ROCKWOOL Foundation Berlin
Institute for the Economy and the Future of Work (RFBerlin)

DISCUSSION PAPER SERIES 086/26

The Gender Wage Gap in Britain: A Meta-Analysis

Patrice Laroche, Alex Bryson, Heather Joshi, David Wilkinson

www.rfberlin.com

MARCH 2026

The Gender Wage Gap in Britain: A Meta-Analysis

Authors

Patrice Laroche, Alex Bryson, Heather Joshi, David Wilkinson

Reference

JEL Codes: J16, J31, J71

Keywords: gender wage gap, meta-analysis, UK

Recommended Citation: Patrice Laroche, Alex Bryson, Heather Joshi, David Wilkinson (2026): The Gender Wage Gap in Britain: A Meta-Analysis. RFBerlin Discussion Paper No. 086/26

Access

Disclaimer

Imprint

RFBerlin
ROCKWOOL Foundation Berlin –
Institute for the Economy
and the Future of Work

Gormannstrasse 22, 10119 Berlin
Tel: +49 (0) 151 143 444 67
E-mail: info@rfberlin.com
Web: www.rfberlin.com

The Gender Wage Gap in Britain: A Meta-Analysis[1]

Patrice Laroche (Université de Lorraine)
Alex Bryson (University College London)
Heather Joshi (University College London)
David Wilkinson (University College London)

Abstract

Ours is the first meta-analysis synthesizing results from econometric studies carried out in the UK to assess the size of the gender wage gap (GWG). Drawing on 90 primary studies published between 1974 and 2024 we assess trends in the gap and identify the substantive and methodological factors that explain variance in results across studies. Expressed relative to men's earnings, the raw GWG averages 25 log points but falls to 13 log points when adjusting for covariates. There has been convergence in the mean wages of men and women at a rate of about 0.3 percentage points per annum, most of which reflects change in the characteristics of workers and their treatment in the labour market rather than differences over time in study characteristics. There is substantial heterogeneity in the size of the GWG by year of observation, worker type and research design, although differences in the size of adjusted GWG by study design are not as large as most economists might imagine.

Key words: gender wage gap; meta-analysis; UK
JEL codes: J16; J31; J71

[1] Bryson, Joshi and Wilkinson acknowledge funding from the Economic and Social Research Council (grant number ES/S012583/1). We are particularly grateful to Katarina Zigova for generously sharing the dataset of primary studies she started to compile for her own research. Her contribution substantially facilitated the construction of our study database.

Introduction

As in much of the advanced industrialised world, women in the UK receive lower pay than their male counterparts, even when they are performing 'like' jobs in terms of their worth to the employer. Although the gap in mean earnings has been declining in the last few decades (ONS 2025a) it remains sizeable and, among some groups of workers such as graduates (Foliano et al 2024) it has been growing. This has prompted renewed calls to look again at the provisions of the 1970 Equal Pay Act and subsequent legislation which are supposed to prevent discriminatory practices including those on grounds of gender.

In this paper, we undertake a meta-analysis of econometric studies on the gender wage gap (GWG) in the UK undertaken between 1974 and 2024. Our objective is to identify long-run trends in the gap and to examine the structural and modeling factors that systematically influence its reported magnitude – patterns that are not readily apparent when studies are assessed individually. By synthesizing results across studies and explicitly modeling study-specific characteristics, meta-analysis enables the identification of general regularities that emerge from the accumulated empirical record (Stanley & Doucouliagos, 2012).

Although several meta-analyses have examined gender wage differentials from international or large-country perspectives (e.g., Stanley & Jarrell, 1998; Jarrell & Stanley, 2004; Weichselbaumer & Winter-Ebmer, 2005; Iwasaki & Ma, 2020), none provides a systematic synthesis focused exclusively on the UK. International meta-analyses offer valuable insights but they pool evidence across heterogeneous institutional and labor market regimes complicating the identification of country-specific dynamics and specification effects.

The UK represents a particularly informative case for a focused synthesis. It is a country with a long-standing tradition in econometric analyses of microdata and continuity in wage-setting arrangements including legislative efforts to promote equal pay between men and women. This context allows us to examine temporal evolution, specification choices and publication patterns within a comparatively stable institutional environment. By restricting attention to a single national setting, we reduce cross-country institutional variance and sharpen inference regarding the determinants of heterogeneity in reported GWG estimates.

Beyond its country-specific focus, the study also contributes methodologically. We employ multilevel Restricted Maximum Likelihood (REML) and Bayesian hierarchical estimators that explicitly account for clustering and dependence among multiple estimates drawn from the same primary study. This framework allows us to model hierarchical variance structures transparently and to obtain more reliable inferences when effect sizes are non-independent, an issue that conventional fixed-effects or standard random-effects approaches address only imperfectly.

Within this multi-level framework, we consider sources of variance in GWG estimates including model specification, sample composition and research design. Differences in estimation samples, specification strategies and modeling conventions may all contribute to dispersion in reported wage gap estimates. By explicitly modeling these moderators, we can assess their relative importance while isolating the underlying magnitude and temporal evolution of the UK GWG. In addition, the meta-analytic approach allows us to test and adjust for potential publication bias, thereby distinguishing between genuine structural patterns and selective reporting effects.

We show that a sizeable residual gap persists after accounting for observed human capital (and other) controls, specification heterogeneity and publication bias. This has important implications for interpreting the sources of wage inequality and for policy design.

The paper is organized as follows. In section 1, we present the theoretical/empirical framework that underpins the research field. In section 2, we describe how we collect the data from the literature, and we provide summary statistics of the dataset. In section 3, we test the presence of publication selection bias and adjust our estimates of the size of the GWG accordingly. In section 4 we examine the heterogeneity in the reported estimates. Section 5 presents the results of the meta-analysis and section 6 concludes by highlighting the implications of findings for policy and future research.

Background to the Gender Wage Gap in the UK

In neo-classical analysis wages are set not by employers but by the market at the intersection of the labour demand and supply curves. The wage an individual receives should reflect their productivity which, in turn, relates to their acquisition of human capital through schooling and via the accumulation of skills and experience in the labour market. This framework informed Mincer's (1974) model of wage determination whereby the wage received by an individual – usually expressed as the log of hourly earnings – reflects one's educational qualifications and labour market experience captured by time elapsed since leaving school and firm specific human capital captured by tenure with one's current employer. Diminishing returns to additional experience are usually captured by squared terms with regards to experience (sometimes proxied by age) and tenure.

In this setting there should be no role for sex in wage determination. And yet, if one enters a female dummy variable into a wage equation it often returns a negative and statistically significant coefficient. Some years ago this might have been anticipated given what Becker (1985) referred to as the 'gains from trade' by which men and women in couples specialised in market and non-market production respectively. Social norms dictated that women took the majority of household production and, as such, would invest less than their male partners in both education and labour market experience (Mincer and Polachek, 1974, Polachek, 1975). The implication is that conditioning on human capital would lead to attenuation in the negative coefficient attached to being a woman, since the wage gap reflected differential human capital investments. Nevertheless, as we shall see, a negative and statistically significant female coefficient persists even after conditioning on human capital traits of men and women. This 'residual gap', as it was often termed, was traditionally interpreted as 'discrimination' reflecting employer discrimination against women either on grounds of taste (Becker, 1971) or statistical discrimination (Arrow, 1973). In the latter employers, faced with a lack of precise information regarding the productivity of individual women, inferr their productivity based on what they expect to be the average productivity of women. In contrast to the view that women's low pay is due to rational choices of participation and occupation to accommodate their domestic roles, it can also be suggested that the undervaluation of women's work could arise from the attitudes of other workers and members of society more generally through gendered stereotypes, misogyny, or harassment of women breaking the norms (Rubery, 2017; Folke and Rickne, 2022).

There is evidence that discrimination against women persists, both in hiring (Goldin and Rouse, 2000) and wage determination, whether for taste or statistical reasons, and that stereotypes of women persist such that those in occupations dominated by women continue to be paid less than they might be in more gender-equal occupations by virtue of the systematic undervaluation of what is perceived as ‘women’s work’, even when it requires high- skilled labour inputs. Women may end up with lower pay than men because of lower bargaining power or less assertive negotiation styles (Babcock and Laschever, 2003) Nevertheless, analysts in recent years have been less quick to attribute all the residual gender wage gap to discrimination alone. They recognise that the residual also captures other unobserved traits in the data which may be correlated with sex and productivity. These might include other aspects of human capital not traditionally observed in data. One such omitted factor might be non-cognitive skills (although evidence indicates that women score more highly on these traits than men (Blau and Kahn, 2017: 836). They may also include personality traits and other traits which the researcher may find harder to observe than the employer and which may impact earnings. Manning and Swaffield (2008) investigate a range of such traits, measured in late childhood, and examine their consequences for earnings by age 30, in a cohort born in 1970. They point to factors such as differences in career orientations which later impact men’s and women’s propensities to invest in human capital. The inclusion of these orientations, and indeed other traits, does reduce the size of the residual gender wage gap, though only modestly.

Developments in terms of schooling and family formation have had important implications for earnings formation for both men and women in the UK, as in other advanced industrialised societies. First, women’s participation in the labour market has grown rapidly in the last half century, such that the positive selection of women into employment has become less important over time (Bryson et al., 2020). Second, recent cohorts of young women are out-performing men in terms of educational qualifications in all but a few science subjects. Other things equal, this trend would result in a closure of the raw gender wage gap. Third, women give birth to fewer children, time to first birth has lengthened and the time to re-entering the labour market post-birth has shortened. These trends help explain the closure in the gap in labour market experience between men and women which, again, other things equal, should close the raw gender wage gap. Furthermore, when women do return to paid work they are more likely than they used to be to return to full-time employment, thus minimising their exposure to the part-time pay penalty that is still evident in the UK. (For a fuller rendition of these developments and their relevance in the UK context see Bryson et al., 2020).

Notwithstanding these trends, it appears that women continue to suffer a wage penalty attached to motherhood, and that this penalty is in varying degrees common across the Western world (Kleven et al, 2019). By contrast, a pay premium is often apparent for new fathers, reflecting either increased labour market commitment on the part of men when they become fathers, or at least a perception of increased commitment on the part of their employers.

Some have argued that motherhood remains the sole reason for a GWG in Britain today. But this is not the case. Childless women continue to face a wage penalty as compared to ‘like’ childless men (Joshi et al 2021) and a wage gap persists among workers in their mid 20s prior to the moment most become parents (Foliano et al 2024). The simulations made by Adda et al (2017) of the emergence of pay penalty for motherhood through endogenous influences on choices regarding education, employment and family-friendly occupations in data for the Netherlands also find an unexplained gap between the wages of men and women who never have children.

A further development is the persistence of large GWGs at the top of the earnings distribution in the UK, the US and some European labour markets. Although women are increasingly breaking the ‘glass ceiling’ by entering occupations in the top echelons of the pay distribution, these gaps are often within-occupation and partly reflect women's propensity to take time off work around the time of childbirth, and a desire to combine work and family life through more flexible work schedules which often come at the price of lower wages than they might otherwise have received (Goldin, 2014). Policy efforts to address the wage penalty attached to gendered household responsibilities through incentives for men to spend time at home with their children have only been partially successful (Kleven et al., 2024; Kleven et al., 2026).

The raw GWG – the difference between men's and women's hourly earnings expressed as a percentage of men's – was sizeable at around 40% and persisted for several decades after the Second World war. It fell markedly in the five years or so after the passing of the Equal Pay Act which came into full force in 1975, reaching around 32 percent in 1977 (Zalbalza and Tzannotos, 1985; Bryson et al 2020).[2] There has been further progress since then with the legislation being strengthened by the Equal Pay Amendement Act in 1983 and the Equalities Act of 2010. The headline gap has fallen according to current statistics by a further 20 percentage points. On the latest figures, the GWG in median gross hourly earnings was 12.8 percent among all employees in April 2025 (ONS, 2025a). In terms of log points as featured below, the raw GWG fell from .39 in 1977 to .14 in 2025 – roughly 0.5 percentage points per year.

Methods and Data

We collect data from the econometric literature estimating the GWG in the UK.[3] Following Kitagawa (1955), Blinder (1973) and Oaxaca (1973), the standard practice in the literature is to use regression analyses to decompose the observed average gender gap into two components: one attributable to differences in endowments (or skills) and the remainder, which is often characterized as a difference in coefficients (or return to skills). Once equated with wage discrimination,. it has more recently, been recognized that the residual may be picking up a number of factors, unobserved to the analyst, which might result in wage differences by sex – including discrimination. The effects of differing skills and returns are usually estimated with separate male and female log hourly wage regression models from samples of individual male and female wage earners as in equation (1):

$$(1) \quad W_f = X_f \beta_f + \varepsilon_f \quad \text{and} \quad W_m = X_m \beta_m + \varepsilon_m$$

Sometimes, these two equations are combined into one (equation (2)):

$$(2) \quad W_i = \beta_0 + \Sigma\beta_k X_{ki} + \gamma S_i + \varepsilon_i$$

where S is the sex of the worker *i*. In more flexible specifications sex may be interacted with other terms in the model. In models where sex is fully interacted with all other terms in the model estimates converge towards those recovered from separate equations (as in (1)).

Setting all covariates to zero in equation (2) one recovers the raw GWG, namely the unconditional gap in log wages between men and women. This is informative in its own right

[2] The gap is measured at the median, of gross hourly earnings (excluding overtime) and includes part-time workers.

[3] We treat the UK and Great Britain as synonymous. Many of the datasets used, listed in Appendix A2, apply to the latter (England, Scotland and Wales) eg BHPS, GHS and Birth Cohort Studes. Some sources (eg the LFS, UKHLS and New Earnings Survey) cover the whole United Kingdom, adding a relatively small sample from Northern Ireland.

but studies always report covariate-adjusted GWGs too in an effort to make *ceteris paribus* comparisons across the sexes.

From equations (1) and (2), researchers estimate the magnitude of the GWG found in the j^{th} study by:

$$(3)\ Gj = \bar{X}f\beta m - \bar{X}f\beta f = \bar{X}f(\beta m - \beta f)$$

(Kitagawa, 1955; Blinder, 1973; Oaxaca, 1973) or by γ of equation (2). The above wage gap, G_j, *measures the difference between the logarithm of female wages from those of male workers due to the fact their worker traits are valued differently*. It is thus an estimate of the difference of the logarithm of wages relative to what it would have been in the absence of discrimination, or with equal treatment of the traits.[4] Sometimes studies report Oaxaca's *D* as a measure of the magnitude of discrimination. It may be interpreted as the amount of wage discrimination as a proportion of the male wage which expresses the gap in terms of the proportion of the male wage rather than log points. It is related to G_j by:

$$(4)\ D = e^{Gj} - 1$$

In our meta-analysis, a study's estimate of G_j is the dependent variable. In some instances this is a raw differential, but in most cases it is covariate-adjusted. Variation in G_j both within-study and from one study to another is explained by variations in the estimation sample, model specification, alternative measurements of wages, alternative sub-populations of workers (their ages, occupation, employment sectors), the passage of time, data quality, the gender of the researchers and other factors. The choice, or availability, of conditioning variables can be particularly important since they will affect the size of any residual GWG. There is controversy in the literature regarding the potential endogeneity of variables such as family structure and job traits which may inadvertently partial out some of the wage gap which might otherwise be attributed to sex. Conversely, the absence of certain wage determinants may bias estimates of the GWG. For instance, a recent literature on the GWG indicates that where you work can have a substantial impact on the size of the GWG (Theodoropoulos et al., 2022). When the employer identifier is absent, these differences will contribute to the residual gap.

Collection and coding of studies

To build our dataset, we searched for empirical studies that provide estimates analyzing the gender wage gap G_j. We undertook a database search for relevant studies in ISI Web of Science, Business Source Premier, and Google scholar and identified 129 papers using the following broad keywords: 'gender wage gap in UK', 'gender wage discrimination in UK', 'gender pay gap in UK'. Our search also included examination of references in empirical studies to other studies that might report gender wage differences. The search was completed in April 2023.

[4] Usually, parameters for males are taken as representing the returns to traits one would expect in a market setting and in the absence of discriminatory factors. However, there is potential value using pooled estimates of male and female parameters to proxy the counterfactual non-discrimination differential (Jann, 2008). In our meta-regression analyses (MRA) we do not differentiate the weighting schemes used within Oaxaca-type decompositions.

We read the 129 potential studies to see whether they included a GWG estimate as shown in equations (2) or (3).[5] We include peer-reviewed and non-peer reviewed studies controlling for their publication status in the MRA. [6]

All studies reported an empirical estimate of the GWG or sufficient information to calculate it. That is, studies needed to report regression coefficients, sample size, standard errors and/or *t*-statistics or GWG estimates from a decomposition approach.

Several studies could not be included in the final analysis because they did not satisfy our criteria for inclusion (see Table A1 in Appendix which lists the studies excluded from the meta-analysis and the reasons for exclusion). The final data set draws on 870 estimates of the gender wage gap from 90 studies. All these studies are listed in Appendix Table A2, together with the number of estimates, the period covered by the data and method of analysis. Because most studies report multiple estimates obtained from different specifications or sub-samples it is difficult to select a representative estimate for each study. Therefore, we follow good practice in recent meta-analyses by collecting all estimates from the relevant studies. These include estimates of the raw (unadjusted) GWG which is derived using estimates as per equations (1) and (2) but omitting the Xs. We also codify variables reflecting study characteristics that may influence the reported GWGs. These variables are described in Section 4.

To summarize and compare the results from existing studies, we need a common metric to measure the GWG. We denote the GWG as G_j as it is commonly used in this field of research. The GWG can be derived from the *t*-statistics of the reported regression estimates and residual degrees of freedom or directly from the decomposition of equations (1) and (2) above. From the individual studies we collect both the *raw wage gap and the regression-adjusted wage gap – sometimes several sets of gaps and estimates – all in natural logs*. The *raw* (or *unadjusted)* *wage gap* is simply the difference between the log mean of men's and women's wages (usually measured hourly).The *adjusted wage gap* uses regression methods to measure the difference in pay between women and men after accounting for factors that determine pay, for example, but not exclusively, education and experience. This type of analysis accounts for relevant factors that influence pay and then shows whether a significant GWG remains and how big that gap is. However, while the adjusted GWG is potentially more informative than the raw gap because it seeks to make *ceteris paribus* comparisons, it is not always straightforward to interpret the adjusted gap because analysts often condition on variables which are themselves partly a function of gender and, as such, partial out some of the gender-specific reasons for a gender wage gap. Model specifications can vary quite markedly across studies. The point of our meta-analysis is to identify the variables that explain most of the variation in the gender wage gap. It is therefore interesting to combine the results of all the studies, whether they present a raw gap or an adjusted gap, in order to determine which variables best explain the wage gap between men and women.

Data screening and Descriptive Statistics

[5] It is essential to adopt an explicit set of selection criteria to create a dataset suitable for meta-analysis. The studies included in the meta-analysis must be comparable so that their differences can be coded. The coding procedure serves as the basis for the transparency and replicability of the meta-analysis.

[6] Two-thirds (66%) of the studies are refereed journal articles or book chapters, with the remaining 34% consisting of PhD dissertations or working papers.

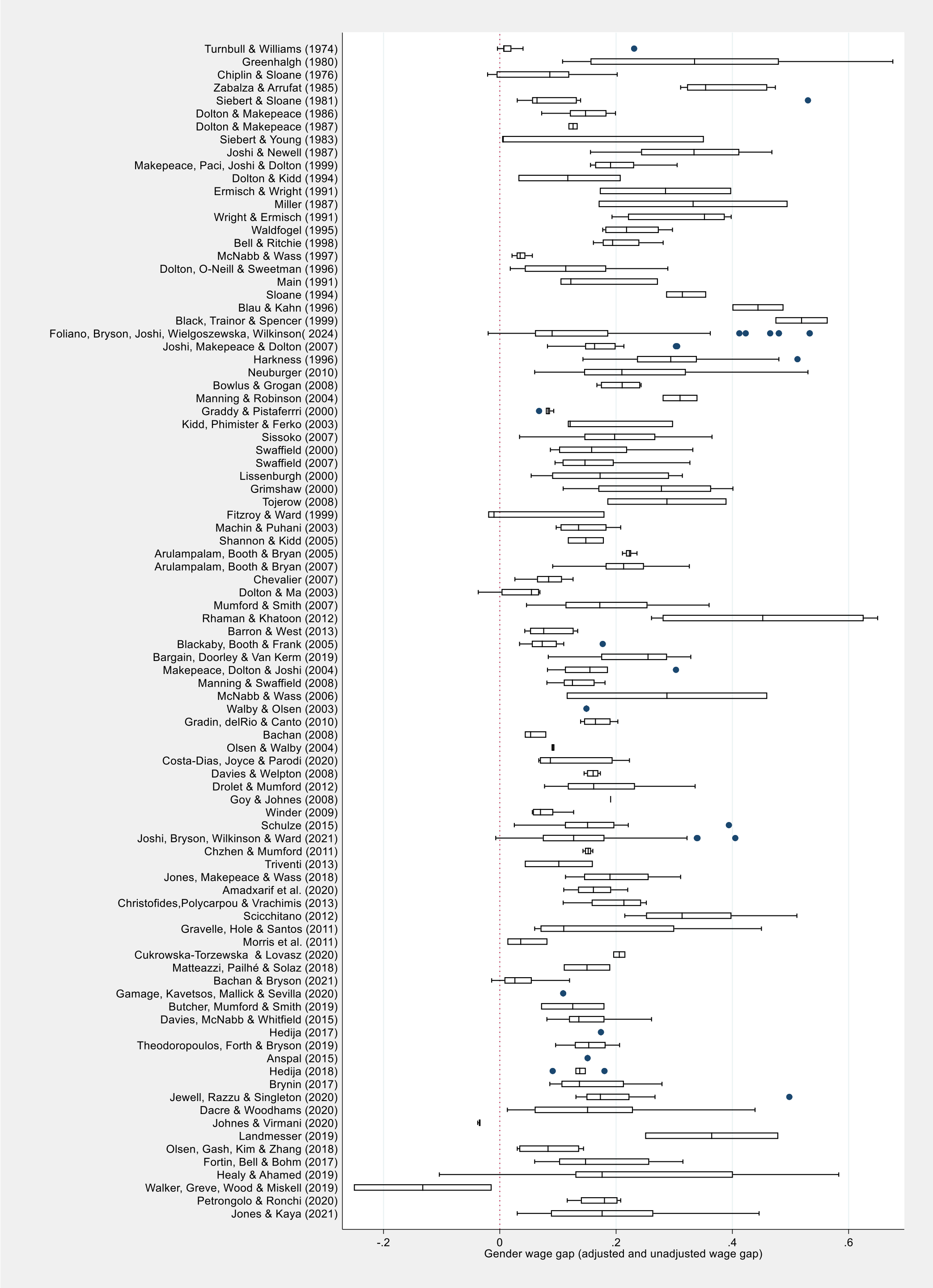
Figure 1 shows the median and the inter-quartile range of the GWG estimates in each study in our dataset. The estimates are presented in chronological order of the median observation in each study, but one cannot easily infer the trend in the GWG from the figure because the observations in some studies span a number of years. Positive coefficients represent gaps in favour of men. Three stylized facts emerge from Figure 1. First, most studies report positive estimates suggesting a gap in favour of men. The gender wage gap appears to favour women unambiguously in only two studies (Walker et al., 2019; Jones & Virmani, 2019) which are each concerned with the elite labour market for university vice chancellors[7]. Second, the average gap is sizeable at +.183. Third, more recent estimates of the GWG appear to be a little smaller than earlier estimates, based on the average year of observations they use.

Table 1 shows summary statistics for all the studies in our data, together with those for four sub-groups, (which are not mutually exclusive), namely those using Kitigawa/ Blinder/Oaxaca decomposition techniques, using a dummy for gender in a regression analysis, recent studies, and unpublished studies. Across all studies the weighted average of the GWG estimate are, respectively, .246 for the raw gap and .127 for the adjusted gap. [8] Adjusting for potential confounders tends to reduce the raw GWG quite considerably, regardless of study type, often by as much as one-half. A simple descriptive sub-group analysis (Table 1, columns 2 to 4) indicates differences in the GWG according to the methodological approach used (Kitigawa/Blinder/Oaxaca vs others), the period under investigation (before vs after 2010) and the type of the outlet (published vs unpublished studies).

[7] The pay of vice chancellors, not typical of the academic sector as a whole, has received attention as being one of the few top earning groups with accessible pay data.

[8] The simple average (unweighted average) suffers from several shortcomings. First, it does not consider the estimate's precision, as each estimate is ascribed the same weight regardless of the sample size from which it is derived. Second, the simple unweighted average does not consider possible publication selection which can bias the average effect. We return to these issues later. Note that some studies only report raw or adjusted GWGs.

Figure 1. Raw and Adjusted Gender Wage Gaps Within and Across Study



Notes: The figure depicts a box plot of the unadjusted (raw) and adjusted gender wage gap corresponding to the estimates reported in individual studies. The studies are sorted by the age of the data they use from the oldest to the youngest. The length of each box represents the interquartile range (P25-P75), and the dividing line inside the box is the median value. The whiskers represent the highest and lowest data points within 1.5 times the range between the upper and lower quartiles. The points represent extreme values.The vertical red line denotes zero difference between women's and men's pay. The x-axis is measured in natural logarithms.

For instance, studies centered on more recent years (post-2010) exhibit smaller weighted gaps overall, suggesting a modest narrowing over time. Moreover, unpublished studies tend to report slightly smaller weighted gaps than the full sample, although differences remain moderate.

Table 1. Descriptive statistics of estimated Gender Wage Gaps

	All studies	Studies using Kitigawa/Blinder/Oaxaca decomposition	Studies using dummies in regression analysis	Studies centred on recent years (after 2010)	Unpublished studies
Gender Wage Gap					
Number of studies	90	67	47	27	20
Number of estimates	870	335	277	249	181
Unweighted average	.186 (.177 to .194)	.200 (.186 to .214)	.140 (.128 to .153)	.178 (.161 to .195)	.178 (.162 to .195)
Weighted average	.178 (.169 to .187)	.201 (.187 to .214)	.117 (.103 to .130)	.148 (.133 to .164)	.158 (.143 to .174)
Raw Wage Gap					
Number of studies	71	27	12	25	15
Number of estimates	315	103	56	109	72
Unweighted average	.251 (.236 to .266)	.278 (.252 to .306)	.216 (.181 to .252)	.254 (.227 to .281)	.244 (.216 to .272)
Weighted average	.250 (.235 to .265)	.265 (.239 to .292)	.239 (.201 to .278)	.208 (.185 to .230)	.217 (.188 to .245)
Adjusted Wage Gap					
Number of studies	83	40	37	25	16
Number of estimates	555	232	221	140	109
Unweighted average	.149 (.140 to .158)	.166 (.151 to .179)	.121 (.109 to .134)	.119 (.102 to .135)	.135 (.118 to .152)
Weighted average	.137 (.127 to .146)	.167 (.154 to .180)	.103 (.089 to .117)	.097 (.079 to .116)	.126 (.111 to .141)

Notes: The unweighted average is the simple average of observations. The weighted average is calculated by using the inverse of number of estimates per study as weights. These weights are connected to the number of estimates in each study and not the size of the samples or the population group concerned.

To provide a first visual overview of the data, Figure 2 presents the density distributions of raw and adjusted gender wage gaps (in log points). The figure allows us to assess differences in magnitude and dispersion of raw and adjusted gaps before turning to multivariate analysis. The raw gaps are more lightly shaded than the adjusted gaps.

The figure shows raw gaps are centered at higher values and display a heavier upper tail, indicating that uncontrolled specifications tend to report larger GWGs. The adjusted estimates are more tightly concentrated and peak at lower log-point values, consistent with the reduction in the gap once observed characteristics such as education, experience or occupation are taken into account. The substantial overlap between the two distributions nonetheless suggests that adjustment attenuates but does not eliminate the reported wage gap.

Figure 2. Raw gap and Adjusted gap estimates distribution

Overall, the descriptive evidence highlights substantial heterogeneity across methodological approaches, time periods and publication status, thereby justifying a multivariate meta-regression framework. Part of this heterogeneity may stem not only from methodological or contextual differences but also from selective publication mechanisms, which makes it essential to assess the presence of publication bias.

Publication Bias and Corrected Estimates

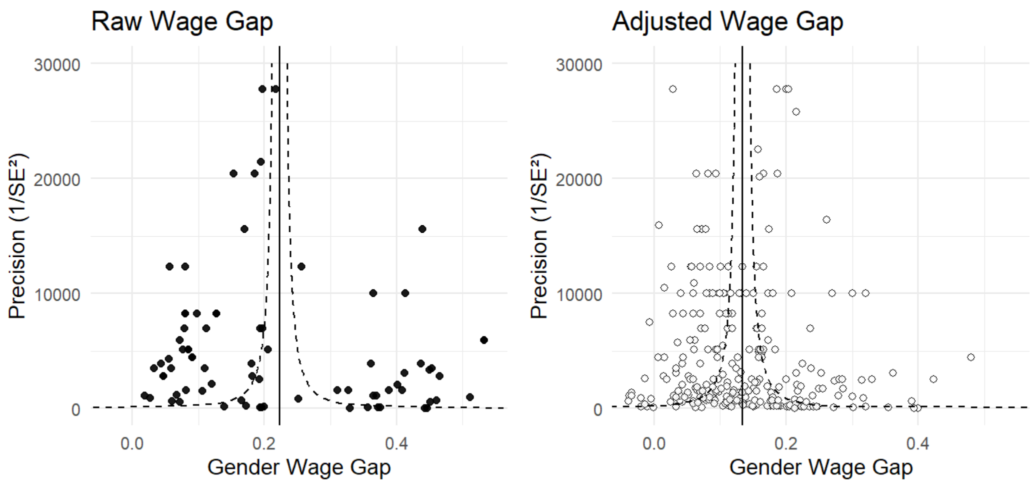
Publication bias arises when researchers, referees or editors have a preference for publishing results that either support a particularly theory or are statistically significant. Publication bias is particularly strong in research fields that show little disagreement concerning the sign of the effect (Doucouliagos et al. 2005). As a consequence, estimates supporting the prevailing theoretical view are more likely to be published, whereas results showing an effect inconsistent with the theoretical arguments tend to be underrepresented. In the case of literature on the GWG, for example, it is hard to find contributions that show no difference in the pay between men and women or cases where men are paid less than women. We do have studies like this in our sample, such as those relating to Vice Chancellors' pay, but they are unusual.

The funnel plot provides a simple tool for visualizing possible publication bias (Stanley & Doucouliagos 2010). A funnel plot is a scatter diagram of all empirical estimates of a given phenomenon in which the size of the estimated effect is plotted on the x-axis against a measure of the estimate's precision (i.e. the inverse of the estimates' standard errors, 1/SE) plotted on the y-axis.[9] In the absence of publication bias, the effect size should be symmetrically

[9] Standard errors are only available on a subsample, 385 estimates out of 870. . Our publication bias tests should thus be interpreted with caution. That said, given this limitation, we rely on a series of both linear and nonlinear tests to enhance the robustness of the findings.

distributed around the 'true' value of the effect. Empirical estimates with less precision are more widely spread at the bottom of the graph, while more precise estimates are found at the top of the funnel.

Figure 3. The funnel plot is consistent with publication bias



Notes: The figure shows adjusted and raw wage gap estimates reported in primary studies. In the absence of publication bias the funnel should be symmetrical around the most precise estimates. *N*=385. For graphical clarity, the precision axis is truncated at 30,000.

Figure 3 shows that, in the case of UK literature on the GWG the scatter plot resembles the theoretically predicted inverted funnel. The funnel is asymmetrical as there are more estimates of the GWG on the left-hand side of the figure than on the right (especially for the adjusted GWG). More precise estimates of the raw gap tend to be below the mean for all estimates but this is not the case for the adjusted GWG where, if anything, more precise estimates are to the right of the line denoting the mean estimate. It is difficult to infer anything about potential publication bias from a purely visual inspection of the funnel plot. However, visual methods are subjective and it is preferable to focus on formal methods of detection of and correction for publication bias which we come onto next.

In Table 2, we test the asymmetry of the funnel plot by regressing estimates on their standard errors (Egger et al., 1997; Stanley, 2005). If publication bias is a linear function of the standard error and if there is no correlation between estimates and standard errors in the absence of publication bias, then the slope coefficient in the MRA identifies the degree of publication bias and the constant determines the mean GWG corrected for the bias. The linearity assumption is motivated by the Lombard effect: an increase in noise (captured by standard errors) indicates that researchers are increasing their efforts to produce larger estimates so that they obtain a statistically significant result. Because statistical significance, measured by the t-statistics, is given by the ratio of the estimates to its standard error, there is hope that the selection effort will increase proportionally with the standard error in order to achieve the same t-statistic. The non-correlation assumption is motivated by the fact that the ratio of estimates and standard errors is assumed to have a symmetrical distribution, which means that estimates and standard errors are statistically independent quantities which is a property implied by most empirical techniques.

Table 2. Publication Bias Tests, Unconditional Estimates

	OLS (1)	**RE PET (2)**	**RE PEESE (3)**	**BE (4)**	**FE (5)**	**IV (6)**
Publication bias	0.566**	1.702***	9.949***	0.181	1.227***	1.442*
SE/SE²	(0.214)	(0.138)	(1.764)	(0.272)	(0.260)	(0.572)
Effect beyond bias	0.135***	0.094***	0.121***	0.146***	0.143*	0.109***
Constant	(0.009)	(0.014)	(0.013)	(0.010)	(0.063)	(0.018)
Observations	385	385	385	385	337	385
Studies	48	48	48	48	48	48

	Top 10 (7)	**WAAP (8)**	**Stem (9)**
Publication bias			
SE	-	-	-
Effect beyond bias	0.151***	0.141***	0.214***
Constant	(0.042)	(0.013)	(0.019)
Observations	40	343	385
Studies	9	45	48

Note: The table reports results of regression $GWG_{is} = GWG_0 + \gamma SE(GWG_{is}) + \varepsilon_{is}$, (where GWG denotes the gender wag gap of the *i*-th estimate from the *s*-th study and SE (GWG) denotes its standard error) using different linear regression specifications along with non-linear methods. OLS = Ordinary Least Square. Precision = inverse standard error weights. FE= Fixed Effects. BE = Between Effects. RE = Random Effects. IV = instrumental variable regression with the inverse of the square of root of sample size as the instrumental variable WAAP = Weighted Average of the Adequately Powered. Stem = Stem-based method. . *p<0.1, ** p<0.05, ***p<0.01. Given the presence of multiple effect sizes per study, we estimate multilevel meta-regressions using *rma.mv*, allowing for random intercepts at the study level to account for within-study dependence, especially for PET, PEESE, and WAAP. A small number of effect sizes exhibited extremely small standard errors, leading to disproportionate weights and numerical instability in multilevel estimation. To ensure stability, we trim the lower 1% of standard errors. Results remain robust to alternative trimming thresholds.

To evaluate the presence of publication bias and small-study effects, we estimate a series of precision-effect tests (PET), Precision-Effect Estimate with Standard Error (PEESE) regressions and alternative bias-correction procedures. PET, PEESE, WAAP, and STEM are different approaches used in meta-analysis to detect and correct publication bias. PET estimates a meta-regression of effect sizes on their standard errors and is mainly used to test whether a genuine effect exists once small-study bias is accounted for. PEESE extends this approach by regressing effect sizes on their variances and is typically used to obtain a more accurate estimate of the true effect when PET indicates that an underlying effect is present. WAAP (Weighted Average of Adequately Powered studies) follows a different logic: it limits the estimation to studies that have sufficient statistical power, under the assumption that publication bias primarily affects small, underpowered studies. Finally, STEM (Selection-Test Estimate Method) belongs to the family of selection models and attempts to correct publication bias by explicitly modeling the probability that studies are published depending on the statistical significance of their results.

Table 2 reports unconditional estimates. Across specifications, the coefficient on the standard error is positive and statistically significant. In the random-effects PET model (column 2), the SE coefficient equals 1.702 ($p < 0.001$), indicating substantial funnel asymmetry. Less precise studies systematically report larger GWGs, a pattern consistent with small-study effects and potential selective reporting. The magnitude of the standard error coefficient suggests economically meaningful distortion such that moving from a highly precise study to a moderately imprecise one corresponds to a sizeable increase in the reported gap. The intercept in column 2 implies a bias-corrected GWG of approximately 0.094 log-points, corresponding to about 10 percent. This estimate is substantially smaller than the uncorrected mean gap (around 0.18 log-points in Table 1 column 1) implying that roughly half of the observed average differential may be attributable to small-study effects. However, PET is known to be conservative in the presence of substantial heterogeneity, potentially understating the underlying effect. Because the PET bias coefficient is statistically significant, we follow the decision rule proposed by Stanley and Doucouliagos (2012) and report PEESE estimates. The PEESE intercept (column 5) equals 0.121 ($p < 0.001$), implying a corrected wage gap of approximately 12 percent. The difference between PET (≈10%) and PEESE (≈13%) suggests that the true effect likely lies within this range. Notably, the PEESE correction reduces the estimated bias relative to PET while still indicating substantial funnel asymmetry.

The *Weighted Average of Adequately Powered* (WAAP) estimator, based on studies with statistical power exceeding 80 percent, yields an effect of 0.141 log-points (≈15%).[10] Because WAAP relies exclusively on sufficiently powered studies, it provides a complementary robustness check less sensitive to small-study bias. The WAAP estimate lies between the uncorrected mean and the PET-adjusted estimate, reinforcing the conclusion that the true GWG remains economically meaningful even after accounting for publication bias.

Across the nine methods the mean corrected adjusted GWG is around .130 (between .094 to .151) compared to the uncorrected mean of .186 presented in Table 1.[11] The finding of a publication bias seems to be robust across different methods, but some of the evidence may be contaminated by the differences in the data and methods used to identify the GWG. Here we assume that all heterogeneity in the results is caused only by publication bias and sampling error but this assumption is not realistic. Therefore we turn to a multivariate Meta-Regression Analysis (MRA) to take into account all the factors that may explain the differences in results observed in the literature.

A Multi-Level Meta-Regression Approach (MRA)

MRA seeks to account for variance in estimates of the raw and adjusted GWG taken from our 90 studies. The moderator variables used to explore heterogeneity in the GWG estimates are listed in Table 3. Different types of moderator variables are hypothesized to influence the estimated GWG, namely variables capturing measurement and definitions, variables capturing model specification and estimation, and study-specific variables.

[10] Statistical power refers to the probability that a study will correctly detect a real effect if it truly exists. The WAAP estimator whose statistical power is greater than 80% to detect the estimated effect. In othjer words, the WAAP estimator computes the average effect using only studies that had at least an 80% chance of detecting the effect if it were real.

[11] We also inspect a stem-and-leaf display of reported t-statistics (Stem in Table 2) to identify potential clustering just above conventional significance thresholds (e.g. $|t| \geq 1.96$). However, this device is particularly fragile in meta-analyses dominated by small-sample studies, where imprecise estimates naturally generate erratic and asymmetric distributions of test statistics. In such contexts, apparent "bunching" near significance cutoffs may reflect sampling variability rather than selective reporting. The stem-and-leaf display therefore provides, at best, a rough visual screening and must be interpreted with considerable caution.

Table 3. Description of Moderator Variables (51 variables)

Moderator variables	Description	*N*	*Mean*	*SD*	*WM*
SE	Standard error of the estimates	393	.031	.039	.033
Data characteristics					
Data year (midpoint)	The year in which wages are earned (Year=0 in 1970) ((Average)Year of the data - 1970)	850	34.246	13.221	34.051
Nbyear	Year span	853	3.306	4.642	3.639
Private sector	= 1 if the study investigated the wages of workers in the private sector only	870	.121	.326	.152
Comparison	= 1 if the study used a country comparison (mostly EU database)	870	.113	.316	.156
Full time only	= 1 if a study included only full-time workers	866	.234	.424	.265
White collars	= 1 if a study investigated only medium and high-prestige occupations (e.g. white collars, college graduate and academics)	870	.318	.466	.317
Marital status	= 1 if a study included only married workers	41	1.268	.449	1.264
Narrow occupation	= 1 if a study used narrow occupation/sector data only	870	.214	.410	.181
Academia (VCs)	= 1 if a study used Vice-Chancelor's data	870	.068	.252	.100
Age_22	=1 if a study included workers under 22 years old	870	.411	0.492	.533
Age_22-29	=1 if a study included workers between 22 and 29 years old	870	.847	.360	.846
Age_30-39	=1 if a study included workers between 30 and 39 years old	870	.836	.371	.888
Age_40-49	=1 if a study included workers between 40 and 49 years old	870	.726	.432	.830
Age_50	=1 if a study included workers over 50 years old	870	.733	.442	.782
Perc10	=1 if the study included the 10% lowest-paid workers	870	.837	.370	.878
Perc10-30	=1 if the study included workers in the bottom 10 to 30 percent of the pay scale	870	.891	.312	.896
Perc30-50	=1 if the study included workers in the bottom 30 to 50 percent of the pay scale	867	.840	.367	.886
Perc50	= 1 if the study included workers in the middle of the pay scale	870	.838	.369	.881
Perc50-80	=1 if the study included workers in the bottom 50 to 80 percent of the pay scale	867	.893	.310	.901
Perc80	=1 if the study included workers in the bottom 80 to 90 percent of the pay scale	870	.838	.369	.881
Perc90	=1 if the study included the 10% highest-paid workers	870	.872	.334	.945
Alternative measures of wages					
Annual	= 1 if a study used annual salary as its measure of wages	868	.146	.354	.186
Week	= 1 if a study used weekly salary as its measure of wages	868	.008	.089	.010
Month	= 1 if a study used monthly salary as its measure of wages	868	.038	.191	.023
Hourly_d	=1 if a study used hourly salary directly as its mesasure of wages	870	.638	.481	.584
Hourly_c	= 1 if a study used hourly wages computed from daily, weekly, monthly or annual salary	868	.179	.38	.212
Econometric approach and Specification					
Selection	= 1 if a study did correct for selection bias	869	.099	.299	.079
Dummy	= 1 if a dummy for gender is used (to investigate the GWG and no blinder-Oaxaca decomposition)	863	.320	.467	.341
RE	= 1 if a random-effects approach is used	870	.030	.170	.033
OLS	= 1 if an OLS estimator is used for estimation	870	.421	.494	.393
IV	= 1 if a study used instrumental variables	870	.010	.101	.021
Panel	= 1 if a study used panel data	870	.102	.303	.131

Pooled	= 1 if a study used pooled cross-sectional data	870	.209	.407	.150
Blinder-Oaxaca	= 1 if Blinder-Oaxaca decomposition was used	853	.393	.489	.343
Neumark	= 1 if Neumark decomposition was used	853	.018	.132	.028
Variables for worker's characteristics					
Immig	= 1 if s study failed to account for immigrants status	870	.977	.150	.981
Training	= 1 if a study omitted the worker's training	870	.823	.382	.864
Exp/Age	= 1 if a study omitted the worker's job experience or worker's age	870	.461	.499	.456
Ind	= 1 if a study omitted the worker's industry of employment	870	.798	.402	.763
Govt	= 1 if a study omitted a government/private employment distinction	869	.877	.329	.805
Union status	= 1 if a study omitted worker's union/non-union status	870	.828	.378	.812
Occupation	= 1 if a study omitted worker's occupation	870	.671	.470	.661
Education	= 1 if a study omitted worker's education	870	.475	.500	.525
Race	= 1 if a study failed to account for race	870	.826	.379	.840
Marital status	= 1 if a study omitted worker's marital status	870	.614	.487	.675
Kids	= 1 if a study omitted whether or not worker has children	870	.747	.435	.807
Tenure	= 1 if a study omitted tenure	870	.708	.455	.742
Share of females in occupation	= 1 if a study omitted the percentage of women in the workers' job	869	.936	.246	.904
Health	= 1 if a study omitted to control for the health condition of workers	870	.898	.303	.938
Unemployment	= 1 if a study omitted to account for worker's unemployment experience in the past	870	.889	.315	.924
Urban	= 1 if a study omitted city size	870	.941	.235	.898
Working time	= 1 if a study omitted worker's working time	870	.970	.170	.974
Region	= 1 if a study omitted geographical area of employment	870	.734	.441	.703
Firm size/employer size	= 1 if the estimation simultaneously controls for the size of the firm to which workers belong	870	.798	.402	.779
Year dummies	= 1 if a study omitted to account for time (year dummies)	870	.883	.322	.904
Temporary	= 1 if a study omitted to account for temporary work	870	.844	.362	.880
Part-time	= 1 if a study omitted to account for part-time work	861	.603	.490	.591
Publication					
Citations	= The number of per-year citations of the study since its first appearance on Google Scholar (derived from Google scholar citations)	870	8.880	17.247	6.245
JIF	= 5 years Journal Impact factor in 2020 (Source: JCR)	870	2.239	1.929	2.011
Unpublished	= 1 if the study is not published in a peer-review journal	870	.210	.408	.222
Book	= 1 if the study is published in a book	870	.014	.117	.033
Male	= 1 if a study was authored solely by men	870	.274	.446	.378
Dependent Variables					
G_j	= the j^{th} estimate of the log gender gap (raw or adjusted)	870	.186	.126	.178

Notes: SD= Standard deviation. WM = Mean weighted by the inverse of the number of estimates reported per study.

The variables collected from the primary studies that are used in the MRA are subdivided into six groups: data characteristics; different measures of the dependent variable; econometric approaches used; sample composition; model specification; and publication and citation characteristics.

Data characteristics: we control for the vintage of the data by creating a variable that reflects the midpoint of the sample used in a particular estimate. We differentiate between public and private sector studies. Another dummy distinguishes studies that are UK-only from those using data from a number of countries including the UK - although we only use results for UK in these cases. Two further dummies distinguish between studies using multi-sector versus single sector data, and a dummy identifying data solely from the academic sector (and Vice Chancellors in particular). In addition, dummy variables are used to identify studies that focus on a single occupation or on the level of jobs held by employees, differentiating between highly-skilled and less-skilled positions. Finally, dummy variables identify studies focusing solely on full-time employees and another identifies studies covering only married workers.

Measurement of the dependent variable: we include five dummy variables to control for the way the authors of the primary studies derive their wage measure. Studies differ in terms of the period over which earnings are collected (daily, weekly, monthly or annual salary) and according to the hours used to derive hourly earnings.

Econometric approach: Analysts only observe wages for those in employment with valid wage data. If one wishes to extrapolate about the GWG as it might look having accounted for non-random selection into employment one must devise a method to adjust results for non-random selection into employment. Only a subset of studies attempt this.[12] We therefore derive a dummy variable scoring '1' if a study adjusts for non-random selection into employment. Another dummy variable identifies those studies that pool men and women in the same earnings equation and recover the GWG with a sex dummy. Thirteen percent of the estimates come from specifications using panel data and twenty-three percent of the studies rely on pooled cross-section data. To see whether cross-sectional and panel data yield systematically different GWG estimates we include a corresponding dummy variable. Next, we construct a dummy variable identifying those studies using a Kitigawa/Blinder/Oaxaca decomposition approach to recover the GWG and the factors that contribute to it. We also introduce a series of dummies to distinguish the type of estimates used to measure GWGs (OLS, Random effects, Fixed-Effects, instrumental variables. Finally, a dummy was created to control for the use of a quantile regression approach in the primary studies.

Model specification: estimates of the adjusted GWG typically control for a series of variables such as age, education, experience, marital status and so on. We examine whether the inclusion (or exclusion) of such variables has a systematic influence on the estimated GWG. We code the variables such that "1" means that the control variable is *omitted*. The full list of the types of control variables used in the primary studies are presented in Table 3.

Publication characteristics: To see whether studies yield different results even when all the main aspects of methodology are controlled for, we include a dummy variable that equals one if the study is published in a peer-reviewed journal. To account for the different quality of publication outlets, we include the Web of Science impact factor. We also control for the number of Google Scholar citations of the study. Finally, we include a dummy variable that equals one if the study is authored solely by men.

[12] For a discussion of the methods used to adjust GWGs for non-random selection into employment see Bryson et al (2020).

The Multi-Level Meta-Regression Methodology

The conventional meta-regression approach in applied economics – often associated with the weighted least squares (WLS) MRA framework developed by Stanley and Doucouliagos (2012) – models reported effect sizes as independent observations and estimates:

$$y_i = x_i'\beta + \varepsilon_i,\ \ \varepsilon_i \sim N(0, v_i)$$

using inverse-variance weights $w_i = 1/v_i$ where v_i is typically constructed from reported standard errors. This approach relies critically on the availability of standard errors for all effect sizes in the sample. In our dataset, however, only a subset of studies reports sufficient information to compute precise sampling variances. Restricting the analysis to this subsample would substantially reduce statistical power and potentially introduce selection bias. For this reason, we do not rely on WLS-MRA as our primary specification. Instead, we adopt a multilevel meta-regression framework that does not require the full availability of reported standard errors. Specifically, we estimate:

$$y_{ij} = x_{ij}'\beta + u_j + \varepsilon_{ij}$$

with

$$u_j \sim N(0, \tau^2)$$

where effect sizes y_{ij} are nested within studies j. This specification explicitly models within-study dependence through the random effect u_j, rather than treating clustered estimates as independent observations. The between-study variance τ^2 is estimated via Restricted Maximum Likelihood (REML), which provides less biased variance-component estimates in finite samples and aligns with contemporary multilevel meta-analytic practice.

Because sampling variances are not consistently available, we approximate precision using sample-size-based proxies, exploiting the relationship between standard errors and effective sample size. As a robustness check, we re-estimate key specifications on the subsample for which standard errors are available and apply inverse-variance weighting. In addition, we compute cluster-robust variance estimators at the study level (reported in Appendix A3) to verify that inference is not driven by within-study correlation.

Finally, given the relatively large number of potentially correlated moderators compared to the number of independent studies, we complement REML estimation with a Bayesian multilevel regression incorporating shrinkage priors on moderator coefficients:

$$\beta_k \sim N\left(0, \tau_\beta^2 \gamma_k^2\right)$$

This approach regularizes weakly supported moderators toward zero while preserving substantively meaningful effects, mitigating overparameterization risk without relying on discrete model selection. All models are implemented in R. Multilevel REML estimations are conducted using the *metafor* package, while Bayesian hierarchical models are estimated using the brms package, which relies on Hamiltonian Monte Carlo sampling via Stan.

By combining multilevel REML estimation, shrinkage-based Bayesian modeling and robustness checks using precision proxies and cluster-robust inference, our approach aligns the econometric specification with the hierarchical structure of the data while accommodating incomplete reporting of standard errors.

Results

Modeling Between-Study Heterogeneity in Gender Wage Gap Estimates

To examine heterogeneity in reported GWG estimates Table 4 reports the results of three alternative specifications – fixed-effects weighted least squares (FE-WLS), multilevel REML and Bayesian multilevel models. Each coefficient represents the marginal change in the reported GWG associated with the presence of the corresponding study characteristic. Since moderators are coded as binary indicators, coefficients indicate the difference in the estimated wage gap between studies that include that feature and those that do not, holding other moderators constant. The magnitude of the coefficients can be interpreted as changes in the reported GWG measured in log points.

Table 4. Why do reported gender wage gap vary?

	Frequentist Model Averaging		Bayesian Model Averaging
Response variable: Gender Wage gap	Fixed-Effects Weighted Least Squares (FE-WLS)	Restricted Maximum Likelihood (REML)	Bayesian Multilevel Regression with Shrinkage (BMRS)
Constant	**0.270***(6.70)**	**0.212***(3.88)**	**0.185 [0.169 ;0.201]**
Study design			
Select	-0.024 (-1.53)	0.005 (0.16)	0.001 [-0.008 ; 0.011]
Dummy	0.002 (0.12)	-0.005 (-0.3)	-0.002 [-0.018 ; 0.014]
Panel	-0.03* (-1.89)	-0.011 (-1.02)	-0.004 [-0.015 ; 0.008]
Pooled	0.008 (0.65)	0.017 (0.91)	0.007 [-0.011 ; 0.025]
Estimation techniques			
Method: OLS	-0.014 (-1.61)	0.006 (0.5)	0.003 [-0.01 ; 0.016]
Method: Random-effect	0.043** (2.48)	0.032** (2.23)	0.005 [-0.003 ; 0.014]
Method: Instrumental Variable	-0.029 (-0.7)	0.049 (0.68)	0.005 [-0.002 ; 0.012]
Method: Blinder/Oaxaca	0.004 (0.41)	0.01 (0.8)	0.005 [-0.006 ; 0.016]
Method: Neumark	-0.013 (-0.22)	-0.023 (-0.53)	-0.003 [-0.013 ; 0.007]
Structural variation			
Data year	**-0.003*** (-3.65)**	**-0.003** (-3.31)**	**-0.042 [-0.053 ; -0.031]**
Nb_year	0.002 (1.53)	0 (-0.1)	-0.001 [-0.015 ; 0.013]
Narrow occupations only	**0.137*** (3.05)**	**0.121** (2.44)**	**0.049 [0.022 ; 0.074]**
White collar	**-0.108*** (-4.42)**	**-0.106*** (-5.24)**	**-0.049 [-0.068 ; -0.03]**
Private sector only	0.021** (2.43)	0.014 (1.17)	0.004 [-0.005 ; 0.014]
Country comparison	0.05** (2.12)	0.027 (1.09)	0.008 [-0.008 ; 0.025]
Academia (VCs)	**-0.144*** (-3.48)**	**-0.141*** (-3.29)**	**-0.026 [-0.041 ; -0.01]**
Full-time worker	**-0.055*** (-3.94)**	**-0.048** (-1.99)**	**-0.02 [-0.032 ; -0.009]**
Measures of wages			
Annual Salary	-0.01 (-0.3)	0.004 (0.12)	0.001 [-0.013 ; 0.016]
Monthly	0.099*** (3.25)	0.037 (0.77)	0.007 [-0.014 ; 0.027]
Weekly	**0.378*** (7.59)**	**0.293* (1.78)**	**0.026 [0.018 ; 0.034]**
Hourly constructed	0.024 (1.11)	0.017 (0.75)	0.006 [-0.009 ; 0.022]
Omitted Control variables			
Race	-0.015 (-0.86)	0.013 (0.51)	0.005 [-0.007 ; 0.017]
Immigrant	0.019 (0.81)	0.064 (1.57)	0.009 [-0.001 ; 0.02]
Marital	**0.023 (1.22)**	**0.044** (2.23)**	**0.021 [0.008 ; 0.035]**
Kids	-0.02 (-1.11)	-0.025 (-1.47)	-0.01 [-0.021 ; 0]
Experience/Age	**0.056*** (3.15)**	**0.035** (2.2)**	**0.017 [0.005 ; 0.03]**
Training	0.037** (2.23)	0.012 (0.75)	0.005 [-0.009 ; 0.018]
Tenure	0.019 (0.93)	0.02 (1.14)	0.009 [-0.005 ; 0.023]
Occupation	**0.021 (1.33)**	**0.029* (1.97)**	**0.014 [0.001 ; 0.027]**
Industry	0.021 (1.02)	0 (0)	0 [-0.011 ; 0.011]
Government	-0.009 (-0.36)	-0.006 (-0.27)	-0.002 [-0.015 ; 0.01]
Union	-0.005 (-0.18)	0.01 (0.36)	0.004 [-0.011 ; 0.018]
Education	**0.021 (0.97)**	**0.028* (1.75)**	**0.014 [0.001 ; 0.026]**
Female Share	0.028* (1.69)	0.01 (0.61)	0.002 [-0.007 ; 0.012]

Part-time	0.01 (0.9)	0.003 (0.21)	0.001 [-0.011 ; 0.013]
Worktime	0.011 (0.32)	-0.001 (-0.04)	0 [-0.011 ; 0.011]
Health status	-0.029 (-1.3)	-0.024 (-0.81)	-0.007 [-0.02 ; 0.005]
Unemployment experience	-0.029 (-1.2)	-0.027 (-1.01)	-0.009 [-0.02 ; 0.003]
Urban	-0.022 (-0.78)	0.003 (0.12)	0.001 [-0.01 ; 0.012]
Region	0.011 (0.68)	0.013 (0.78)	0.006 [-0.007 ; 0.018]
Employer size	-0.008 (-0.47)	0.017 (0.9)	0.007 [-0.004 ; 0.018]
Year dummies	-0.014 (-0.64)	-0.008 (-0.36)	-0.003 [-0.016 ; 0.01]
Temporary workers	**-0.017 (-0.97)**	**-0.031* (-1.79)**	**-0.011 [-0.023 ; 0]**
Publication characteristics			
Male author	0.016 (0.86)	-0.002 (-0.12)	-0.001 [-0.018 ; 0.015]
Journal Impact	-0.004 (-1.06)	-0.008 (-1.47)	-0.015 [-0.035 ; 0.003]
Study Citations,	-0.001** (-2.52)	0 (-0.84)	-0.005 [-0.023 ; 0.013]
Book chapter	**0.081** (2.56)**	**0.101** (2.4)**	**0.012 [0.001 ; 0.023]**
Unpublished	-0.043** (-2.45)	-0.031 (-1.51)	-0.013 [-0.033 ; 0.007]
SD Residual (REML)		0.083	
SD Study Intercept (REML)		0.058	
Observations	750	834	834
R^2	0.668	0.424	0.440

Note: (1) the coefficients are log points (2) the numbers in parentheses in columns 1 and 2 are t-statistics. *=statistically significant at a 90% confidence interval; **=statistically significant at a 95% confidence interval; ***=statistically significant at a 99% confidence interval (2) the numbers in parentheses in column 3 are the 95% confidence bands for the point estimate.

Several findings emerge. First, the estimated average GWG remains substantial across specifications, ranging from about 0.19 (Bayesian) to 0.27 log points. The smaller GWGs when estimated via REML and Bayesian methods compared with the FE-WLS suggests that ignoring between-study heterogeneity inflates the estimated mean GWG. Once study-level variation is explicitly modeled, the central tendency stabilizes around .19 to .27 log points, which likely provides a more reliable benchmark.

Second, methodological characteristics such as estimation technique (OLS, IV, random effects), study design (panel versus pooled data) and decomposition methods, explain relatively little of the cross-study variation once study-level clustering of estimates is accounted for. This pattern indicates that part of the apparent methodological heterogeneity is attributable to unmodeled between-study variation rather than systematic differences in econometric strategy.

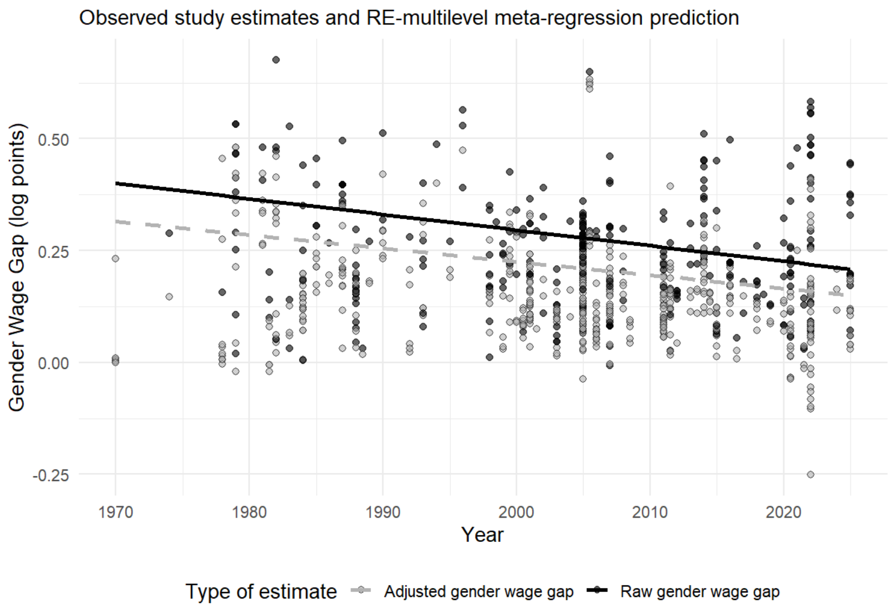
Third, structural characteristics display stronger and more robust effects. For example, the coefficient on data year is consistently negative and statistically significant across all specifications. In the Bayesian specification, predictors are standardized for numerical stability. Converting the midpoint coefficient back to its original scale indicates that the GWG declines by approximately 0.003 log-points per year (i.e. 0.3 percentage points per year):

$$\beta_{raw} = \frac{\beta_{std}}{SD} = \frac{-0.042}{13.221} = -0.00303 \approx -0.003\ log - points\ per\ year$$

This corresponds to a reduction of roughly 0.03 log-points per decade, closely mirroring the frequentist estimate and providing strong confirmation of a gradual long-run convergence in gender wage differentials. Figure 4 represents the temporal evolution of the gender wage gap, all other things equal.

Figure 4 includes both raw and adjusted wage gap estimates. This is intentional, as the meta-regression explicitly examines how specification choices – particularly the inclusion of control variables – affect the reported magnitude of the gender wage gap. This graph therefore reflects the temporal evolution of estimates in the literature rather than a direct estimate of the underlying population wage gap. Since adjusted gaps typically control for worker and job characteristics and are therefore smaller, the observed downward trend may partly reflect changes in estimation practices over time rather than solely changes in the true gender wage gap. As a result, Figure 4 should be interpreted as documenting shifts in empirical approaches as well as potential changes in the underlying phenomenon.

Figure 4. The gender wage gap over time



Similarly, studies restricted to narrow occupational groups report larger gaps while samples focusing on white-collar workers or full-time employees tend to find smaller gaps. Estimates based on academic labor markets (e.g. vice-chancellors) also report significantly smaller wage differentials. These patterns suggest that occupational structure and labor market segmentation play a central role in shaping reported gender wage gaps.

Fourth, measurement choices, particularly wage periodicity, appear influential under the FE-WLS model but are substantially attenuated in the REML and Bayesian frameworks. The large weekly wage effect observed in FE-WLS shrinks considerably once between-study heterogeneity is modeled, suggesting that fixed-effects specifications may overstate measurement-driven variation.

Fifth, publication characteristics provide limited evidence of systematic bias. While citation counts and publication status appear significant in the FE-WLS model, these effects weaken or disappear in the multilevel and Bayesian models. This suggests that publication-related heterogeneity is modest once study-level clustering is properly addressed.

The comparison of model fit further reinforces these conclusions. The FE-WLS model yields a relatively high R^2 (0.668) but this likely overstates explanatory power because it ignores the hierarchical structure of the data. In contrast, the REML marginal R^2 (0.424) and Bayesian R^2 (0.440) provide more conservative and arguably more appropriate measures of explained heterogeneity. The estimated standard deviation of the study-level intercept (0.058) confirms substantial between-study variation, justifying the use of multilevel modeling.

Overall, the results indicate that structural and contextual characteristics, rather than methodological differences, account for most of the systematic heterogeneity in reported GWG estimates. Once between-study variance is properly modeled, the core findings become more stable and conservative, with the long-term decline in the GWG emerging as the most robust and consistent pattern across specifications.

We now turn to the specific study-level determinants driving this heterogeneity. In particular, our meta-regression sheds light on how key modeling choices and omitted covariates shape reported GWG estimates.

Specification Choices and Systematic Variation in Reported GWG Estimates

One of the clearest findings is that omitting **marital status** is associated with a larger reported gender wage gap (β=0.044, p<0.05 in the REML estimates). In other words, studies that control for marital status tend to report smaller wage differentials. This pattern is theoretically coherent. Marriage is strongly associated with divergent labor market trajectories for men and women. Becker's (1985) household specialization framework predicts that married women are more likely to reduce labor supply or invest less in market-oriented human capital. If marital status is omitted, part of the wage differential reflecting household specialization will be mechanically attributed to gender itself. Thus, the positive association between omission of marital status and the reported GWG suggests that some portion of the raw GWG reflects differences in the family-related division of labour rather than direct labor market discrimination. Including marital status absorbs part of this structural variation. Importantly, this does not imply that the gap is "explained away." Rather, it indicates that family structure operates as a mediating channel through which gendered labor market outcomes emerge.

The meta-regression shows that omitting controls for **experience or age** is associated with an increase in the estimated gap (β=0.035, p<0.05). This is entirely consistent with human capital theory. Women, on average, experience more career interruptions and shorter tenure due to childbearing and caregiving responsibilities. When experience is omitted, part of the wage differential reflecting accumulated human capital differences is captured by the gender coefficient. The persistence of a non-trivial gap even when experience is included, however, suggests that differential returns to experience – or discrimination conditional on experience – remain relevant. To complete the analysis, we carried out a meta-regression on sub-samples of primary studies, enabling us to test the relationship between worker's age and the GWG. The results are presented in Table 5 and Figure 5 which also present GWGs across the pay distribution, discussed below.

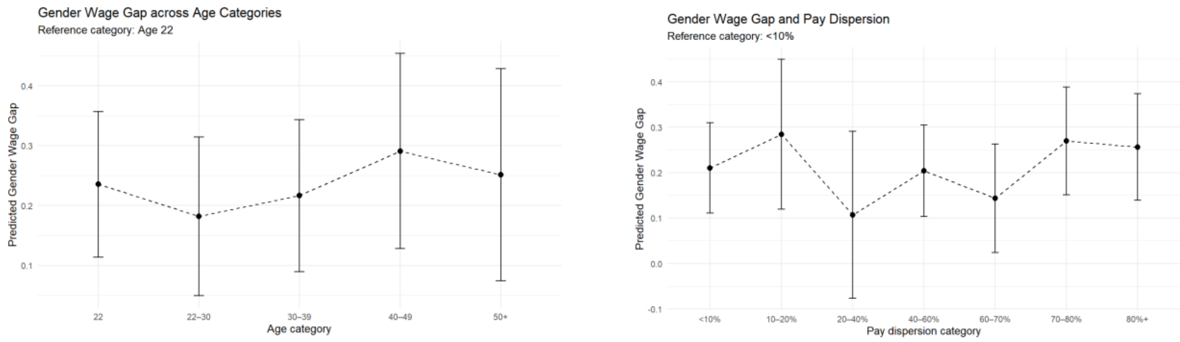
Table 5. The effects of worker's age and wage distribution on the gender wage gap

	MRA (REML)					
	Age (1)			Pay distribution (2)		
	Coef.	SE	p-value	Coef.	SE	p-value
Age						
Constant	0.235***	0.063	0.002			
<22 years old (ref.)	-	-	-			
22-29	-0.054*	0.027	0.084			
30-39	-0.019	0.018	0.332			
40-49	0.056	0.056	0.379			
>50 years old	0.016	0.066	0.822			
Other control variables		Yes				

Pay distribution				
Constant		0.210***	0.051	0.001
<20% (ref.)		-		
20%		0.074	0.067	0.352
30-40%		-0.103	0.079	0.317
50%		-0.006	0.001	0.390
60-70%		-0.067	0.034	0.144
80%		0.059	0.033	0.211
>80%		0.046	0.032	0.241
Other control variables			Yes	
Observations	834		831	

Note: *p<.10 ** p<.05 *** p<0.01. The model includes the full set of control variables employed in the previous frequentist and Bayesian specifications (see Table 4).

Figure 5. Worker's age, wage distribution and gender wage gap



Although there is a well-established hump-shape in the GWG over the life-course (Bryson et al., 2020) the meta-regression results provide limited support for a life-cycle pattern in the GWG. Indeed, the GWG appears to fall a little when workers are aged 22-29 years and remain fairly flat thereafter. The findings suggest that age does not systematically moderate the gender wage gap once other study characteristics are taken into account.

Including **occupation** controls reduces the estimated GWG. This reflects the role of occupational segregation. Gender differences in occupational sorting are one of the strongest empirical regularities in labor economics. If occupational controls are omitted, between-occupation wage differences are attributed to gender. When included, the estimated gap becomes conditional on within-occupation wage differences. The fact that occupation controls attenuate but do not eliminate the gap is theoretically important: it suggests that both segregation (between-job sorting) and within-job wage differentials contribute to observed gender inequality. In particular, it appears that studies focusing on GWG within academia (mainly vice chancellors) report smaller differentials than those observed in other industries (β=-0.141, p<0.01)..

Studies that restrict samples to full-time workers report smaller gender wage gaps (β=-0.048, p<0.10). This is because women are disproportionately represented in part-time and flexible arrangements, which often carry wage penalties. Restricting the sample to full-time workers effectively holds constant one major channel of wage differentiation. Consequently, the conditional GWG declines. In addition, the negative coefficient on the indicator for studies that do not control for temporary employment (β=-0.031, p<0.10) suggests that omitting contract type is associated with slightly smaller reported GWGs. This pattern is consistent with women being overrepresented in temporary positions. Once contract type is controlled for, the conditional GWG tends to be marginally larger. The effect is small but robust across specifications. These findings align with models emphasizing labor supply constraints and institutional rigidities rather than pure wage discrimination.

The inclusion of education as a control results in modest attenuation of the GWG. This is expected under human capital theory, as educational attainment differences historically contributed to wage disparities. However, in many advanced economies women now exceed men in formal educational attainment. The persistence of a GWG conditional on education in these studies partly reflects the vintage of the underlying data they use. It is plausible that analyses based on more recent cohorts of workers will mean education will close the GWG still further due to women outperforming men in education (Goldin, 2014). Any remaining gap will reflect differential returns to schooling rather than differences in endowments.

The weaker and less robust effects of training and tenure controls suggest that while these variables matter in specific contexts, they do not systematically drive cross-study heterogeneity. This may reflect measurement error in primary studies or the fact that tenure effects are partially captured by experience variables.

Finally, we also examine whether the GWG varies systematically across the wage distribution as several theoretical perspectives predict distributional heterogeneity. Glass ceiling and sticky floor hypotheses suggest that gender disparities may be particularly pronounced at the top or bottom of the wage distribution, reflecting barriers to promotion on the one hand and concentration in low-paid positions on the other. Models of discrimination and tournament theory further imply that gender differences may widen in higher-paying jobs where promotion dynamics and incentive structures amplify early career disparities. Investigating the GWG at different parts of the wage distribution allows us to assess whether the factors affecting wage gaps differ at different wage levels.

The results presented in Table 5 and Figure 5 do not provide strong evidence that the GWG systematically varies across individuals' positions in the wage distribution. Although some coefficients suggest larger gaps at higher percentiles and smaller gaps in intermediate categories, these differences are not statistically significant. This indicates that neither a robust "glass ceiling" nor a clear "sticky floor" pattern emerges once study characteristics are taken into account.

Discussion

Our meta-regressions provide new insights into the sources of heterogeneity in reported GWG estimates for the UK. We find differences in estimation methods play a relatively small role in explaining variance in GWGs across studies. Instead, our findings suggest they reflect what we describe as 'structural' variation emanating from the nature of the estimation sample which, in turn, reflects deeper structural mechanisms operating through labor supply decisions, occupational sorting and institutional context. However, the persistence of a sizable conditional GWG across specifications irrespective of model specification also indicates that these structural mechanisms cannot fully account for the GWG.

We demonstrate that roughly half the raw GWG tends to be accounted when adjusting for potential confounders in regression analyses. However, one must be mindful of the difficulties interpreting regression-adjusted estimates where analysts condition on potentially endogenous variables which inadvertently partial out some of the gender differences driving the GWG. As noted earlier, classic human capital models (Mincer and Polachek, 1974; Polachek, 1975) emphasize that wage differentials may reflect forward-looking investment decisions shaped by expected lifetime labor force attachment. In this framework, women anticipating intermittent participation due to fertility or household specialization may rationally invest less in market-oriented human capital. Crucially, these expectations are not directly observed in most cross-sectional datasets. Instead, researchers typically rely on proxies such as age, tenure, experience, marital status, occupation or current employment status.

Our meta-results show that the inclusion of such controls systematically attenuates reported GWGs. In particular, conditioning on marital status, occupational structure, full-time employment and accumulated experience reduces the magnitude of the estimated gap. This pattern is consistent with expectation-driven human capital models: when variables capturing household specialization or labor supply constraints are omitted, part of their effect is absorbed by the gender coefficient itself.

Importantly, however, these controls are imperfect proxies for future work expectations. Observed experience at a single point in time does not capture anticipated career interruptions; marital status does not fully measure expected fertility trajectories and occupational controls may reflect both voluntary sorting and constrained opportunity sets. Thus, the attenuation of the gap when such variables are included should not be interpreted as fully isolating discrimination from structural differences. Rather, it indicates that a meaningful share of cross-study heterogeneity reflects differences in how studies model labor supply and occupational selection channels.

The persistent negative time trend we document further supports a structural interpretation of the GWG and is consistent with the downward impetus given to unequal treatment by legislation, changes in social norms and other factors. The estimated decline of roughly 0.3 percentage points per year is consistent with declining fertility, rising female labor force attachment and institutional change documented in recent work (e.g. Goldin, 2014 for USA). As anticipated lifetime participation converges across genders, expectation-driven human capital differences likely diminish, contributing to the observed narrowing of the gap, while the persistence of social expectations may contribute to its persistence.

Nevertheless, a substantial conditional wage differential remains even in studies incorporating extensive controls. Under our most conservative Bayesian multilevel specification, the average gap is log points (approximately 17 percent of mens pay). This residual differential cannot be fully attributed to observable human capital or occupational sorting proxies. It may reflect unobserved productivity characteristics, institutional rigidities, bargaining differences or discriminatory wage-setting practices within occupations. Our analysis does not permit a clean decomposition among these mechanisms but it does suggest that structural explanations attenuate rather than eliminate the gap.

Our findings also clarify the role of methodological variation. Once study-level heterogeneity is explicitly modeled, differences in estimation technique (OLS, IV, random effects) contribute relatively little to cross-study variation. This suggests that reported differences in GWG estimates are less a product of econometric strategy and more a reflection of underlying structural and institutional conditions.

Taken together, the evidence supports a layered interpretation of the GWG. A first layer reflects structural differences in labor supply and occupational sorting shaped by household expectations and institutional constraints. A second layer captures within-occupation wage differentials that persist conditional on observable characteristics. The coexistence of these layers cautions against interpreting the residual component mechanically as discrimination while equally cautioning against viewing the gap as entirely structural.

By situating methodological heterogeneity within this broader theoretical framework, the meta-analysis advances the literature beyond the binary opposition between "discrimination" and "human capital". Instead, it highlights how forward-looking investment decisions, institutional context and wage-setting practices jointly generate the observed distribution of gender wage gap estimates.

Conclusion

Ours is the first meta-analysis that examines variance in the GWG in the UK. It draws on 870 estimates from 90 studies conducted over the period 1974 to 2024. In contrast to early meta-analyses, we deploy multiple meta-regression techniques to show how our results differ when we account for the non-independence of multiple estimates nested within the same study.

In some respects our results reflect what we thought we knew from the myriad separate studies that have been conducted.

First, and most notably, we observe a raw GWG of around 25 log points which falls to around half that size when adjusting for covariates. That leaves a very substantial residual gender wage gap. It is not straightforward interpreting what lies behind this residual gap. Some of it may indeed reflect persistent discriminatory practices on the part of employers (and perhaps co-workers and customers) which works to the disadvantage of women in the labor market. Evidence supporting this proposition comes from the law courts where women continue to win equal pay for work of equal value claims (Labour Research Department, 2026). But it is also likely to reflect omitted variables which differ systematically across men and women which also reflect earnings, some of which, in turn, reflect differences in preferences, responses to social norms, and personality traits.

Second, we see convergence in the GWG at a rate of around 0.3 percentage points per annum. The rate of convergence is not constant over time, but at this rate of convergence it would take women many years to reach wage parity with men.

Third, we show that there is substantial heterogeneity in estimates of the GWG which reflect sample composition. This is hardly surprising since we know that experiences of men and women and the factors determining their pay vary markedly across the labor market according to contractual arrangements, occupation and working hours.

Fourth, what the analyst conditions on really matters. Model specifications differ in notable ways across studies making inferences about the GWG and its origins difficult unless one undertakes a meta-analysis which brings this analytical choice into focus. Much of the recent literature has been focused on trying to close the residual GWG by loading in control variables which capture some aspects of what is usually unobserved, such as personality traits. It turns out that these additional controls matter, but only at the margins. What matters most are the treatment of human capital variables and their interaction with family commitments. This is a tricky area because, as has been demonstrated so eloquently in the literature – and examined very carefully in some studies for other countries (eg. Adda et al., 2017) – a number of key 'controls' are, in fact, not independent of gender. Quite the opposite: they reflect choices made endogenously by men and women as they consider what weight to attach to progression in the labor market. This does not mean that analysts should set to one side such controls. Rather, they must exercise caution when interpreting GWG estimates which adjust for such variables.

Our meta-analyses have also revealed some surprises. We find some evidence consistent with publication biases, something that is not possible to establish in the absence of meta-analyses. However, a sizeable GWG is apparent having netted out the effects of publication bias. Contrary to expectations, we find limited evidence of age-related variance in the GWG. One must treat this finding with some caution because it is not easy to distinguish between age and cohort effects with our data. A second surprise is that methodological differences in estimating the size of the GWG are not large. This might come as a particular surprise to economists who are schooled in the importance of carefully designing estimation strategies in the knowledge that they can often produce markedly different results. This is not to say that estimation methods do not matter. Rather, most of the across-study variance in the GWG is accounted for by other factors.

Although the meta-analyses presented here draw on a well-established body of econometric research, many of these studies are based on studies using data collected some time ago. We need to bear in mind the ongoing changes in the experiences of men and women in the labour market over the years. These include women's advances in education such that, in the most recent cohorts women are outperforming men (Bryson et al., 2020); changes in fertility which have partially closed the work experience gap between men and women (ONS, 2025b); and changes in social norms which, other things equal, reduce barriers to women's progression in the labor market. And yet, the ONS estimates the GWG officially to be 12.8 percent in 2025 (ONS, 2025a), convergence in the GWG has slowed in the USA (Fry and Arragao, 2025). In the case of graduates in the UK, the GWG has even reversed recently (Foliano et al., 2024), and big issues such as sexual harassment in the workplace are largely unexplored in the literature as potential causes of the GWG. We anticipate that recent legislative changes in the UK including requirements for large employers to report their raw GWGs and a new statutory duty on employers to prevent sexual harassment and create a safe working environment could impact the size of the GWG found in future meta-analyses for the UK.

Reference list (* denotes studies included in the analysis)

Adda, J., Dustmann, C. and Stevens, K (2017) The Career Costs of Children. *Journal of Political Economy,* vol. 125, no. 2, , pp. 293-337.

*Amadxarif, Z., Angeli, M., Haldane, A., & Zemaityte, G. (2020). *Understanding pay gaps*. Staff Working Paper No. 877, Bank of England.

*Anspal, S. (2015). Non-parametric decomposition and the choice of reference group: A study of the gender wage gap in 15 OECD countries. *International Journal of Manpower*, *36*(8), 1266-1280.

Arrow, K. J. (1973). The theory of discrimination. In O. Ashenfelter & A. Rees (Eds.), *Discrimination in labor markets* (pp. 3–33). Princeton University Press.

*Arulampalam, Wiji, Booth, Alison L. and Bryan, Mark L. (2007) Is there a glass ceiling over Europe? Exploring the gender pay gap across the wage distribution. *Industrial & Labor Relations Review*, Volume 60 (Number 2). pp. 163-186

Azmat, G., & Petrongolo, B. (2014). Gender and the labor market: What have we learned from field and lab experiments? *Labour Economics*, 30, 32–40.

Babcock L and Laschever, S (2003) *Women Don't Ask: Negotiation and the Gender Divide*, Princeton: Princeton University Press.

*Bachan, R. (2008). On the determinants of pay of CEOs in UK public sector higher education institutions. IZA Discussion Paper, No 3858, IZA Bonn.

*Bachan, R., & Bryson, A. (2022). The gender wage gap among university vice chancellors in the UK. *Labour Economics*, *78*, 102230.

*Bargain, O., Doorley, K., & Van Kerm, P. (2019). Minimum wages and the gender gap in pay: new evidence from the United Kingdom and Ireland. *Review of Income and Wealth*, *65*(3), 514-539.

*Barron, D. N., & West, E. (2013). The financial costs of caring in the British labour market: Is there a wage penalty for workers in caring occupations?. *British Journal of Industrial Relations*, *51*(1), 104-123.

Becker, G. S. (1971). *The economics of discrimination* (2nd ed.). University of Chicago Press.

Becker, G. S. (1985) Human capital, effort and the sexual division of labor, *Journal of Labor Economics,* 3, 1, Part 2: S33-S58

*Bell, D., & Ritchie, F. (1998). Female earnings and gender differentials in Great Britain 1977–1994. *Labour Economics*, *5*(3), 331-357.

*Black, B., Trainor, M., & Spencer, J. E. (1999). Wage protection systems, segregation and gender pay inequalities: West Germany, the Netherlands and Great Britain. *Cambridge Journal of Economics*, *23*(4), 449-464.

*Blackaby, D., Booth, A. L., & Frank, J. (2005). Outside offers and the gender pay gap: Empirical evidence from the UK academic labour market. *The Economic Journal*, *115*(501), F81-F107.

*Blau, F. D., & Kahn, L. M. (1996). Wage structure and gender earnings differentials: an international comparison. *Economica*, S29-S62.

Blau, F. D. & Kahn, L. M. (2017) The Gender Wage Gap: Extent, Trends and Explanations. *Journal of Economic Literature*, 55, 3: 789-865

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, 8(4), 436–455.

*Bowlus, A. J., & Grogan, L. (2009). Gender wage differentials, job search, and part-time employment in the UK. *Oxford Economic Papers*, *61*(2), 275-303.

*Brynin, M. (2017). *The gender pay gap*. Research Report 109, Equality and Human Rights Commission (EHRC).

Bryson, A, Joshi, H, Wielgoszewska, B, Wilkinson D, (2020)A short history of the gender wage gap in Britain. *Oxford Review of Economic Policy*, Volume 36, Issue 4, Winter 2020, Pages 836- 854, https://doi.org/10.1093/oxrep/graa046

*Butcher, T., Mumford, K. A., & Smith, P. N. (2019). *The Gender Earnings Gap in British Workplaces: A Knowledge Exchange Report*. DP in Economics, No 19/10, University of York.

*Chevalier, A. (2007). Education, occupation and career expectations: determinants of the gender pay gap for UK graduates. *Oxford Bulletin of Economics and Statistics*, *69*(6), 819-842.

*Chiplin, B., & Sloane, P. J. (1976). Personal characteristics and sex differentials in professional employment. *The Economic Journal*, *86*(344), 729-745.

*Christofides, L. N., Polycarpou, A., & Vrachimis, K. (2013). Gender wage gaps,'sticky floors' and 'glass ceilings' in Europe. *Labour Economics*, *21*, 86-102.

*Chzhen, Y., & Mumford, K. (2011). Gender gaps across the earnings distribution for full-time employees in Britain: Allowing for sample selection. *Labour Economics*, *18*(6), 837-844.

*Costa Dias, M., Joyce, R., & Parodi, F. (2020). The gender pay gap in the UK: children and experience in work. *Oxford Review of Economic Policy*, *36*(4), 855-881.

*Cukrowska-Torzewska, E., & Lovasz, A. (2020). The role of parenthood in shaping the gender wage gap–A comparative analysis of 26 European countries. *Social Science Research*, *85*, 102355.

*Dacre, J., Woodhams, C., Atkinson, C., Laliotis, I., Williams, M., Blanden, J., ... & Brown, D. (2020). *Mend the gap: the independent review into gender pay gaps in medicine in England*, University of Surrey, December.

*Davies, R., & Welpton, R. (2008). How does workplace monitoring affect the gender wage differential? Analysis of the annual survey of hours and earnings and the 2004 workplace employment relations survey—A research note. *British Journal of Industrial Relations*, *46*(4), 732-749.

*Davies, R., McNabb, R., & Whitfield, K. (2015). Do high-performance work practices exacerbate or mitigate the gender pay gap?. *Cambridge Journal of Economics*, *39*(2), 537-564.

*Dolton, P. J., & Kidd, M. P. (1994). Occupational Access and Wage Discrimination. *Oxford Bulletin of Economics & Statistics*, *56*(4).

*Dolton, P. J., & Makepeace, G. H. (1986). Sample selection and male-female earnings differentials in the graduate labour market. *Oxford Economic Papers*, *38*(2), 317-341.

*Dolton, P. J., & Makepeace, G. H. (1987). Marital status, child rearing and earnings differentials in the graduate labour market. *The Economic Journal*, *97*(388), 897-922.

*Dolton, P., & Ma, A. (2003). CEO pay in the public sector: The case of vice chancellors in UK universities. *Retrieved July*, *16*, 2011.

*Dolton, P., O'Neill, D., & Sweetman, O. (1996). Gender differences in the changing labor market: the role of legislation and inequality in changing the wage gap for qualified workers in the United Kingdom. *Journal of Human Resources*, 549-565.

*Drolet, M., & Mumford, K. (2012). The gender pay gap for private-sector employees in Canada and Britain. *British Journal of Industrial Relations*, *50*(3), 529-553.

*Ermisch, J. F., & Wright, R. E. (1991). Differential returns to human capital in full-time and part-time employment In *Issues in contemporary economics: proceedings of the ninth World Congress of the International Economic Association*, Athens, Greece, 195-213.

*Fitzroy, F., & Ward, M. (1999). Discrimination, segregation and firm effects: some UK microevidence. *Applied Economics Letters*, *6*(11), 711-715.

*Foliano, F., Bryson, A., Joshi, H., Wielgoszewska, B., & Wilkinson, D. (2024). Gender wage gap among young adults: a comparison across British cohorts. *Labour Economics*, 91, 102614.
Folke, O. and Rickne, J. (2022). Sexual harassment and gender inequality in the labor market. *The Quarterly Journal of Economics*, 137 (4), 2163–2212
*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*, 107-123.
Fry, R and Arrago, C (2025) Gender pay gap in U.S. has narrowed slightly over 2 decades, Pew Research, https://www.pewresearch.org/short-reads/2025/03/04/gender-pay-gap-in-us-has-narrowed-slightly-over-2-decades/
*Gamage, D. D. K., Kavetsos, G., Mallick, S., & Sevilla, A. (2020). Pay transparency initiative and gender pay gap: Evidence from research-intensive universities in the UK, IZA DP No 13635.
Goldin, C. (2014). A grand gender convergence: Its last chapter. American Economic Review, 104(4), 1091-1119.
Goldin, C. (2000) Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians, American Economic Review, 90, 4: 715-741
*Goy, S. C., & Johnes, G. (2008). That's what makes the difference today: an international analysis of the determinants of discrimination. *Lancaster University Management School, Economics Department Working Papers*, (4470).
*Graddy, K., & Pistaferri, L. (2000). Wage differences by gender: Evidence from recently graduated MBAs. *Oxford Bulletin of Economics and Statistics*, *62*, 837-854.
*Gradín, C., Del Río, C., & Cantó, O. (2010). Gender wage discrimination and poverty in the EU. *Feminist Economics*, *16*(2), 73-109.
*Gravelle, H., Hole, A. R., & Santos, R. (2011). Measuring and testing for gender discrimination in physician pay: English family doctors. *Journal of Health Economics*, *30*(4), 660-674.
*Greenhalgh, C. (1980). Male-female wage differentials in Great Britain: Is marriage an equal opportunity?. *The Economic Journal*, *90*(360), 751-775.
*Grimshaw, D. (2000). Public sector employment, wage inequality and the gender pay ratio in the UK. *International Review of Applied Economics*, *14*(4), 427-448.
*Harkness, S. (1996). The gender earnings gap: evidence from the UK. *Fiscal Studies*, *17*(2), 1-36.
Havránek, T; Horváth, R.; Iršová, Z; Rusnak, M. (2013) : « Crosscountry heterogeneity in intertemporal substitution », IES Working Paper, No. 11/2013, Charles University in Prague, Institute of Economic Studies (IES), Prague
*Healy, G., & Ahamed, M. M. (2019). Gender pay gap, voluntary interventions and recession: the case of the British financial services sector. *British Journal of Industrial Relations*, *57*(2), 302-327.
*Hedija, V. (2017). Sector-specific gender pay gap: evidence from the European Union Countries. *Economic research-Ekonomska istraživanja*, *30*(1), 1804-1819.
*Hedija, V. (2018). Is the Rule of Law Significant for the Explanation of the Differences in the Gender Pay Gap?. *Prague Economic Papers*, *27*(6), 704-722.
Hinne M, Gronau QF, van den Bergh D, Wagenmakers E-J. A Conceptual Introduction to Bayesian Model Averaging. Advances in Methods and Practices in *Psychological Science.* 2020;3(2):200-215
Irsova Z. & Havranek, T (2013), "Determinants of Horizontal Spillovers from FDI: Evidence from a Large Meta-Analysis." *World Development* 42, 1-15.

Iwasaki, I. & Ma, X, (2020). "Gender Wage Gap in China: A Large Meta-Analysis," CEI Research Paper Series 2020-5, Center for Economic Institutions, Institute of Economic Research, Hitotsubashi University.

Jann, B. (2008). The Blinder-Oaxaca decomposition for linear regression models. *The Stata Journal* 8(4): 453-479.

Jarrell, S. B &. Stanley, T. D (2004). "Declining Bias and Gender Wage Discrimination? A Meta-Regression Analysis," *Journal of Human Resources*, vol. 39(3) : *828-838*

*Jewell, S. L., Razzu, G., & Singleton, C. (2020). Who works for whom and the UK gender pay gap. *British Journal of Industrial Relations*, *58*(1), 50-81.

*Johnes, J., & Virmani, S. (2020). Chief executive pay in UK higher education: the role of university performance. *Annals of Operations Research*, *288*(2), 547-576.

*Jones, M. K., & Kaya, E. (2021). *The gender pay gap in UK medicine*, IZA DP No14177.

*Jones, M., Makepeace, G., & Wass, V. (2018). The UK gender pay gap 1997–2015: What is the role of the public sector?. *Industrial Relations: A Journal of Economy and Society*, *57*(2), 296-319.

*Joshi H , Bryson, A, Wilkinson D and Ward, K (2021) The Gender Gap in Wages over the Life Course: Evidence from a British Cohort Born in 1958 . *Gender Work and Organization,* 28(1) pp 397-415

*Joshi, H., & Newell, M. L. (1987). *Pay differences between men and women: longitudinal evidence from the 1946 birth cohort* (No. 156). CEPR Discussion Papers.

*Joshi, H., Makepeace, G., & Dolton, P. (2007). More or less unequal? Evidence on the pay of men and women from the British Birth Cohort Studies. *Gender, Work & Organization*, *14*(1), 37-55.

*Kidd, M. P., Phimister, E., & Ferko, I. (2003). Are employment effects of gender discrimination important? Some evidence from Great Britain. *The Manchester School*, *71*(6), 593-610.

Kitagawa, E. M. (1955). Components of a difference between two rates. *Journal of the American Statistical Association*, 50 (272), 1168–1194.

Kleven, H., Landais, C., Posch, J., Steinhauer, A. and Zweimuller, J. (2019) Child Penalties across Countries: Evidence and Explanations. *American Economic Association Papers and Proceedings*, 109: 122-126

Kleven, H., Landais, C., Posch, J., Steinhauer, A. and Zweimuller, J. (2024) Do Family Policies Reduce Gender Inequality? Evidence from 60 Years of Policy Experimentation. *American Economic Journal: Economic Policy*, 16, 2: 110-149

Kleven, H., Landais, C., Lassen, A. S. S., Rosenbaum, P., Steingrimsdottir, H. and Sogaard, J. E. (2026) Expanding Paternity Leave: Effects on Beliefs, Norms and Gender Gaps, *NBER Working Paper No. 34862*

Labour Research Dept, (2026) Payouts agreed in landmark equal pay claim, https://www.lrd.org.uk/free-read/payouts-agreed-landmark-equal-pay-claim

*Landmesser, J. M. (2019). Differences in income distributions for men and women in the European Union countries. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *14*(1), 81-98.

Lippens, L., Vermeiren, S. and Baert, S. (2023) The state of hiring discrimination: a meta-analysis of (almost) all recent correspondence experiments, *European Economic Review,* 151: 14315

*Lissenburgh, S. (2000). *Gender discrimination in the labour market* (No. 3). Policy Studies Institute, UK.

*Machin, S., & Puhani, P. A. (2003). Subject of degree and the gender wage differential: evidence from the UK and Germany. *Economics Letters*, *79*(3), 393-400.

*Main, B. G. (1991). Gender Differences in Wage Rates, Work Histories, and Occupational Segregation. *Journal of Economic Studies*, *18*(2).

*Makepeace, G., Dolton, P., & Joshi, H. (2004). Gender earnings differentials across individuals over time in British cohort studies. *International Journal of Manpower*, *25*(3/4), 251-263.

*Makepeace, G., Paci, P., Joshi, H., & Dolton, P. (1999). How unequally has equal pay progressed since the 1970s? A study of two British cohorts. *Journal of Human Resources*, 534-556.

*Manning, A., & Robinson, H. (2004). Something in the way she moves: a fresh look at an old gap. *Oxford economic papers*, *56*(2), 169-188.

*Manning, A., & Swaffield, J. (2008). The gender gap in early-career wage growth. *The Economic Journal*, *118*(530), 983-1024.

*Matteazzi, E., Pailhé, A., & Solaz, A. (2018). Part-time employment, the gender wage gap and the role of wage-setting institutions: Evidence from 11 European countries. *European Journal of Industrial Relations*, *24*(3), 221-241.

*McNabb, R., & Wass, V. (1997). Male-female salary differentials in British universities. *Oxford Economic Papers*, *49*(3), 328-343.

*McNabb, R., & Wass, V. (2006). Male-female earnings differentials among lawyers in Britain: a legacy of the law or a current practice?. *Labour Economics*, *13*(2), 219-235.

*Miller, P. W. (1987). The wage effect of the occupational segregation of women in Britain. *The Economic Journal*, *97*(388), 885-896.

Mincer, J. (1974). *Schooling, experience, and earnings*. NY: Columbia University Press

Mincer, J. and Polachek, S. (1974) "Family Investment in Human Capital: Earnings of Women," *Journal of Political Economy*, 82 : S76 - S108

*Morris, Stephen, Rosalind Goudie, Matt Sutton, Hugh Gravelle, Robert Elliott, Arne Risa Hole, Ada Ma, Bonnie Sibbald, and Diane Skåtun (2011). *Determinants of general practitioners' wages in England*. *Health Economics* 20, no. 2, 147-160.

*Mumford, K., & Smith, P. N. (2007). The gender earnings gap in Britain: including the workplace. *The Manchester School*, *75*(6), 653-672.

*Neuburger, J. (2010). *Trends in the unequal pay of women and men across three British generations* (Doctoral dissertation, Institute of Education, University of London).

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. International Economic Review, 14(3), 693–709.

ONS (Office for National Statistics (ONS) (2025a), Gender pay gap in the UK: 2025: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/bulletins/genderpaygapintheuk/2025

ONS (2025b) Childbearing for women born in different years, England and Wales: 2023, Statistical Bulletin. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/conceptionandfertilityrates/bulletins/childbearingforwomenbornindifferentyearsenglandandwales/2023

*Olsen, W. K., & Walby, S. (2004). Modelling gender pay gaps. *EOC Working Paper Series*.

*Olsen, W., Gash, V., Sook, K., & Zhang, M. (2018). *The gender pay gap in the UK: evidence from the UKHLS*. Government Equalities Office.

*Petrongolo, B., & Ronchi, M. (2020). Gender gaps and the structure of local labor markets. *Labour Economics*, *64*, 101819.

Polachek,S, (1975) "Expected Post-School Investment as a Determinant of Market Wage Differentials," *International Economic Review*, 16 (2) , 451-470

*Rahman, M. M., & Khatoon, R. (2012). Why do men earn more than women? An analysis using British Household Panel Survey. *The Bangladesh Development Studies*, 27-42.

Rubery, J. (2017) *Why is women's work low-paid*?, Oxfam Discussion Papers, Oxfam: Oxford
*Schulze, U. (2015). The gender wage gap among PhDs in the UK. *Cambridge Journal of Economics*, *39*(2), 599-629.
*Scicchitano, S. (2012). The male–female pay gap across the managerial workforce in the United Kingdom: a semi-parametric decomposition approach. *Applied Economics Letters*, *19*(13), 1293-1297.
*Shannon, M., & Kidd, M. P. (2005). Projections of the future path of the gender wage gap in Great Britain. *International Journal of Manpower*, *26*(4), 350-363.
*Siebert, W. S., & Sloane, P. J. (1981). The measurement of sex and marital status discrimination at the workplace. *Economica*, *48* (190), 125-141.
*Siebert, W. S., & Young, A. (1983). Sex and family status differentials in professional earnings: the case of librarians. *Scottish Journal of Political Economy*, *30*(1), 18-41.
*Sissoko, S. (2007). Wage inequalities in Europe: influence of gender and family status: a series of empirical essays. PhD Thesis, ULB--Universite Libre de Bruxelles.
*Sloane, Peter J. (1994): The Gender Wage Differential and Discrimination in the Six SCELI Local Labour Markets, in: *Gender Segregation and Social Change, Men and Women in Changing Labour Markets*, edited by MacEwen, A., Oxford University Press, 157-204
Stanley, T. D., & Doucouliagos, H. (2012). *Meta-regression analysis in economics and business*. London: Routledge.
Stanley T,D & Jarrell, S B (1998.) "Gender Wage Discrimination Bias? A Meta-Regression Analysis," *Journal of Human Resources*, vol. 33(4): 947-973.
*Swaffield, J. (2000). *Gender, motivation, experience and wages* (No. 457). London School of Economics and Political Science. Centre for Economic Performance.
*Swaffield, J. K. (2007). Estimates of the impact of labour market attachment and attitudes on the female wage. *The Manchester School*, *75*(3), 349-371.
*Theodoropoulos, N., Forth, J. and Bryson, A. (2022) "Are Women Doing It For Themselves? Gender Segregation and the Gender Pay Gap", *Oxford Bulletin of Economics and Statistics*, 84, 6: 1329-1355
*Tojerow, I. (2008). Industry wage differentials, rent sharing and gender: three empirical essays. ULB--Universite Libre de Bruxelles.
*Triventi, M. (2013). The gender wage gap and its institutional context: a comparative analysis of European graduates. *Work, Employment and Society*, *27*(4), 563-580.
*Turnbull, P., & Williams, G. (1974). Sex differentials in teachers' pay. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *137*(2), 245-258.
*Walby, S., & Olsen, W. (2003, July). The UK gender wage gap and gendered work histories. In Conference of the British Household Panel Survey, Institute for Social and Economic Research, University of Essex.
*Waldfogel, J. (1995). The price of motherhood: family status and women's pay in a young British cohort. *Oxford Economic Papers*, *47*(4), 584-610.
*Walker, J., Greve, P., Wood, G., & Miskell, P. (2019). Because you're worth it? Determinants of Vice Chancellor pay in the UK. *Industrial Relations Journal*, *50*(5-6), 450-467.
Weichselbaumer, D., & Winter-Ebmer, R. (2005). A meta-analysis of the international gender wage gap. *Journal of economic surveys*, *19*(3), 479-511.
*Winder, K. L. (2009). Flexible scheduling and the gender wage gap. *The BE Journal of Economic Analysis & Policy*, *9*(1). "T
*Wright, R. E., & Ermisch, J. F. (1991). Gender discrimination in the British labour market: a reassessment. *The Economic Journal*, *101*(406), 508-522.
*Zabalza, A, and. Arrufat J L. (1985) In Zabalza< A and Tzannatos "The extent of sex discrimination in Great Britain. Pp 70-96 -in " *Women and Equal Pay: The effects of*

legislation on female employment and wages in Britain . Cambridge: Cambridge University Press.

Appendix

TABLE A1
Studies excluded from the meta-analysis organized by the four reasons for exclusion

Articles	*Country*	*Outlets*	*Remarks/Comments*
Reviews and essays excluded			
Altonji & Blank (1999)	USA	Handbook of Labor Economics	Literature review
Azmat & Petrongolo (2014)	UK	Working Paper	A literature review on gender wage gap (field and Lab experiments)
Bryson, Joshi, Wielgoszewska &Wilkinson (2020)	UK	Oxford Review of Economic Policy	A short history of the gender wage gap in Britain No data on se of estimates Estimates not in logs and presented graphically
Kunze (2017)	OECD countries	IZA Discussion Paper	The gender wage gap in developed countries
Kunze (2018)	OECD countries	The Oxford Handbook of women and the Economy	The gender wage gap in developed countries
McGuiness & Pyper (2018)	UK	Briefing Paper	The gender pay gap
No quantifiable effect size available			
Ahamed, Wen & Gupta (2019)	UK	Economics Letters	Board composition and gender pay gap
Apergis, Hayat & Kadasah (2019)	UK	Applied Economics Letters	Gender pay gap and monetary policy shocks
Bagguley (1990)	UK	Service Industries Journal	Gender and Labour flexibility
Blackaby, Clark, Leslie & Murphy (1997)	UK	Oxford Economic Papers	The distribution of male and female earnings 1973-91
Blundell, Gosling, Ichimura & Meghir (2007)	UK	Econometrica	Distribution of male and female wages accounting for employment composition using bounds
Booth, Francesconi & Frank (2003)	UK	European Economic Review	A sticky floor model of promotion, pay and gender
Costa-Dias, Elming & Joyce (2016)	UK	IFS Briefing note	The gender wage gap
Costa-Dias, Joyce & Parodi (2018b)	UK	IFS Briefing note	Wage progression and the gender wage gap
Dex, Sutherland & Joshi (2000)	UK	NIER	Effects of minimum wages on the gender pay gap
Duchini, Simion & Turrell (2020)	UK	Warwick working paper	Pay transparency and cracks in the glass ceiling (to check again)
England, Budig & Folbre (2002)	UK	Social Problems	Wage of virtue: The relative pay of care work
Fortin, Bell & Böhm (2017)	UK	Labour Economics	Top earnings inequality (to check again)
Gomulka & Stern (1990)	UK	Economica	The employment of married women in the UK
Harkness (2016)	UK	European Sociological Review	The effect of motherhood on earnings
Joshi, Layard & Owen (1985)	UK	Journal of Labor Economics	Why are more women working in Britain?
Joshi, Paci & Waldfogel (1999)	UK	Cambridge Journal of Economics	The wages of motherhood
Kidd & Goninon (2000)	UK	Applied Economics Letters	Female concentration and the gender wage differential
Manning & Petrongolo (2008)	UK	The Economic Journal	The part-time pay penalty for women in Britain
Nightingale (2019)	UK	Community, work and Family	Part-time employment and the gender gap in low pay for UK employees
Neuburger, Kuh & Joshi (2011)	UK	Longitudinal and Life course studies	Cross-cohort changes in gender pay differences in Britain
Perales (2013)	UK	Work, Employment and Society	Occupational sex segregation and wages
Zabalza & Tzannatos (1985)	UK	The Economic Journal	The effect of Britain's anti-discriminatory legislation on relative pay and employment

Wage gap is not clearly measured/manipulated			
Dex et al. (2008)	UK	Working paper	Wage growth of men and women as a measure of pay gap
Dolton, Makepeace & Marcenaro-Gutierrez (2010)	UK	Education Economics	Differences in occupational wages (1991-2000)
Data included through other study			
Amadxarif, Angeli, Haldane & Zemaityte (2020)	UK	Bank of England Working paper N°877	Understanding pay gap
Arulampalam, Booth & Bryan (2005)	UK	ISER Working Paper	Is there a glass-ceiling over Europe?
Blackaby & Frank (2000)	UK	The Economic Journal	Ethnic and other minority representation in UK Academic economics
Costa-Dias, Joyce & Parodi (2018a)	UK	Working paper	The gender pay gap in the UK
Drolet & Mumford (2009)	UK/Can	IZA Discussion Paper	Gender pay gap for private sector employees in Canada and Britain
Drolet & Mumford (2008)	UK	Working paper	Through the snow and across the pond
Joshi & Paci (1998)	UK	Book	Unequal Pay for a women and men:evidence from the British Birth Cohor Studies. MIT
Waldfogel (1998)	UK	Journal of Labor Economics	The Family Gap for Young Women in the United States and Britain: Can Maternity Leave Make a Difference?

TABLE A2
Studies included in the meta-analysis

#	Study	*Sample size*	*# of est.*	*Period*	*Data*	*Average adjusted Gj*
1	Amadxarif, Angeli, Haldane, & Zemaityte (2020)	537,177	4	1994-2019	Labour Force Survey (LFS)	0.144
2	Anspal (2015)	3,175	1	2011-2012	PIAC	-
3	Arulampalam, Booth & Bryan (2007)	6,017 to 21,931	48	1995-2001	European Community Household Panel	0.193
4	Bachan (2008)	1,476	3	1997-2006	UK Higher Education Institutions (HEI)	0.059
5	Bachan & Bryson (2021)	2,300	9	2000-2019	115 UK Universities	0.015
6	Bargain, Doorley & Van Kerm (2019)	7,732	26	1998, 2000	British Household Panel Survey (BHPS)	0.163
7	Barron & West (2013)	987 to 3,368	6	1991-2007	British Household Panel Survey (BHPS)	0.085
8	Bell & Ritchie (1998)	-	18	1977-1994	New Earnings Survey Panel Dataset (NESPD)	0.185
9	Black, Trainor & Spencer (1999)	307	2	1989	International Social Survey Programme (ISSP)	0.474
10	Blackaby, Booth & Frank (2005)	88 to 351	11	1999	UK Survey of academic economists	0.081
11	Blau & Kahn (1996)	N/A	1	1985-1989	International Social Survey Programme (ISSP)	-
12	Bowlus & Grogan (2008)	600 to 874	4	1992	British Household Panel Survey (BHPS)	-
13	Brynin (2017)	744,335	4	1993-2014	Labour Force Survey (LFS)	0.117
14	Butcher, Mumford & Smith (2019)	17,763	2	2011	Workplace Employment Relations Survey (WERS) 2011	0.071
15	Chevalier (2007)	5,058	11	1995	Graduates from 33 UK higher education institutions	0.072
16	Chiplin & Sloane (1976)	198	9	1974	Multi-plant organization	0.104
17	Christofides, Polycarpou &Vrachimis (2013)	49,559	14	2007	EU-SILC	0.158
18	Chzhen & Mumford (2011)	274 to 3,030	5	2005	British Household Panel Survey (BHPS)	-
19	Costa-Dias, Joyce & Parodi (2020)	75,482	5	1991-2017	UK Household Longitudinal Survey (HLS) (BHPS+Understanding Society)	0.128
20	Cukrowska-Torzewska & Lovasz (2020)	39,207	2	2005-2013	EU-SILC (Eurostat)	0.206
21	Dacre & Woodhams (2020)	10,539,635	26	2009-2018	NHS Electronic Staff Record	0.045
22	Davies, McNabb & Whitfield (2015)	12,739 to 14,542	11	2004, 2011	WERS	0.132
23	Davies & Welpton (2008)	5,586 to 18,420	4	2004	WERS and Annual Survey of Hours and Earnings (ASHE)	0.160
24	Dolton & Kidd (1994)	4,470	3	1980	UK cohort	0.070

25	Dolton & Ma (2003)	357 to 1,007	4	1995-2001	Higher Education Statistics Agency (HESA)	0.035
26	Dolton & Makepeace (1986)	335	20	1977	DEUMS Survey	0.167
27	Dolton & Makepeace (1987)	919	2	1977	DEUMS Survey	-
28	Dolton, O'Neill & Sweetman (1996)	4,309 to 8,965	7	1967-1986	Graduate cohort data	0.031
29	Drolet & Mumford (2012)	3,515 to 14,272	12	2004	WERS	0.143
30	Ermisch & Wright (1991)	3,380	2	1980	WES	0.172
31	Fitzroy & Ward (1999)	34 to 74	3	?	PLC Scotland	0.050
32	Foliano, Bryson, Joshi, Wielgoszewska, Wilkinson (2024)	528 to 11,546	56	1972-2015	NSHD, NCDS, BCS, Next Steps	0.136
33	Fortin, Bell & Bohm (2017)	31,136 to 55,199	6	1997, 2015	ASHE	0.011
34	Gamage, Kavetsos, Mallick & Sevilla (2020)	64,772	1	2004-2016	HESA	0.109
35	Goy & Johnes (2008)	-	1	2004	ISSP	-
36	Graddy & Pistaferri (2000)	234	5	1992-1995	LBS	0.082
37	Gradin, del Rio & Canto (2010)	2,556	2	2001	European Community Household Panel (ECHP)	-
38	Gravelle, Hole & Santos (2011)	1,902	6	2008	NPCR	0.135
39	Greenhalgh (1980)	756, 772	6	1971, 1975	General Household Survey (GHS)	-
40	Grimshaw (2000)	N/A	4	1986, 1995	New Earnings Survey (NES)	-
41	Harkness (1996)	2,763 to 8,057	30	1973-1993	Family Expenditure Survey (FES), GHS, BHPS	0.280
42	Healy & Ahamed (2019)	2,882 to 13,997	45	2003-2017	LFS	0.149
43	Hedija (2017)	5,459	1	2011	EU-SILC	-
44	Hedija (2018)	?	6	2010-2012	EU-SILC	0.123
45	Jewell, Razzu & Singleton (2020)	541,346 to 1,708,132	17	2002-2016	ASHE	0.158
46	Johnes & Virmani (2020)	603	4	2009-2017	THE and HESA	-0.035
47	Jones & Kaya (2021)	243 to 35,560	24	2018	ASHE	0.107
48	Jones, Makepeace &Wass (2018)	204,681 to 684,551	6	1997-2015	Labour Force Survey (LFS)	0.139
49	Joshi, Makepeace & Dolton (2007)	5,363 to 6,850	12	1978, 1991, 2000	NSHD, NCDS, BCS	0.160
50	Joshi, Bryson, Wilkinson & Ward (2021)	801 to 7,848	40	1981-2013	NCDS	0.120
51	Joshi & Newell (1987)	255 to 1,946	6	1972, 1977	NSHD	0.267
52	Kidd, Phimister & Ferko (2003)	20,038	3	1991-1996	BHPS	0.119
53	Landmesser (2019)	8,179	2	2014	EU-SILC	0.250
54	Lissenburgh (2000)	1,208 to 4,267	13	1991-1995	BHPS	0.093
55	Machin & Puhani (2003)	5,166	4	1996	UK LFS	0.123
56	Main (1991)	2,092	2	1986	SCELI	0.114
57	Makepeace, Dolton & Joshi (2004)	5,363 to 6,850	6	1991, 2000	NCDS, BCS	0.166
58	Makepeace, Paci, Joshi & Dolton (1999)	263 to 1,421	6	1978, 1991	NCDS, NSHD	0.191
59	Manning & Swaffield (2008)	2,899 to 6,962	9	1991-2002	BHPS and BCS70	0.133

60	Manning & Robinson (2004)	14,559 to 43,8399	2	1991-2000	BHPS	0.280
61	Matteazzi, Pailhé & Solaz (2018)	N/A	2	2009	EU-SILC	0.110
62	McNabb & Wass (1997)	3,303 to 21,036	9	1975, 1985, 1992	USR	0.037
63	McNabb & Wass (2006)	834	2	1999, 2001	Law Society Survey of Solicitors (LSSS)	0.115
64	Miller (1987)	6,885	2	1980	BHS	0.170
65	Morris, Goudie, Sutton, Gravelle, Elliott, Hole, Ma, Sibbald & Skatun (2011)	2,271	3	2008	NPCR	0.024
66	Mumford & Smith (2007)	667	25	1998	WERS	0.128
67	Neuburger (2010)	1,488 to 8,060	18	1972, 2004	British Cohort Studies, LFS	0.250
68	Olsen, Gash, Kim & Zhang (2018)	N/A	4	2014/2015	UK HLS	0.037
69	Olsen & Walby (2004)	33,688	2	2001/2002	BHPS, LFS	0.092
70	Petrongolo & Ronchi (2020)	8,937	4	2017	LFS	0.162
71	Rahman & Khatoon (2012)		10	1991-2006	BHPS	
72	Schulze (2015)	84 to 1,048	16	2004-2005	DLHE Survey (HESA)	0.168
73	Scicchitano (2012)	2,413	20	2007	EU-SILC	0.254
74	Shannon & Kidd (2005)	358	2	1996	BHPS	0.117
75	Siebert & Sloane (1981)	73 to 151	4	1976	5 establishments	0.079
76	Siebert & Young (1983)	930	1	1977	Professional librarian Survey	0.006
77	Sissoko (2007)	19,593	8	1994, 2001	ECHP	0.160
78	Sloane (1994)	2,715	3	1986	SCELI	0.318
79	Swaffield (2000)	10,067 to 12,884	18	1991-1997	BHPS 91-97	0.161
80	Swaffield (2007)	10,064 to 12,979	14	1991-1997	BHPS 91-97	0.155
81	Theodoropoulos, Forth & Bryson (2019)	19,269 to 39,966	11	2004, 2011	WERS	0.134
82	Tojerow (2008)	79,563	2	1995	ESES	0.185
83	Triventi (2013)	NA	2	2005-2006	REFLEX survey	0.043
84	Turnbull & Williams (1974)	1,327 to 3,414	14	1963, 1971	School teacher surveys	0.027
85	Walby & Olsen (2003)	3,741	1	2000	BHPS	0.149
86	Waldfogel (1995)	7,639	2	1981-1991	NCDS	0.203
87	Walker, Greve, Wood & Miskell (2019)	146, 365	2	2013-2017	THE, HESA	-0.133
88	Winder (2009)	17,336 to 18,684	12	2004	WERS	0.078
89	Wright & Ermisch (1991)	3,962	12	1980	WES	0.249
90	Zabalza & Arrufat (1985)	6,319	7	1975	GHS	0.290

Table A3. Test of publication bias (Adjusted wage gap)

Linear techniques (Adjusted wage gap)	OLS	WAAP	RE	BE	Precision
Publication bias (Standard Error)	.304	-	.255	.142*	7.538***
	(.234)		(.281)	(.269)	(2.225)
Effect beyond bias (Constant)	.125***	.129***	.126***	.130***	.111***
	(.009)	(.012)	(.009)	(.010)	(.013)
Observations	306	268	306	306	306

Notes: Results of regression $GWG_{is} = GWG_0 + \gamma SE(GWG_{is}) + \varepsilon_{is}$, where GWG denotes the **adjusted wage gap** of the *i*-th estimate from the *s*-th study and SE (GWG) denotes its standard error. The standard errors of the regression parameters are clustered at the study level and shown in parentheses. OLS = ordinary least squares, FE = study-level fixed effects, RE = study-level random effects, Study = weighted by the inverse of the number of estimates reported per study, Precision = weighted by the inverse of the estimate's standard error.

Table A4. Test of publication bias (Raw wage gap)

Linear techniques (Raw wage gap)	OLS	WAAP	RE	BE	Precision
Publication bias (Standard Error)	.974*	-	.896	.720	11.163***
	(.440)		(.584)	(0.732)	(3.151)
Effect beyond bias (Constant)	.190***	.213***	.193***	.201***	.194***
	(.021)	(.030)	(.023)	(.026)	(.031)
Observations	79	75	79	79	79

Notes: Results of regression $GWG_{is} = GWG_0 + \gamma SE(GWG_{is}) + \varepsilon_{is}$, where GWG denotes the **raw wage gap** of the *i*-th estimate from the *s*-th study and SE (GWG) denotes its standard error. The standard errors of the regression parameters are clustered at the study level and shown in parentheses. OLS = ordinary least squares, FE = study-level fixed effects, RE = study-level random effects, Study = weighted by the inverse of the number of estimates reported per study, Precision = weighted by the inverse of the estimate's standard error.

Appendix: Robustness test

Using a Meta-regression Analysis (MRA), we try to find out which characteristics of study design systematically affect the reported gender wage gap by estimating the gender wage gap after corrections for publication bias. This step is to estimate an MRA model to investigate the heterogeneity and selection of reported research results:

$$(1)\ G_j = \beta_0 + \beta_1 SE_{ij} + \sum \beta_k Z_{ki} + \varepsilon_{ij}$$

where G_j denotes the gender wage gap, SE is the standard error of the gender wage gap, Z is a vector of moderator variables and ε is the disturbance term. Equation (1) is used to identify the factors that create heterogeneity in reported estimates. Some of this heterogeneity will reflect real moderators, but some will be created by research design choices. Heterogeneity can be identified and quantified by the Z vector in Equation (5).

However, as it is not possible to obtain a SE for all studies, several MRA models are proposed, using the inverse of the number of estimates per study as a weighting. The MRA results are presented in Table A5. Columns 1 and 2 present the results of applying a general to specific modelling strategy whereby we started with a pool of explanatory variables and then sequentially eliminated any that were not statistically significant at the 10% level of significance., For comparison purposes only, column 1 and 2 report the baseline OLS results. Column 3 and 4 present the results of models that are estimated by weighted least square (WLS). Here, all observations are weighted by the inverse of the number of estimates per study. In columns 5 to 6, we use multi-level and random-effects models as a method of addressing within-study dependence (Stanley & Doucouliagos 2012). Multilevel and Random Effects data analysis adjust standard errors for data dependence arising from multiple estimates reporting within studies. The results obtained are highly consistent with those derived from the REML and BMRS models in Table 4.

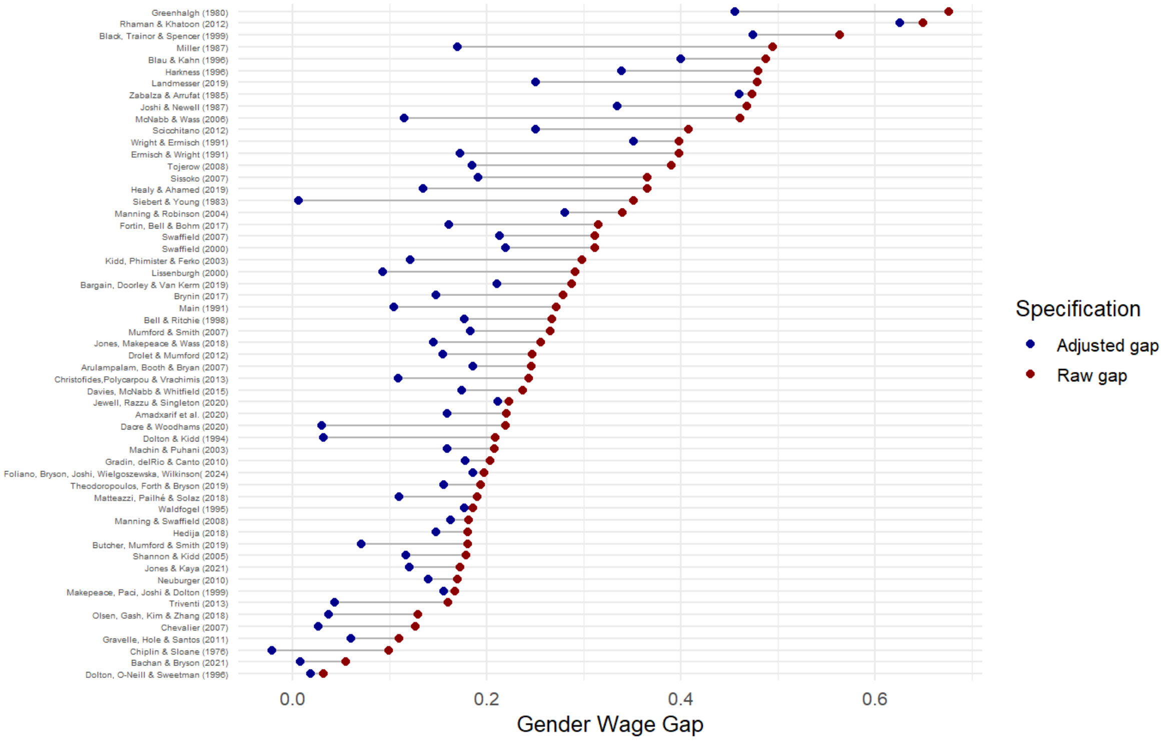
Table A5. Meta-regression Analysis (cluster robust)

Dependent variable : Gender wage gap estimates	(1) **Cluster Robust OLS (General)**	(2) **Cluster Robust OLS (Specific)**	(3) **Cluster Robust WLS (General)**	(4) **Cluster Robust WLS (Specific)**	(5) **Multi-level Mixed Effect (RM)**	(6) **Cluster Robust Random-Effect Panel GLS**
Study design						
Select	-0.0107		-0.00203		0.00412	0.00434
	(-0.71)		(-0.14)		(0.26)	(0.18)
Dummy	-0.00205		-0.0379***	-0.0496***	-0.00854	-0.00545
	(-0.17)		(-2.98)	(-5.37)	(-0.55)	(-0.38)
Panel	-0.0145		-0.0197	-0.0229*	-0.0121	-0.0112
	(-1.02)		(-1.56)	(-1.96)	(-0.69)	(-1.12)
Pooled	0.0172		0.0256**	0.0192**	0.0155	0.0164
	(1.22)		(2.18)	(1.97)	(0.79)	(0.98)
Estimation techniques						
Method: OLS	-0.0155	-0.0219***	0.0189*	0.0245***	0.00661	0.00662
	(-1.43)	(-2.88)	(1.92)	(2.95)	(0.53)	(0.55)
Method: Random-effect	0.0551**		0.0437**	0.0430**	0.0355	0.0322***
	(2.22)		(2.01)	(2.22)	(1.47)	(2.66)
Method: Instrumental Variable	0.0140		0.0573**	0.0611***	0.0471	0.0484
	(0.41)		(2.42)	(2.68)	(1.39)	(0.77)
Method: Blinder/Oaxaca	0.0149		0.00849		0.0106	0.00990
	(1.51)		(0.96)		(0.99)	(0.88)
Method: Neumark	-0.0613*	-0.0476*	-0.0554**	-0.0579***	-0.0289	-0.0218
	(-1.89)	(-1.76)	(-2.37)	(-2.81)	(-0.81)	(-0.65)
Structural variation						
Data year	**-0.00280*****	**-0.00282*****	**-0.00343*****	**-0.00348*****	**-0.00316*****	**-0.00320*****
	(-7.36)	**(-7.96)**	**(-9.35)**	**(-10.36)**	**(-7.65)**	**(-3.96)**
Nb_year	0.000919	0.00187**	-0.000652		0.0000632	-0.000148
	(0.87)	(2.34)	(-0.79)		(0.05)	(-0.10)
Narrow occupations only	**0.175*****	**0.166*****	**0.142*****	**0.138*****	**0.132*****	**0.120*****
	(9.14)	**(10.57)**	**(8.72)**	**(9.77)**	**(4.83)**	**(2.69)**
White collars	**-0.104*****	**-0.0941*****	**-0.131*****	**-0.132*****	**-0.108*****	**-0.106*****
	(-7.14)	**(-7.42)**	**(-10.98)**	**(-12.84)**	**(-5.80)**	**(-5.54)**
Private sector only	0.00861		0.0267**	0.0251**	0.0132	0.0140
	(0.69)		(2.44)	(2.48)	(0.93)	(1.26)
Country comparison	0.0221	0.0229**	0.0524***	0.0516***	0.0267	0.0276
	(1.24)	(2.04)	(4.48)	(4.87)	(1.17)	(1.20)
Academia (VCs)	**-0.179*****	**-0.186*****	**-0.125*****	**-0.123*****	**-0.145*****	**-0.140*****
	(-7.06)	**(-9.01)**	**(-6.06)**	**(-6.76)**	**(-4.00)**	**(-3.63)**
Full-time workers	**-0.0500*****	**-0.0382*****	**-0.0738*****	**-0.0736*****	**-0.0485*****	**-0.0477****
	(-4.31)	**(-4.60)**	**(-7.06)**	**(-9.51)**	**(-3.72)**	**(-2.15)**
Measures of wages						
Annual Salary	-0.0463***	-0.0423***	-0.00666		-0.00315	0.00419
	(-3.50)	(-3.71)	(-0.59)		(-0.17)	(0.15)
Monthly	0.0860***	0.104***	0.0197		0.0396	0.0371
	(3.43)	(5.27)	(0.79)		(0.88)	(1.08)
Weekly	**0.371*****	**0.367*****	**0.273*****	**0.276*****	**0.302*****	**0.293*****
	(8.84)	**(9.20)**	**(7.76)**	**(8.23)**	**(7.10)**	**(3.05)**
Hourly constructed	0.0124		0.00269		0.0151	0.0171
	(0.92)		(0.26)		(0.83)	(0.82)
Control variables						
Race	-0.0112		-0.0320***	-0.0302***	0.00999	0.0129
	(-0.79)		(-2.59)	(-2.69)	(0.64)	(0.63)
Immigrant	0.0485		**0.0684****	**0.0629****	**0.0622***	**0.0638****
	(1.51)		**(2.38)**	**(2.49)**	**(1.90)**	**(2.03)**
Marital	**0.0334*****		**0.0474*****	**0.0501*****	**0.0439*****	**0.0436****
	(2.70)		**(4.60)**	**(5.68)**	**(3.29)**	**(2.44)**
Kids	-0.0193		0.000686		-0.0246**	-0.0250*
	(-1.59)		(0.06)		(-1.99)	(-1.73)
Experience/Age	**0.0436*****	**0.0397*****	**0.0612*****	**0.0592*****	**0.0362*****	**0.0345****
	(3.43)	**(3.89)**	**(5.35)**	**(6.35)**	**(2.89)**	**(2.39)**
Training	0.00142		-0.00577		0.0111	0.0126
	(0.09)		(-0.44)		(0.65)	(0.95)

Tenure	**0.0339****	**0.0230****	**0.0420*****	**0.0389*****	0.0239	0.0206
	(2.48)	**(2.26)**	**(3.63)**	**(3.81)**	(1.62)	(1.33)
Occupation	**-0.00367**		**0.00660**		**0.0257****	**0.0290****
	(-0.31)		**(0.61)**		**(2.01)**	**(2.17)**
Industry	0.0332***	0.0295***	0.0146	0.0177**	0.00406	-0.000102
	(2.70)	(3.14)	(1.36)	(2.04)	(0.30)	(-0.01)
Government	-0.00366		0.00644		-0.00601	-0.00572
	(-0.23)		(0.50)		(-0.34)	(-0.29)
Union	0.00722		-0.00882		0.00791	0.00961
	(0.43)		(-0.57)		(0.43)	(0.42)
Education	**0.0423*****	**0.0518*****	**-0.00350**		**0.0276****	**0.0278****
	(3.65)	**(5.80)**	**(-0.34)**		**(2.31)**	**(2.03)**
Female Share	0.0215		0.00983		0.0127	0.0109
	(1.20)		(0.66)		(0.66)	(0.78)
Part-time	-0.00750		0.00315		0.00178	0.00268
	(-0.68)		(0.33)		(0.15)	(0.22)
Worktime	0.0481*	0.0424**	0.0674***	0.0656***	0.00345	-0.00261
	(1.66)	(2.10)	(2.64)	(2.95)	(0.11)	(-0.10)
Health status	-0.0454**	-0.0487***	-0.0194	-0.0278*	-0.0279	-0.0237
	(-2.51)	(-3.46)	(-1.14)	(-1.81)	(-1.42)	(-0.98)
Unemployment experience	-0.0233		0.0105		-0.0249	-0.0281
	(-1.46)		(0.65)		(-1.40)	(-1.17)
Urban	-0.00763		0.00163		0.00475	0.00509
	(-0.36)		(0.10)		(0.20)	(0.21)
Region	-0.00382		0.0120	0.0159*	0.0110	0.0134
	(-0.30)		(1.08)	(1.74)	(0.81)	(0.95)
Employer size	0.0141		0.00353		0.0166	0.0171
	(1.13)		(0.33)		(1.25)	(1.08)
Year dummies	-0.0212		-0.0588***	-0.0506***	-0.0117	-0.00873
	(-1.19)		(-3.93)	(-3.96)	(-0.61)	(-0.45)
Temporary workers	-0.0203		**-0.0487*****	**-0.0438*****	**-0.0315****	**-0.0313****
	(-1.37)		**(-3.56)**	**(-3.55)**	**(-2.08)**	**(-2.11)**
Publication characteristics						
Male author	0.0285***	0.0170*	0.00602		0.000809	-0.00258
	(2.71)	(1.93)	(0.71)		(0.05)	(-0.14)
Journal Impact	-0.00625**	-0.00609***	-0.00696***	-0.00583***	-0.00740*	-0.00768
	(-2.24)	(-2.70)	(-3.05)	(-2.94)	(-1.76)	(-1.61)
Study Citations	-0.000500		-0.000926**	-0.000897**	-0.000335	-0.000302
	(-1.26)		(-2.18)	(-2.44)	(-0.68)	(-0.99)
Book chapter	**0.0800****	**0.0729****	**0.0516****	**0.0572*****	**0.0992****	**0.103*****
	(2.55)	**(2.48)**	**(2.40)**	**(2.99)**	**(2.30)**	**(2.72)**
Unpublished	-0.0466***	-0.0374***	-0.0487***	-0.0440***	-0.0324	-0.0297
	(-3.21)	(-3.28)	(-4.35)	(-4.49)	(-1.56)	(-1.54)
Constant	0.204***	0.228***	0.232***	0.255***	0.210***	0.210***
	(4.62)	(9.11)	(5.50)	(7.42)	(4.49)	(4.67)
Constant					-3.084***	
					(-22.70)	
Constant					-2.502***	
					(-95.43)	
N	834	834	834	834	834	834

t statistics in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Estimates are from 86 studies

Figure A1. Wage Gap differentials (Raw – Adjusted wage gap) by study



Note: For visualization purposes, we retain one raw and one adjusted specification per study (selected based on the largest sample size), ensuring symmetric within-study comparisons in the dumbbell plot. In approximately 89% of studies, adjusted specifications yield smaller wage gaps than raw estimates. A small minority (about 10%) report larger adjusted gaps, likely reflecting compositional effects or specification differences rather than systematic patterns and are not included.

Using sample-size weighted multilevel REML models, we find that raw gender wage gaps are on average 10.3 percentage points larger than adjusted estimates ($\beta = 0.103$, $p < .001$). This implies that observable covariates account for approximately 38% of the raw wage gap, while a substantial residual gap of about 17% remains after adjustment.

Heterogeneity is substantially larger for raw wage gap estimates ($\tau^2 = 0.0173$) than for adjusted specifications ($\tau^2 = 0.0066$). This suggests that observed covariates not only reduce the average magnitude of the gender wage gap, but also account for a significant share of cross-study heterogeneity. Put differently, adjustment procedures substantially compress the dispersion of reported wage gaps across studies as suggested in Figure A1.

The average adjusted wage gap is ($\widehat{\beta_0}$=0.1671), whereas raw specifications are on average 10.3 percentage points larger ($\widehat{\beta_{raw}}$= 0.1033, p = .0000). This suggests that approximately 38% of the raw gender wage gap is accounted for by observed covariates included in primary studies. $Raw = 0.1671 + 0.1033 = 0.2704$. $Explained\ Share = \frac{0.1033}{0.2704} \approx 38\%$. Approximately 60% of the gap remains unexplained. Even after controlling for education, experience, occupation, and other observable characteristics, a substantial differential persists.

Technical Appendix:
Modeling Within-Study Dependence in Meta-Regression

1. REML Multi-Level Meta-regression

Let y_{ij} denote effect size estimates i from study j, with known sampling variance v_{ij}. The standard multilevel meta-regression model is:

$$y_{ij} = x'_{ij}\beta + u_j + \varepsilon_{ij}$$

Where:

x_{ij} is a vector of moderators
β is the parameter vector
$u_j \sim N(0, \tau^2)$ capture between-study heterogeneity
$\varepsilon_{ij} \sim N(0, v_{ij})$ represents sampling error (v_{ij} is the known sampling variance)

REML estimation maximizes the restricted likelihood to obtain consistent estimates of τ^2 and β, correcting for the downward bias in variance component estimation typical of standard ML. Inference proceeds via asymptotic normality of $\hat{\beta}$. τ^2represents the between-study variance, capturing the true heterogeneity in effect sizes across studies beyond sampling error.

In our dataset, multiple effect sizes are reported within the same primary study. Let y_{ij} denote effect size i from study j. Because these estimates are derived from the same sample, research design or estimation strategy, they are correlated:

$$Cov(y_{ij}, y_{i'j}) \neq 0 \text{ for } i \neq i'$$

Ignoring this dependence leads to (i) downward-biased standard errors, (ii) overstated effective sample size, and (iii) inflated statistical significance.

The appropriate unit of replication is the study, not the individual estimate. Accordingly, dependence must be modeled either structurally (multilevel modeling) or through variance adjustment (cluster-robust inference). Under REML, τ^2 is estimated via restricted likelihood, which corrects for finite-sample bias in variance component estimation. This specification explicitly models within-study dependence rather than correcting for it *ex post*. This is clearly the choice we made in our meta-analysis.

As a robustness check (see Table A5), cluster-robust variance estimators (CRVE) can be applied:

$$\widehat{VAR_{CR}}(\hat{\beta}) = (X'WX)^{-1}\left(\sum_j X'_j \widehat{u_j}\widehat{u_j}' X_j\right)(X'WX)^{-1}$$

where clustering occurs at the study level. This approach adjusts standard errors but does not alter point estimates or model structure. CRVE is a useful robustness tool, but it corrects standard errors rather than modeling dependence. It also relies on large-cluster asymptotics and it does not estimate between-study heterogeneity. Moreover, it may be inefficient when hierarchical structure is strong.

In meta-analysis with multiple estimates per study, multilevel random-effects models are generally more appropriate when the goal is structural modeling rather than purely corrected inference.

2. Bayesian Multilevel Regression with Shrinkage (BMRS)

We estimate the same hierarchical structure within a Bayesian framework:

$$y_{ij} \mid \beta, u_j \sim N\left(x'_{ij}\beta + u_j, v_{ij}\right)$$
$$u_j \sim N(0, \tau^2)$$

To regularize moderator coefficients, we impose shrinkage priors:

$$\beta_k \sim N(0, \gamma_k^2)$$

where γ_k^2 may itself follow a hierarchical prior (e.g., global–local shrinkage such as horseshoe):

$$\gamma_k \sim HalfCauchy\left(0, \tau_\gamma\right)$$

This specification induces continuous shrinkage toward zero for weakly supported moderators while allowing substantial effects to remain large.

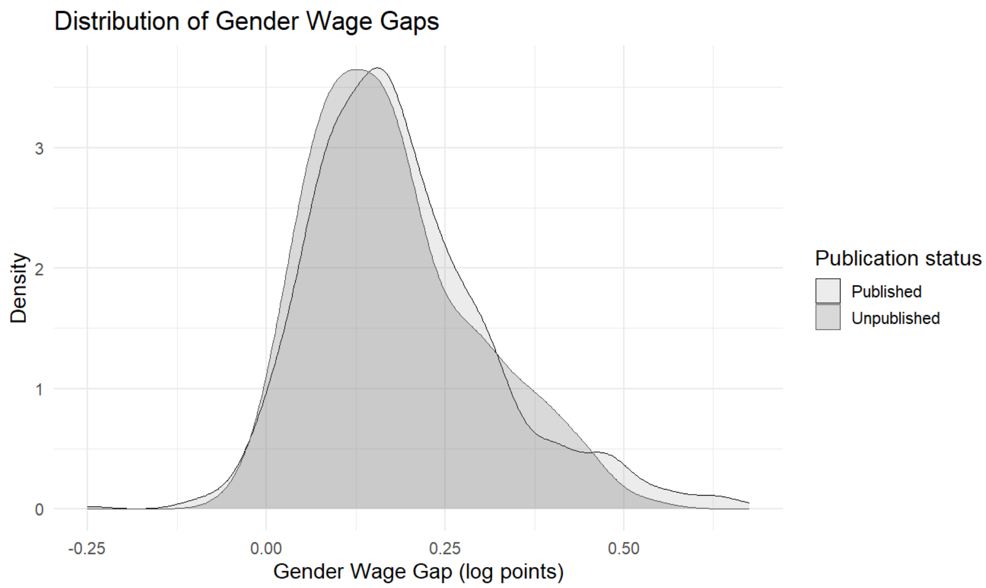
REML provides likelihood-based point estimates and heterogeneity diagnostics (τ^2), whereas BMRS integrates hierarchical modeling with regularization, mitigating overfitting when the moderator dimension is large relative to the number of studies. Concordance across the two frameworks increases confidence that estimated moderator effects – such as institutional moderators of wage-gap estimates – are not artifacts of model complexity.

TABLE A6
Sample Sizes

	Raw Wage Gap	Adjusted Wage Gap	Total estimates
All estimates	315	555	870
Estimates with SE	80	313	393
Estimates with SE + trim 1%	79	306	385

Note: For certain analyses, we trim the data at the 1% level, excluding the most extreme observations in the upper and lower tails of the distribution to mitigate the influence of outliers.

Appendix: Published vs Unpublished

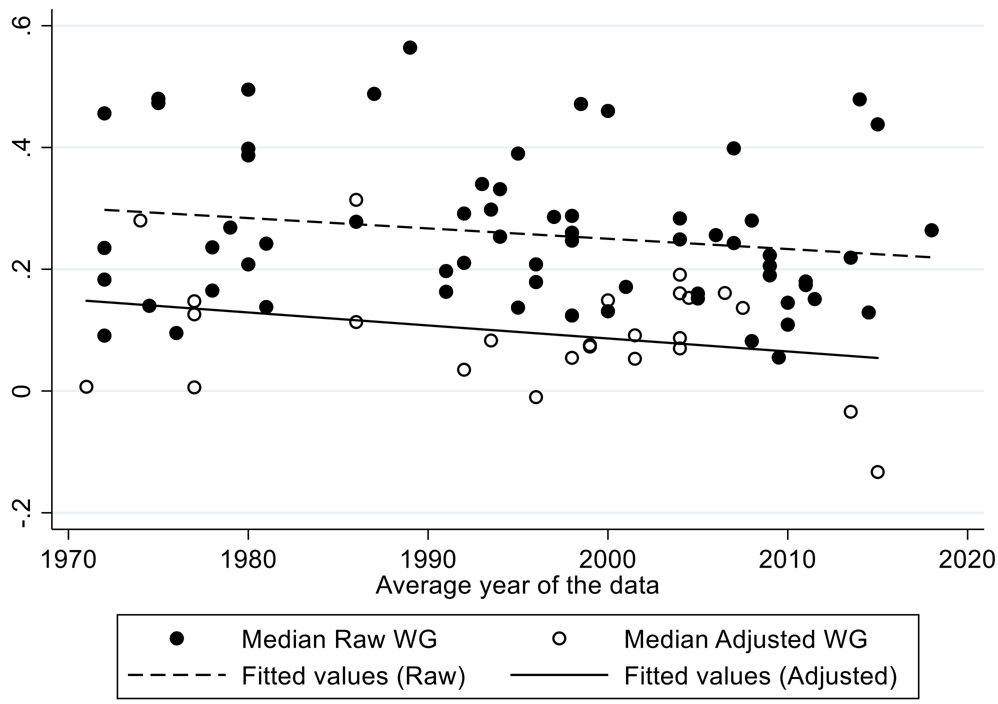


The figure shows that the distributions of published and unpublished gender wage gap estimates are broadly similar in shape and central tendency. Both peak in the range of roughly 0.10–0.20 log points and exhibit right-skewness, with a small number of larger reported gaps. Published estimates appear slightly more concentrated around the central mass, while unpublished results display somewhat greater dispersion in the upper tail. Overall, the substantial overlap between the two distributions provides no strong visual evidence of systematic differences by publication status, although formal tests would be required to assess publication bias more rigorously.

Appendix: Time trend

The Figure presents the evolution of the gender wage gap. The horizontal axis indicates the reported effects in chronological order (of the average date of the estimates in a study, some of which span several decades). The simple linear trend line fitted to these data shows a decline in the reported gender wage gap. Although this might be consistent with a decline in the gender wage gap over time, the decline may be also due to studies and data improving over the years. The publication trend may be driven by changes in research design, the data and/or econometric techniques over the years. The results from our multilevel models confirm this trend while controlling for a range of covariates. That said, we cannot rule out the possibility of omitted variable bias that could call this temporal evolution into question.

Figure 2. The gender wage gap over time



Notes: Each point on the graph represents the median value of the gender wage gap (adjusted + unadjusted wage gap) calculated from all the estimates drawn from each study.