smart_worker	– 0.37***	– 0.31***	– 0.23***	– 0.11***
RegPartTime	0.22***	0.06***	0.27***	0.11***
RegSmartW	– 0.41***	– 0.00	– 0.44***	– 0.01*
RegFlexibility	– 0.68***	– 0.42***	– 0.52***	– 0.18***
Profes_Dev	– 0.41***	– 0.09***	– 0.48***	– 0.15***

The unadjusted GPG was calculated following the Eurostat definition ($\frac{Earnings_m - Earnings_f}{Earnings_m}$). Thus, negative correlations between a determinant and the GPG indicate that the determinant reduces the GPG within the firm

For the analysis, IPF weights were applied to make the sample representative of the labour force in Northern Italy

$^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

[a]Percentage of women in the company, included here only for the company description; in the regressions, we use the dichotomic individual employment indicator in highly feminised occupations which also accounts for worker position (*High_fem*)

with the literature and empirical evidence, indicating that sectors with higher female representation tend to exhibit a more pronounced gender pay gap. This phenomenon can be supported by two different theories: the first suggests that an increased supply of female labour in these occupations drives down wages. The second theory posits that women choose particular occupations requiring low investments in human capital or investments in human capital that do not depreciate over time, due to existing or anticipated family commitments (Centra and Cutillo 2009).

Professional development policies are negatively correlated to the gender pay gap, as are all formal regulations supporting work-life balance tools (except for part-time regulations which are positively correlated with the GPG). While part-time work can be a means of balancing responsibilities, it often sacrifices hours worked. Nevertheless, it provides a way for women with caregiving responsibilities to remain in the labour market, which they would otherwise leave. Flexible working hours are also negatively correlated with the GPG, particularly when GPGs are adjusted for job positions. Furthermore, a high percentage of women in middle management is negatively correlated with the GPG, while, on the other hand, the percentage of women in senior management is positively correlated with the GPG with a higher coefficient of correlation on adjusted GPG. This correlation can be related to the observed positive correlation between the percentage of women in senior management and the extent to which part-time is regulated. Therefore, companies with a higher percentage of women in management tend to be more inclusive regarding work-life balance, allowing women with caregiving responsibilities to remain in the labour market rather than leaving, which contributes to the observed negative correlation of the percentage of women in senior positions with the GPG. Multivariate analyses will be used to investigate the impact of the observed variables on wages and the GPG.

4 Methodology

The starting point for studying labour income is the Mincerian wage equation (Mincer 1974), which states that the impact of certain characteristics, such as level of education, work experience, and age, should be studied in terms of potential productivity. Greater precision of estimation can be achieved by supplementing this equation with additional information about the job individuals perform and the characteristics of the companies in which they work. Finally, if we expand the Mincerian equation with a dichotomous variable indicating gender, we can obtain an initial quantification of the gender wage gap that is not purely descriptive.

Our estimates are based on individual-level data, while incorporating selected corporate-level variables to account for firm-specific characteristics.

First, we look at the wage differential through regressions and afterwards, we quantify the discrimination component of that differential using the Oaxaca decomposition (Oaxaca 1973).

The regression in our analyses is structured as in Eq. 1, where the subscripts, *i* and *f,* denote, respectively, the individual and the firm, Y_{if} denotes the hourly wage and X_{if} represents a vector of the characteristics of the individual, *i,* belonging to a company or firm, *f.* T_{if} represents a dummy variable indicating whether the worker, *i,* belonging to a firm, *f,* is in a highly feminised occupation. γ_f represents a vector of firm-level characteristics for firm, f. Finally, ε_{if} represents the error term.

$$\log Y_{if} = \alpha + X_{if}\beta_1 + T_{if}\beta_3 + \gamma_f\beta_2 + \varepsilon_{1if} \text{with i } (1, \ldots .34.486);\ \text{f } (1, 25) \tag{1}$$

The level of feminisation of an occupation was created by following the Centra and Cutillo (2009) procedure. We used 2022 data from the Italian Social Security Institute (INPS) and we cross-referenced 96 ATECO (two-digit NACE) sectors with gender and 5 worker classifications (*Blue collar, Employee, Middle manager, Manager*). This procedure allowed us to compute the percentage of females in each occupation. Then we created a dummy variable, indicative of whether the individual's occupation is a typically female-type job or not, identifying the threshold of 62.4% of the percentage of females in the occupation. This value was obtained by multiplying the percentage of female employment in the entire market (about 41.6%) by 1.5, thus inflating the standard line by an additional 50%.

Equation 2 reports the Oaxaca Decomposition formula in its reduced form estimated at the mean values of the variables included in the model[6]:

$$\overline{\Delta \log Y} = (\widehat{\alpha}_m - \widehat{\alpha}_f) + (\widehat{\beta}_m - \widehat{\beta}_f) \cdot \overline{X}_f + \widehat{\beta}_m \cdot (\overline{X}_m - \overline{X}_f) + \varepsilon \tag{2}$$

where $\overline{X}$ indicates the vector of the mean values of the characteristics used, $\widehat{\beta}$ indicates the vector of estimated coefficients, and the subscripts$m$ and f indicate, respectively, the collective of males and the collective of females.

The first part of Eq. 2 (the discriminatory component) is composed by the difference in the wages intercept which corresponds to direct discrimination (gap due exclusively to gender), meanwhile $(\widehat{\beta}_m - \widehat{\beta}_f) \cdot \overline{X}_f$ relates to the possibility of men receiving higher pay for their characteristics.

Finally, the latter part of the equation $\widehat{\beta}_m \cdot (\overline{X}_m - \overline{X}_f)$ is called the "endowment effect", which calculates the component of the wage gap arising from actual differences in the characteristics between genders (such as seniority, contract, company position, …).

It is important to point out that this decomposition is applied to a sub-sample of the population that is not randomly selected as we are omitting everyone who is not employed. However, the higher women's employment rates in Northern Italy than in any other part of the country, in part, reduces the problem, though the different composition by education with regards to the total population would still suggest correcting for the bias. However, the available dataset does not allow any control for the whole sample variables that are crucial to correct for self-selection (like the level of education and presence of children in the family).

To improve the significance of our results we adjusted our dataset, aligning it with the composition of the workforce in the area where the organisations analysed are located by using the Italian labour force survey – quarterly cross-sectional open

[6] To simplify the notation, all characteristic vectors of individuals, companies, and years are collected in X, since they can all be processed as individual identifiers. Subscripts have also been removed to focus on the algebraic gender designation, but the initial specification remains implicit.

access microdata from the National Statistics Institute data for the same reference period. We included only employed individuals and focused exclusively on employees. Additionally, we restricted our analysis to the age group selected in our database and excluded employees working very few hours.

To adjust the sampling weights, we employed the iterative proportional fitting (IPF) algorithm, commonly known as raking, which was initially proposed by Deming and Stephan in 1940 (Deming and Stephan 1940).

Raking has long been used to benchmark sample distributions against external reference distributions (Kalton and Flores-Cervantes 2003). In their 2023 study, Bick et al. (2023) applied iterative proportional fitting to address potential concerns about the representativeness of their sample (Real-time population survey). This technique helped align their data with the U.S. labour force statistics from the Current Population Survey, ensuring that their research on working from home reflected the broader population. Similarly, Sauer et al. (2023) employed the Deming and Stephan weighting scheme to adjust their survey sample to match the German population on variables such as age, gender, education, and region. However, in their case, the weights did not significantly alter the results, as their sample was already representative of the German population. Szusecki et al. (2023) exploit the IPF technique adjusting their sample of 13,104 active workers to represent Hungary's workforce of 3.87 million, ensuring their analysis of workplace behaviour and mental health was grounded in a representative sample.

The applied algorithm enables a gradual adjustment of weights to match known population margins.

In our case, we calculated the known population margins, simultaneously considering these key variables: gender, permanent contract, job classification, and ATECO sector. These values served as a reference in our model to generate new weights. The IPF algorithm continues to adjust the weights until the difference between the weighted margins of the specified variables and the known population margins falls below a specified tolerance level or until a maximum number of iterations is reached. We set this maximum number to 25 as we found it sufficient to achieve congruence in the distributions while generating a substantial amount of information supported by Stata17.

The descriptive statistics on the distribution of characteristics, available upon request, show that the marginal totals of our sample—both with and without weights—are now perfectly aligned with the workforce distribution in the reference regions. This confirms the success of the applied methodology in aligning our sample with the worker population in the area. However, this method does not account for firm characteristics such as size, policies, practices and, more importantly, for the unobserved factors related to the firms' decisions to participate in the certification process.

5 Multivariate Analyses

The gender wage gap is defined by Eurostat as "the difference in average gross hourly earnings between women and men employed in companies with more than 10 employees" and is expressed as a percentage of the corresponding male wage. In our inference sample, by following the Eurostat definition, women earn 87% of a man's hourly wage. The average hourly income for women is €16.59/hour, while for men it is €19.09/hour: with a difference of €2.50/hour, which, according to the t-test, has a high level of significance. From this initial exploratory analysis, it appears that there are gender wage differentials (both considering time spent working, but also considering equal hours worked). The wage gap calculated, however, is a raw statistic and does not consider many other factors that may be simultaneously correlated with gender and wages.

The gender variable (*Female*), alone, is only able to explain around 1.6% of the variation in the hourly wage. By adding working time per week in the regression, we can see that, together with gender, it could explain the 10.6% variation of the logarithm of hourly income. The sign of the coefficient of working time per week is positive and highly significant, which is not obvious since, by having hourly wages as the dependent variable, we are already cleaning up economic income by the number of hours worked. If the employees' number of hours is an important variable explaining earnings we are facing indirect gender discrimination: indeed women are more likely to work fewer hours, taking into account their role as the main caregiver in the family, therefore, they would be the most affected by this dynamic. We then, simultaneously, control for all observable worker characteristics.

Table 3 shows the estimation of Eq. 1 in the methodology section. The regressions are replicated by gender on the whole sample (also including part-timers, columns 1–3) and on employees with a full-time contract (columns 4–6). The adjusted R-square reports a very good fit of the model, indicating that the variables used can explain about 80% of the variations in hourly wages.

Regardless of the weekly hours worked, the gender coefficient is negative and statistically significant (p-value < 0.01). Age and seniority have a positive impact on wages: age has a higher coefficient for the sub-sample of female employees, probably linked to a decrease in care work connected to the presence of young children in the household as women age. Having a permanent contract has a significant and positive influence on hourly wages, especially for women. As expected, improvements in working positions are correlated with increases in wages for both men and women. The dummy variable related to "smart working" is associated with higher wages which could be because the types of work and tasks that can be done in "smart working" usually start at a medium–high level (no manual, blue collar or front office tasks). It is also important to highlight that this effect has a greater impact on men's than on women's wages.

Working on a part-time basis has a negative effect on hourly wages and we note here that the dependent variable of interest is the hourly wage which is already cleaned up by the number of hours worked. The dummy variable *High_fem* shows that being in a highly feminised occupation leads to significantly lower wages (especially for women) and this is in line with previous studies and the *"Annual Labour Force Survey—2020 Results"* published by ISTAT in September 2021 (ISTAT 2021a).

When analysing corporate fixed effects, we found that companies which have policies on professional development demonstrate a significantly positive impact on wages for all groups regardless of gender and working hours. In addition, the implementation of formal regulations on part-time work positively affects the hourly wages of the sample, also including part-timers, but has a negative impact on the wages of full-time

Table 3 Wage equations by gender and hours of work

	All workers (part-time included)			Only full-time workers		
	(1)	(2)	(3)	(4)	(5)	(6)
	All	M	F	All	M	F
Female	– 0.080*** (0.002)			– 0.069*** (0.003)		
Age	0.026*** (0.001)	0.024*** (0.001)	0.033*** (0.001)	0.028*** (0.001)	0.026*** (0.001)	0.037*** (0.001)
Age squared	– 0.000*** (0.000)	– 0.000*** (0.000)	– 0.000*** (0.000)	– 0.000*** (0.000)	– 0.000*** (0.000)	– 0.000*** (0.000)
Seniority	0.003*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	0.002*** (0.000)	0.002*** (0.000)	0.002*** (0.000)
Permanent	0.090*** (0.004)	0.069*** (0.005)	0.115*** (0.005)	0.083*** (0.004)	0.066*** (0.005)	0.109*** (0.006)
White_collar	0.250*** (0.003)	0.252*** (0.004)	0.246*** (0.005)	0.256*** (0.003)	0.252*** (0.004)	0.266*** (0.006)
Middle_manager	0.773*** (0.005)	0.781*** (0.007)	0.761*** (0.008)	0.779*** (0.005)	0.780*** (0.007)	0.770*** (0.009)
Manager	1.489*** (0.008)	1.489*** (0.009)	1.488*** (0.014)	1.490*** (0.008)	1.486*** (0.009)	1.488*** (0.015)
Smart_worker	0.039*** (0.003)	0.038*** (0.005)	0.022*** (0.005)	0.023*** (0.003)	0.030*** (0.005)	0.006 (0.005)
PartTime_worker	– 0.082*** (0.004)	– 0.049*** (0.010)	– 0.079*** (0.004)			
Duncan	– 0.076*** (0.007)	– 0.048*** (0.012)	– 0.078*** (0.009)	– 0.102*** (0.008)	– 0.069*** (0.012)	– 0.131*** (0.011)
%F_manager	0.300** (0.103)	0.387* (0.173)	0.215 (0.120)	– 0.000* (0.000)	– 0.000 (0.000)	0.000 (0.000)
%F_Middle_man	– 0.001*** (0.000)	– 0.001*** (0.000)	– 0.001*** (0.000)	– 0.002*** (0.000)	– 0.003*** (0.000)	– 0.000 (0.000)
FirmAge	– 0.001*** (0.000)	– 0.003*** (0.000)	0.001*** (0.000)	0.002*** (0.000)	0.002*** (0.000)	0.001*** (0.000)
Ln_revenue	0.003*** (0.000)	0.004*** (0.000)	0.002*** (0.000)	– 0.008*** (0.002)	– 0.002 (0.003)	– 0.009*** (0.003)
Cultural_Pract	0.014*** (0.002)	0.016*** (0.003)	0.016*** (0.002)	0.002*** (0.000)	0.002*** (0.000)	0.002*** (0.000)
Flex_hours	0.001*** (0.000)	0.001*** (0.000)	0.000** (0.000)	– 0.000*** (0.000)	– 0.000* (0.000)	0.000 (0.000)
Av.corp_overtime	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	– 0.012*** (0.002)	– 0.010*** (0.002)	– 0.012*** (0.002)
%_PT_worker	– 0.015*** (0.001)	– 0.013*** (0.002)	– 0.015*** (0.002)	– 0.001*** (0.000)	– 0.002*** (0.000)	0.000 (0.000)

Table 3 (continued)

	All workers (part-time included)			Only full-time workers		
	(1)	(2)	(3)	(4)	(5)	(6)
	All	M	F	All	M	F
%_Smart_workers	− 0.004*** (0.000)	− 0.004*** (0.000)	− 0.004*** (0.000)	0.000 (0.000)	0.000 (0.000)	0.001* (0.000)
RegPartTime	− 0.001*** (0.000)	− 0.001*** (0.000)	− 0.001** (0.000)	− 0.040*** (0.004)	− 0.015* (0.006)	− 0.049*** (0.006)
RegSmartW	0.022*** (0.003)	0.027*** (0.005)	0.039*** (0.004)	− 0.075*** (0.006)	− 0.044*** (0.009)	− 0.086*** (0.008)
RegFlexibility	− 0.020*** (0.005)	0.007 (0.008)	− 0.021** (0.006)	0.033** (0.011)	− 0.008 (0.016)	0.102*** (0.016)
Profes_Dev	− 0.099*** (0.009)	− 0.107*** (0.014)	− 0.053*** (0.012)	0.078*** (0.008)	0.095*** (0.012)	0.037** (0.011)
Constant	0.182*** (0.007)	0.184*** (0.010)	0.154*** (0.008)	2.070*** (0.047)	2.008*** (0.068)	1.713*** (0.064)
Number of obs	34,634	18,296	16,338	28,154	17,453	10,701
Adjusted R^2	0.797	0.801	0.796	0.788	0.805	0.758

Blue collar variable omitted for collinearity in position. For the analysis, IPF weights were applied to make the sample representative of the labour force in the geographical area of Italy where the organisations are located

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

employees. As expected, higher corporate revenues are associated with higher wages. Conversely, high segregation (noting that the Duncan index in our analysis is given by the interaction between horizontal and vertical segregation Duncan indices) is negatively associated with all wages, especially those of women. It is crucial to highlight that, apart from a few variables, the other variables have a negligible magnitude. This suggests that these factors do not always exhibit direct linear relationships with wage outcomes.

Table 4 shows the results of the Oaxaca decomposition presented in Eq. 2 of the methodology section.

Here, the most important finding that arose from the different sub-samples is that, when we exclude part-time workers, the gender gap due to explained characteristics, which initially favoured males, changes direction and becomes advantageous to females.

In fact, the coefficient of being a part-time worker in the explained part of the wage gap is the one with a higher magnitude in favour of males across all regressors. This is consistent with the indirect discrimination rationales we raised at the beginning of the paragraph.

The difference in earnings in men's favour amounts to 10.8% of hourly wages (2.8% of the component explained by characteristics and 8.0% the unexplained component) for the sample that also includes part-time employees. The same gap is lowered to 1.3% and loses significance when we restrict the analysis to only those who work full-time. The explained part of the differential is in favour of women (− 5.7%) while the unexplained part slightly decreases to 6.9%.

The explained part in estimates gives results which are extremely in line with descriptive statistics. Although women are older they have lower corporate seniority (possible causes could be higher level of education and discontinuous careers). They are under-represented in higher positions (managers and middle managers) and this, in turn, is a feature that negatively affects wages. At the same time, they are slightly overrepresented in permanent contracts, a characteristic that helps reduce the gender pay gap.

The positive sign on part-time characteristics means that this variable is in men's favour: more women work part-time and part-time is a characteristic associated with lower salaries, so the part-time characteristic is in men's favour due to their lower take-up. The same interpretation can be given to the high feminisation of occupations having a positive and significant coefficient. On the unexplained part, having a permanent contract is more remunerative for women.

When we shift to the corporate fixed effects, we notice that being in a company with high horizontal and/or vertical segregation is a characteristic that amplifies the gender pay gap, but only for the sample which also includes part-time workers.

The percentage of women among managers is a characteristic that has a slightly negative impact on female wages when considering part-time workers. This can be explained by the reasoning presented in the descriptive section, which indicates that the percentage of female managers is also associated with policies on greater corporate work-life balance, potentially attracting women with lower earnings due to caregiving responsibilities. Consistent with the literature, the presence of jobs that can improve

Table 4 Oaxaca decomposition

	All workers (Part-time included)		Only full-time workers	
PANEL A – Oaxaca decomposition into explained and unexplained part				
Male	2.828*** (0.005)		2.850*** (0.005)	
Female	2.720*** (0.004)		2.837*** (0.005)	
Difference	0.108*** (0.006)		0.013 (0.007)	
Explained	0.028*** (0.005)		– 0.057*** (0.005)	
Unexplained	0.080*** (0.005)		0.069*** (0.006)	
PANEL B - Breakdown of explained and unexplained parts into variables' contribution				
	Explained	Unexplained	Explained	Unexplained
Age	– 0.038*** (0.004)	– 0.397* (0.177)	0.019*** (0.004)	– 0.465** (0.160)
Age squared	0.026*** (0.003)	0.236** (0.090)	– 0.018*** (0.003)	0.256** (0.082)
Seniority	0.003*** (0.000)	0.002 (0.006)	0.003*** (0.001)	0.004 (0.007)
Permanent	– 0.005*** (0.001)	– 0.040* (0.018)	– 0.002*** (0.000)	– 0.037* (0.018)
White_collar	– 0.032*** (0.002)	0.003 (0.008)	– 0.062*** (0.003)	– 0.009 (0.010)
Middle_manager	0.017*** (0.002)	0.001 (0.001)	0.002 (0.003)	0.001 (0.002)
Manager	0.026*** (0.002)	0.000 (0.001)	0.020*** (0.003)	– 0.000 (0.001)
Smart_worker	– 0.004*** (0.001)	0.006 (0.005)	– 0.004** (0.002)	0.010 (0.006)
PartTime_worker	0.010*** (0.002)	0.000 (0.002)		
High_fem	0.005*** (0.001)	0.001 (0.002)	0.008*** (0.001)	0.004* (0.002)
Duncan	0.006 (0.003)	0.010 (0.022)	– 0.009** (0.003)	0.022 (0.025)
%F_manager	0.005** (0.002)	– 0.006 (0.012)	0.002 (0.002)	– 0.010 (0.012)
%F_Middle_man.	0.008** (0.003)	– 0.117*** (0.019)	0.012*** (0.002)	– 0.070*** (0.019)
FirmAge	0.004*** (0.001)	0.029 (0.016)	– 0.002*** (0.001)	0.018 (0.019)
Ln_revenue	– 0.008*** (0.001)	0.007 (0.086)	0.004* (0.002)	0.123 (0.114)
Cultural_Practices	– 0.017*** (0.004)	0.047** (0.017)	– 0.030*** (0.004)	0.004 (0.017)

Table 4 (continued)

	All workers (Part-time included)		Only full-time workers	
Flex_hours	– 0.001* (0.000)	– 0.014* (0.007)	0.001 (0.000)	– 0.010 (0.008)
Av.corporate_overtime	0.017*** (0.003)	0.006 (0.021)	– 0.000 (0.001)	0.006 (0.017)
%_PT_worker	0.043*** (0.003)	0.005 (0.012)	0.001* (0.000)	– 0.024* (0.010)
%_Smart_worker	– 0.001* (0.000)	– 0.008 (0.018)	– 0.002 (0.002)	– 0.012 (0.021)
RegPartTime	– 0.003** (0.001)	– 0.013 (0.015)	0.006*** (0.001)	0.033 (0.019)
RegSmartW	0.002** (0.001)	0.039 (0.020)	0.001 (0.001)	0.059* (0.025)
RegFlexibility	– 0.009*** (0.001)	– 0.076 (0.047)	– 0.000 (0.000)	– 0.168** (0.065)
Profes_Dev	– 0.027*** (0.002)	0.022 (0.017)	– 0.008*** (0.002)	0.039 (0.021)
Constant		0.337* (0.141)		0.295 (0.159)
Number of observations	34,634		28,154	

work-life balance comes at the cost of a wage penalty for the whole sample but the impact loses significance when the estimation is restricted only to full-timers.

A significant proportion of women among middle managers has a statistically significant positive impact on women's wages in the unexplained component, suggesting progress in breaking the glass ceiling. At the same time, the explained part shows a slightly negative effect. Additional checks reveal that—similar to the percentage of female managers—a higher percentage of female middle managers is positively associated with the percentage of part-time and remote workers in the company. The negative relationship with wages did not appear in the simple correlations shown in Table 2, as the overall positive effect is more pronounced, being driven by a higher chance of breaking the glass ceiling.

A high level of inclusive cultural practices within a company has an ambiguous impact when part-time workers are included, nevertheless, the highest significance of the explained component (characteristics) is in favour of women. Moreover, we can see that a high level of inclusive cultural practices within a company is definitely beneficial for women when considering only full-time workers while it is ambiguous for the whole sample with a reduction of the explained part of the gender gap in hourly wages and an increase in the unexplained part.

Additionally, a high percentage of overtime hours of work within the firm undoubtedly increases the gender pay gap to the disadvantage of women in the sample that also includes part-time workers, and this can be taken as a signal of a higher evaluation of overtime work and hours of work by the organisation thus penalising women and part-timers who are more likely not to work overtime.

Furthermore, a substantial percentage of part-time workers within a company adversely affects female wages when examining the overall sample, including part-time employees. Conversely, among full-time workers, there is an unexplained component that benefits women. This may indicate that women employed full-time in an environment with a significant percentage of part-time workers enjoy a 'premium' or competitive advantage.

Having formalised policies regarding part-time and time flexibility does contribute to reducing the gender pay gap at the disadvantage of women for the whole sample. Time flexibility also has the advantage of enabling women to achieve better work-life balance without reducing their working hours or changing their workplace.

The presence of a policy of professional development is a corporate characteristic that contributes to the reduction of the gender wage gap especially when considering the whole sample.

6 Conclusions and Policy Implications

The company-level data that have been used in this study confirm, for a sample of big companies located in Northern Italy, the existence of a consistent gender wage gap: women earn, on average, 87% of men's wages. To provide a more nuanced understanding of this gap, we employed multivariate analyses that control for a range of worker-level and corporate-level variables. Preliminary analyses have found a negative effect of reduced working hours on hourly wages, which can be considered a form of indirect discrimination since this condition particularly affects women in our sample[7] and also in the national population. This trend is consistent with the negative and highly statistically significant ($P < 0.000$) effect of part-time employment on hourly wages, which is found in our regression analysis and is in line with previous research.

Other findings from the regression analysis, which are expected and consistent with previous research on Italian data, include the presence of a gender wage gap in favour of men, observed in both the full sample (including part-time workers) and the full-time subsample. Additionally, age and seniority are found to have a positive impact on wages.

Moreover, in agreement with Cirillo and Ricci (2022), having a permanent contract has a significant and positive impact on hourly wages, especially for women and, as expected, hourly wages increase with a rise in job responsibility (from blue-collar to manager). Moreover, as observed by Centra and Cutillo (2009), both the regression models and the Oaxaca decomposition confirm that being in a highly feminised occupation not only results in lower overall wages but also reduces women's wage returns. Higher wages are found to occur both for men and women when there are higher company revenues.

Finally, regarding the results of firm determinants in the regression analysis, it is important to note that, apart from a few variables, most have a negligible effect, while they become more significant in the Oaxaca decomposition.

[7] 15% of them are part-time workers against 2% of men.

Turning to the Oaxaca decomposition, the results show that the percentage of women among managers slightly negatively impacts female wages when considering part-time workers. This can be explained by the percentage of female managers being associated with greater corporate work-life balance policies, which potentially attract women with lower earnings due to caregiving responsibilities. Notably, this negative effect disappears when focusing only on full-time workers.

Regarding regulatory policies aimed at reaching a better work-life balance, the main results are heterogeneous.

Firstly a formalised part-time policy has a slightly positive effect on reducing the gender wage gap for the whole sample. This suggests that any negative impact of a part-time arrangement is likely captured by the individual variable indicating whether a worker is working part-time, rather than by the fixed effect related to the formalisation of part-time work. A substantial percentage of part-time workers within the company increases the gender wage gap to the disadvantage of women when examining the overall sample, including part-time employees. This highlights how much of what happens within (and access to) the labour market is determined by choices driven by a gender-biased distribution of unpaid work within the family and in society, or by an involuntary part-time status (information not provided by the firms' data). This should lead organisations not only to formalise part-time rights for workers but to put in place policies promoting a less biased distribution of unpaid care work by gender, promoting part-time work, as well as parental and paternity leaves for fathers. The regulation and protection of part-time workers do not negatively impact the gender pay gap, indeed they provide positive safeguards. Instead, the composition of the workforce (the intra-firm proportion of part-time workers) does. This result is consistent with a strand of literature on the gender wage gap pointing to a pay penalty for other dimensions of working conditions, in this case, part-time as a means to achieve higher work-life balance. Formalised hourly flexibility, on the other hand, positively impacts women's wages. The effect is felt even more on full-time workers. This policy enables women to balance work and life better without reducing their working hours or changing their workplace.

Thirdly, professional development policies within companies significantly enhance women's wages, especially for those in part-time roles. The fact that these organisations self-selected into the gender equality certification process suggests that they may be more proactive in adopting gender equality measures compared to other firms that have not chosen to undergo this certification.

Based on the evidence presented, different policy recommendations could be suggested at both the corporate and national levels to address the gender gap in terms of wages but also in terms of the composition of (and participation in) the labour force.

At the corporate level, career development plays a role in reducing the gender wage gap. As our results show, a more balanced management composition is positively related to higher work-life balance policies and practices in the firm and this could establish a virtuous cycle toward gender equality within the company.

In addition, companies should overcome the problem of the high percentage of women in part-time positions by offering corporate welfare, such as contributions or agreements to access childcare centres at reduced rates; contributions or agreements for extracurricular activities for the children of employees; and parenting support services (corporate day-care, study space, home babysitting, etc.). This can also be achieved without any additional expense on the part of the company by simply converting the welfare services/bonuses already offered into others aimed at work-life balance. This would also have a positive return on the company in terms of hours worked by employees and less tiredness in the workplace. However, to be effective, these policies should be offered to the entire labour force and promote the caregiver role of fathers.

Companies should strongly emphasise time flexibility, as it is the only work-life balance measure that does not alter either the workplace or the number of hours worked.

Finally, a focus on professional development is recommended, as evidence suggests that having formalised professional development policies has a positive impact on women's wages.

At both the national and regional levels, public administrations should provide information and awareness-raising policies to mitigate gender bias in the distribution of unpaid domestic and care work, specifically by informing women about the long-term cost (lower career prospects, lower future pensions, etc.) of discontinuous careers and reduced working hours and times and mitigating the still persistent strong familist culture by encouraging the greater involvement of men in unpaid care and domestic work. Expanding access to full-time childcare services and providing full-time primary and secondary schools can help increase female labour force participation and retention, especially in a context where women still bear the majority of unpaid care responsibilities. In addition, as part of outreach policies, it is also important to include guidance for young students, so that they can choose their academic career more consciously and avoid those gender stereotypes that lead to horizontal segregation.

Policies should be implemented to promote gender diversity in managerial positions, such as setting quotas or targets, as seen in Italy for certain businesses. Turning to measures that are more expensive for the State, the government could allow tax incentives for companies providing childcare services or implementing policies to retain or rehire mothers after childbirth (State-business action). Regarding families and individuals, the State could increase paternity leave, daycare services and child support, as well as offer tax deductions, bonuses or financial aid for families who hire babysitters.

While the database used in this study offers valuable insights, it also presents limitations. The sample is composed of 25 large legal entities, primarily located in Northern Italy, self-selected by participating in a gender equality evaluation. As demonstrated by our weighting procedure, this self-selection likely leads to an underestimation of the gender wage gap and contributes to a more favourable view of corporate gender equality. At the same time, however, another potential implication of not accounting for sample selection is that these organisations may retain mothers or women with higher caregiving responsibilities (and reduced working hours), or have lower-skilled female workers, resulting in less self-selection of female employees.

The results we are examining thus suggest that these organisations successfully retain a larger proportion of women who might otherwise exit the labour force, without causing significant increases in the gender pay gap compared to other firms.

Moreover, while the companies exhibit heterogeneity in their NACE codes, they do not encompass the full spectrum of the Italian manufacturing sector. Nevertheless, the applied weighting procedure allows us to infer that the results can be generalised to the Northern Italian workforce, specifically across six[8] major categories in the ISTAT ATECO classification in 12 classes, which account for over 63% of the labour force in the geographic area analysed.

Although the IPF adjustment improved the representativeness of the sample in terms of workers, it does not fully address the non-random nature of the firms. Consequently, the results should be interpreted with caution and may not be fully generalised to the entire corporate population in Italy.

We plan to extend the analysis by collecting more corporate observations, obtaining more information on the workers' levels of education and estimating quantile regressions to analyse the path of wage discrimination over the wage distribution and its relationship with the observed evidence of glass ceiling and sticky floor effects.

Appendices

Appendix A: Sample Characteristics and Nature of Self-Selection

Our sample includes 25 large legal entities located in Northern Italy, belonging to the following ATECO (NACE) classifications: Agriculture, Forestry, and Fishing (two-digit NACE 01–03); Industry (two-digit NACE 05–39); Trade (two-digit NACE 45–47); Hotels and Restaurants (two-digit NACE 55–56); Financial and Insurance Activities (two-digit NACE 64–66); and Real Estate, Business Services, and Other Professional and Entrepreneurial Activities (two-digit NACE 68–82). So, firms are representative only of 6 ATECO codes classes (NACE) out of the 12 ISTAT classes and some sections are not represented (e.g.: extraction of minerals from quarries and mines; construction; etc.). However, these sectors represent 63% of the workforce in the geographical area considered and they encompass heterogeneous physiological characteristics (such as the percentage of females, required tasks, duties, products, and sectors).

The size of the enterprises is relatively big if compared to the national average in which more than 60% of enterprises have only one employee (usually sole proprietorships with the owner self-employed), 30% are microenterprises, small enterprises account for about 5%, and medium and large enterprises together are less than 0.7% (ISTAT 2021b).[9] Also, all the companies are headquartered in Northern Italy, which we know has different characteristics than the rest of the peninsula, especially in terms of female labour force participation.

By using an IPF weighting system, we have aligned the marginal totals of worker characteristics with those from the ISTAT Labour Force Survey for the corresponding

[8] Agriculture, forestry, and fishing (two-digit NACE 01–03); Industry (two-digit NACE 05–39); Trade (two-digit NACE 45–47); hotels and restaurants (two-digit NACE 55–56); financial and insurance activities (two-digit NACE 64–66); real estate, business services, and other professional and entrepreneurial activities (two-digit NACE 68–82).

[9] Size definition of the Italian Decree of 18 April 2005 Adjustment of the criteria for identifying small and medium-sized enterprises to the Community Framework. (GU Serie Generale no. 238 of 12–10-2005). https://www.gazzettaufficiale.it/atto/serie_generale/caricaDettaglioAtto/originario?atto.dataPubblicazioneGazzetta=2005-10-12&atto.codiceRedazionale=05A09671&elenco30giorni=false.

geographical area and period, allowing our results to be considered representative of large firms in Northern Italy across the six reported ATECO categories.

Regarding the firm size, in addition, the firm determinants considered are not typically present in small and medium-sized enterprises (SMEs), which often lack the more structured organisational framework required for such policies (e.g.: the existence of a whistleblowing system, corporate practices for the protection of the work environment; the existence of climate/satisfaction/well-being surveys from a gender perspective, the existence of high formalisation of work-life balance policies).

Concerning workforce composition, the comparison between descriptive statistics (Table 5) and inferential statistics (Table 6) shows that our weighting procedure aligned the sample composition to the workforce in the reference area (Table 7). The inferential correction reveals that our sample overrepresented permanent contracts and middle manager positions, while it underrepresented blue-collar positions. The application of weights allows for corrected multivariate analyses.

Finally, the firms in our sample are companies that show interest in gender equality and assume they can get certified. This self-selection could lead us to a virtuous picture of corporate practices regarding gender equality and to underestimate the phenomenon. This is supported by comparing descriptive and inferential statistics, which account for weighting procedures.

Our correction through weighting reveals an interesting picture of changes in firm determinants. Firstly, when we compare our inferential findings (Table 10) with the sample descriptive (Tables 8 and 9), we observe a slight decrease in the gender pay gap (GPG). However, this is accompanied by more pronounced declines across nearly every indicator of work-life balance. Specifically, there is a significant reduction in the percentage of smart workers within companies, a marked decrease in flexible working hours, and a notable drop in the proportion of women in both management and middle management roles. Additionally, we see an increase in overtime hours and a slight uptick in company profits.

These findings suggest that, as we have assumed and expected, the companies in our sample are more virtuous in terms of gender equality practices and policies compared to the general ones, as well as in their organisational culture regarding gender equality. As a result, they are likely better equipped to retain a higher percentage of women with caregiving responsibilities who might otherwise leave the labour market. Interestingly, this occurs with minimal change in the wage differential, even though workers from other companies likely include more self-selected women—a higher percentage of those with fewer caregiving burdens—potentially leading to higher earnings profiles. This underscores the importance of creating inclusive work environments that support employees' diverse needs, ultimately benefiting both the workforce and the companies themselves.

Appendix B: Employee Descriptive Statistics and Selection Bias Correction

See Tables 5, 6 and 7.

Table 5 Descriptive statistics—employees

Variable	Obs	Mean	Std. Dev	Min	Max
Yearly earnings	34,634	44,479.57	26,644.36	7,894.74	885,264.5
Yhour	34,634	23.3	13.28	7.01	457.5
Female	34,634	0.47	0.50	0	1
Age (in classes)[a]					
≤ 24	34,634	0.01	0.08	0	1
$> 24 \leq 29$	34,634	0.04	0.20	0	1
$> 29 \leq 34$	34,634	0.08	0.28	0	1
$> 34 \leq 39$	34,634	0.09	0.29	0	1
$> 39 \leq 44$	34,634	0.14	0.35	0	1
$> 44 \leq 49$	34,634	0.17	0.37	0	1
$> 49 \leq 54$	34,634	0.19	0.39	0	1
$> 54 \leq 59$	34,634	0.18	0.38	0	1
$> 59 \leq 64$	34,634	0.10	0.30	0	1
Seniority	34,634	8.68	8.69	0	47.49
Permanent	34,634	0.96	0.21	0	1
Blue collar	34,634	0.25	0.43	0	1
White collar	34,634	0.51	0.50	0	1
Middle manager	34,634	0.22	0.42	0	1
Manager	34,634	0.02	0.13	0	1
Smart_worker	34,634	0.33	0.47	0	1
PartTtime_worker	34,634	0.05	0.22	0	1
NACE classification					
Agriculture, forestry and fishing	34,634	0.04	0.2	0	1
Manufacturing	34,634	0.07	0.25	0	1
Trade	34,634	0.16	0.36	0	1
Hotels and restaurants	34,634	0.05	0.22	0	1
Financial and insurance activities	34,634	0.59	0.49	0	1
Real estate, business services other prof. and entr. activities	34,634	0.10	0.30	0	1

[a] Age is a continuous variable in the regression model, here in classes for comparison with the labour force survey—quarterly cross-sectional open-access microdata

Table 6 Inferential statistics—employees

Variable	Mean	Std. Dev	[95% conf. interval]	
Yearly earnings	35,951.55	304.69	35,354.34	36,548.77
Yhour	18.06	0.15	17.77	18.35
Female	0.41	0.01	0.40	0.43
Age (in classes)[a]				
≤ 24	0.05	0.01	0.04	0.07
$> 24 \leq 29$	0.10	0.00	0.09	0.11
$> 29 \leq 34$	0.11	0.00	0.10	0.12
$> 34 \leq 39$	0.12	0.00	0.11	0.13
$> 39 \leq 44$	0.14	0.00	0.13	0.15
$> 44 \leq 49$	0.16	0.00	0.15	0.17
$> 49 \leq 54$	0.16	0.00	0.15	0.16
$> 54 \leq 59$	0.12	0.00	0.12	0.13
$> 59 \leq 64$	0.05	0.00	0.04	0.05
Seniority	9.93	0.12	9.70	10.15
Permanent	0.87	0.01	0.86	0.88
Blue collar	0.44	0.01	0.43	0.45
White collar	0.48	0.01	0.46	0.49
Middle manager	0.06	0.00	0.06	0.07
Manager	0.02	0.00	0.02	0.02
Smart_worker	0.31	0.01	0.30	0.32
PartTtime_worker	0.07	0.00	0.07	0.08
NACE classification				
Agriculture, forestry and fishing	0.02	0.00	0.02	0.02
Manufacturing	0.54	0.00	0.54	0.55
Trade	0.19	0.00	0.19	0.19
Hotels and restaurants	0.06	0.00	0.06	0.06
Financial and insurance activities	0.06	0.00	0.06	0.06
Real estate, business services and other prof. and entrepr. activities	0.14	0.00	0.13	0.14

[a] Age is a continuous variable in the regression model, here in classes for comparison with the labour force survey—quarterly cross-sectional open-access microdata

Table 7 Employees descriptive statistics calculated on the Italian population[a, b]

Variable	Mean	Std. Dev	[95% conf. interval]	
Yearly earnings	35,951.55	304.69	35,354.34	36,548.77
Yhour	18.06	0.15	17.77	18.35
Female	0.41	0.01	0.40	0.43
Age (in classes)[a, c]				
≤ 24	0.05	0.01	0.04	0.07
$> 24 \leq 29$	0.10	0.00	0.09	0.11
$> 29 \leq 34$	0.11	0.00	0.10	0.12
$> 34 \leq 39$	0.12	0.00	0.11	0.13
$> 39 \leq 44$	0.14	0.00	0.13	0.15
$> 44 \leq 49$	0.16	0.00	0.15	0.17
$> 49 \leq 54$	0.16	0.00	0.15	0.16
$> 54 \leq 59$	0.12	0.00	0.12	0.13
$> 59 \leq 64$	0.05	0.00	0.04	0.05
Seniority	9.93	0.12	9.70	10.15
Permanent	0.87	0.01	0.86	0.88
Blue collar	0.44	0.01	0.43	0.45
White collar	0.48	0.01	0.46	0.49
Middle manager	0.06	0.00	0.06	0.07
Manager	0.02	0.00	0.02	0.02
Smart_worker	0.31	0.01	0.30	0.32
PartTtime_worker	0.07	0.00	0.07	0.08
NACE classification[d]				
Agriculture, forestry and fishing	0.02	0.00	0.02	0.02
Manufacturing	0.54	0.00	0.54	0.55
Trade	0.19	0.00	0.19	0.19
Hotels and restaurants	0.06	0.00	0.06	0.06
Financial and insurance activities	0.06	0.00	0.06	0.06
Real estate, business services and other prof. and entrepr. activities	0.14	0.00	0.13	0.14

[a]The variables reported are those available from the Italian labour force survey – quarterly cross-sectional open-access microdata

[b]Descriptive statistics were applied to workers in the same geographical area as our sample and corporate units with at least 50 employees

[c]Age is a continuous variable in the regression model, here in classes for comparison with the labour force survey—quarterly cross-sectional open-access microdata

[d]For comparative purposes with the previous tables, the average (which corresponds to the percentage by classification) is calculated by taking as the total only the six NACE codes present in our sample. The respective national population percentages of big companies for the same geographical area and reference period as our sample, when including the classes absent from our sample, are: (Agriculture, forestry, and fishing 0.5%; Manufacturing 44.1%; Construction 2%; Trade 8.1%; Hotels and restaurants0.96%; Transportation and storage 6.87%; Information and communication services 4.22%; Financial and insurance activities 2.87%; Real estate activities, business services, and other professional and entrepreneurial activities 6.82%; Public administration, defence, and compulsory social security 4.88%; Education, healthcare, and other social services; 17.48%; Other collective and personal services 1.20%). These values were used as reference marginal totals in our weighting procedures

Appendix C: Firm Determinants Descriptive Statistics, Selection Bias Correction and Correlations with the Outcome

See Tables 8, 9 and 10.

Table 8 Descriptive statistics—firms (employee weights: bigger firms weigh more)

Variable	Obs	Mean	Std. Dev	Min	Max
Unadj. GPG (income)	34,634	9091.7	4639.82	− 3110.36	22,388.54
Unadj. GPG (hourly Inc.)	34,634	3.43	2.19	− 1.8	10.75
%F_company[a]	34,634	49.12	13.51	16.39	70.88
%F_Managers	34,634	22.44	11.49	0	100
%F_Middle_man	34,634	33.97	8.57	0	57.45
FirmAge	34,634	30.71	12.81	7	152
Ln_revenue	34,634	20.66	2.25	10.31	22.43
Cultural_Practices	34,634	48.23	18.62	0	100
Flex_hours	34,634	68.36	83.26	15	750
Av.corporate_overtime	34,634	2.7	3.42	0.04	9.74
%_PT_worker	34,634	21.21	24.07	0.77	81.9
%_Smart_worker	34,634	32.4	20.37	0	100
RegPartTime	34,634	1.46	0.76	0	2
RegSmartW	34,634	0.74	0.92	0	2
RegFlexibility	34,634	1.74	0.44	1	2
Profes_Dev	34,634	0.97	0.3	0	2

[a]Percentage of females in the company, included here only for the company description; in the regressions, we use the dichotomic individual employment indicator in highly feminised occupations which also account for worker position (*High_fem*)

Table 9 Descriptive statistics—firms (firm weights: each firm has the same weight)

Variable	Mean	Std. Dev	Min	Max
Unadj. GPG (income)	6775.82	6369.6	− 3110.36	22,388.54
Unadj. GPG (hourly Inc.)	2.47	3.12	− 1.8	10.75
%F_company[a]	46.49	17.29	16.39	70.88
%F_Managers	25.74	24.9	0	100
%F_Middle_man	31.8	14.01	0	57.45
FirmAge	30.32	29.48	7	152
Ln_revenue	17.69	2.63	10.31	22.43
Cultural_Practices	47.83	32.52	0	100

Table 9 (continued)

Variable	Mean	Std. Dev	Min	Max
Flex_hours	86.8	140.51	15	750
Av.corporate_overtime	3.49	3.54	0.04	9.74
%_PT_worker	19.54	27.47	0.77	81.9
%_Smart_worker	44.33	35.89	0	100
RegPartTime	1.16	0.85	0	2
RegSmartW	1.28	0.79	0	2
RegFlexibility	1.52	0.51	1	2
Profes_Dev	0.84	0.55	0	2

[a]Percentage of females in the company, included here only for the company description; in the regressions, we use the dichotomic individual employment indicator in highly feminised occupations which also account for worker position (*High_fem*)

Table 10 Inferential statistics—firms (labour force weights: we apply the IPF weights to make our sample representative of the labour force in Northern Italy)

	Mean	Std. Dev	[95% conf. interval]	
Unadj. GPG (income)	6729.18	29.92	6670.54	6787.82
Unadj. GPG (hourly In.)	2.33	0.01	2.30	2.35
%F_company[1]	40.79	0.11	40.58	41.00
%F_Managers	13.93	0.08	13.77	14.09
%F_Middle_man	26.65	0.10	26.46	26.84
FirmAge	24.26	0.15	23.98	24.55
Ln_revenue	18.30	0.03	18.25	18.35
Cultural_Practices	40.35	0.21	39.93	40.77
Flex_hours	74.15	0.44	73.30	75.01
Av.corporate_overtime	3.80	0.02	3.76	3.83
%_PT_worker	20.78	0.09	20.59	20.96
%_Smart_worker	28.63	0.15	28.34	28.92
RegPartTime	1.00	0.01	0.99	1.01
RegSmartW	1.44	0.00	1.44	1.45
RegFlexibility	1.47	0.00	1.47	1.48
Profes_Dev	0.67	0.00	0.66	0.67

For the analysis, IPF weights were applied to make the sample representative of the labour force in the geographical area of Northern Italy

Acknowledgements The views expressed in this paper are those of the authors only and do not involve the responsibility of the University of Modena and Reggio Emilia or any other affiliated institution. The authors are grateful to the session chairs and participants of the XVII Espanet-Italy Conference "Accounting for Inequalities and Reframing Social Policies" (Department of Social Sciences, University of Naples Federico II, September 4-6, 2024) and to anonymous referees for their stimulating comments on previous versions of the paper.

Funding Open access funding provided by Università degli Studi di Modena e Reggio Emilia within the CRUI-CARE Agreement. The open access publication was made under the national CARE-CRUI contract, signed by the University of Modena and Reggio Emilia University Library System. The authors did not receive support from any organization on the research activities leading to the submitted work.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Addabbo T (2018) Disuguaglianze di genere nel lavoro e nelle retribuzioni in Italia. In: Franzini M, Raitano M (eds) Ch. 8: Il mercato rende diseguali? La disuguaglianza dei redditi in Italia. Il Mulino, Bologna, pp 173–192

Addison JT, Ozturk OD, Wang S (2018) The occupational feminization of wages. ILR Rev 71(1):208–241

Bertrand M (2018) Coase lecture—the glass ceiling. Economica 85(338):205–231. https://research.ebsco.com/linkprocessor/plink?id=915e9956-64b3-3418-9301-c5b8e25dc307

Bick A, Blandin A, Mertens K (2023) Work from home before and after the COVID-19 outbreak. Am Econ J Macroecon 15(4):1–39. https://doi.org/10.1257/mac.20210061

Blau DF, Kahn LM (2017) The gender wage gap: extent, trends, and explanations. J Econ Lit 55(3):789–865

Busch F (2020) Gender segregation, occupational sorting, and growth of wage disparities between women. Demography (Springer Nature) 57(3):1063–1088

Busch A, Holst E (2011) Gender-specific occupational segregation, glass ceiling effects, and earnings in managerial positions: results of a fixed effects model. IZA Discussion Paper No. 5448

Card D, Cardoso AR, Kline P (2016) Bargaining, sorting, and the gender wage gap: quantifying the impact of firms on the relative pay of women. Q J Econ 131(2):633–686

Cardoso AR, Guimarãs P, Portugal P (2016) What drives the gender wage gap? A look at the role of firm and job-title heterogeneity. Oxf Econ Pap 68(2):506–524

Casarico A, Lattanzio S (2024) What firms do: gender inequality in linked employer-employee data. J Law Econ 42(2):325–355. https://doi.org/10.1086/723177

Centra M, Cutillo A (2009) Differenziale salariale di genere e lavori tipicamente femminili. Studi Isfol, no. 2

Cirillo V, Ricci A (2022) Heterogeneity matters: temporary employment, productivity and wages in Italian firms. Econ Polit 39(2):567–593. https://doi.org/10.1007/s40888-020-00197-2

Deming WE, Stephan FF (1940) On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Ann Math Stat 11(4):427–444

Destefanis S, Mazzotta F, Parisi L (2024) Goldin's last chapter on the gender pay gap: an exploratory analysis using Italian data. Work Employ Soc 38(2):549–572

Duncan OD, Duncan B (1955) A methodological analysis of segregation indexes. Am Sociol Rev 20(2):210–217. https://research.ebsco.com/linkprocessor/plink?id=e77fe2d2-0be4-3a87-80b8-ee78554e99b1

Goldin C (2014) A grand gender convergence: its last chapter. Am Econ Rev 104(4):1091–1119. https://doi.org/10.1257/aer.104.4.1091

Hotz VJ, Johansson P, Karimi A (2018) Parenthood, family friendly workplaces, and the gender gaps in early work careers. What firms do: gender inequality, 353, NBER Working Paper no. 24173. National Bureau of Economic Research, Cambridge

Istituto Nazionale di Statistica (ISTAT) (2021a) Lavoro e retribuzioni. Indagine sulle forze di lavoro [Work and earnings. Labor force survey]. https://www.istat.it/it/files/2021/08/Lavoro-e-retribuzioni-II-trimestre-2021.pdf

Istituto Nazionale di Statistica (ISTAT) (2021b) Rapporto sulle imprese 2021. Edited by Andrea de Panizza. 1st ed. IT: Istat. https://doi.org/10.1481/Istat.Rapportoimprese.2021

Kalton G, Flores-Cervantes I (2003) Weighting methods. J off Stat 19(2):81

McKay S, Mussida C (2018) Wages and employment chances for women: the role of expected and actual fertility. Rivista Internazionale di Scienze Sociali 126(3):255–282

Mincer J (1974) Schooling, experience, and earnings. Human Behavior & Social Institutions No. 2. National Bureau of Economic Research, Inc

Morchio I, Moser C (2024) The gender pay gap: micro sources and macro consequences. NBER Working Paper No. 32408. http://www.nber.org/papers/w32408

Mussida C, Picchio M (2014) The gender wage gap by education in Italy. J Econ Inequal 12(1):117–147

Oaxaca R (1973) Male-female wage differentials in urban labour markets. Int Econ Rev 14:693–709

Olivetti C, Petrongolo B (2008) Unequal pay or unequal employment? A cross-country analysis of gender gaps. J Labour Econ 26(4):621–654

Sauer S, Schasching M, Wohlrabe K (2023) Handbook of ifo Surveys (No. 100). ifo Beiträge zur Wirtschaftsforschung

Szusecki T, Konkolÿ Thege B, Stauder A (2023) The prevalence and mental health correlates of exposure to offensive behaviours at work in Hungary: results of a national representative survey. BMC Public Health 23(1):78

Zizza R (2013) The gender wage gap in Italy. Questioni di Economia e Finanza (Occasional Papers) 172, Bank of Italy