

GENDER WAGE GAP FALSE POSITIVES: A MONTE CARLO SIMULATION

Péter Juhász¹, Yuetong Meng¹, Kata Váradi¹

¹Institute of Finance, Corvinus University of Budapest

e-mail: peter.juhasz@uni-corvinus.hu ,

yuetong.meng@stud.uni-corvinus.hu ,

kata.varadi@uni-corvinus.hu

KEYWORDS

Gender wage gap, Monte Carlo simulation, Measurement errors, EU Pay Transparency Directive.

ABSTRACT

This study examines whether and how often the standard regression-based model would falsely detect gender wage discrimination in a “discrimination-free” world. In the current context of the EU Pay Transparency Directive (EU) 2023/970, we build a Monte Carlo simulation that generates individual wages under a zero-discrimination benchmark, driven by observable skills and realistic wage-setting frictions. The results present non-trivial false-positive rates that rise with multiple gender-related tests, while goodness-of-fit (e.g., adjusted R^2) does not reliably indicate detection risk. The findings highlight the limitations of identification and the sensitivity of specifications in gender wage gap calculation, indicating that enforcement and compliance assessments must consider multiple testing and robustness checks to prevent erroneous conclusions.

INTRODUCTION

Wage inequality is an important dimension of broader economic inequality across the EU and remains closely linked to persistent pay opacity. Directive (EU) 2023/970 was adopted by the European Parliament and the Council on May 10, 2023, and published in the Official Journal on May 17, 2023. Member states are required to complete the transposition by June 7, 2026. This directive aims to strengthen the principle of equal pay for equal work, or work of equal value, between women and men through pay-transparency and enforcement mechanisms (EUR-Lex, 2023).

Implementing pay-transparency legislation is challenging not only because labour-market conditions vary substantially across countries, but also because compliance may impose considerable administrative burdens on firms. The limited research suggests that pay transparency reporting with sanctions can be helpful in narrowing the gender wage gap, but the understanding of the effects of pay reporting obligations on wage outcomes across different policy designs remains insufficient (OECD, 2024). Real-world wage-setting features, such as measurement error, unobserved

productivity, minimum-wage truncation, and upper-tail premia, can distort inference and generate false positive gender effects even when no true discrimination exists.

Thus, this paper evaluates the reliability of regression-based gender wage gap detection under realistic wage-setting frictions using a Monte Carlo simulation under a zero-discrimination benchmark. We quantify the frequency of spurious significance and identify the conditions that amplify detection risk, thereby offering both methodological and policy implications for pay-transparency enforcement. The research question is formulated accordingly: How often do gender-related coefficients become statistically significant when true discrimination is zero?

LITERATURE REVIEW

The evolution of the gender wage gap

The equal pay for equal work principle was already incorporated into the 1957 Treaty of Rome (First Article 119 EEC; later Article 141 EC; currently Article 157 TFEU) (European Commission, 2015). Prior research shows that the gender wage gap is shaped not only by individual characteristics, but also by wage structures, labour-market institutions, and the distinction between explained and unexplained components (Blau & Kahn, 1992, 1996, 1999). Although the gap narrowed substantially after the late 1970s, progress slowed in the 1990s as the improvement in favourable gender-specific factors weakened (Blau & Kahn, 2007).

Gender wage gap research in the EU

Recent European evidence shows that the gender wage gap may be amplified by departmental wage-setting mechanisms, early-career divergence, and glass-ceiling dynamics at the upper end of the wage distribution (Bergmann et al., 2019; Ferri et al., 2022; Segovia-Pérez et al., 2019). High female labour-force participation also does not guarantee equal representation among top earners (Fortin et al., 2017). In addition, observed wage-gap changes after pay-transparency reform may reflect changes in men’s earnings as well as women’s wages (Bennedsen et al., 2023).

Rationale for Monte Carlo Detection-Risk Simulation

Omitted and mismeasured variables. Economists often emphasise that excluding relevant productivity-related variables biases the estimated gender effect. Edgeworth Economics (2023) explain that if important controls such as tenure, job level or prior experience are missing, and these differ systematically by gender, the regression will wrongly attribute part of their effect to gender itself (omitted variable bias). Similarly, measurement error in key controls (for example, using a noisy proxy for actual experience) reduces the estimated return to that variable and shifts some of its explanatory power to the gender dummy.

Sample selection and labour force participation. Eurostat's methodological work (2009) on the Structure of Earnings Survey emphasises that wage-gap estimates can be biased by non-random selection into employment. If women with the highest earnings potential are more likely to participate in the labour market than low-potential women, whereas male participation is less selective, simple comparisons of observed wages can understate or overstate true underlying gaps. Demographic research on the gender wage gap (GWG), accounting for selection into full-time work or particular occupations, similarly shows that ignoring selection can make the estimated wage equations inconsistent.

Model specification and decomposition choices. Eurostat's report (2009) on econometric methods notes that different model specifications, such as functional form, interaction terms, and the choice of reference group, yield different decompositions of the wage gap into "explained" and "unexplained" components. Averkamp et al. (2020) show that standard Oaxaca-Blinder decompositions that omit family-context variables (such as partner characteristics) may misattribute part of the gap to unexplained discrimination, even when it results from household career choices. Incorrect functional forms (for example, linear when returns are non-linear) or missing interactions (such as gender-specific returns to experience) can also distort the estimated gap.

Inadequate control of job structure and sorting. Eurostat (2009) emphasises that the unadjusted indicator conjoins within-job pay differences by sorting across firms, sectors, and occupations. If women are over-represented in low-pay firms or sectors, estimates that do not properly account for employer and job heterogeneity can overstate firm-level discrimination and understate the role of sorting, or vice versa. Recent family-economics decompositions reveal that neglecting how couples decide whose career to prioritise also leads to misinterpretation of the contribution of occupational and sectoral sorting to the gap.

Statistical uncertainty and power issues. Eurostat's methodological study (2009) notes that small samples, noisy data, and an unbalanced gender composition can lead to imprecise estimates and unstable standard errors,

which affect the statistical significance of the estimated gap. In such settings, researchers may fail to detect a true (but modest) wage gap or, symmetrically, may find a statistically significant coefficient driven largely by random variation rather than systematic discrimination. In addition, improper handling of the sampling design and weights can further bias both point estimates and their standard errors.

Interpretation: discrimination versus "unexplained". Averkamp et al. (2020) argue that a substantial part of what is usually labelled the "unexplained" component in wage decompositions may not be pure discrimination but instead originates from omitted household-level characteristics and joint career decisions. Eurostat likewise cautions that even carefully adjusted wage gaps remain composite indicators and should not be interpreted mechanically as direct measures of unequal pay for equal work. Misinterpreting the unexplained residual as discrimination is, therefore, itself a conceptual mistake that can mislead both empirical work and policy debates.

Thus, this study applies Monte Carlo simulation to assess the reliability of regression-based gender wage gap detection under the enforcement context of Directive (EU) 2023/970. Existing studies discuss specific sources of bias in gender wage gap estimation, but fewer studies quantify how often a standard regression framework may generate false-positive gender effects under a zero-discrimination benchmark with multiple realistic frictions combined. By simulating wage-setting environments with realistic frictions and imposing a zero-discrimination benchmark, we estimate how frequently standard econometric specifications produce statistically significant gender effects purely due to model misspecification and distributional features. This approach helps inform both methodological choices and enforcement design, where incorrect classification may otherwise lead to misguided compliance actions. The following methodology therefore distinguishes between the true data-generating process and the estimated regression model used for inference.

METHODOLOGY

To evaluate how standard regression-based gender wage gap models may misidentify discrimination, we develop a Monte Carlo simulation under a zero-discrimination benchmark. Each simulation generates a cross-sectional dataset of N individuals.

Gender is represented by the indicator $F_i \in \{0,1\}$, where $F_i = 1$ denotes female employees and $F_i = 0$ denotes male employees. The parameter w denotes the share of female employees in the simulated sample. This coding is used purely for interpretive convenience: reversing the coding would reverse the sign of the estimated coefficient but would not alter the substantive inference.

In the baseline design, gender and all wage-relevant traits are generated independently by construction. Accordingly, any statistically significant gender effect detected in the estimation stage cannot reflect true structural discrimination, but instead arises from sampling variation, multiple testing, omitted information, or functional-form mismatch.

Each individual is characterized by four wage-influencing traits: Education, Experience, Commitment, and OtherSkills:

$$S_i = (Edu_i, Exp_i, Com_i, OtherSkills_i) \quad (1)$$

Each skill is drawn from a normal distribution:

$$S_{i,k} \sim N(\mu_k, \sigma_k^2) \quad (2)$$

$$\widehat{S}_{i,k} = S_{i,k} + \varepsilon_{i,k} \quad (3)$$

where the measurement error term satisfies:

$$\varepsilon_{i,k} \sim N(0, \sigma_{k,me}^2) \quad (4)$$

independently across individuals and traits.

Equivalently, conditional on the true trait, the measured value follows:

$$\widehat{S}_{i,k} | S_{i,k} \sim N(S_{i,k}, \sigma_{k,me}^2) \quad (5)$$

Thus, the employer does not observe the true productivity component directly, but an unbiased signal centred on the true value, with normally distributed noise. This corresponds to a classical measurement error structure, where the signal is unbiased and independent of the latent trait and other regressors.

The wage paid to an individual is determined by the following Structural Wage Equation:

$$W_i = (\sum_k \beta_k \cdot \widehat{S}_{i,k} + \beta_j \cdot g(Edu_i, Exp_i) - F_i \cdot \delta_A) \cdot (1 - F_i \cdot \delta_R) \quad (6)$$

where absolute discrimination δ_A represents an additive penalty, while δ_R represents proportional discrimination. $g(E_i, X_i)$ is a non-linear “joint premium” term capturing an interaction between education and experience only in the upper tail where we consider the actual trait value without measurement error:

$$g(Edu_i, Exp_i) = \max(0, Edu_i - \mu_{Edu} - \sigma_{Edu} \cdot d) \cdot \max(0, Exp_i - \mu_{Exp} - \sigma_{Exp} \cdot d) \quad (7)$$

The interaction term is a thresholded product: it remains zero for most workers and becomes positive only when both education and experience exceed a high threshold. This reflects a convexity aligned with “superstar” or

“high-flyer” pay mechanisms, where the benefits of combined credentials and experience are realised only above a certain cutoff. The parameter β_j controls the magnitude of this upper-tail complementarity. At the same time, if the calculated wage is less than a pre-specified *Minimum Wage* (mw), the employee receives the Minimum Wage. Thus, these features define the true wage-generating process used in the simulation. The simulation is designed to test whether a standard empirical GWG regression may falsely detect gender effects when the true data-generating process contains no gender discrimination.

As a next step, we estimate a linear wage regression to examine whether discrimination would be inferred from the simulated data. We assume the econometrician can perfectly track each employee's Education, Experience, and Commitment values (no measurement error). However, the *OtherSkills* level is unobservable, and the “superstar” interaction and the minimum-wage effect are also neglected by the estimator. Thus, our estimated econometric specification is as follows:

$$w_i = \alpha + \gamma F_i + \sum_{k \in \{Edu, Exp, Com\}} \beta_k \cdot S_{i,k} + \sum_{k \in \{Edu, Exp, Com\}} \theta_k \cdot (F_i \cdot S_{k,i}) + \varepsilon_i \quad (8)$$

The regression includes a gender dummy and interaction terms, allowing gender-specific returns to skills. Equation (8) reflects a standard reduced-form wage regression used in GWG research. The gender dummy captures conditional mean wage differences, while the interaction terms allow the returns to observed characteristics to differ by gender. This specification is consistent with the broader wage-gap literature, in which estimated gender effects depend on the set of observed controls, the treatment of interactions, and the assumed functional form.

Omitting *OtherSkills* creates an omitted-variable component in the estimation stage, which is useful for studying bias and inference distortions when relevant productivity determinants are unobserved by the econometrician. The non-linear joint premium term and the minimum-wage effect are also excluded from the linear regression, creating functional-form misspecification and particularly affecting the upper tail. This deliberate mismatch between the true data-generating process and the estimated model allows us to quantify how often standard empirical practice may falsely attribute misspecification-induced patterns to gender discrimination.

Statistical significance of gender-related parameters is evaluated using t-tests at $\alpha = 0.05$. We keep separate records of significance levels for all wage-gap-related variables and also track cases in which at least one of those variables is significant.

The simulation is repeated M times. For each replication, coefficient estimates, standard errors, and p-values are

recorded. The empirical frequency of detecting gender effects estimates Type I error rate under the zero-discrimination benchmark.

Parameters varied across simulations include sample size, number of skill variables, measurement error variance, distributional assumptions, correlation structure, and functional form of discrimination. Because true structural parameters are known, deviations between estimated and true discrimination reflect statistical identification limitations rather than data uncertainty.

RESULTS

Baseline – zero discrimination

We first examine the baseline zero-discrimination case, in which no gender wage gap is built into the wage-generating process. Even under this benchmark, however, false positives may arise because the estimated regression omits *OtherSkills*, ignores the superstar effect and the minimum-wage floor, and is evaluated using multiple gender-related terms. We therefore record both the significance rate of the female dummy and the probability that at least one gender-related coefficient becomes significant.

As for a single variable to show up as significant when it should not, the significance rate used (for this paper 5%) gives a proper estimate. For any of multiple variables to be significant, based on An et al. (2013), we may estimate the Type I error using familywise error rate (FWER). To estimate FWER we use the following equation:

$$FWER = 1 - (1 - \alpha)^n \quad (9)$$

where α marks the level of significance applied for individual tests, and n stands for the number of hypothesis tests. Thus, our expected FWER value is

$$FWER = 1 - (1 - 0.05)^4 = 18.55\% \quad (10)$$

When several gender-related terms are tested simultaneously, however, the probability that at least one of them appears significant is higher because of multiple-testing risk.

Table 1 presents the ratio of falsely significant gender wage gap effect. All estimates are in line with expectations: the gender dummy would not appear significantly more often than 5.27% of the cases, and the simulations never showed at least one of the four gender-related variables significant in more than 16.27% of cases, well below the expected 18.55%. More surprisingly, these values showed little difference across adjustments in the number of employees (firm size), trait measurement precision during wage allocation, or the ratio of female employees. The only articulated effect is in the adjusted R-squared: it falls sharply as estimation precision declines. The relatively small number of employees also has a slight adverse effect.

Note that the very high adjusted R-squared observed in most cases is expected, as our random wage-generating process is linear. However, an extensive literature (Chen & Qi, 2023; Spiess & Neumeyer, 2010) exists warning researchers of such high values. As an explanation, authors usually claim overfitting, which occurs when the in-sample fit is strong but the function performs poorly out of sample. In our case, however, we assume that we include the whole population in the regression and test how the significance varies with the randomness of the distribution within the population. As we do not have a time effect, data trends are also irrelevant.

Table 1: False Positive rates under the baseline zero-discrimination setting (%)

N	Gender variable	Baseline	Estimation precision $\sigma_{Edu} = \sigma_{Exp}$ $= \sigma_{Com}$		Female share in sample (w)		
			0.1	10	25	10	1
500	Female	5.15	5.21	5.28	5.28	5.17	-
	Any	16.11	15.96	15.66	15.18	13.50	-
	adj R ²	98.96	99.99	49.22	98.96	98.97	-
1000	Female	5.04	4.78	5.10	4.48	5.10	-
	Any	15.96	15.85	15.94	15.28	15.80	-
	adj R ²	98.98	99.99	49.36	98.99	98.99	-
10000	Female	5.10	4.91	5.15	5.02	5.06	5.05
	Any	15.92	16.08	16.22	16.36	16.00	15.25
	adj R ²	99.01	99.99	50.07	99.01	99.01	99.01
100000	Female	5.14	5.15	5.07	5.10	5.27	5.25
	Any	16.10	16.10	16.06	15.81	16.03	16.27
	adj R ²	99.01	99.99	49.99	99.0	99.01	99.01

Distributional Frictions Amplify Spurious Gender Significance

We next examine how lower-tail and upper-tail distortions affect false-positive risk. Specifically, we vary the minimum-wage level, the strength of the joint Education–Experience (“superstar”) effect, and the omission of *OtherSkills* from the estimated regression.

The results indicate that omitted balanced traits alone do not materially inflate false gender findings, because these traits are generated with the same distribution across men and women. By contrast, distributional distortions matter more. A minimum-wage floor truncates the lower tail of the wage distribution, while the superstar effect stretches the upper tail. Because the estimated model remains linear and excludes both features, it is more prone to spurious significance when such distortions are present.

This distinction is substantively important. The key problem is not merely that some information is omitted, but that the omitted or neglected features change the shape of the wage distribution in ways that the estimated regression does not model well. In that sense, minimum-wage truncation and upper-tail non-linearity are more consequential for false-positive detection than the omission of an otherwise balanced productivity trait.

Sample Size Effects and Combined Distortions

We then combine these distortions to examine whether they offset or reinforce one another. Table 2 shows that false-positive rates increase markedly in the combined-effects setting. The share of replications in which at least one gender-related term is significant rises from 23.40% at $N = 50$ to 40.08% at $N = 10,000$, while the false-significance rate of the female dummy rises from 10.88% to 25.69%. These values are substantially above the baseline levels reported in Table 1.

Table 2: False Positive rates under the combined-effects setting (%)

N	Gender variable	Minimal Wage=1600, Superstar effect (β_j)=1, Omitted traits effect ($\beta_{Otherskills}$) = 1							
		Estimation precision $\sigma_{Edu} = \sigma_{Exp} = \sigma_{Com}$			Female share in sample (w)				
		σ_{Edu}	σ_{Exp}	σ_{Com}					
		0.1	1*	10	50%*	25%	10%	1%	
50	Female	10.88	8.36	4.36	8.36	7.05	5.80	-	
	Any	23.40	19.70	14.70	19.70	17.60	14.40	-	
	adj R ²	99.06	98.08	48.30	98.08	98.08	98.08	-	
100	Female	13.97	9.85	4.11	9.85	8.78	6.81	-	
	Any	28.30	22.70	14.90	22.70	20.30	16.60	-	
	adj R ²	99.03	98.08	48.60	98.08	98.08	98.08	-	
1000	Female	23.60	14.10	3.95	14.10	13.40	12.30	5.99	
	Any	38.13	28.50	14.70	28.50	27.30	24.00	14.70	
	adj R ²	98.98	98.06	49.20	98.06	98.06	98.06	98.06	
10000	Female	25.69	16.07	4.50	16.10	14.70	15.20	11.60	
	Any	40.08	29.78	15.10	29.80	29.10	28.90	23.30	
	adj R ²	98.97	98.06	49.20	98.06	98.06	98.06	98.06	

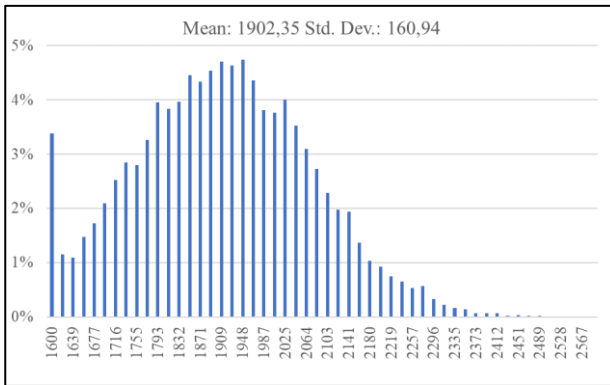


Figure 1: A simulated wage distribution histogram for the combined effects case

The results also show that lower measurement precision and a smaller female share can reduce false-positive rates in some settings, but these mitigating effects are limited relative to the broader impact of distributional misspecification. More importantly, the combined-effects case demonstrates that seemingly modest deviations from the linear estimating model can accumulate into economically and statistically meaningful false-detection risk.

It is also important to note that the lower-tail minimum-wage effect and the upper-tail superstar effect partly

offset each other. This counterbalancing makes the combined case less extreme than a simple additive interpretation might suggest. However, the offset is only partial and does not remove the broader identification problem. Even in the combined setting, false-positive rates remain substantial, especially as sample size increases.

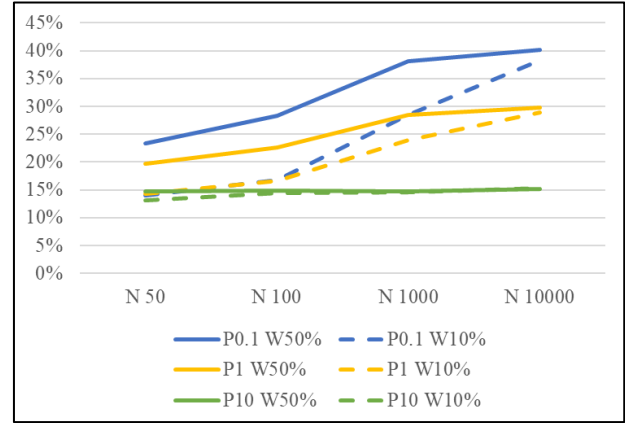


Figure 2: Ratio of cases with any gender wage gap variable (dummy and cross-products) being statistically significant when no wage gap exists, all-effects-combined cases

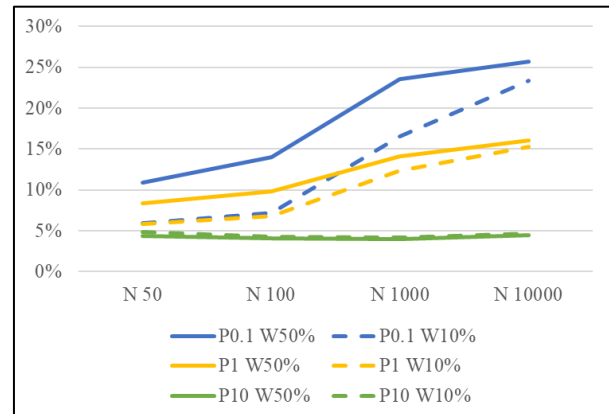


Figure 3: Ratio of cases with a statistically significant gender dummy when no wage gap exists, all-effects-combined cases

CONCLUSION

This study examines whether a standard regression-based gender wage gap model can falsely detect gender discrimination in a zero-discrimination world. The main conclusion is that statistically significant gender-related coefficients do not necessarily indicate structural discrimination. Even when the true data-generating process contains no gender-based wage discrimination, a standard empirical wage regression may still produce significant gender effects because of measurement error, omitted information, distributional misspecification, and multiple-testing risk. Under the baseline zero-discrimination benchmark, the false-significance rate of

the female dummy remains close to the nominal 5% level, but the probability that at least one gender-related term becomes significant is materially higher. Once lower-tail truncation, upper-tail non-linearity, and other distortions are combined, false-positive rates rise substantially. This shows that the practical risk of misclassification is not limited to any single coefficient but extends to the broader interpretation of a regression model that includes multiple gender-related terms. A second important conclusion is that conventional goodness-of-fit measures, such as adjusted R^2 , are not reliable safeguards against false detection. A model may fit the data well and still generate misleading inferences about gender effects. In other words, statistical significance and high in-sample fit should not be interpreted mechanically as evidence of structural discrimination. The results also clarify the distinct roles of different distortions. Omitting a balanced but unobserved productivity trait is less problematic than introducing features that alter the shape of the wage distribution. Minimum-wage truncation compresses the lower tail, whereas the superstar effect stretches the upper tail. Because the estimated regression does not model these mechanisms explicitly, it may misattribute distributional patterns to gender effects. Although these two distortions may partly offset each other, the offset is incomplete and false-positive risk remains substantial.

To conclude, these findings have an important implication for research and policy. Apparent gender wage gaps detected by a single regression specification should be interpreted with caution, especially in regulatory or compliance settings where sanctions, audits, or reputational consequences may follow. More broadly, the results suggest that robustness checks, alternative specifications, and validation across subsamples or out-of-sample settings are essential before statistical significance is interpreted as evidence of structural discrimination.

REFERENCES

- An, Q., Xu, D., & Brooks, G. P. (2013). Type I error rates and power of multiple hypothesis testing procedures in factorial ANOVA. *General Linear Model Journal*, 39(2), 1-16.
- Averkamp, S., Bredemeier, C., & Juessen, F. (2020). *Decomposing gender wage gaps: A family economics perspective* (IZA Discussion Paper No. 13601). IZA Institute of Labor Economics.
- Bergmann, N., Scheele, A., & Sorger, C. (2019). Variations of the same? A sectoral analysis of the gender pay gap in Germany and Austria. *Gender, Work & Organization*, 26(5), 668-687.
- Bennedsen, M., Larsen, B., & Wei, J. (2023). Gender wage transparency and the gender pay gap: A survey. *Journal of Economic Surveys*, 37(5), 1743-1777.
- Blau, F. D., & Kahn, L. M. (1992). The Gender Earnings Gap: Learning from International Comparisons. *American Economic Review*, 82(2), 533-538.
- Blau, F. D., & Kahn, L. M. (1996). International Differences in Male Wage Inequality: Institutions versus Market Forces. *Journal of Political Economy*, 104(4), 791-836.
- Blau, F. D., & Kahn, L. M. (1999). Analyzing the gender pay gap. *The Quarterly Review of Economics and Finance*, 39(5), 625-646.
- Blau, F. D., & Kahn, L. M. (2007). The Gender Pay Gap. *The Economists' Voice*, 4(4), 1-6.
- Blau, F. D., & Kahn, L. M. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, 55(3), 789-865.
- Chen, Q., & Qi, J. (2023). How much should we trust R^2 and adjusted R^2 : Evidence from regressions in top economics journals and Monte Carlo simulations. *Journal of Applied Economics*, 26(1), 2207326.
- Edgeworth Economics. (2023). *Omitted variable bias's cousin: How measurement error can also bias the results of pay equity studies*. Edgeworth Economics.
- European Commission. (2015, March 4). Questions and Answers: What is the EU doing for women's rights and gender equality? (MEMO/15/4563). https://ec.europa.eu/commission/presscorner/detail/en/memo_15_4563
- European Council. (2025, February 4). *Pay Transparency in the EU*. [www.consilium.europa.eu. https://www.consilium.europa.eu/en/policies/pay-transparency/](https://www.consilium.europa.eu/en/policies/pay-transparency/)
- Eurostat. (2024, March). *Gender Pay Gap Statistics - Statistics Explained*. Europa.eu. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Gender_pay_gap_statistics
- Eurostat. (2009). *Development of econometric methods to evaluate the gender pay gap using Structure of Earnings Survey data* (KS-RA-09-011). Publications Office of the European Union. <https://ec.europa.eu/eurostat/web/products-statistical-working-papers>
- EUR-Lex. (2023). Directive (EU) 2023/970 of the European Parliament and of the Council of 10 May 2023 to strengthen the application of the principle of equal pay for equal work or work of equal value between men and women through pay transparency and enforcement mechanisms. *Official Journal of the European Union*, L 132, 21-44.
- Ferri, V., García-Pereiro, T., & Pace, R. (2022). Gender pay-gap: exploring the school-to-work transition of graduates in Italy. *International Journal of Manpower*, 44(6).
- Fortin, N. M., Bell, B., & Böhm, M. (2017). Top Earnings Inequality and the Gender Pay gap: Canada, Sweden, and the United Kingdom. *Labour Economics*, 47(1), 107-123.
- OECD. (2024). *Pay Transparency Tools to Close the Gender Wage Gap*. OECD. https://www.oecd.org/en/publications/pay-transparency-tools-to-close-the-gender-wage-gap_eba5b91d-en.html

Segovia-Pérez, M., Castro Núñez, R. B., Santero Sánchez, R., & Laguna Sánchez, P. (2019). Being a woman in an ICT job: an analysis of the gender pay gap and discrimination in Spain. *New Technology, Work and Employment*, 35(1), 20–39.

Smith, M. (2012). Social regulation of the gender pay gap in the EU. *European Journal of Industrial Relations*, 18(4), 365–380.

Spiess, A.N., & Neumeyer, N. (2010). An evaluation of R^2 as an inadequate measure for nonlinear models in pharmacological and biochemical research. *BMC Pharmacology*, 10, 6. <https://doi.org/10.1186/1471-2210-10-6>

AUTHOR BIOGRAPHIES

PÉTER JUHÁSZ, PHD, CFA serves as associate professor for Business Valuation and Modelling at the Corvinus University of Budapest. His research interest covers ethics in finance, AI in finance, financial programming, and competitiveness. He also regularly

acts as a management consultant and trainer and has been on the board of the CFA Society of Hungary for more than 15 years. His e-mail address is: peter.juhasz@uni-corvinus.hu.

YUETONG MENG is a doctoral student at Corvinus University of Hungary. Her main research area is the gender wage gap, while she is also interested in Islamic banking and biodiversity finance. Her e-mail address is: yuetong.meng@stud.uni-corvinus.hu.

KATA VÁRADI, PHD is an associate professor at Corvinus University of Budapest, at the Institute of Finance, where she serves as the Head of the Institute. She holds a PhD in finance from Corvinus University of Budapest. Her main research areas are market liquidity, risk management in central clearing, and arbitrage opportunities on cryptocurrency markets. Her e-mail address is: kata.varadi@uni-corvinus.hu.

APPENDIX A. Variable definitions

Symbol	Definition	Notes
F_i	Gender dummy, where ($F_i = 1$) for female employees and ($F_i = 0$) for male employees	Individual-level indicator
w	Female share in the simulated sample	Sample composition
S_i	Vector of wage-relevant traits	Includes observed and unobserved traits
Edu_i	Education	Observed trait
$Expi$	Experience	Observed trait
Com_i	Commitment	Observed trait
$OtherSkills_i$	Other unobserved productivity-related skills	Unobserved trait
$\widetilde{S}_{i,k}$	Observed trait with measurement error	Noisy proxy
$\varepsilon_{i,k}$	Measurement error term	Trait-specific error
W_i	Wage generated by the structural wage equation	Simulated wage
$g(Edu_i, Exp_i)$	Upper-tail joint premium (“superstar” effect)	Nonlinear premium component
β_k	Coefficients on wage-relevant traits	Structural coefficients
β_j	Coefficient on the joint premium term	Premium coefficient
mw	Minimum wage floor	Lower truncation
FWER	Familywise error rate	Multiple-testing metric
N	Number of individuals in each simulated dataset	Sample size