Discussion Paper Series

IZA DP No. 18960

September 2026

The Gender Wage Gap Among Those Born in 1958: Conditioning on What and How?

Alex Bryson
University College London
and IZA@LISER

IZA@LISER NETWORK
ISSN: 2365-9793

The Gender Wage Gap Among Those Born in 1958: Conditioning on What and How?*

Abstract

Most of the literature estimating the gender wage gap (GWG) in the UK uses regression techniques adjusting for a range of covariates measured in adulthood, many of which are likely endogenous. In doing so, they typically find that about half the raw gap can be accounted for by gender differences in those conditioning variables. Using data for a birth cohort born in 1958 collected prospectively across ten survey sweeps, I depart from the literature in two ways. First, I condition solely on covariates measured in childhood to minimize potential biases arising from conditioning on endogenous variables. Second, I use a range of matching estimators to recover the GWG among 'like' men and women. I show that the raw GWG rises until cohort members enter their 40s, then falls but remains sizeable until age 63. In contrast to literature, when conditioning solely on childhood variables the covariate-adjusted GWG is similar to the raw GWG. This is the case at all ages regardless of the matching estimator I use and the set of conditioning childhood covariates. The implication is that among those born in 1958 women have lower earnings than men across the life-course which cannot be accounted for by their pre-labour market experiences.

JEL classification

J16, J31, J71

Keywords

gender wage gap, birth cohort, National Child Development Study, propensity score matching, entropy weighting, inverse probability weighted regression-adjustment (IPWRA)

Corresponding author

Alex Bryson
a.bryson@ucl.ac.uk

* *Acknowledgements*: I acknowledge funding from the Economic and Social Research Council (grant number ES/S012583/1). I thank Bernd Fitzenberger, Heather Joshi, Steven Raphael and participants at seminars at UC Berkeley's Institute for Research on Labor and Employment (IRLE), IAB (Nuremberg), UCL's Social Research Institute, the EU's Joint Research Centre's Competence Centre on Microeconomic Evaluation, Ispra, Italy, and the Colloquium of Personnel Economics (COPE) in Amsterdam for comments. I also thank Bozena Wielgoszewska and David Wilkinson for their valuable assistance with the data.

1. Introduction
In recent decades the gap in hourly earnings between men and women in the UK and elsewhere has been closing as women's potential earnings have risen due to increases in work experience and educational attainment, the two cornerstones to human capital (Bryson et al., 2020; Mincer, 1974; Mincer and Polachek, 1974). Changes in social norms governing gendered roles in the home and in the workplace have facilitated rising female labour force participation and entry to high-paid occupations in what Goldin (2014) has described as the "great gender convergence". Legislative efforts to overcome discriminatory hiring and pay practices which penalize women have also played a role (Zabalza and Arrufat, 1985; Zabalza and Tzannatos, 1985). And yet, a sizeable gender wage gap (GWG) persists in the UK (ONS, 2025) and elsewhere, even if one accounts for differential selection into employment over time (Blau et al., 2024; Foliano et al., 2024).

Efforts to account for the UK GWG via covariate-adjustment and related decomposition techniques (Oaxaca, 1973; Kitagawa, 1955) typically find that about half the raw gap can be accounted for by gender differences in those conditioning variables (Laroche et al., 2026). However, many of the controls used in those studies might be considered 'bad' controls in the sense that they are partly determined by gender and, as such, are potentially endogenous. These concerns extend to fertility and educational investment decisions, for example (Blau and Kahn, 2017: 838-839).

Using data for a birth cohort born in 1958 collected prospectively across 10 survey sweeps, I depart from the literature in two ways. First, I condition solely on covariates measured in childhood to minimize potential biases arising from conditioning on endogenous variables. Second, I use a range of matching estimators to recover the GWG among 'like' men and women. I show that the raw GWG rises until cohort members enter their 40s, then falls but remains sizeable until age 63. In contrast to the literature, when conditioning solely on childhood variables the covariate-adjusted GWG is similar to the raw GWG. This is the case regardless of the matching estimator I use. The implication is that among those born in 1958 women have lower earnings than men across the life-course which cannot be accounted for by their pre-labour market experiences.

The remainder of the paper is organized as follows. In Section Two reviews relevant literature regarding estimation of the GWG in the UK and methodological choices in doing so. Section Three introduces our data and estimation techniques. In Section Four we present our findings and Section Five Concludes.

2. Literature

2.1: Estimates of the Gender Wage Gap in the UK
Drawing on 870 estimates published in 90 primary studies examining the GWG in the UK, Laroche et al. (2026) estimate the raw GWG fell by roughly 0.3 percentage points per annum between 1974 – when the earliest study was conducted – and 2024, a continuation of a trend that goes back over a century (Bryson et al., 2020). The covariate-adjusted GWG is also declining over time, in large part because "women have increased their productivity enhancing characteristics and as they 'look' more like men the human capital part of the wage difference has been squeezed out" (Goldin, 2014: 1094).

Laroche et al. (2026) examine potential reasons for heterogeneity in the estimated GWG across UK studies using a multi-level meta-regression approach (MRA). In addition to the time the data were observed, they emphasize two sources of variance. First, they consider the estimation method. Unsurprisingly the most common method used to estimate the GWG is Ordinary Least Squares (OLS), which accounts for 42 percent of the estimates in the meta-analysis. They find methodological choices regarding estimators contribute relatively little to the cross-study variance in the GWG. Matching estimators are not separately identified in the analysis, presumably because none of the papers they reviewed used a matching estimator.

A second source of variance in the estimated GWG that they consider is model specification. Across all studies, the raw GWG averages 25 log points, but it falls by around one-half to 13 log points when adjusting for covariates. The authors find that "conditioning on marital status, occupational structure, full-time employment and accumulated experience reduce the magnitude of the estimated gap" (p. 25). However, as the authors note, "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" (Laroche et al., 2026: 25).

Birth cohort studies such as the one used in this paper track individuals over the life cycle. They find the GWG is small early on in workers' careers, then begins to widen until they are in their early 40s, only to narrow thereafter, even once one accounts for selection into employment (Bryson et al., 2020; Joshi et al., 2021; Neuberger, 2010). Using the birth cohort used in this paper, Joshi et al. (2021) find that conditioning on covariates such as work experience and family formation accounts for around half the raw GWG between ages 42 and 55, and around two-fifths of the raw GWG at age 33.

However, the GWG has been falling across birth cohorts for those of a similar age, such that new entrants to the labour market are liable to face lower GWGs than their counterparts from previous generations (Bryson et al., 2020; Neuberger et al., 2011; Foliano et al., 2024).

2.2: Which variables to condition on to minimise potential endogeneity issues

Aware of the bias that may arise in estimating the GWG when conditioning on potentially endogenous covariates, analysts have adopted one of two approaches. The first is to estimate a dynamic life cycle model accounting for the endogeneity of choices liable to affect earnings, such as those relating to fertility, marital status, savings, labour supply and occupational choice. Adda et al. (2017) construct such a model to estimate the career costs of children in Germany. They emphasize the costs of fertility associated with occupational choice, intermittent labour supply, lower investments in skills, and skill atrophy due to time out of work and reduced hours as central in understanding the life-cycle pattern in the GWG.

The second approach is to condition solely on covariates measured early in life prior to labour market entry. This may help ameliorate the problem of conditioning on endogenous variables but does not eliminate it. In deciding which childhood variables to condition on we follow Heckman et al. (2006) and others who emphasize the importance of both "cognitive and noncognitive abilities [which] are shaped early in the life cycle..and…are crucial to social and economic success" (pp. 413-414).

2.2.1: Cognitive skills

Heckman et al. capture cognitive skills with test scores in childhood warning that later schooling is "a choice variable, and any convincing analysis must account for the endogeneity of schooling…Removing the conditioning on schooling solves the problem of endogeneity of schooling in wage equations and produces estimates of the net effect of abilities on wages (their direct effects plus their net effects through schooling)" (2006: 416).

An early seminal paper examining wage differentials conditioning solely on pre-labour market variables is Neal and Johnson (1996). They examine black-white wage differentials among workers in their late-20s in the 1991 National Longitudinal Survey of Youth (NLSY), controlling for their Armed Forces Qualification Test (AFQT) undertaken before labour market entry. The only other variables they condition on are age and race. In robustness tests they also condition on a wider range of post-compulsory education variables but they prefer their AFQT-only estimates because schooling "beyond the age of compulsory attendance [is] endogenous" (p. 876).

A previous study using the National Child Development Study (NCDS) I use in this paper finds reading and mathematics test scores at age 7 are strong predictors of socio-economic status in midlife, even when controlling for potential confounders like social class. The size of the effects appears similar for men and women (Ritchie and Bates, 2013).

2.2.2: Non-cognitive skills

Heckman et al. (2006) measure noncognitive skills in childhood using validated scales capturing concepts such as locus of control and perceptions of self-worth. They find "important gender differences in the effects of these skills" (p. 412) on an array of labour market outcomes.

Parsons et al. (2024) use both the NCDS and its successor, the British Cohort Survey (BCS) which followed all individuals born in a single week in 1970, to examine the role played by teenage conduct problems on cohort members' subsequent time in employment and in employment, education and training. These conduct problems - captured with a validated Rutter Scale (Rutter et al., 1970) and based on mothers' assessments of their child's behaviour – are a potentially important component of children's noncognitive skills. They find teenage conduct problems have a substantial impact on months spent in employment and months in employment, education or training which persists across the life cycle. Raw gaps between those with and without conduct problems remain sizeable when conditioning on circumstances in childhood and rise with the number of conduct problems. Although girls had fewer conduct problems than boys, teenage conduct problems had the greatest negative impact on women's labour market prospects.

Turner et al. (2022) capture childhood social and emotional skills with another scale, the Bristol Social Adjustment Guide (BSAG) scale (Stott, 1958; Shepherd, 2013; Dennison and Peters, 2022), taking the average of NCDS cohort members' scores at ages 7 and 11. They show this score is predictive of cohort members' accumulated lifetime earnings through to age 63. Gender enters their analysis as a control variable.

2.2.3: Family Background

Heckman et al. (2006) warn that noncognitive abilities "are likely to be affected by family background characteristics" (p. 418) which may also directly impact a child's subsequent labour market prospects through networks and other factors linked to social class. It is therefore important to account for family background in addition to cognitive and noncognitive skills.

Becker and Tomes (1986) present a model in which utility maximizing parents concerned about the welfare of their children invest in their children's futures, transferring advantages subject to the inheritability of endowments and parents' capacity or willingness to invest in their children. Their review of empirical studies indicates a reasonable positive correlation between the earnings of fathers and sons, though there is regression to the mean across generations. They do not examine empirical studies for daughters.

Although there is evidence that rising income differences within social class have become more important over time in explaining intergenerational persistence in earnings, at least among men (Blanden et al., 2007), parental social class continues to play a significant role in children's labour market prospects in adulthood (Goldthorpe and Jackson, 2007). It can do so for a variety of direct and indirect reasons, as illustrated in de Neubourg et al.'s (2018) dynamic lifecycle model which seeks to explain the reproduction of wealth and health over generations.

Family circumstances can also negatively impact children's labour market prospects by exposing them to a range of problems in their early life. In their paper, which uses the same 1958 National Child Development Study (NCDS) used in this paper, Millemaci and Sciulli (2014) use matching estimators to identify the impact of family difficulties in childhood (measured at age 7) on employment and wage outcomes in adulthood (ages 33, 42 and 50).[1] Conditioning solely on childhood circumstances they find experiencing at least one family difficulty in childhood decreases cohort members' employment probabilities and wages in adulthood. This labour market disadvantage increases with the number of problems experienced and the negative effect persists over the life course. The authors do not conduct any analysis by gender. Theirs is also a complete case analysis which does not adjust for panel attrition. Nor do the wage estimates account for selection into employment.

2.2.4: Health

It is plausible that childhood illness, by limiting one's ability to invest directly in human capital, and by limiting a person's mental or physical capacity for work, will impact labour market prospects in adulthood. In these ways, one can think of health as akin to human capital (Becker, 2007). Case et al. (2005) confirm the importance of childhood illness for labour market prospects in their analyses of the NCDS. They show that childhood ill-health continued to impact labour market outcomes in adulthood (employment status and socio-economic status) through to age 42.

2.2.5: Occupational Expectations

Gender differences in occupations account for a sizeable part of the GWG (Blau and Kahn, 2017). These differences can arise for a number of reasons including gendered occupational preferences, barriers to occupational entry such as those linked to hiring discrimination, and gendered social norms which mean women often face considerable difficulties combining high-earning jobs with family responsibilities (Dex, 1987; Goldin and Katz 2016). However, as Adda et al. (2017) point out, occupational choice is endogenous.

The endogeneity issue can be partially addressed by conditioning on the occupational expectations and aspirations expressed in childhood. Examining the occupational expectations of 16-year-olds in the birth cohort used in this paper Brown et al. (2011) show those expectations differed markedly by gender with boys much more likely to expect entry into professional occupations whereas girls were more likely to expect entry to a clerical occupation, indicating that these expectations are likely influenced by gender identity norms prevalent at the time (Akerlof and Kranton, 2000; Bertrand, 2020). Brown et al. find boys were

[1] Difficulties include housing problems, family health problems, economic problems, death of a parent, parental divorce or separation, parental alcoholism.

also more likely to realise their expectations in their first job on leaving school. Below we condition on occupational expectations at age 11.

The conditioning variables used in this paper include, but are not confined to, items discussed above. Section 3.1 describes them in detail.

2.3: Matching estimators

Frölich (2007) first described how matching – specifically propensity score matching – could be used beyond the treatment evaluation literature as an extension of the Oaxaca-Blinder decomposition of earnings differences into their observed and unobserved components. He illustrates the approach by examining the GWG among college graduates in the United Kingdom and points to the importance of university subject choice in helping to explain large GWGs at the top of the earnings distribution.

In recovering the impact of gender on earnings, regression and matching estimators both assume relevant differences between women and men are captured by data available to the analyst (known as the conditional independence assumption in the evaluation literature). The assumption, which is not testable, is violated if the analysis does not incorporate all factors affecting the 'treatment' – in our case being a woman - and the outcome of interest, namely hourly earnings.

Notwithstanding this shared identification assumption, the potential advantages of matching estimators over regression techniques are well-known. Heckman et al. (1998) identify three advantages of propensity score matching (PSM) relative to linear estimators. First, linear estimation (and decompositions based on them) rely on assumptions regarding linearity in the outcome equation, whereas PSM is semi-parametric and so does not require the functional form assumption. Second, there is no assumption of constant additive effects so individual causal effects are completely unrestricted such that heterogeneous treatment effects can be captured. Third, matching highlights the problem of common support. In our setting, this means women must have 'like' counterparts in male population. Matching avoids extrapolating beyond common support in the propensity to be a woman.

Although matching estimators have become increasingly common in the literature since Heckman et al. (1998) they are used relatively infrequently in recovering the GWG. There are a small number of exceptions. Black et al. (2008) were among the early proponents of matching in estimating the GWG. Examining GWGs by race among highly educated workers in the United States they undertake nonparametric matching by simply 'hard matching' on key demographic traits. They conclude that "inferences from familiar regression-based decompositions can be quite misleading" (p. 3).

In motivating his use of matching to decompose the GWG Nopo (2008: 298) argues: "relationships that govern the co-movement of wages and individual characteristics are not necessarily linear". He argues that a lack of common support across women and men can result in an overestimation of the unexplained component of the GWG and attribution of some of the GWG to the wage returns of men who have no reasonable comparators in the female wage distribution. To examine this issue Nopo's decomposition does not discard individuals beyond common support. Instead, they are retained and appear as additional components in the decomposition. In practice, when he compares estimators on his data for workers in urban Peru, he finds no differences in the GWG estimated via matching techniques or standard regression estimates with a Mincerian specification.

Perhaps the paper closest in spirit to the current one is Strittmatter and Wunsch (2025). In it the authors show that the choice of estimator and model flexibility when controlling for wage determinants both matter greatly when estimating the GWG.[2] Using administrative data for 1.7 million employees in Switzerland they point to a substantial common support problem. They show that semi-parametric estimators including variants of propensity score matching and inverse probability weighting, together with the enforcement of common support, substantially reduce the size of the unexplained GWG compared with Oaxaca-Blinder estimates. However, the primary reason for the lack of common support in their paper is occupational sorting leading to job segregation something which, as discussed above, is arguably a function of gender itself and, as such endogenous.[3]

The paper by Goraus et al. (2017) examining the GWG in Poland with cross-sectional Labour Force Survey data for 2012 is also informative from our perspective because it examines the sensitivity of the GWG to decomposition techniques and to the set of conditioning variables. In contrast to the meta-analysis by Laroche et al. (2026) they argue that results are sensitive to estimation technique. They also find estimates vary a great deal with the set of conditioning variables. What is unusual about their results is that, regardless of choice of estimator or conditioning variables, the adjusted GWG is always larger than the raw GWG. However, as they recognise (p. 126) they are conditioning on some variables such as occupation and industry that are plausibly endogenous.

Finally, Meara et al. (2020) use a variety of matching estimators to examine heterogeneity in the GWG in the United Status using the Current Population Survey. In doing so, they explicitly split their sample on potentially endogenous variables to help uncover the heterogeneity.

Throughout this small literature using matching estimators to recover the GWG being female is the 'treatment' variable and the matched males constitute the counterfactual wage for those females. The parameter recovered is average treatment-on-the-treated (ATT) since matches for the treated group (women) are found among the controls (men). If the impacts of being a woman on earnings are heterogeneous, the ATT may differ from the average treatment-on-the-non-treated (ATNT). This can be recovered with matching estimators by simply designating men as the treated group and seeking their comparators among women as the controls. The weighted average of the two will deliver the average treatment effect (ATE) for a random individual. However, the studies confine themselves to ATT. We follow the same approach below.

[2] Bonaccolto-Topfer and Briel's (2022) paper using data from the German Socio-Economic Panel (GSOEP) is also notable in underlining the importance of the way in which conditioning variables are chosen (their paper uses machine learning instead of a theory-driven approach) and underscores the value of incorporating conditioning variables in a flexible fashion.

[3] Djurdjevic and Radyakin (2007) also examined the gender wage gap in Switzerland (for the period 1996-2003). They deployed an exact matching approach similar to Nopo's (2008) and, like Nopo, decompose the GWG into four components as opposed to the two in the standard Oaxaca-Blinder decomposition ("endowment" and "remuneration" effects) by distinguishing between components for those on and off common support. This is a nice feature because it sheds light on that part of the overall GWG which is attributable to the characteristics of men that can not be found in the female population. They also use propensity score matching using a calliper. They find the common support problem to be substantial, but they condition on a range of potentially endogenous variables including family formation and occupation, a point they acknowledge (p. 381).

2.4: Selection into employment and panel attrition

A further issue often overlooked in the GWG literature is the potential for non-random selection into employment and, in panel data, non-random panel attrition. The issue is largely ignored in the matching literature reviewed in Section 2.3. The exception is the paper by Goraus et al. (2017) which undertakes a sample selection adjustment using instrumental variables.[4] Others recognise selection as a limitation (eg. Djurdjevic and Radyakin, 2007: 380) whilst others simply ignore it.[5]

Previous work with the data used in this study indicates that the raw (unadjusted) GWG in median hourly earnings is understated through to cohort members' late 30s if one does not account for selection into employment, indicating that women are more positively selected into employment than men over that period in their lives. Selection into employment has no effect on the estimated differential after age 42 (Bryson et al., 2020). By contrast, adjusting for panel attrition results in a small reduction in the GWG from age 33 onwards.

As Bryson et al. (2020) note in their online appendix, there are a multitude of ways to adjust for non-random selection into employment. They decide to impute earnings to non-participants using matching estimators which are survey sweep and gender-specific. They reweight their data with match weights, combined with weights for sample attrition. Below I describe a similar approach using matching estimators to impute log mean hourly wages by sweep and gender but, unlike Bryson et al (2020), I condition only on pre-labour market variables in the matching and imputation processes.

In their cross-country analysis of the GWG, Olivetti and Petrongolo (2008) emphasize the importance of correction for sample selection due to substantial cross-country differences in the proportion of women participating in the labour market. Using data for the period 1994-2001 they adjust for selection by imputing median earnings to non-participants through out-of-sample prediction of earnings.[6] They find imputed earnings tend to be lower than actual earnings, consistent with positive selection into employment among women. However, the gap between actual and imputed earnings is small in the United Kingdom and the United States but is larger in Southern Europe. I adopt a similar strategy below, though with mean earnings.

3. Data and Estimation

3.1: Data

Our data are the National Child Development Study (NCDS), a nationally representative birth cohort of those born in a single week in 1958 (Pearson, 2016). Data are collected prospectively, initially from parents and teachers, then later from cohort members themselves, from birth and then at interview sweeps when the cohort members are aged 7, 11, 16, 23, 32, 42, 50, 55, 61

---

[4] They rely on a set of exclusion restrictions based on family formation and non-earned income.

[5] Frölich (2007: 363) explains how PSM can be used to reweight data in the presence of non-random attrition and non-response and that, in the case of small samples, it has advantages over the more common approach of using inverse probability weighting.

[6] They also use alternative imputation methods such as interpolation of missing data using panel observations on earnings observed in other survey waves.

and 63.[7] The data are available for download from the UK Data Archive (https://datacatalogue.ukdataservice.ac.uk/series/series/2000032#abstract).[8]

We use data through to age 16 for co-variate adjustment and matching. For this birth cohort age 16 was the end of compulsory education, so all these variables are captured prospectively prior to labour market entry.

The data on childhood circumstances available in NCDS is extensive. The variables we condition on are presented, together with summary statistics, in Appendix Table 1.

Our model specifications are informed by previous studies and the literature reviewed in Section 2.2. We examine the sensitivity of GWGs to conditioning on two sets of controls in a full model specification and a more parsimonious model. The full model incorporates the following[9]:

- Circumstances at birth: birthweight in ounces; sibling birth order; whether the mother smoked prior to pregnancy; father's social class[10]; the country of birth; race.

[7] Data were collected by postal questionnaire at age 46 but response rates were poor so we have excluded this sweep from the paper. Data at ages 61 and 63 were collected in three survey sweeps around the time of COVID (May 2020, September/October 2020 and February/March 2021 (Brown et al., 2024). For details of the survey questionnaire see for more details about the content of the survey, including the questionnaire see https://cls.ucl.ac.uk/data-access-training/covid-19-survey.

[8] Further details are available here: https://cls.ucl.ac.uk/cls-studies/1958-national-child-development-study/

[9] I also experimented with other conditioning variables which do not appear in the final estimates. These include number of days gestation; mum's social class; cohort member was breast-fed; region at birth; housing tenure at age 7; household size at age 7; number of hospital admissions through to age 7; handedness age 7; verbal and non-verbal test scores aged 11; female teacher at age 11; cohort member's intentions at age 11 on leaving school; teacher rating of child at ages 11 and 16; mum's and dad's interest in education of child when child aged 11; financial hardship at ages 7 and 11; receipt of free school meals at ages 7 and 11; type of school attended age 16; mathematics capability age 16; English capability age 16; science capability age 16; foreign language capability age 16; practical subjects capability age 16; social sciences capability age 16.

[10] This is recorded at the child's birth. It codes the father's occupation according to the Register General's social classification

- Cognitive ability captured by the Southgate Group Reading Score at age 7, a score between 0 and 30[11]; the Problem Arithmetic Test at age 7[12]; standardized reading comprehension test score at age 11[13]; standardized mathematics score at age 11[14]

- Non-cognitive skills: Following Parsons et al. (2024) we capture conduct problems using the Rutter Scale (Rutter et al., 1970) based on mothers' assessments of their child. Whereas Parsons et al (2024) use the measures at age 16 we use mothers' assessments when their child was aged 7. The scale sums up to 23 conduct problems.[15] Following Turner et al. (2022) we also use the Bristol Social Adjustment Guide (BSAG) scored by teachers to capture behavioural problems. Unlike Turner et al. who take the average of the scores at ages 7 and 11 we take the overall BSAG score when the child was aged 7.[16]

- Health: we count the number of illnesses and health problems the mother says the child had until age 7[17]; parent's report on whether child is disabled age 16; mother's assessment as to whether the child is over- or under-weight at age 16.

[11] On 16 (of 30) occasions, the child was given a picture of an object and had to ring the word, from 5 different options describing that object in the picture. On the other 14 occasions, the teacher read out a word and the child had to circle the correct one. For a full description see: https://closer.ac.uk/our-resources/cross-study-data-guides/cognitive-measures/information-closer-studies/ncds/age7/7-southgate-group-reading-test/

[12] Developed by Pringle et al. (1966), this standardized test evaluates early numeracy, logical reasoning, and basic mathematical capability. Teachers explained ten arithmetic problems graded in level of difficulty. In order to avoid penalizing the poor readers, the teachers were asked to read the problems to the children if necessary. The test was discontinued after three successive incorrect answers. One mark was awarded for each correct answer, giving a score between 0 and 10. https://closer.ac.uk/our-resources/cross-study-data-guides/cognitive-measures/information-closer-studies/ncds/age7/problem-arithmetic-test/?_gl=1*1ezlvb4*_up*MQ..*_ga*MTA3NjA3NjAwNy4xNzg4MDE1ODg0*_ga_FRV5V1J9CD*czE3ODgwMTU4ODQkbzEkZzAkdDE3ODgwMTU4ODQkajYwJGwwJGgw

[13] This reading test, administered by the teacher, was constructed by the National Foundation for Educational Research (NFER) for NCDS. The raw score runs from 0 to 35 but is standardized into standard deviation units (z-scores) relative to the sample mean. https://closer.ac.uk/our-resources/cross-study-data-guides/cognitive-measures/information-closer-studies/ncds/age11/reading-comprehension/?_gl=1*i3sg3f*_up*MQ..*_ga*MzUyMjA5MjQ4LjE3ODgwMTgwNDI.*_ga_FRV5V1J9CD*czE3ODgwMTgwNDIkbzEkZzAkdDE3ODgwMTgwNDIkajYwJGwwJGgw

[14] The test, administered by the teacher, consisted of 40 items and included number skills, fractions, measures and geometry. Most questions were calculated directly, with a few involving multiple-choice answers. The raw score ranges between 0 and 40 but is standardized into z-scores https://closer.ac.uk/our-resources/cross-study-data-guides/cognitive-measures/information-closer-studies/ncds/age11/mathematics-test/?_gl=1*1qzu71f*_up*MQ..*_ga*MzUyMjA5MjQ4LjE3ODgwMTgwNDI.*_ga_FRV5V1J9CD*czE3ODgwMTgwNDIkbzEkZzEkdDE3ODgwMTgwNTAkajUyJGwwJGgw

[15] The scale sums the conduct problems a mother identifies in her 7-year-old child. There are 23 items, so the maximum potential score is 23. In fact, the maximum score in the sample is 21 with a mean of 5.7. The items are: headaches; temper tantrums; reluctant to go to school; bad dreams, night terrors; difficulty getting to sleep; sleepwalks; food fads and dislikes; poor appetite; over eats; difficulty concentrating; prefers doing things alone; bullied by other kids; generally destructive; miserable or tearful; squirmy or fidgety; continually worried; irritable ; sucks thumb or fingers; upset by new situations; mannerisms or twitches; fights other children; bites nails; disobedient.

[16] This is an additive scale running from zero to 64 based on 12 syndromes at age 7 scored by a teacher: unforthcomingness; withdrawal; depression; anxiety for acceptance by adults; hostility towards adults; 'writing off' of adults and adult standards; anxiety for acceptance by children; hostility towards children; restlessness; inconsequential behaviour; miscellaneous symptoms and miscellaneous nervous symptoms. https://cls.ucl.ac.uk/wp-content/uploads/2017/07/NCDS-BSAG.pdf?_gl=1*siiaq9*_up*MQ..*_ga*MTQ1NDc0MDg5MS4xNzg3OTMzODA1*_ga_EYRQV4V0KV*czE3ODc5MzM4MDUkbzEkZzAkdDE3ODc5MzM4MDUkajYwJGwwJGgw

[17] This is a count measure from a battery of questions identifying up to 46 health problems and illnesses reported by the mother up to age 7.

- Occupational expectations: children aged 11 were told "Imagine now that you are 25 years old. Write about the life you are leading, your interests, your home life, and your work at the age of 25…." Their imagined occupation was then coded.[18] As noted in Section 2.2.5 earlier work found these expectations differed markedly by sex. I recoded the 'housewife' response to 'not classified' because, of the 500 coding this, only 1 was a boy.[19] The distributions for girls and boys are presented in Table 1.

- Behaviours in the last year of compulsory education: self-reported whether smokes cigarettes; self-reported how long since last drank alcohol; teacher report of child getting into trouble with the police.

The dependent variable is log hourly wages derived from gross hourly earnings divided through by hours worked. These were collected from those in paid work at each survey wave. I deflated them by the Consumer Price Index (CPI) to their real value in January 2000.

There are dangers in over-specifying or over-fitting a propensity score model which can include increasing covariate imbalance and model-dependence of matched pairs (King and Nielsen, 2019). We therefore test the sensitivity of our results to a more parsimonious set of covariates. The parsimonious model only contains father's social class; the reading and mathematics scores; and the Rutter and BSAG scores. (There are starred with an asterisk * in Appendix Table A1).

We deal with item non-response by incorporating missing categories for categorical variables and imputing values at the wave-sex mean for categorical variables, incorporating imputation dummies in our estimation to identify those cases when we do so. This can cause a multi-collinearity problem (as indicated by a high Variance Inflation Factor (VIF) value) so in our final estimates we replace these single missing dummies with an overall count of the number of times a respondent is missing on relevant control variables.

3.2: Estimation

The full NCDS sample consists of 18,558 cohort members. Our analysis focuses on the 15,540 individuals who were issued for interview at age 23 – the first survey interview on leaving compulsory education – thus excluding 960 who had died by that point, 1,196 who had emigrated and 862 who were not issued for interview for other reasons. At each survey sweep we confine our analyses to those who were issued for interview, thus excluding those who had died or emigrated, or were not issued for interview for other reasons.[20]

---

[18] For more information on the way they were coded see Morrow and Elliott (2020) and Elliott and Morrow (2007).

[19] Results might have looked very different if I'd retained the original coding either in the estimation of a propensity score, or in 'hard matching' girls and boys, since this might have led to a substantial common support issue. This is something I will return to.

[20] The survey data include outcome codes at each survey sweep which identify those who were issued for interview and those who were not, and why. Our analyses are based on those who were issued for interview. These include 'productive' cases who actually provided an interview; refusals; non-contacts; and 'other unproductives'.

Throughout we recover the GWG in log hourly mean earnings for those in paid employment with valid hourly wages at the time of the survey at ages 23, 33, 42, 50, 55, 61 and 63.[21]

We recover the raw unconditional GWG in log points by subtracting the female log hourly mean wage from the male log hourly mean wage and, as an alternative, by introducing a female dummy variable into a log hourly wage equation with no controls. We present GWGs on unrestricted wage distributions but in our preferred estimates we windsorize log hourly mean earnings at the 99th and 1st percentiles of the wage distribution.

We recover the adjusted GWG by introducing our controls measured prior to the end of compulsory education at age 16. Our preferred models are those conditioning on a full set childhood of controls but we also test the sensitivity of our results to a more parsimonious set of controls, as described in Section 3.1 and Appendix Table A1. We establish the impact of these controls on the female dummy in pooled sex equations but also recover the residual difference in log hourly earnings from separate sex earnings equations. We account for missingness in the way described in Section 3.1.

The birth cohort numbers deplete over time due to death, emigration and for other reasons recorded by the survey agency. In addition, cohort members who do respond to a given survey do not necessarily provide hourly earnings, either because they are not in employment or, if they are, they have not provided valid earnings or hours data. We tackle the joint issues of panel attrition and non-random selection into employment by imputing log hourly mean earnings for the population who are issued for the survey in a given sweep. We do so with sex and survey sweep specific wage equations that condition on the same set of pre-labour market variables used in our GWG regressions and use the out-of-sample predicted wages under those models to impute hourly earnings where they are missing.[22]

Throughout we run two sets of analyses. The first recovers the GWG among those in paid work at the time of the survey who reported valid earnings and hours. At age 23, the sample consists of 8,011 individuals but fell to 2,117 individuals by age 63 (Table 2). The second set of analyses incorporates those with missing earnings data arising either from panel attrition, being out of paid work at the time of the survey, or failure to report accurate hours and earnings data among those who are in paid employment. At age 23, these models are run on 15,540 cohort members – almost twice the number observed with valid hourly earnings, falling to 11,448 cohort members issued for interview at age 63 (Table 3). In the models which include those with imputed earnings we incorporate a dummy variable which equals 1 when the cohort member is among those with imputed earnings.

The tables and graphs in Section 4 present six estimates of the GWG for each of the seven survey sweeps providing insights into the size of the GWG across the life course for those born in 1958. The first estimate is the raw GWG expressed as the difference in log points between men's and women's log hourly mean earnings.

[21] Earnings at age 61 were collected retrospectively at age 63. Respondents were asked what they were earning just prior to the onset of COVID.

[22] When we run log hourly earnings models at ages 23, 33, 42, 50, 55, 61 and 63 these childhood variables account for a sizeable proportion of the variance in earnings in adulthood – the adjusted r-squared ranging between 0.12 and 0.15 for men and 0.05 and 0.14 for women (apart from in the case of men aged 23 when the adjusted r-squared is 0.03). The full models are available on request.

The second estimate is based on the difference in log hourly earnings between women and their male matched comparators. I use propensity score matching (PSM) to match women to men on a single index (the propensity score) derived from a probit model for being a woman. The variables used in the model are the pre-labour market characteristics also used in the regression analyses for earnings. We recover matches using PSM with replacement. The comparator earnings for each woman are the male earnings taken from the women's five nearest neighbours within a 0.005 calliper enforcing common support.[23] In doing so, among those in paid employment, we lose 27 women who are off common support at age 23 (see Appendix Chart 1), but enforcement of common support leads to the loss of no women at later ages. (When we incorporate those with imputed earnings we lose 23 observations at age 23 and, again, nobody at later ages). This procedure recovers the average treatment on the treated (ATT). We examine the quality of the matching with balancing tests pre- and post-PSM. The matching performs well according to the standard tests (see Appendix Table 2 and accompanying text). This is also the case when we run estimates based on the parsimonious set of control variables, and when we incorporate those with imputed earnings.

The third estimate is the log hourly pay differential between women and men captured with a female dummy variable in an unweighted OLS regression with controls, run on the common support. The fourth estimate is similar but weights the OLS with the weight derived from the PSM.[24] The fifth estimate is also based on an OLS regression with controls, but this time the regression is weighted with entropy weights to obtain balance between women and men on covariates (Hainmueller, 2012).[25]

The sixth estimate uses Inverse Probability Weighted Regression Adjustment (IPWRA). It is a three-step procedure. Step one (IPW) estimates the probability of exposure to treatment (being a woman). The second step (RA) weights models for the outcome (log hourly wages) separately for treated and non-treated groups (women and men). The weights in the treatment group are based on the inverse of the probability of being treated conditional on covariates. Weights in the control group are based on the inverse of the probability of not being treated. To recover the ATT in the third step predicted means are calculated based on treated cases only with counterfactuals based on parameters from the control group estimation. A particularly attractive feature of IPWRA is that it is doubly robust in the sense that it is sufficient that either the model for the conditional mean outcome or the model for the propensity for treatment are correctly specified. Correct specification is not required for both (Wooldridge, 2010: 930-931).

4. Results

We begin by presenting the raw GWG over the life-course for the NCDS birth cohort. We observe their earnings at seven points between the ages of 23 (in 1981) and forty years later (2021) when they were aged 63. Chart 1 and Table 2 map the log hourly mean earnings for women and men in employment at the time of the survey. Women's and men's earnings follow a hump-shape over the life-course. At every age men have higher hourly wages than women,

[23] The use of 5 nearest neighbours balances the tradeoff between efficiency and bias. For discussion regarding the performance of PSM variants see Bryson et al. (2002).
[24] For nearest neighbor matching, the weight variable holds the frequency with which the observation is used as a match. The weight for treated cases (women) is equal to 1.
[25] We implement entropy balancing with the Stata command *ebalance* (Hainmueller and Xu, 2013). This is a data pre-processing method that reweights observations in the control group, such that the mean of the variables is equal to those in the treatment group. In this sense it is similar to PSM. However, there is no need to enforce common support.

but the gap is smallest at age 23 (17 log points), rising to 45 log points at age 42 before falling later in life. However, at age 63 the raw GWG remains almost the size it was at age 33. This pattern has been remarked upon previously by researchers who followed cohort members through to age 55 (Bryson et al, 2020), but we are adding two data points at ages 61 and 63.

Chart 2 and Table 3 perform the same exercise but this time the analysis includes imputed earnings for those who did not respond to a given sweep, responded but were not employed at the time of the survey, or were employed but did not provide valid responses on either earnings or hours to permit the derivation of hourly earnings. This leads to an estimation sample at age 23 that is almost twice as large as the sample providing hourly earnings. By age 63, the estimation sample with imputed earnings is nearly five-and-a-half times larger than the sample with hourly earnings. The broad pattern in earnings for women and for men over the life-cycle is hump-shaped and similar to that observed for those with valid hourly earnings data.

Comparing the log hourly earnings with and without imputations in Table 3 and Table 2 it is apparent that the means are marginally lower for both women and men when we incorporate imputed earnings suggesting there is positive selection into employment for both women and men. This is the case across the whole life-cycle. However, the differences are relatively small. The raw GWG is smaller when incorporating imputed earnings – apart from at age 23 when the raw GWG is larger when including imputed earnings (19.6 log points compared to 16.8 log points for those in paid work with valid earnings). The implication is that positive selection into employment is stronger among men than it is among women, but the differences are marginal until the cohort reach their 60s at which point positive selection into employment is more pronounced among men.

Chart 3 and Appendix Table 3 present baseline estimates of the log hourly GWG at the mean by age for those in paid work. I replicate the raw GWG in log points (the gray line and row 1 in Appendix Table 3) and then present the five different estimates using the estimation techniques described in Section 3.2. There are two striking findings. First, the GWG follows a similar pattern across the life-course irrespective of the estimation technique used. It is smallest early on in one's career, peaks when workers reach their early 40s, shrinks during their 40s before flattening out between age 50 and 63. Second, the adjusted GWGs are very similar to the raw GWG across the life course, irrespective of the estimation technique deployed. The range of estimates in any given survey year is not large. For example, when aged 23 the GWG ranges between 16.8 and 18.2 log points; at age 42 it ranges between 42.2 and 44.6 log points; and at age 63 it ranges between 31.5 and 35.4 log points. The implication is that although childhood co-variates are predictive of earnings in adulthood (see footnote 21), they account for very little of raw GWG.

A similar picture emerges when I use a more parsimonious set of childhood controls, as described in Section 3.2, to match women and men and to adjust for the GWG (Chart 4 and Appendix Table 4). The implication is that these results are not particularly sensitive to the set of variables used to construct counterfactual wages.

Chart 5 and Appendix Table 5 replicate the analyses in Chart 3 and Appendix Table 3 but incorporate those with imputed earnings to account for non-random selection into employment and panel attrition. Once again, the same life-cycle pattern in the GWG emerges – peaking at age 42, shrinking through one's 40s before flattening out after age 50. Estimates from alternative estimators are similar to one another, although in this instance it is apparent that the IPWRA estimates return a lower GWG than other methods except at age 23. Rerunning the analyses with the more parsimonious set of conditioning variables produces a similar set of results, although the IPWRA estimates come closer to those using other estimators (Appendix Chart 2 and Appendix Table 6).

5. Conclusion

This paper examines raw and adjusted gender wage gaps (GWGs) across the life-course for a birth cohort born in 1958. Following them through to age 63 we show a sizeable GWG early on it their careers rises until they reach their 40s, then falls but remains sizeable to age 63. Concerned about the potential endogeneity of conditioning on, or matching on covariates post-labour market entry, we confine our attention to gender differences in childhood. We know from the literature, and confirm in this study, that childhood traits and experiences account for a sizeable proportion of the variance in earnings in adulthood. We also know that, using data for those born in the early 1960s, Neal and Johnson (1996) were able to explain the whole black-white wage gap in the United States by controlling for a test score administered to teenagers. However, in contrast to findings in the literature in which the regression-adjusted GWG is considerably smaller than the raw GWG, differences in log hourly mean earnings between men and women having conditioned on pre-labour market variables are similar to the raw GWG. This is the case at all ages, and whether we use matching or linear estimation techniques. Patterns are similar when accounting for non-random selection into employment and panel attrition. Results are not particularly sensitive to the covariates used in the analysis.

The implication is that among those born in 1958 women have lower earnings than men across the life-course which cannot be accounted for by their pre-labour market experiences. Instead GWGs are related to differential treatment by sex in childhood and adulthood (eg. as in the literature on "premarket discrimination"; (unobserved) differences between boys and girls in childhood; and experiences in adulthood.

References

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

Akerlof, G.A. and Kranton, R. E. (2000) Economics and Identity, *The Quarterly Journal of Economics*, 115, 3: 715–753

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

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

Becker, G. S. (2007) Health as human capital: synthesis and extensions, *Oxford Economic Papers*, 59: 379-410

Becker, G. S. and Tomes, N. (1986) Human Capital and the Rise and Fall of Families. *Journal of Labor Economics*, 4, 3: Part 2 S1-S39

Bertrand, M. (2020) Gender in the Twenty-First Century. *American Economic Association Papers and Proceedings*, 110: 1-24

Black, D. A., Haviland, A. M., Sanders, S. G., and Taylor, L. J. (2008) Gender Wage Disparities among the Highly Educated, *Journal of Human Resources*, XLIII, 3: 630-659

Blanden, J., Gregg, P. and Macmillan, L. (2007) Accounting for Intergenerational Income Persistence: Noncognitive Skills, Ability and Education. *The Economic Journal*, 117, 519: C43-C60

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

Blau, F. D., Kahn, L. M., Boboshko, N., and Comey, M. (2024) The Impact of Selection into the Labor Force on the Gender Wage Gap, *Journal of Labor Economics*, 42, 4: 1093-1133

Blinder, A. S. (1973) "Wage Discrimination: Reduced Form and Structural Estimates", *The Journal of Human Resources*, 8, 4: 436-455

Bonaccolto-Topfer, M. and Briel, S. (2022) The gender pay gap revisited: Does machine learning offer new insights?, *Labour Economics*, 78: 102223

Brown, S., Ortiz-Nunez, A., Taylor, K. (2011) What will I be when I grow up? An analysis of childhood expectations and career outcomes, *Economics of Education Review*, 30: 492-506

Brown, M., Fitzsimons, E., Goodman, A., Peters, A., Ploubidis, G., Sanchez, A., Silverwood, R. and Smith, K. (2024) *COVID-19 survey in five national longitudinal studies: Waves 1, 2* *and 3: User guide (version 4)* available at https://cls.ucl.ac.uk/wp-content/uploads/2017/02/UCL-Cohorts-COVID-19-Survey-user-guide.pdf

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

Bryson, A., Dorsett, R. and Purdon, S. (2002) *The use of propensity score matching in the evaluation of active labour market policies*, Department for Work and Pensions Working Paper No.4

Bryson, A., Joshi, H., Wielgoszewska, B., and 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

Case, A., Fertig, A., and Paxson, C. (2005) The lasting impact of childhood health and circumstance. *Journal of Health Economics*, 24, 365–389

Dennison, K. and Peters, A. (2022) *National Child Development Study – Bristol Social Adjustment Guide Forms at ages 7 and 11*. London: UCL Centre for Longitudinal Studies https://cls.ucl.ac.uk/wp-content/uploads/2022/03/NCDS-Bristol-Social-Adjustment-User-Guide.pdf?_gl=1*15fvz4d*_up*MQ..*_ga*MTcxMTQ1ODAyNS4xNzg3OTM2NTA3*_ga_EYRQV4V0KV*czE3ODc5MzY1MDYkbzEkZzAkdDE3ODc5MzY1MDYkajYwJGwwJGgw

Dex, S. (1987) *Women's Occupational Mobility: A Lifetime Perspective*, Palgrave Macmillan

Djurdjevic, D. and Radyakin, S. (2007) Decomposition of the Gender Wage Gap Using Matching: An Application for Switzerland, *Swiss Journal of Economics and Statistics*, 143, 4: 365-396

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

Elliott, J. and Morrow,V. (2007) Imagining the Future: Preliminary analysis of NCDS essays written by children at age 11, *CLS Cohort Studies Working Paper 2007/1*, Centre for Longitudinal Studies: London

Foliano, F., Bryson, A., Joshi, H., Wielgoszewska, B., and Wilkinson, D. (2024). Gender wage gap among young adults: a comparison across British cohorts. *Labour Economics*, 91, 102614

Frölich, M. (2007) Propensity score matching without conditional independence assumption – with an application to the gender wage gap in the United Kingdom, *Econometrics Journal*, 10: 359-407

Goldin, C. (2014) A grand gender convergence: Its last chapter. *American Economic Review*, 104, 4: 1091-1119

Goldin, C. and Katz, L. F. (2016) A Most Egalitarian Profession: Pharmacy and the Evolution of a Family-Friendly Occupation. *Journal of Labor Economics*, 34, 3: 705–746

Goldthorpe, J. H. and M. Jackson (2007). Intergenerational Class Mobility in Contemporary Britain: Political Concerns and Empirical Findings. *British Journal of Sociology*, 58, 525-546

Goraus, K., Tyrowicz, J. and Van der Velde, L. (2017) Which gender wage gap estimates to trust? A comparative analysis, *The Review of Income and Wealth*, 63, 1: 118-146

Hainmueller, J. (2012) Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies, *Political Analysis*, 20: 20-46

Hainmueller, J. and Xu, J. (2013) Ebalance: a Stata package for entropy balancing. *Journal of Statistical Software*, 54(7), 1–18

Harkness, S. (1996). The gender earnings gap: evidence from the UK. *Fiscal Studies*, *17*(2), 1-36

Heckman, J., Ichimura, H., Smith, J. and Todd, P. (1998) Characterizing selection bias using experimental data. *Econometrica,* 66, 5: 1017-1098

Heckman, J. J., Stixrud, J. and Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior, *Journal of Labor Economics*, 24, 3: 411-482

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

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

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: 397-415

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

King, G. and Nielsen, R. (2019) Why Propensity Scores Should Not be Used for Matching, *Political Analysis*, 27, 4: 435-454

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

Laroche, P. Bryson, A., Joshi, H. and Wilkinson, D. (2026) ”The gender wage gap in Britain: A Meta-analysis”, *IZA Discussion Paper No. 18484*

Makepeace, G., Dolton, P., and 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., and 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., and 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., and Swaffield, J. (2008). The gender gap in early-career wage growth. *The Economic Journal*, *118*(530), 983-1024

Meara, K., Pastore, F. and Webster, A. (2020) The gender pay gap in the USA: a matching study, *Journal of Population Economics*, 33: 271-305

Millemaci, E. and Sciulli, D. (2014). The long-term impact of family difficulties during childhood on labor market outcomes. *Review of Economics of the Household*, 12: 663-687

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

Morrow, V. and Elliott, J. (2020) ‘I'd keep them tidy’: domesticity, work and nostalgia in girls' imagined futures Described in Essays Written by 11 Year Olds in 1969, Chapter 12 in L. Moran et al. (eds.), *Narrating Childhood with Children and Young People*, Studies in Childhood and Youth, pp. 283-306

Neal, D. A. and Johnson, W. R. (1996) “The Role of Premarket Factors in Black-White Wage Differences”, *Journal of Political Economy*, 104, 5: 869-895

De Neubourg, E., Borghans, L., Coppens, K. and Jansen, M. (2018) Explaining Children's Life Outcomes: Parental Socioeconomic Status, Intelligence and Neurocognitive Factors in a Dynamic Life Cycle Model. *Child Indicators Research*, 11: 1495-1513

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

Neuberger, J., Kuh, D. and Joshi, H. (2011) Cross-cohort changes in gender pay differentials in Britain: accounting for selection into employment using wage imputation, 1972-2004, *Longitudinal and Life Course Studies*, 2, 3: 260-285

Nopo, H. (2008) Matching as a Tool to Decompose Wage Gaps, *The Review of Economics and Statistics*, 90, 2: 290-299

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

Office for National Statistics (2025) *Gender Pay Gap in the UK: 2025*. https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/bulletins/genderpaygapintheuk/2025

Olivetti, C. and Petrongolo, B. (2008) Unqual Pay or Unequal Employment? A Cross-country Analysis of Gender Gaps, *Journal of Labor Economics*, 26, 4: 621-654

Parsons, S., Bryson, A. and Sullivan, A. (2024) "Teenage conduct problems: a lifetime of disadvantage in the labour market?", *Oxford Economic Papers*, 76, 1: 60-80

Pearson, H. (2016). *The Life Project: The extraordinary story of our ordinary lives*. Allen Lane: Penguin Random House UK.

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

Pringle, M. K., Butler, N. and Davie, R. (1966) *11,000 seven year olds*, London, England: Longman

Ritchie, S. J. and Bates, T.C. (2013) Enduring Links From Childhood Mathematics and Reading Achievement to Adult Socioeconomic Status. *Psychological Science*, 24, 7: 1301-1308

Rutter, M., Tizard, J., and Whitmore, K. (1970) *Education, Health and Behaviour*, Longman, London

Shepherd, P. (2013) *Bristol Social Adjustment Guides at 7 and 11 Years*, Centre for Longitudinal Studies, London https://cls.ucl.ac.uk/wp-content/uploads/2017/07/NCDS-Bristol-Social-Adjustment-Guides-final.pdf

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

Strittmatter, A. and Wunsch, C. (2025) Labor market sorting and the gender pay gap revisited, *Journal of Population Economics*, 38: 59, 1-41

Stott, D. (1958) *The Social Adjustment of Children: Manual to the Bristol Social Adjustment Guides*. University of London Press

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

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

Wooldridge, J. M. (2010) Estimating Average Treatment Effects, Chapter 21 in *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, Cambridge: Mass

Wright, R. E., and 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, Z. (eds.) 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

Zabalza, A., and Tzannatos, Z. (1985) The Effect of Britain's Anti-discriminatory Legislation on Relative Pay and Employment, *The Economic Journal*, 95: 679–99

Table 1: Occupational Expectations by Age 25, Asked at Age 11 (column %)

	Male	Female
Professional	10	5
Other non-manual, scientific	7	4
Typist, clerical	2	12
Shop assistant	1	8
Junior non-managerial	4	1
Personal services	1	9
Foreman, manual	<1	<1
Skilled manual	20	2
Semi-skilled manual	3	1
Unskilled manual	1	1
Self-employed	2	2
Farm worker	2	2
HM Forces	7	<1
Sports man/woman	10	<1
Student	1	1
Teacher/nurse	2	22
Unclassifiable	27	32
Unweighted N	7,894	7,556

Table 2: Log Hourly Earnings by Age Among Those in Paid Work

	23	33	42	50	55	61	63
Men	1.704	2.209	2.354	2.435	2.359	2.314	2.253
Women	1.536	1.843	1.908	2.08	2.022	1.973	1.899
Difference	-0.168	-0.366	-0.446	-0.355	-0.337	-0.341	-0.354
N	8011	6881	7175	6031	4992	2047	2117

Table 3: Log Hourly Earnings by Age Including Imputed Earnings

	23	33	42	50	55	61	63
Men	1.697	2.168	2.302	2.392	2.321	2.247	2.181
Women	1.501	1.814	1.89	2.053	1.991	1.934	1.863
Difference	-0.196	-0.354	-0.412	-0.339	-0.33	-0.313	-0.318
N	15540	15181	14675	12252	11070	11448	11448

Chart 1: Log Real Hourly Earnings Among Those in Paid Work

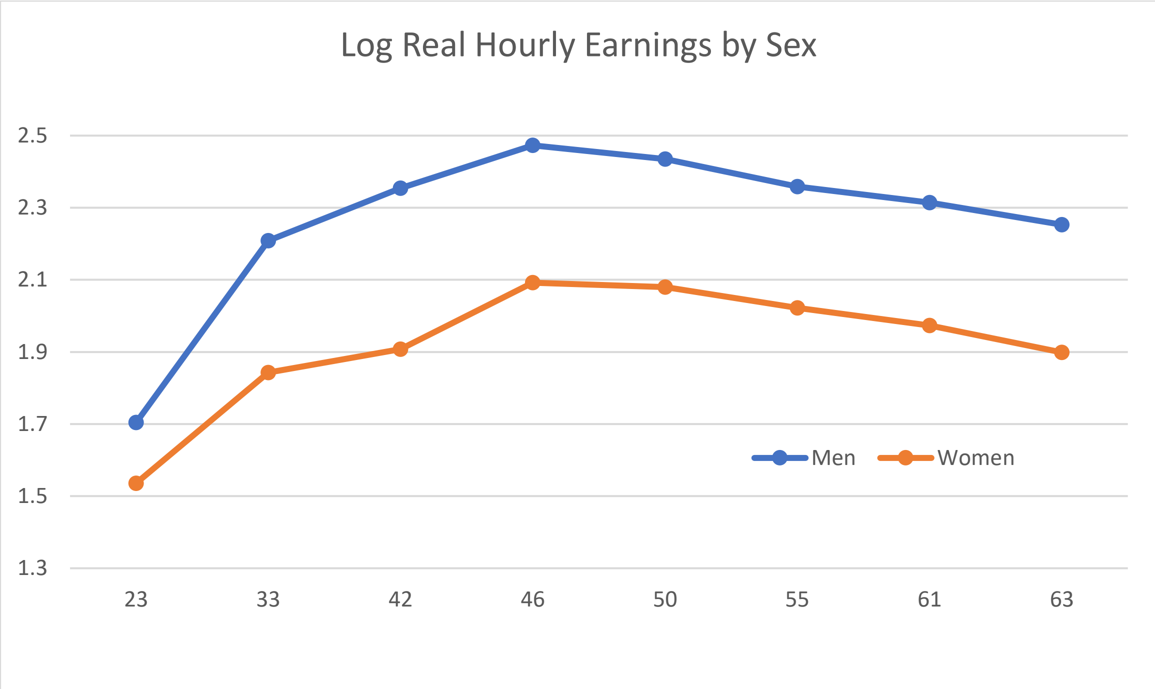


Chart 2: Log Real Hourly Earnings Including Imputed Earnings

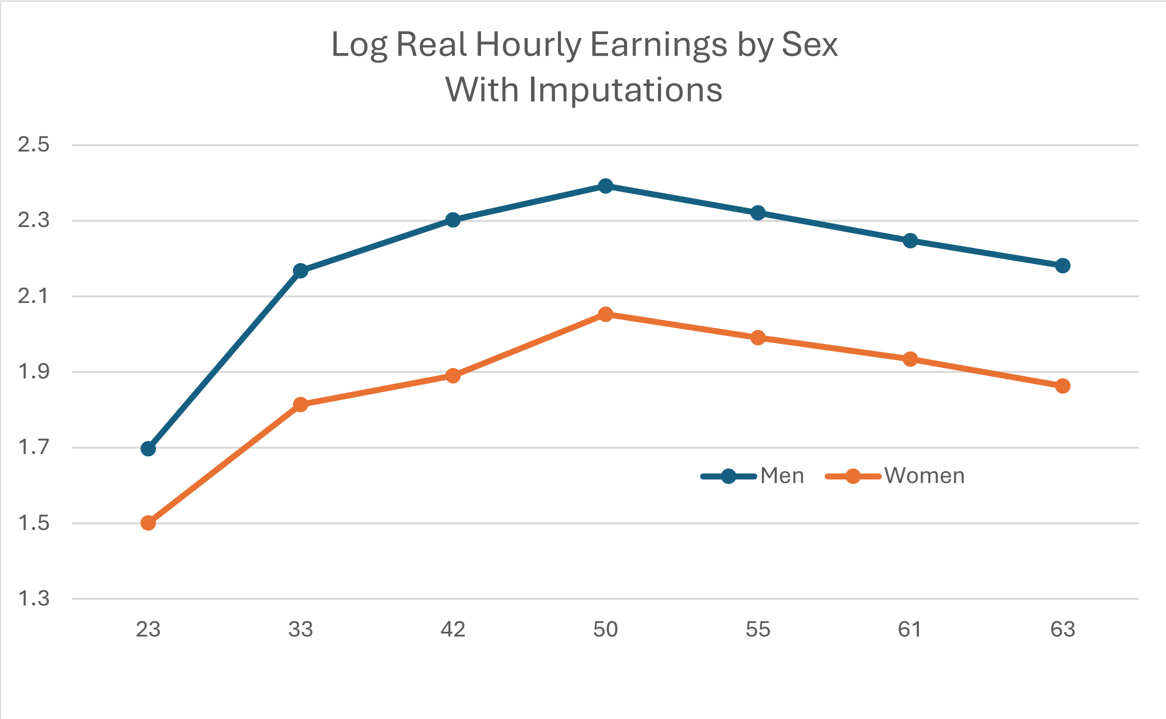


Chart 3: Log Hourly Gender Wage Gap by Age, Workers Only, Full Controls

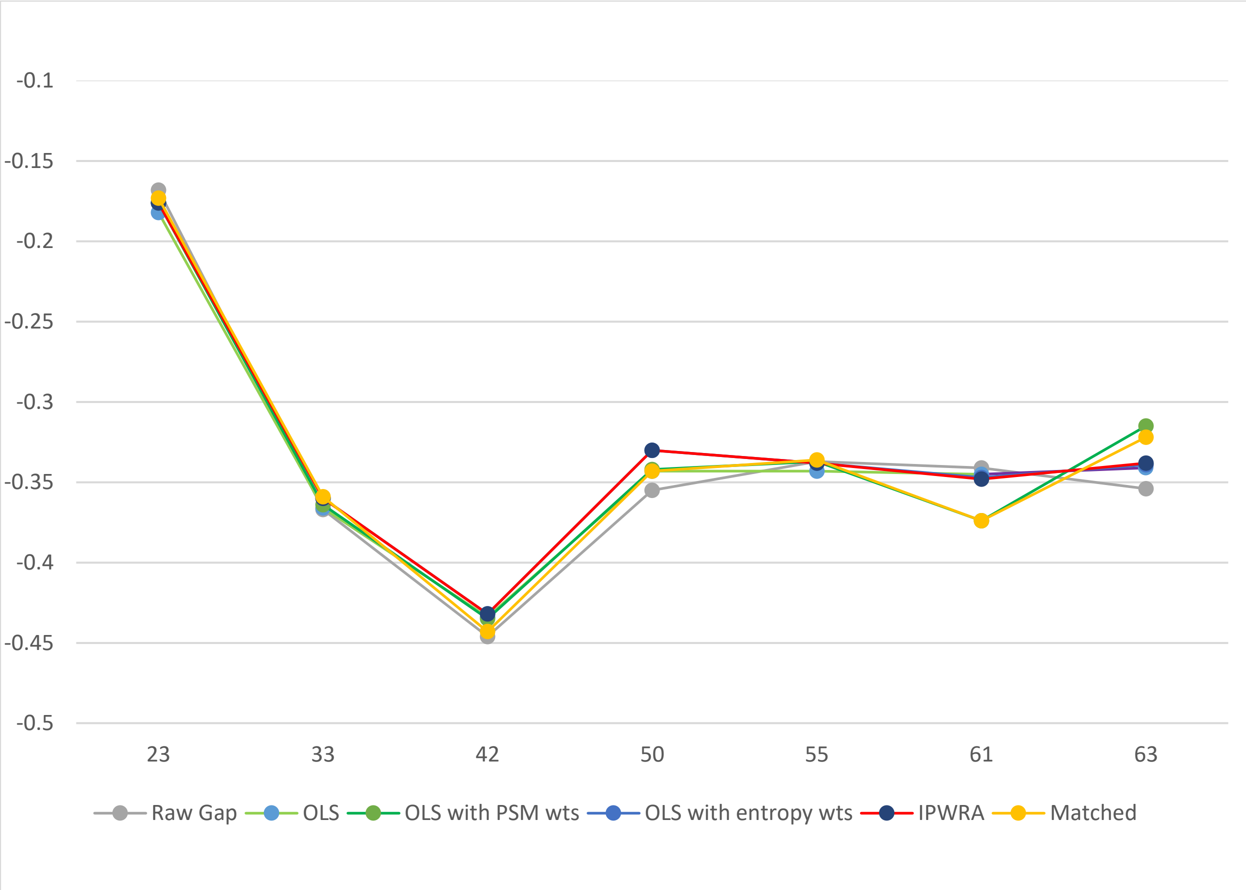


Chart 4: Log Hourly Gender Wage Gap by Age, Workers Only, Parsimonious Controls

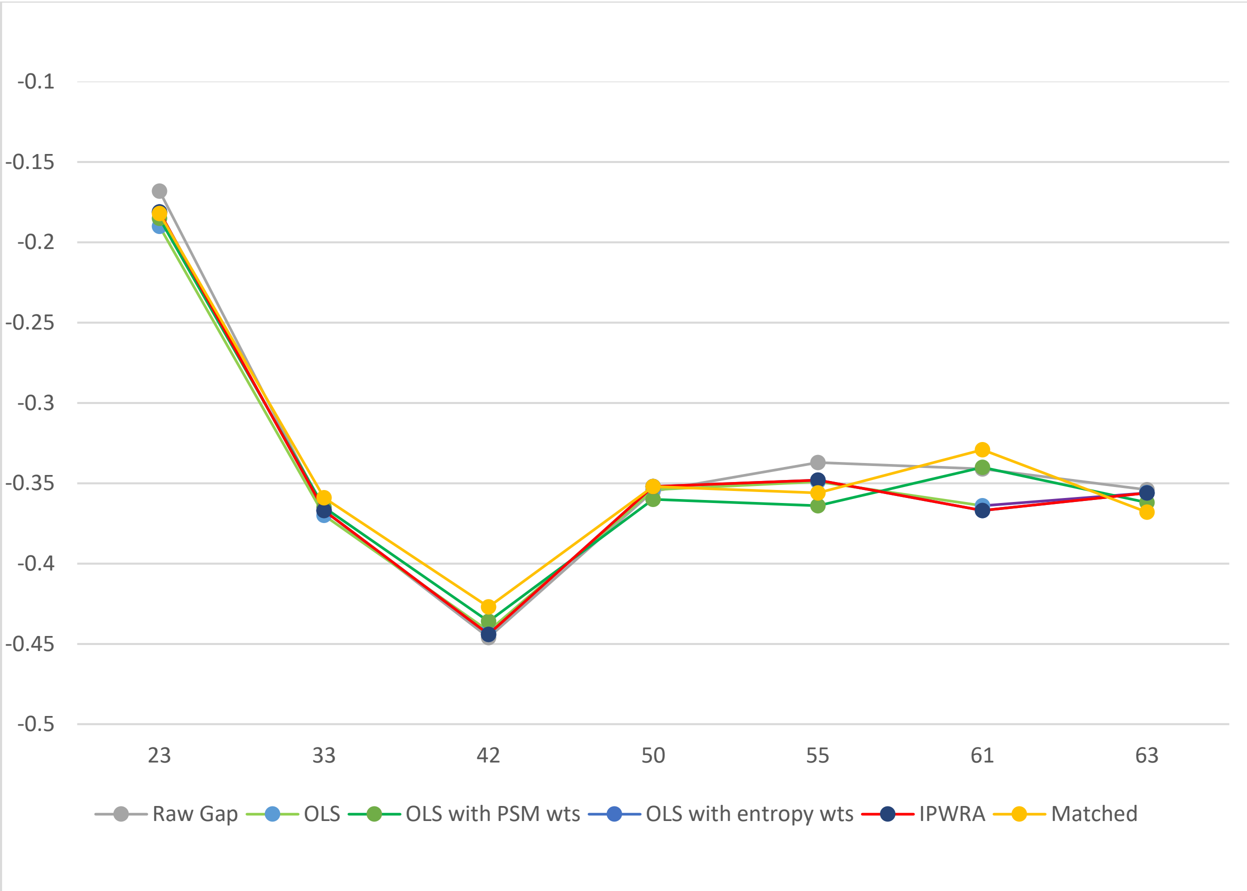
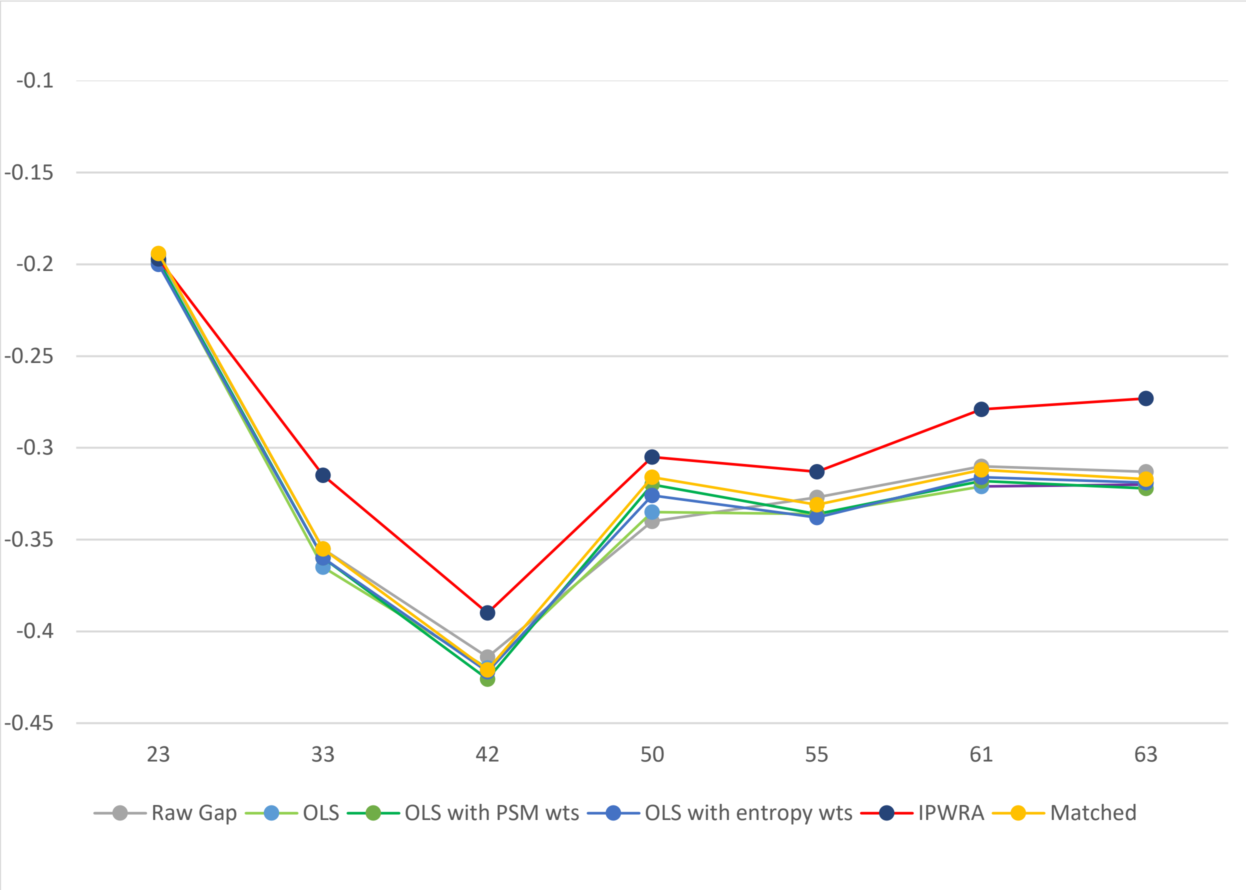


Chart 5: Log Hourly Gender Wage Gap by Age, With Imputed Earnings, Full Controls



Appendix Table 1: Summary Statistics for Conditioning Variables

Variable	Mean	Min	Max
At birth:			
White	0.97	0	1
Birth weight (ounces)	117.4	36	204
Birth weight missing	0.09	0	1
Country of Birth (col %):			
England	78.4	0	100
Wales	5.1	0	100
Scotland	10.5	0	100
Other	1.9	0	100
Not in GB	4	0	100
Siblings at Birth (col %):			
None	23	0	100
Older Sibs only	19.5	0	100
Younger Sibs only	22.3	0	100
Older and Younger Sibs	23.4	0	100
Missing	11.8	0	100
Father's social class: (col %)*			
I Professional	3.9	0	100
II Managerial and Technical	11.6	0	100
III Skilled Non-manual	8.7	0	100
III Skilled Manual	45.2	0	100
IV Partly Skilled	11	0	100
V Unskilled	8.7	0	100
Missing (inc. no father present)	10.9	0	100
Mother smoked before pregnancy (col %):			
Yes	55.1	0	100
No	38.6	0	100
Missing	6.4	0	100
At Age 7:			
Problem Arithmetic Test	5.1	0	10
Arithmetic problems missing	13	0	1
*Southgate reading test score**	23.3	0	30
Southgate reading score missing*	12.8	0	1
*N Rutter type symptoms in childhood reported by mum**	5.7	0	23
N Rutter items missing*	3.5	0	23
*Bristol Social Adjustment Guide score**	7.7	0	64
Missing on overall BSAG*	12.8	0	1
N child illnesses or health difficulties	6.1	0	46
N missing on child illness/difficulties	2.2	0	46
At Age 11:			
*Standardized reading score**	0.002	-2.54	2.26
Missing on reading score*	16	0	1
*Standardized maths score**	0.002	-1.61	2.26

			Continued…
Missing on maths score*	16	0	1
Occupational expectations at 25(col %):			
Professional	7.2	0	100
Teacher/Nurse	11.8	0	100
Other non-manual	5.6	0	100
Typist/clerical	6.9	0	100
Shop assistant	4.5	0	100
Junior non-manual	2.3	0	100
Personal service	4.9	0	100
Skilled manual	11.3	0	100
Foreman, manual	0.3	0	100
Semi-skilled manual	2	0	100
Unskilled manual	0.6	0	100
Self-employed	1.6	0	100
Farm worker	2.1	0	100
Armed Forces	3.9	0	100
Sportsperson	5.2	0	100
Student	0.7	0	100
Unclassifiable	20.3	0	100
At Age 16:			
in trouble with police, teacher (col %):			
Yes	6.4	0	100
No	62.2	0	100
Missing	31.5	0	100
Mother's perception of child's weight (col %):			
Normal	51.9	0	100
Thin	8.4	0	100
Obese	7.3	0	100
Don't know	32.4	0	100
Disabled (col %):			
Yes	4.9	0	100
No	60.9	0	100
Don't know	34.2	0	100
How long since drank alcohol (col. %):			
Never	4.7	0	100
< 1 week	33.1	0	100
> 1 week, < 13 weeks	20.8	0	100
>12 weeks	4.7	0	100
Don't know	36.7	0	100
If smoke (col. %):			
Yes	25.7	0	100
No	46.5	0	100
Don't know	27.4	0	100
N missings on continuous variables	1.14	0	8

Note: variables with * are those appearing in the parsimonious estimation.

Appendix Table 2: Match Bias Pre- and Post-PSM

	23	33	42	50	55	61	63
Pseudo r-squared:							
Unmatched	0.108	0.088	0.095	0.092	0.098	0.113	0.112
Matched	0.001	0.001	0.001	0.001	0.002	0.004	0.006
Rubin's B	8.60	7.80	8.70	8.50	9.60	15.60	17.90
Rubin's R	1.04	0.98	0.99	1.03	0.97	1.06	1.10

The table presents balancing tests pre- and post-propensity score matching. The first row shows the pseudo r-squared for probit models estimating a (0,1) dummy where 1=being a woman. The model is not statistically significant once the matching weights have been applied, indicating that childhood co-variates are balanced across males and females. Row 2 reports Rubin's B which is the absolute standardised difference of means of linear index of propensity score in treated and match non-treated groups. Values above 25 are a cause for concern (Wooldridge, 2010: 917). Rubin's R is the ratio of treated to matched non-treated variances of propensity score index. An R between 0.5 and 2 is deemed balanced. We see that we achieve balance according to these criteria. The same is true when we introduce those with imputed earnings.

Appendix Table 3: Log Hourly Gender Wage Gap by Age, Workers Only, Full Controls

Age	23	33	42	50	55	61	63
Raw Gap	-0.168	-0.367	-0.446	-0.355	-0.337	-0.341	-0.354
Matched	-0.173	-0.359	-0.443	-0.343	-0.336	-0.374	-0.322
OLS	-0.182	-0.366	-0.434	-0.343	-0.343	-0.345	-0.341
OLS with PSM wts	-0.176	-0.364	-0.435	-0.342	-0.337	-0.374	-0.315
OLS with entropy wts	-0.176	-0.36	-0.432	-0.33	-0.338	-0.347	-0.339
IPWRA	-0.176	-0.36	-0.432	-0.33	-0.338	-0.348	-0.338
N	8011	6881	7175	6031	4992	2074	2177

Notes: wages are deflated using the CPI to their real value in January 2000. Earnings are windsorized.

Appendix Table 4: Log Hourly Gender Wage Gap by Age, Workers Only, Parsimonious Controls

Age	23	33	42	50	55	61	63
Raw Gap	-0.168	-0.367	-0.446	-0.355	-0.337	-0.341	-0.354
Matched	-0.182	-0.359	-0.427	-0.352	-0.356	-0.329	-0.368
OLS	-0.19	-0.37	-0.442	-0.354	-0.349	-0.364	-0.356
OLS with PSM wts	-0.185	-0.365	-0.436	-0.36	-0.364	-0.34	-0.362
OLS with entropy wts	-0.181	-0.367	-0.444	-0.352	-0.348	-0.367	-0.356
IPWRA	-0.181	-0.367	-0.444	-0.352	-0.348	-0.367	-0.356
N	8011	6881	7175	6031	4992	2074	2177

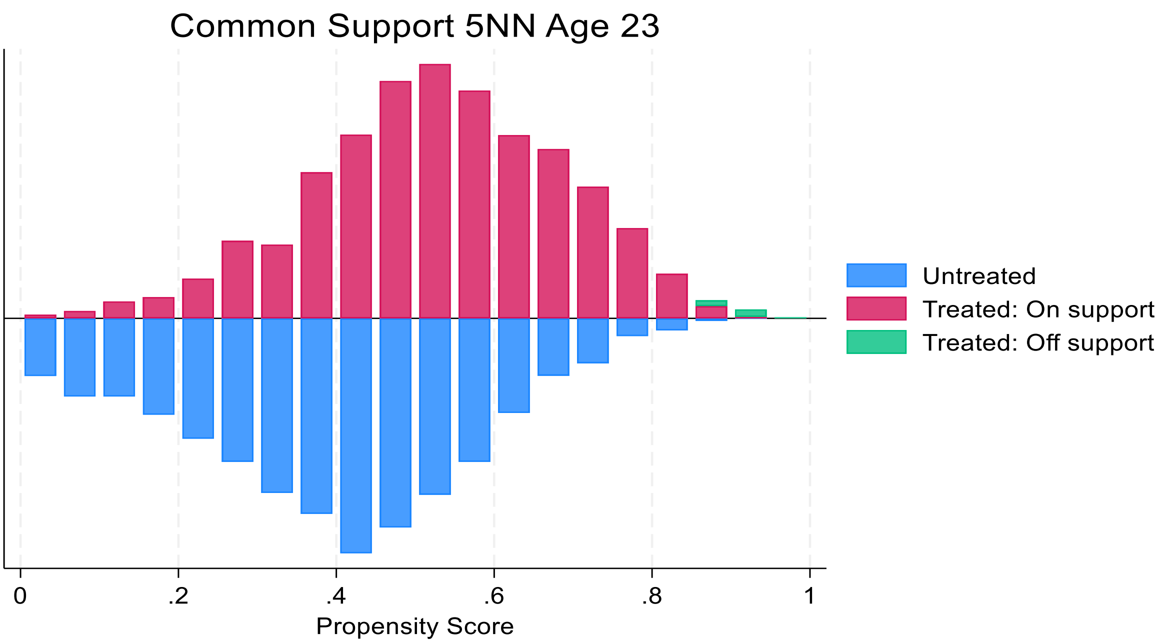
Appendix Table 5: Log Hourly Gender Wage Gap by Age, With Imputed Earnings, Full Controls

Age	23	33	42	50	55	61	63
Raw Gap	-0.195	-0.355	-0.414	-0.34	-0.327	-0.31	-0.313
Matched	-0.194	-0.355	-0.421	-0.316	-0.331	-0.312	-0.317
OLS	-0.199	-0.365	-0.42	-0.335	-0.336	-0.321	-0.32
OLS with PSM wts	-0.198	-0.36	-0.426	-0.32	-0.336	-0.318	-0.322
OLS with entropy wts	-0.2	-0.36	-0.422	-0.326	-0.338	-0.316	-0.319
IPWRA	-0.197	-0.315	-0.39	-0.305	-0.313	-0.279	-0.273
N	15540	15181	14675	12252	11070	11448	11448

Appendix Table 6: Log Hourly Gender Wage Gap by Age, With Imputed Earnings, Parsimonious Controls

Age	23	33	42	50	55	61	63
Raw Gap	-0.193	-0.361	-0.42	-0.346	-0.331	-0.322	-0.322
Matched	-0.203	-0.363	-0.422	-0.344	-0.326	-0.332	-0.324
OLS	-0.204	-0.373	-0.432	-0.348	-0.346	-0.347	-0.336
OLS with PSM wts	-0.204	-0.37	-0.43	-0.35	-0.337	-0.34	-0.337
OLS with entropy wts	-0.201	-0.372	-0.436	-0.348	-0.346	-0.344	-0.342
IPWRA	-0.201	-0.327	-0.406	-0.333	-0.324	-0.319	-0.299
N	15540	15181	14675	12252	11070	11448	11448

Appendix Chart 1: Common Support Age 23



Appendix Chart 2: Log Hourly Gender Wage Gap by Age, With Imputed Earnings, Parsimonious Controls