Journal of
Economic Studies

Estimating the gender wage gap using stochastic frontiers: some modelling issues

Graziella Bonanno
Department of Environmental Engineering, University of Calabria, Rende, Italy, and
Antonio Alvarez
Department of Economics, University of Oviedo, Oviedo, Spain

Received 1 September 2025
Revised 13 March 2026
5 June 2026
27 July 2026
Accepted 27 July 2026

Abstract

Purpose – The purpose of this study is to analyse the gender wage gap in Italy by adopting a Stochastic Frontier Approach, with the aim of assessing whether gender differences persist after controlling for individual and job-related characteristics and how modelling choices affect the estimated gap.

Design/methodology/approach – The analysis is based on microdata from the 2021 wave of the European Union Statistics on Income and Living Conditions (EU-SILC) for Italy. A wage frontier model is estimated to decompose observed wage differentials into productivity-related components and wage inefficiency, with particular attention to the role played by gender in the specification of the model.

Findings – The results provide clear evidence of a gender wage gap in Italy, with women earning significantly less than men even after accounting for observable characteristics. Two main findings emerge. First, women display lower levels of wage efficiency, suggesting structural disadvantages in converting human capital into equivalent earnings. Second, the magnitude of the estimated gender wage gap is sensitive to modelling choices, particularly with respect to whether the gender dummy variable is included in the wage frontier or in the inefficiency term.

Originality/value – This study contributes to the literature on gender wage disparities by applying the Stochastic Frontier Approach to recent Italian data and by highlighting the importance of model specification in shaping conclusions about the size and sources of the gender wage gap.

Keywords Wage stochastic frontiers, Gender wage gap, Italy

Paper type Research article

1. Introduction

The gender wage gap remains a central issue in labour economics and policy debate (Caparrós Ruiz, 2025; González *et al.*, 2022; Nchor and Náplava, 2025; Nguyen *et al.*, 2026). Although the gap has declined in many advanced economies since the 1980s, substantial earnings differences between men and women persist (Blau and Kahn, 2017). These differences are attributed to observable factors—such as education, experience, occupation, and sector—and unobserved mechanisms including wage-setting institutions, bargaining power, firm heterogeneity, and discrimination.

A large empirical literature decomposes the gender wage gap into explained and unexplained components. The most common method is the Oaxaca–Blinder decomposition, which estimates separate wage equations for men and women and attributes part of the gap to observable characteristics and the remainder to unexplained factors. However, this approach focuses on mean differences and does not model the upper bound of attainable wages.

JEL Classification — J31

Funding: This work was supported by Università della Calabria; Fondo sostegno aree socio-umanistiche – Cda del 26.03.2021 – quota DESF.

Journal of Economic Studies
Vol. 53 No. 9, 2026
pp. 289-305
Emerald Publishing Limited
e-ISSN: 1758-7387
p-ISSN: 0144-3585
DOI 10.1108/JES-09-2025-0667

An alternative is provided by wage frontier models, first proposed by Herzog *et al.* (1985). In this framework, wages are bounded by a stochastic frontier representing the maximum wage attainable given observable characteristics. Deviations from the frontier reflect wage-setting frictions, institutional constraints, or other unobserved factors preventing workers from reaching potential earnings.

Stochastic Frontier Analysis (SFA), originally developed in production economics (Aigner *et al.*, 1977; Meeusen and van den Broeck, 1977), provides a suitable econometric framework to estimate wage frontiers. Recent studies apply SFA to labour market outcomes, including wage determination and gender differences (e.g. Díaz and Sánchez, 2011; García-Prieto and Gómez-Costilla, 2017; Pérez-Villadóniga and Rodríguez-Álvarez, 2017a, b). Frontier models are useful because they capture wage dispersion and unobserved constraints beyond mean regressions.

A key modelling issue concerns the role of gender in the stochastic frontier model. In pooled models, the gender dummy may enter the deterministic frontier, the inefficiency term, or both. Including gender in the frontier allows different potential wage ceilings for men and women, while including it in the inefficiency component allows gender to influence dispersion around the frontier. Existing studies typically adopt one specification but rarely compare them systematically or examine their implications for measuring the gender wage gap.

The first objective of this paper is to estimate the gender wage gap in Italy using a stochastic wage frontier approach. Italy is an interesting case because it shows a relatively small observed wage gap despite persistent gender differences in labour market participation and career progression. Using data from the European Union Statistics on Income and Living Conditions (EU-SILC), we estimate wage frontiers for Italian workers and analyse gender differences in both potential wages and distance from the frontier.

The second objective is methodological: to examine how different modelling choices regarding the placement of the gender dummy affect the estimated gender wage gap. We compare three specifications: (1) gender shifts the wage frontier, (2) gender affects the variance of the inefficiency term, and (3) gender enters both components. This comparison shows that the estimated gender wage gap is sensitive to modelling choices.

A further contribution is the explicit computation of the gender wage gap in monetary terms within the stochastic frontier framework. Previous studies often measure efficiency or wage indices but rarely express the gap in euros. We derive the expected wage differential attributable to gender by combining differences in the estimated frontier with differences in expected inefficiency, evaluated at representative values of observable characteristics.

The results show a persistent gender wage gap in Italy. On average, women face both a lower estimated wage frontier and different wage dispersion around the frontier, indicating that gender differences reflect not only observable characteristics but also heterogeneous wage-setting frictions. Moreover, the estimated gap varies depending on how gender is incorporated into the stochastic frontier model, highlighting the importance of modelling choices.

The remainder of the paper is organized as follows. Section 2 presents the stochastic frontier framework. Section 3 reviews the literature. Section 4 describes the data. Section 5 outlines the econometric methodology and the computation of the gender wage gap. Section 6 discusses the results, and Section 7 concludes.

2. Stochastic frontier analysis

Aigner *et al.* (1977) and Meeusen and Van den Broeck (1977) independently proposed the estimation of stochastic production frontiers. These models consider that deviations from the production frontier can be decomposed, allowing to separate the random effects, such as climatic events, from the effects of changes in technical efficiency [1]. Since these pathbreaking contributions, stochastic frontiers have been used in contexts different from production functions, such as earnings equations (Herzog *et al.*, 1985) or demand functions (Algieri and Alvarez, 2023).

In the framework of panel data, a general stochastic production frontier model can be given by:

$$y_{it} = \beta' x_{it} + v_{it} - u_{it} \tag{1}$$

where subscript i indexes individuals and subscript t indexes time, y_{it} represents output produced by firm i at time t (in our case, wage), x_{it} is a vector of inputs (in our case, characteristics of the workers), β is a vector of unknown parameters to be estimated, v_{it} is a symmetric random disturbance which captures the effect of statistical noise, whereas u_{it} is a non-negative stochastic term that is assumed to be independent from v and to capture distance from the stochastic frontier. When $u = 0$, the observation lies on the frontier and is therefore efficient. When $u > 0$, the observation is below the frontier, indicating that it is technically inefficient, i.e. in our case indicates the existence of a wage gap (the observed wage is different form the wage given by the frontier after taking into account all the x variables included in the wage frontier).

Since we are interested in finding which variables explain the efficiency of the workers, i.e. which variables are behind the fact that some workers do not achieve the potential wage they could obtain given their characteristics, we estimate several models that modify equation (1) by allowing the inefficiency term u to be a function of some exogenous variables z. The general form of this type of models is:

$$y_{it} = \beta' x_{it} + v_{it} - u_{it}(z_{it}) \tag{2}$$

There are two possible alternative specifications of u(z), depending on the way that the variables z affect the distribution of u. In particular, they can affect the mean or the variance of the distribution of u.

Battese and Coelli (1995) is the most popular model among practitioners in order to allow technical inefficiency to be a function of some exogenous variables. In this model some variables z explain the mean of the pre-truncated distribution of u. The inefficiency term can be expressed in the following way:

$$u_{it} = z_{it}\delta + w_{it} \tag{3}$$

where z are the explanatory variables associated with technical inefficiency and w_{it} is defined by the truncation of a normal distribution with zero mean and variance σ^2, such that the point of truncation is $-z_{it}\delta$.

The other alternative is to model the variance of u. Reifschneider and Stevenson (1991) was the first paper to incorporate heteroskedasticity in the stochastic frontier model. Caudill *et al.* (1995) (from now on referred to as CFG95) assumed that u exhibits multiplicative heteroskedasticity, a choice that we will use in this paper. In particular, the CFG95 model suggests an exponential function:

$$u_{it} \sim N^{+}\left(0, \sigma_{it}^{2}\right), \sigma_{it} = exp\left(\delta z_{it}\right) \tag{4}$$

where the + sign indicates truncation of the distribution at zero.

Modelling the variance of the one-sided error term is very important since the presence of heteroskedasticity in u will yield biased estimates of both the frontier parameters and the efficiency scores. This result differs markedly from the typical effect of heteroskedasticity in the two-sided error term v, which causes the variances of the parameter estimates to be biased. For this reason, the heteroskedastic model (CFG95) will be our preferred specification.

The parameters of the stochastic frontier and the model for the technical inefficiency effects in Equation (2) are estimated simultaneously by maximum likelihood. If the dependent variable is measured in logs, the technical efficiency (TE) of unit i in period t can be calculated as:

$$TE_{it} = e^{-u_{it}} \tag{5}$$

Given that u is non-negative, the formula in (5) ensures that the TE index is bounded between 0 and 1.

3. Wage stochastic frontiers and the gender wage gap

The literature on the gender wage gap is vast and spans several decades. Early contributions emphasized differences in observable characteristics such as education, experience, occupation, and sector as primary drivers of wage differentials (Blau and Kahn, 2007). Over time, however, it became clear that these factors explain only a fraction of the observed gap, shifting attention toward unobserved mechanisms including discrimination, institutional features of wage setting, bargaining power, and firm-level heterogeneity (Goldin, 2014; Blau and Kahn, 2017).

A more integrated conceptual framework helps explain why individuals with similar human capital endowments differ in their ability to convert this capital into realized earnings. Beyond education and work experience, the gender literature points to several additional determinants of this "conversion efficiency", including age and life-cycle effects, health status, managerial responsibility and organizational position, and firm-level characteristics such as corporate social responsibility (CSR) practices and equality-oriented workplace policies (Card *et al.*, 2018; Goldin and Katz, 2016). Within our stochastic frontier framework, such factors can be conceptualized as shifting either the potential wage frontier itself or the extent to which workers are able to approach it. Structural disadvantages, such as institutional barriers and "sticky floor" effects, may prevent women from reaching the same wage ceiling as men with identical human capital. Relatedly, the inefficiency component can be interpreted as capturing market-driven earnings shortfalls, i.e. wage-setting frictions rather than differences in inherent productivity, reflecting the market's failure to reward female human capital equitably. Finally, unobserved constraints, such as gender differences in bargaining power and unequal access to firm-level pay premia, may generate systematic deviations from the potential wage that are not fully captured by observable characteristics.

Within this broad literature, stochastic frontier models offer a distinctive perspective by explicitly modelling the upper bound of attainable wages conditional on observed characteristics. Rather than focussing on mean wage differences, this approach allows researchers to examine how far individual wages lie from a potential wage frontier and to study the dispersion of wages around that frontier.

The application of stochastic frontier analysis to earnings equations dates back to Herzog *et al.* (1985), who introduced the concept of an earnings frontier to study labour market outcomes. Subsequent studies extended this approach to various contexts, including migration (Lang, 2005), labour market information (Polachek and Robst, 1998), and occupational settings (Jane, 2013).

While these contributions established the usefulness of frontier methods in detecting wage disparities, they largely relied on homoscedastic inefficiency specifications and provided limited discussion of the mechanisms underlying gender differences in wage dispersion.

A central modelling issue in the wage frontier literature concerns how gender should be incorporated into the stochastic frontier model. Two broad empirical strategies have emerged. While some studies estimate gender-specific frontiers, allowing for entirely different wage-setting processes for men and women (e.g. Watson, 2000; Ogloblin and Brock, 2005), others opt for a pooled frontier, introducing a gender dummy variable to capture gender-specific effects.

Within pooled models, gender can enter the deterministic frontier, the inefficiency component, or both. Introducing gender in the frontier allows for different potential wage ceilings, whereas introducing gender in the inefficiency term allows for differences in the dispersion of wages around the frontier. Díaz and Sánchez (2011) were among the first to model gender as a determinant of inefficiency in a European context, showing that women tend to exhibit larger distances from the wage frontier. Subsequent studies by Pérez-Villadóniga and Rodríguez-Álvarez (2017a, b) and Bashford-Fernández and Rodríguez-Álvarez (2019) further developed this approach.

More recent contributions have emphasized the importance of allowing for heteroskedasticity in the inefficiency term. Modelling the variance of inefficiency as a function of observable characteristics enables researchers to capture differences in wage dispersion. This is particularly relevant in the context of gender wage gaps, where institutional constraints, bargaining outcomes, and firm-level pay policies may generate asymmetric wage distributions for men and women. Pérez-Villadóniga *et al.* (2025) extend this line of research using latent class stochastic frontier models to analyse gender wage gaps among managers, revealing substantial heterogeneity across unobserved labour market segments.

Recent advances in empirical labour economics highlight that gender disparities are driven by complex interactions between firm-level heterogeneity and individual bargaining power. Furthermore, "sticky floor" phenomena suggest that women are disproportionately concentrated in lower-paid positions within wage distributions, leading to greater dispersion at the lower tail (Said *et al.*, 2022). Our SFA approach complements these perspectives by modelling the dispersion of wage outcomes relative to a potential ceiling, capturing the heterogeneous constraints that mean regressions often obscure.

Parallel evidence from non-frontier approaches also underscores the relevance of wage dispersion. Some recent studies have linked gender wage inequality to firm-level pay premia, occupational sorting within job titles, and differential bargaining power (e.g. Card *et al.*, 2018; Blau and Kahn, 2017; Goldin and Katz, 2016). These findings are consistent with interpretations of frontier inefficiency as capturing wage-setting frictions rather than productivity shortfalls.

Despite these advances, several gaps remain in the existing literature. First, few studies systematically compare alternative modelling choices regarding the placement of the gender dummy within a stochastic frontier framework. Second, while many papers assess the statistical significance of gender effects, far fewer explicitly compute the gender wage gap in monetary terms using frontier-based estimates.

This paper contributes to the literature by addressing these gaps. We compare three alternative stochastic frontier specifications that differ in how gender enters the model, explicitly quantify the gender wage gap in euros. In doing so, we provide new evidence on both the magnitude of the gender wage gap in Italy and the methodological implications of modelling choices in stochastic wage frontier analyses.

4. Data and variables

Our analysis is based on a sample of 6,346 individuals from the European Union Statistics on Income and Living Conditions (EU-SILC) for Italy in 2021. This survey provides a wide range of details concerning the labour market and individual characteristics of the respondents [2]. Specifically, our analysis focuses on individuals who have permanent, full-time contracts, and excludes those employed in the agricultural and public sectors. In line with previous studies (e.g. Zveglich *et al.*, 2019; Díaz and Sánchez, 2011) we further restrict the sample to individuals aged between 25 and 65 years old.

We estimate a Mincer equation (Mincer, 1974) with the dependent variable being the logarithm of the annual in-work income [3]. This is calculated using respondents' self-reported earnings from their main job in the previous year.

As explanatory variables, we account for the number of hours worked during the year, since our dependent variable is not wage per hour. *Worked hours* represents the number of hours that the individuals declare to work in their main job. As suggested by economic theory, we also include a set of human capital proxies, such as education and work experience. *Work experience* is the number of years spent in paid work. As for education, EU-SILC contains information on the highest level of education attained according to the International Standard Classification of Education (ISCED) level successfully completed. While some papers convert this discrete variable into a continuous one by assigning years to each schooling category (e.g. Pérez-Villadóniga and Rodríguez-Álvarez, 2017a), we opt for binary variables. Specifically, we include four binary variables capturing the attainment of primary (*D_edu_primary)*, lower secondary (*D_edu_lowsec)*, upper secondary (*D_edu_upsec)*, vocational and training education (*D_edu_training)*, with tertiary or post-tertiary education serving as the reference category.

To control for job type, we include a set of dummies reflecting categories of the International Standard Classification of Occupations (ISCO). In detail, *D_manager* is equal to 1 if the respondent is a manager, and 0 otherwise; *D_professionals* is 1 if respondent is a professional. *D_admin* is 1 for clerical support workers. *D_service_workers* is 1 for employees in service and sales areas. *D_craft_workers* is 1 if the respondent works in craft and related trades. The category of plant and machine operators and assemblers serves as the control group.

Another relevant aspect of the data refers to the sector of activity (NACE Rev. 2). We include 3 dummies: *D_manufacturing*, *D_trade* and *D_construction*. The reference group is *Services*.

Other control variables included in the estimation are marital status, as well as regional dummies. *D_ever_married* takes value 1 if the individual has ever been married. Moreover, we employ the interaction between *D_Female* and *D_ever_married* to capture the simultaneous effect of gender and marriage. *Single* is the control group. Four dummy variables, *D_North-East*, *D_Centre*, *D_South* and *D_Islands*, account for the macro-regions of residence according with the NUTS 1 classification. *North_West* is the omitted group. Moreover, we add *D_access_Internet*, which is equal to 1 if the respondent declares to have an Internet connection at home for personal use; and *D_rural_areas*, which is equal to 1 for individuals having residence in rural zones, 0 for individuals with residence in towns, cities and suburbs.

Although the gender literature also points to health status, managerial responsibility, organizational position, and firm-level characteristics as potentially relevant determinants of wage conversion efficiency, indicators for these dimensions are not available in the EU-SILC dataset. We return to this limitation, and its implications for future research, in the Conclusions.

Finally, we include a dummy variable for our main variable of interest, gender. *D_Female* is equal to 1 if the respondent is female, and 0 for males. As explained in the introduction, we will include it not only in the frontier but also in the inefficiency term when the model allows for that possibility.

Table 1 presents descriptive statistics for the 6,347 individuals in our sample. Men comprise 57.5% and women 42.5%. Average annual earnings are €24,342 for men and €21,116 for women, so female wages are 86.7% of male wages—high relative to similar countries, reflecting Italy's low gender wage gap (Blau and Kahn, 1996).

Women work fewer hours per week (36 vs. 40) and have slightly less work experience (19 vs. 21 years). Education levels show 1.3% with primary or less, 20% lower secondary, 46.4% upper secondary, 2% vocational, and 30% bachelor's or postgraduate degrees.

Job composition is 1.8% managers, 49% professionals, 16.8% clerical support, 12.4% service/sales, 11.4% craft workers, and 8.7% machine operators. By sector, 53.5% work in services, 29.7% in manufacturing, 10.9% in trade, and 5.9% in construction.

Marital status is relevant to the gender pay gap. Our sample is 54.4% married, 7.7% separated, and 37.8% single; combining married and separated gives 62.2% ever-married.

Table 1. Descriptive statistics of the sample

Variables	All sample			Males			Females		
Wage (In-work income)	6,346	22970.69	13103	3,648	24341.78	15109.96	2,698	21116.32	9443.56
Worked hours	6,346	38.284	6.225	3,648	39.822	5.204	2,698	36.203	6.857
D_Female	6,346	0.425	0.494						
Work experience	6,346	20.390	10.534	3,648	21.134	10.628	2,698	19.384	10.321
D_edu_primary	6,346	0.013	0.113	3,648	0.016	0.126	2,698	0.009	0.092
D_edu_lowsec	6,346	0.203	0.403	3,648	0.255	0.436	2,698	0.134	0.341
D_edu_upsec	6,346	0.464	0.499	3,648	0.478	0.500	2,698	0.445	0.497
D_edu_training	6,346	0.021	0.144	3,648	0.018	0.132	2,698	0.026	0.158
D_manager	6,346	0.018	0.133	3,648	0.023	0.151	2,698	0.011	0.105
D_professional	6,346	0.490	0.500	3,648	0.428	0.495	2,698	0.573	0.495
D_admin	6,346	0.168	0.374	3,648	0.145	0.352	2,698	0.198	0.399
D_service_worker	6,346	0.124	0.330	3,648	0.103	0.304	2,698	0.152	0.359
D_craft_workers	6,346	0.114	0.318	3,648	0.179	0.383	2,698	0.026	0.160
D_manufacturing	6,346	0.297	0.457	3,648	0.387	0.487	2,698	0.174	0.379
D_trade	6,346	0.109	0.312	3,648	0.106	0.308	2,698	0.113	0.316
D_construction	6,346	0.059	0.236	3,648	0.095	0.294	2,698	0.011	0.103
D_ever_married	6,346	0.622	0.485	3,648	0.621	0.485	2,698	0.622	0.485
D_rural_areas	6,346	0.209	0.406	3,648	0.215	0.411	2,698	0.200	0.400
D_access_internet	6,346	0.919	0.274	3,648	0.913	0.282	2,698	0.926	0.262
D_NorthEast	6,346	0.266	0.442	3,648	0.267	0.442	2,698	0.266	0.442
D_Centre	6,346	0.270	0.444	3,648	0.265	0.442	2,698	0.275	0.447
D_South	6,346	0.168	0.374	3,648	0.174	0.379	2,698	0.159	0.365
D_Islands	6,346	0.062	0.241	3,648	0.059	0.236	2,698	0.066	0.248

Source(s): Authors' own elaborations on data from EU-SILC for Italy

Geographically, 23.4% live in North-West Italy (control), 26.6% North-East, 27% Central, 16.8% South, and 6.2% Islands.

5. Empirical strategy

5.1 Econometric model

Our objective is to measure the differences between the potential income of each worker, given their socio-economic characteristics, and the income received. To achieve this, we estimate a Mincer equation within a stochastic frontier framework. Our more general model is the following:

$$lnW = \mathrm{f}(\mathrm{X},\mathrm{G}) + \mathrm{v} - \mathrm{u}(G) \tag{6}$$

where the dependent variable is the natural logarithm of the annual wage, X are a set of explanatory variables (worked hours, education, experience) along with control variables for sector, occupation, region and some personal characteristics (place of residence, marital status, access to Internet). G is a dummy variable for gender (1 if female). The error term consists of two components: v which is a random variable accounting for noise, and u, which is a one-sided random variable that accounts for inefficiency (the distance to the frontier) and that we interpret as the wage gap. In two of the models estimated below, we will allow the inefficiency term to be a function of gender.

An important feature of our specification is the use of annual wage as the dependent variable instead of the more common hourly wage (e.g. Garcia-Prieto and Gómez-Costilla, 2017). Using log hourly wages implicitly imposes a coefficient of 1 on the log of hours worked when the equation is rearranged, a restriction that may not be supported by the data. For this reason, we follow the argument of Blau and Kahn (1996), who discuss this issue while estimating models with both wage measures.

Including the gender dummy in the frontier allows men and women to have different wage frontiers. A negative coefficient for women would indicate a lower potential wage for women with the same observed characteristics. In addition, gender is included in the inefficiency model to capture possible gender differences in the ability to reach potential earnings.

Specifically, we estimate the heteroskedastic model proposed by Caudill *et al*. (1995), where u_i is distributed as Half-Normal, $u_i \sim N^+(0, \sigma_{ui}^2)$. In our case, the variance of the pre-truncated distribution of u is a function of gender, expressed as:

$$\sigma_{ui}^2 = \exp(\delta_0 + \delta_1 G_i) \tag{7}$$

It is worth noticing that the inefficiency term, u, is interpreted as capturing wage-setting frictions and unobserved constraints that prevent individuals from attaining their potential wage, conditional on observable characteristics. Rather than reflecting inherent productivity, we consider u as a measure of market-driven earnings shortfalls (Hwu *et al.*, 2021).

The inefficiency term, u, is interpreted here as capturing wage-setting frictions and unobserved constraints that prevent individuals from attaining their potential wage. However, we must acknowledge a fundamental identification challenge: u is a composite statistical residual. It cannot perfectly disentangle structural market frictions from unobserved individual productivity differences, or idiosyncratic labour supply preferences. Therefore, we interpret our results as an estimation of “conversion efficiency” – the relative success of groups in translating observable human capital into earnings – while remaining cautious about attributing the entire residual to institutional discrimination.

As stated in the Introduction, we will also estimate two restricted models to examine the role of gender in wage determination. First, we will consider a model where gender influences only the frontier, introducing the gender dummy only as a frontier shifter. In this model, we assume that the inefficiency term u follows a half-normal distribution, which is the same for men and women. In the second model, we introduce the gender dummy only in the inefficiency term, using the same specification as in Equation (7).

In the Online Appendix A, we report the analytical derivation of the calculation of the gender wage gap.

5.2 Further issues: robustness check and addressing potential sample selection

To ensure results are not driven by labour supply differences, we estimate models using ln (Hourly Wage) as a dependent variable. Both approaches yield qualitatively similar results, confirming robustness of our main results. In doing this and like in our main regressions, we include a rich set of controls for observable characteristics affecting labour force participation (age, marital status, children, part-time status) [4]. This ensures that observed gender differences in wage inefficiency primarily reflect frictions in wage setting rather than sample composition or labour supply constraints.

An important issue in studies measuring the gender wage gap is sample selection. Since wage equations are estimated only for individuals who work, data from non-working individuals are excluded, creating an econometric sample selection problem. If women who choose to work differ systematically from those who do not (for example, being more educated), OLS estimates will be biased.

The literature generally suggests the presence of positive selection in female labour market participation: women who work tend to have above-average unobserved characteristics associated with higher wages. As a result, the uncorrected wage frontier may appear artificially high, making the wage gap or inefficiency seem larger than it actually is for the average woman.

The standard method to address this issue is the Heckman Two-Step Selection Model, which estimates a wage equation together with a participation equation that includes both workers and non-workers. This allows the estimation of the Inverse Mills Ratio, which is then included as an explanatory variable in the wage equation.

However, we cannot apply this method in our study. As shown by Greene (2010), the Heckman procedure cannot be directly applied in stochastic frontier models. Heckman assumes normally distributed errors in the wage equation, while SFA models use an asymmetric composed error term. Simply adding the Inverse Mills Ratio does not correctly capture how selection affects the potential frontier and leads to inconsistent estimates.

Greene (2010) proposed correcting sample selection in a stochastic frontier framework using Maximum Simulated Likelihood, deriving a selection-adjusted likelihood function for a frontier model. However, this specification was developed only for the standard homoscedastic model (ALS77). To our knowledge, a likelihood function for the heteroskedastic models we estimate, where the variance of inefficiency can depend on gender, has not yet been developed.

While a formal selection-adjusted likelihood function for heteroskedastic SFA models has not yet been developed, we have applied the "manual" procedure by employing two steps of estimations [5]. We find that the IMR coefficient is not significant, suggesting that the selection is not significantly biasing the results.

6. Econometric results

Table 2 reports estimated models for Italian individuals from EU-SILC 2021. We present two estimations: one using the full sample and another restricted to individuals working 40 h to avoid issues related to hours-worked specifications (Blau and Kahn, 2017; Goldin and Katz, 2016). Model diagnostics show a significant λ parameter, indicating the relevance of inefficiency effects and supporting the use of stochastic frontier models over ordinary least squares [6]. Finally, the CFG95 model with *D_Female* on both the frontier and the variance of

Table 2. Estimation of the wage stochastic frontier models

	Full sample (1) ALS77	(2) CFG95 - U	(3) CFG95 - both	Only 40 h (4) ALS77	(5) CFG95 - U	(6) CFG95 - both
Intercept	7.7704*** (0.4600)	7.1267*** (0.4921)	7.6301*** (0.4947)	10.1047*** (0.0308)	10.0661*** (0.0292)	10.0908*** (0.0328)
Log(Worked hours)	0.3174*** (0.0574)	0.3837*** (0.0605)	0.3216*** (0.0616)			
D_Female	−0.1236*** (0.0141)		−0.0376* (0.0198)	−0.1133*** (0.0181)		−0.0252 (0.0196)
Log(Work experience)	0.1382*** (0.0076)	0.1428*** (0.0070)	0.1374*** (0.0078)	0.1362*** (0.0041)	0.1384*** (0.0035)	0.1369*** (0.0039)
D_edu_ primary	−0.4281*** (0.0327)	−0.4238*** (0.0370)	−0.4259*** (0.0376)	−0.5034*** (0.0707)	−0.4920*** (0.0687)	−0.4997*** (0.0694)
D_edu_lowsec	−0.3435*** (0.0251)	−0.3428*** (0.0252)	−0.3462*** (0.0259)	−0.3545*** (0.0256)	−0.3511*** (0.0255)	−0.3584*** (0.0259)
D_edu_upsec	−0.2263*** (0.0016)	−0.2228*** (0.0062)	−0.2257*** (0.0036)	−0.2458*** (0.0056)	−0.2382*** (0.0057)	−0.2471*** (0.0062)
D_edu_ training	−0.1972*** (0.0179)	−0.2072*** (0.0136)	−0.2019*** (0.0166)	−0.1964*** (0.0190)	−0.2000*** (0.0100)	−0.2029*** (0.0160)
D_manager	0.6203*** (0.0624)	0.5804*** (0.0647)	0.6026*** (0.0649)	0.6038*** (0.0743)	0.5801*** (0.0681)	0.5858*** (0.0738)
D_ professional	0.1824*** (0.0105)	0.1582*** (0.0124)	0.1798*** (0.0095)	0.1657*** (0.0267)	0.1490*** (0.0276)	0.1614*** (0.0276)
D_admin	0.0952*** (0.0075)	0.0687*** (0.0096)	0.0935*** (0.0090)	0.0680** (0.0276)	0.0454 (0.0278)	0.0653** (0.0294)
D_service_ worker	−0.0075 (0.0067)	−0.0319*** (0.0049)	−0.0060 (0.0054)	−0.0493** (0.0237)	−0.0695*** (0.0236)	−0.0503** (0.0247)
D_craft_ workers	−0.0592*** (0.0203)	−0.0625*** (0.0223)	−0.0573*** (0.0195)	−0.0535*** (0.0190)	−0.0548*** (0.0207)	−0.0526** (0.0205)
D_ manufacturing	0.0412*** (0.0086)	0.0513*** (0.0073)	0.0383*** (0.0093)	0.0179* (0.0105)	0.0232* (0.0136)	0.0152 (0.0106)
D_trade	−0.0484*** (0.0185)	−0.0449** (0.0196)	−0.0540*** (0.0194)	−0.0399 (0.0271)	−0.0413 (0.0299)	−0.0457 (0.0284)
D_ construction	−0.0514*** (0.0144)	−0.0304** (0.0145)	−0.0548*** (0.0135)	−0.0678*** (0.0214)	−0.0537** (0.0214)	−0.0714*** (0.0210)
D_ever_ married	0.0971*** (0.0152)	0.0522*** (0.0197)	0.0971*** (0.0150)	0.0818*** (0.0110)	0.0497*** (0.0178)	0.0813*** (0.0108)
D_Female*D_ ever_married	−0.0608*** (0.0126)		−0.1614*** (0.0325)	−0.0800*** (0.0211)		−0.1711*** (0.0502)
D_rural_areas	−0.0289* (0.0150)	−0.0260* (0.0143)	−0.0279** (0.0135)	−0.0288* (0.0158)	−0.0223 (0.0136)	−0.0288** (0.0135)
D_access_ internet	0.0614*** (0.0122)	0.0601*** (0.0099)	0.0599*** (0.0097)	0.0759*** (0.0171)	0.0756*** (0.0149)	0.0784*** (0.0152)
D_NorthEast	−0.0160*** (0.0021)	−0.0160*** (0.0020)	−0.0156*** (0.0019)	−0.0070*** (0.0021)	−0.0036** (0.0014)	−0.0049*** (0.0015)
D_Centre	−0.1029*** (0.0028)	−0.1037*** (0.0025)	−0.1027*** (0.0027)	−0.0726*** (0.0025)	−0.0715*** (0.0030)	−0.0722*** (0.0023)
D_South	−0.1442*** (0.0085)	−0.1410*** (0.0070)	−0.1438*** (0.0080)	−0.1439*** (0.0030)	−0.1357*** (0.0040)	−0.1455*** (0.0029)
D_Islands	−0.1210*** (0.0069)	−0.1273*** (0.0061)	−0.1239*** (0.0068)	−0.1383*** (0.0024)	−0.1446*** (0.0048)	−0.1389*** (0.0048)
Var(u)						
Intercept	−1.0594*** (0.1282)	−1.3065*** (0.1632)	−1.1547*** (0.1361)	−1.1305*** (0.1459)	−1.3561*** (0.1834)	−1.2379*** (0.1690)

(*continued*)

Table 2. Continued

	Full sample (1) ALS77	(2) CFG95 - U	(3) CFG95 - both	Only 40 h (4) ALS77	(5) CFG95 - U	(6) CFG95 - both
D_Female		0.6944***	0.5481***		0.6771***	0.5669***
		(0.1175)	(0.1369)		(0.1153)	(0.1164)
D_Female*D_ ever_married		−0.3272	−0.6546**		−0.1917	−0.5853*
		(0.2050)	(0.2710)		(0.2074)	(0.3372)
N. of obs.	6,346	6,346	6,346	3,415	3,415	3,415
*N*_clust	5	5	5	5	5	5
Log-likelihood	−3,422	−3,449	−3,382	−1,726	−1,740	−1,708
AIC	6,940	6,994	6,864	3,548	3,576	3,516
sigma_u	0.589	.	.	0.568	.	.
sigma_v	0.242	0.251	0.244	0.235	0.241	0.236
Lambda	2.429	.	.	2.418	.	.

Note(s): Significance levels: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Clustered standard errors by macro-regions are in parentheses

Source(s): Authors' own elaborations on data from EU-SILC for Italy

u is the preferred specification for both samples, as documented by the lowest value of the Akaike Information Criterion (AIC) statistic (Burnham and Anderson, 2004) [7].

To explore the impact of gender, we estimate three models: the first one is the standard specification of Aigner *et al*. (1977) (hereafter, ALS77), where the gender dummy variable (*D_Female*) enters just in the frontier. The two other models follow the CFG95 specification. We first treat *D_Female* as a determinant of the variance of the inefficiency term, and then, we include the gender dummy in both the frontier and the variance of u.

Starting with the results from the models that use the full sample, all three specifications provide very similar findings. The estimates of the parameters of all the control variables are significant and align with our expectations in terms of sign. The estimated parameters for both worked hours and experience are positive and significant across all models.

Regarding the impact of education, our findings align predictably with existing research (e.g. Oglobin and Brock, 2005). Indeed, the estimated coefficients for the dummies capturing the ISCED of individuals are significant and negative, supporting the well-established notion that individuals with tertiary education earn higher wages compared to those with lower educational attainment.

Interesting findings emerge when examining the estimated coefficients for job types. These effects are always significant but mixed in terms of sign. Specifically, positive coefficients were found for managers, professionals and clerical support workers. Conversely, negative coefficients were observed for service and craft workers, suggesting lower wages compared to the control group of machine operators. These results reinforce the evidence found by previous literature, such as Perez-Villadóniga *et al*. (2025), which documents higher salaries for individuals in decision-making positions.

As for the sector of activity is concerned, workers in Manufacturing receive higher wages than those employed in Services (the estimated parameter is always significant and positive). The opposite happens for Trade and Construction, for which we estimate negative coefficients.

In addition, while we estimate positive coefficients for both *D_ever_married* and *D_Internet*, we find a negative effect of living in a rural area. Finally, not surprisingly, the levels of wage achieved in the North-West of Italy are higher than in other regions. In this regard, we report in the Online Appendix B the results referred to the sensitivity analysis by geographical macro-regions (North *vs* South), which typically characterizes the Italian economy duality. We find evidence of robustness in the main results.

6.1 The role of gender

We examine the role of gender in wage outcomes within the stochastic frontier framework. As discussed in Sections 3 and 5, gender may influence wages by affecting both the level of potential wages and the dispersion of wages around the frontier [8].

6.1.1 Gender and the wage frontier. Across all specifications, the coefficient of the female dummy in the wage frontier is negative and statistically significant, confirming the persistent gender penalty found in traditional literature (Blau and Kahn, 1992; Malkiel and Malkiel, 1973). This implies that, given the same observable characteristics, the estimated wage frontier for women is lower than for men. In other words, women face a lower potential wage even with similar human capital, occupation, and job characteristics.

This result aligns with extensive literature showing gender differences in wage setting, promotions, and access to high-paying jobs (MacPherson and Hirsch, 1995; Adamchik and King, 2007; Blau and Kahn, 2017). While this result is consistent with the presence of gender-based wage-setting frictions (Hwu *et al.*, 2021), it must be interpreted with caution as it may also capture gender-specific unobserved productivity or different labour supply preferences for non-pecuniary job attributes.

6.1.2 Gender and wage dispersion around the frontier. A key feature of our analysis is modelling gender differences in the variance of the inefficiency term. When gender affects the variance of *u*, women show significantly different wage dispersion around the frontier compared with men. In the baseline specification, inefficiency variance is higher for women, indicating a wider distribution of wage shortfalls relative to their frontier.

It is important to stress that the inefficiency term u is an empirical construct, estimated directly from the model, that captures the systematic distance between observed and potential wages. What the model formally identifies is therefore the magnitude and statistical properties of this distance, and how they vary by gender; the attribution of this distance to specific mechanisms, such as discrimination, bargaining frictions, or institutional constraints, remains an economic interpretation rather than a direct empirical finding. We adopt this interpretation because it is consistent with the broader literature on gender wage inequality, but we cannot rule out that part of the estimated inefficiency reflects unobserved productivity differences or heterogeneous preferences for non-pecuniary job attributes that are not captured by our control variables.

This finding is consistent with evidence highlighting the distributional dimension of the gender wage gap (Perez-Villadóniga and Rodríguez-Alvarez, 2017a). In particular, it supports “sticky floor” mechanisms, where women are more concentrated at the lower end of the wage distribution, increasing dispersion below the frontier.

6.1.3 Gender and marital status. We also analyse the interaction between gender and marital status. When included in the frontier, the interaction between being female and ever married is negative, meaning marriage reduces women’s potential wages relative to men. This aligns with evidence of marriage penalties for women and premiums for men (Blau and Kahn, 1992; Malkiel and Malkiel, 1973).

When the interaction term *D_Female*D_ever_married* is introduced in the inefficiency component, results show a more complex pattern. Although women generally display greater dispersion, the interaction reduces dispersion for married women. This occurs because marriage lowers the estimated frontier for women, reducing the distance between observed wages and the frontier.

6.1.4 The calculation of GWG. Tables 3 and 4 present estimates of the gender wage gap (GWG) for the three models, using both the full sample and a subsample of individuals working 40 h per week, as described in the Online Appendix A.

Previous studies using wage frontier models rarely compute the GWG in euros. For example, Perez-Villadóniga and Rodríguez-Alvarez (2017a, b) calculate a wage gap index measuring the percentage of potential wage achieved by workers, but not the GWG in euros.

Our results show notable differences across models, indicating that the placement of the gender dummy is an important modelling choice. This pattern holds in both samples. For the

Table 3. Estimation of the gender wage gap: full sample (€/year)

	ALS77 D_Female on frontier only	CFG95 D_Female on Var(u) only	D_Female on both
$E[lnW \mid X = \overline{X}, G = 0]$	20,230	20,031	20,427
$E[lnW \mid X = \overline{X}, G = 1]$	17,593	17,601	18,171
Gender Wage Gap	*2,637*	*2,430*	*2,256*

Source(s): Authors' own elaborations on data from EU-SILC for Italy

Table 4. Estimation of the gender wage gap: individuals working 40 hours (€/year)

	ALS77 D_Female on frontier only	CFG95 D_Female on Var(u) only	D_Female on both
$E[lnW \mid X = \overline{X}, G = 0]$	20,036	19,942	20,240
$E[lnW \mid X = \overline{X}, G = 1]$	17,602	17,128	17,731
Gender Wage Gap	*2,435*	*2,814*	*2,510*

Source(s): Authors' own elaborations on data from EU-SILC for Italy

full sample, the GWG estimated with the CFG95 specification is lower than that from the ALS77 model. However, for individuals working 40 h per week, the CFG95 models produce higher GWGs than the ALS77 specification. These differences are not only statistically significant but also economically meaningful: depending on the specification, the estimated GWG varies by several percentage points of average earnings, a magnitude comparable to other structural determinants of wages such as sector of activity or occupational category. This underscores that model specification is not a purely technical issue but has tangible implications for assessing the economic relevance of the gender wage gap. The differences in magnitude across specifications arise because the ALS77 model attributes gender differences exclusively to a shift in the wage frontier, whereas the CFG95 specifications also allow gender to affect the dispersion of the inefficiency term; when women display both a lower frontier and higher inefficiency variance, models that incorporate the latter tend to produce larger estimated gaps, particularly among subgroups such as full-time workers for whom dispersion around the frontier is more pronounced.

Overall, gender plays a complex role in wage determination. Women face both a lower estimated wage frontier and different wage dispersion around that frontier. Furthermore, the estimated gender wage gap depends strongly on how gender is modelled in the stochastic frontier framework.

7. Conclusions

This study examines gender differences in wage inefficiency in the Italian labour market, using a stochastic frontier approach that accounts for heteroskedasticity across genders. Our results reveal that women, on average, experience higher wage inefficiency than men, and the variance of this inefficiency is also larger. Importantly, these findings do not imply lower female productivity; rather, they reflect structural frictions that prevent women from fully converting their skills and contributions into earnings. The observed pattern appears consistent with the presence of "sticky floor" effects at the lower end of the wage distribution and "glass ceiling" effects at the upper end.

Our findings suggest that gender differences in Italy may be driven by "conversion efficiency" rather than inherent productivity. Policymakers could prioritize measures that reduce information and bargaining frictions. Specifically, implementing wage transparency laws could help women navigate the high-variance "sticky floor" sectors identified in our results. Additionally, addressing the "marriage penalty", which we find lowers women's potential wage ceilings, requires structural reforms in childcare and flexible work arrangements to ensure human capital is consistently rewarded regardless of marital status.

Beyond broad policy recommendations, it is useful to identify more explicitly the mechanisms through which employers, organizational practices, and institutional arrangements may influence wage conversion efficiency. At the firm level, transparent and standardized pay-setting procedures, structured promotion criteria, and formal salary negotiation protocols can reduce the scope for discretionary, bargaining-based wage determination that tends to disadvantage women. Organizational practices such as flexible working arrangements, parental leave policies, and managerial accountability for pay-equity outcomes can further help women translate their human capital into earnings without being penalized for career interruptions or reduced availability. At the institutional level, wage transparency legislation, sector-level collective bargaining, and equality-oriented certification schemes (e.g. linked to corporate social responsibility standards) can strengthen incentives for firms to close conversion efficiency gaps. Making these channels explicit helps bridge the econometric evidence presented here with concrete levers for policy and organizational practice.

The analysis has important implications for labour market policies. Reducing barriers to wage transparency, implementing mentorship and promotion programs for women in high-variance sectors, and reforming part-time and flexible work arrangements can help mitigate inefficiencies and narrow the gender wage gap. Policymakers should focus not only on equal pay for equal work but also on structural measures that enhance the ability of women to translate productivity into earnings across all levels of the labour market.

Finally, future research could extend this analysis by using multi-year datasets or exploring intersectional factors such as age and life-cycle effects, health status, ethnicity, managerial responsibility and organizational position, and occupational sector. Data limitations prevented us from directly controlling for health status and firm-level characteristics such as corporate social responsibility and equality-oriented workplace practices in the present analysis; incorporating such indicators, where available, represents a promising avenue for deepening the intersectional analysis of conversion efficiency. Despite data limitations, our findings provide clear evidence that addressing structural and institutional frictions is key to fostering a more equitable and efficient labour market.

Acknowledgments

The authors are grateful to David Roibás and Filippo Domma for their valuable discussions and support. They also thank Kai Sun and William Greene for their helpful comments and suggestions. The authors acknowledge the useful feedback received from participants at the 66th Annual Meeting of the Italian Economic Association (Società Italiana degli Economisti) and from the anonymous referees. Graziella Bonanno gratefully acknowledges financial support from the Fondo sostegno aree socio-umanistiche – Cda del 26.03.2021 – quota DESF. Any remaining errors are the sole responsibility of the authors.

Notes

1. See Álvarez and Arias (2014) for a survey on stochastic frontier modelling.
2. The guidelines and the description of EU-SILC target variables (2021 version), can be found in https://circabc.europa.eu/sd/a/f8853fb3-58b3-43ce-b4c6-a81fe68f2e50/Methodological%20guidelines%202021%20operation%20v4%2009.12.2020.pdf
3. The in-work income includes both cash and non-cash income from work. The tax at source and the social insurance contributions are deducted. Data on wage are in national currency (euro).

4. Untabulated results are available upon request from the authors.
5. Standard Heckman procedures are incompatible with the asymmetric error structure of heteroskedastic SFA models. However, we have decided to find data for testing the presence of sample selection bias in our basic ALS77 model, by employing two steps of estimations. We, first, run the selection equation, which means a probit model where a dummy equal to 1 for workers (and 0 for individuals not perceiving any income and not working) is the dependent variable. As regressors for this ancillary equation, we use unemployment transfers, personal characteristics already in your data, such as education and region. From this probit model, we calculate the Inverse Mills Ratio (IMR) for each individual. Second, the corrected wage equation was estimated, where we include the IMR as an additional regressor in the ALS77 model. Untabulated results are available upon request from the authors.
6. The parameter λ is equal to $\frac{\sigma_u}{\sigma_v}$, where a zero value indicates that deviations from the frontier are only due to random error, while values greater than 1 indicate that the distance from the frontier is mostly due to inefficiency.
7. AIC is equal to [2*k-2*Log-likelihood], where k is the number of all estimated parameters.
8. We run the same estimations by using ln(Hourly wage) as the dependent variable, as mentioned in Section 5.2. Untabulated results confirm that the patterns of gender differences in wage inefficiency persist, indicating that labour supply differences do not drive the main findings.

Supplementary material
The supplementary material for this article can be found online.

References

Adamchik, V. and King, A. (2007), "Labor market efficiency in Poland: a stochastic wage Frontier analysis", *The International Journal of Business and Finance Research*, Vol. 1 No. 2, pp. 41-51.

Aigner, D., Lovell, C.K. and Schmidt, P. (1977), "Formulation and estimation of stochastic Frontier production function models", *Journal of Econometrics*, Vol. 6 No. 1, pp. 21-37, doi: 10.1016/0304-4076(77)90052-5.

Algieri, B. and Alvarez, A. (2023), "Assessing the ability of regions to attract foreign tourists: the case of Italy", *Tourism Economics*, Vol. 29 No. 3, pp. 788-811, doi: 10.1177/13548166211068669.

Álvarez, A.M. and Arias, C. (2014), "A selection of relevant issues in applied stochastic frontier analysis", *Economics and Business Letters*, Vol. 3 No. 1, pp. 3-11.

Bashford-Fernández, J.M. and Rodríguez-Álvarez, A. (2019), "Wage frontiers in pre and post-crisis Spain: implications for welfare and inequality", *Social Indicators Research*, Vol. 143 No. 2, pp. 579-608, doi: 10.1007/s11205-018-1990-4.

Battese, G.E. and Coelli, T.J. (1995), "A model for technical inefficiency effects in a stochastic Frontier production function for panel data", *Empirical Economics*, Vol. 20 No. 2, pp. 325-332, doi: 10.1007/bf01205442.

Blau, F.D. and Kahn, L.M. (1992), "The gender earnings gap: learning from international comparisons", *The American Economic Review*, Vol. 82 No. 2, pp. 533-538.

Blau, F.D. and Kahn, L.M. (1996), "Wage structure and gender earnings differentials: an international comparison", *Economica*, Vol. 3, Supp, pp. S29-S62, doi: 10.2307/2554808.

Blau, F.D. and Kahn, L.M. (2007), "The gender pay gap: have women gone as far as they can?", *Academy of Management Perspectives*, Vol. 21 No. 1, pp. 7-23.

Blau, F.D. and Kahn, L.M. (2017), "The gender wage gap: extent, trends, and explanations", *Journal of Economic Literature*, Vol. 55 No. 3, pp. 789-865, doi: 10.1257/jel.20160995.

Burnham, K.P. and Anderson, D.R. (2004), "Multimodel inference: understanding AIC and BIC in model selection", *Sociological Methods & Research*, Vol. 33 No. 2, pp. 261-304.

Caparrós Ruiz, A. (2025), "Gender wage gap among highly educated workers: some evidence from Spain", *Journal of Economic Studies*, Vol. 52 No. 6, pp. 1171-1188, doi: 10.1108/jes-06-2024-0371.

Card, D., Cardoso, A.R., Heining, J. and Kline, P. (2018), "Firms and labor market inequality: evidence and some theory", *Journal of Labor Economics*, Vol. 36 No. S1, pp. S13-S70, doi: 10.1086/694153.

Caudill, S.B., Ford, J.M. and Gropper, D.M. (1995), "Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity", *Journal of Business and Economic Statistics*, Vol. 13 No. 1, pp. 105-111, doi: 10.1080/07350015.1995.10524583.

Díaz, M.A. and Sánchez, R. (2011), "Gender and potential wage in Europe: a stochastic Frontier approach", *International Journal of Manpower*, Vol. 32 No. 4, pp. 410-425, doi: 10.1108/01437721111148531.

Garcia-Prieto, C. and Gómez-Costilla, P. (2017), "Gender wage gap and education: a stochastic Frontier approach", *International Journal of Manpower*, Vol. 38 No. 3, pp. 504-516, doi: 10.1108/ijm-11-2015-0186.

Goldin, C. (2014), "A grand gender convergence: its last chapter", *American Economic Review*, Vol. 104 No. 4, pp. 1091-1119, doi: 10.1257/aer.104.4.1091.

Goldin, C. and Katz, L.F. (2016), "A most egalitarian profession: pharmacy and the evolution of a family-friendly occupation", *Journal of Labor Economics*, Vol. 34 No. 3, pp. 705-746, doi: 10.1086/685505.

González, I., Seo, B. and Floro, M.S. (2022), "Gender wage gap, gender norms, and long-term care: a theoretical framework", *Feminist Economics*, Vol. 28 No. 3, pp. 84-113, doi: 10.1080/13545701.2022.2081353.

Greene, W. (2010), "A stochastic Frontier model with correction for sample selection", *Journal of Productivity Analysis*, Vol. 34 No. 1, pp. 15-24, doi: 10.1007/s11123-009-0159-1.

Herzog, H.W. Jr, Hofler, R.A. and Schlottmann, A.M. (1985), "Life on the Frontier: migrant information, earnings and past mobility", *The Review of Economics and Statistics*, Vol. 67 No. 3, pp. 373-382, doi: 10.2307/1925965.

Hwu, S.T., Fu, T.T. and Tsay, W.J. (2021), "Estimation and efficiency evaluation of stochastic Frontier models with interval dependent variables", *Journal of Productivity Analysis*, Vol. 56 No. 1, pp. 33-44, doi: 10.1007/s11123-021-00609-w.

Jane, W.J. (2013), "Overpayment and reservation salary in the nippon professional baseball League: a stochastic Frontier analysis", *Journal of Sports Economics*, Vol. 14 No. 6, pp. 563-583, doi: 10.1177/1527002511433857.

Lang, G. (2005), "The difference between wages and wage potentials: earnings disadvantages of immigrants in Germany", *The Journal of Economic Inequality*, Vol. 3 No. 1, pp. 21-42, doi: 10.1007/s10888-004-7581-4.

MacPherson, D.A. and Hirsch, B.T. (1995), "Wages and gender composition: why do women's jobs pay less?", *Journal of Labor Economics*, Vol. 13 No. 3, pp. 426-471, doi: 10.1086/298381.

Malkiel, B.G. and Malkiel, J.A. (1973), "Male-female pay differentials in professional employment", *The American Economic Review*, Vol. 63 No. 4, pp. 693-705.

Meeusen, W. and van Den Broeck, J. (1977), "Efficiency estimation from cobb-douglas production functions with composed error", *International Economic Review*, Vol. 18 No. 2, pp. 435-444, doi: 10.2307/2525757.

Mincer, J. (1974), *Schooling, Experience, and Earnings*, Columbia University Press, New York, NY.

Nchor, D. and Náplava, R. (2025), "Labour market inequality among central and eastern European countries: an emphasis on the gender employment and pay gap", *Empirical Economics*, Vol. 69 No. 6, pp. 3935-3966, doi: 10.1007/s00181-025-02831-z.

Nguyen, N.T., Phung, V.A., Digdowiseiso, K. and Afni, N. (2026), "Bridging the gender wage gap: the role of education in developing economies", *Journal of Economic Studies*, Vol. 53 No. 3, pp. 474-492.

Ogloblin, C. and Brock, G. (2005), "Wage determination in urban Russia: underpayment and the gender differential", *Economic Systems*, Vol. 29 No. 3, pp. 325-343, doi: 10.1016/j.ecosys.2005.05.002.

Pérez-Villadóniga, M.J. and Rodriguez-Alvarez, A. (2017a), "Analysing wage differentials when workers maximize the return to human capital investment", *Applied Economics*, Vol. 49 No. 42, pp. 4196-4208, doi: 10.1080/00036846.2017.1279267.

Pérez-Villadóniga, M.J. and Rodriguez-Alvarez, A. (2017b), "Comparing the gender gap in gross and base wages", *International Journal of Manpower*, Vol. 38 No. 5, pp. 646-660, doi: 10.1108/ijm-10-2015-0162.

Pérez-Villadóniga, M.J., Rodriguez-Alvarez, A. and Roibas, D. (2025), "An alternative approach for analysing the gender wage gap in managerial positions: a stochastic latent class model", *Applied Economics*, Vol. 58 No. 9, pp. 1732-1745, doi: 10.1080/00036846.2025.2471037.

Polachek, S.W. and Robst, J. (1998), "Employee labor market information: comparing direct world of work measures of workers' knowledge to stochastic Frontier estimates", *Labour Economics*, Vol. 5 No. 2, pp. 231-242, doi: 10.1016/s0927-5371(97)00030-4.

Reifschneider, D. and Stevenson, R. (1991), "Systematic departures from the Frontier: a framework for the analysis of firm inefficiency", *International Economic Review*, Vol. 32 No. 3, pp. 715-723, doi: 10.2307/2527115.

Said, M., Majbouri, M. and Barsoum, G. (2022), "Sticky floors and glass ceilings: gender wage gap in Egypt", *Feminist Economics*, Vol. 28 No. 4, pp. 137-165, doi: 10.1080/13545701.2022.2078499.

Watson, D. (2000), "UK wage underpayment: implications for the minimum wage", *Applied Economics*, Vol. 32 No. 4, pp. 429-440, doi: 10.1080/000368400322606.

Zveglich, J.E. Jr, van der Meulen Rodgers, Y. and Laviña, E.A. (2019), "Expected work experience and the gender wage gap: a new human capital measure", *Economic Modelling*, Vol. 83, pp. 372-383, doi: 10.1016/j.econmod.2019.09.003.

Further reading

Gunderson, M. (1989), "Male-female wage differentials and policy responses", *Journal of Economic Literature*, Vol. 27 No. 1, pp. 46-72.

Kumbhakar, S.C. and Lovell, C.K. (2000), *Stochastic Frontier Analysis*, Cambridge University Press.

O'Neill, J. and Solomon, P. (1993), "Why the gender gap in wages narrowed in the 1980s", *Journal of Labor Economics*, Vol. 11 No. 1, pp. 205-228, doi: 10.1086/298323.

Polachek, S.W. and Yoon, B.J. (1987), "A two-tiered earnings Frontier estimation of employer and employee information in the labor market", *The Review of Economics and Statistics*, Vol. 69 No. 2, pp. 296-302, doi: 10.2307/1927237.

Polachek, S.W. and Yoon, B.J. (1996), "Panel estimates of a two-tiered earnings Frontier", *Journal of Applied Econometrics*, Vol. 11 No. 2, pp. 169-178, doi: 10.1002/(sici)1099-1255(199603)11:2<169::aid-jae373>3.0.co;2-#.

Robinson, M.D. and Wunnava, P.V. (1989), "Measuring direct discrimination in labor markets using a Frontier approach: evidence from CPS female earnings data", *Southern Economic Journal*, Vol. 56 No. 1, pp. 212-218, doi: 10.2307/1059067.

Slottje, D.J., Hirschberg, J.G., Hayes, K.J. and Scully, G.W. (1994), "A new method for detecting individual and group labor market discrimination", *Journal of Econometrics*, Vol. 61 No. 1, pp. 43-64, doi: 10.1016/0304-4076(94)90076-0.

Zhao, X., Jiao, Y. and Wu, D. (2022), "The impact of internet use on labor wage distortions: empirical evidence from China", *Sage Open*, Vol. 12 No. 2, doi: 10.1177/21582440221099290.

Corresponding author
Graziella Bonanno can be contacted at: graziella.bonanno@unical.it