DISCUSSION PAPER SERIES

Gender Wage Gap among Young Adults: A Comparison across British Cohorts

Francesca Foliano
Alex Bryson
Heather Joshi
Bożena Wielgoszewska
David Wilkinson

Gender Wage Gap among Young Adults: A Comparison across British Cohorts

Francesca Foliano
UCL

Alex Bryson
UCL and IZA

Heather Joshi
UCL

Bożena Wielgoszewska
UCL

David Wilkinson
UCL

ABSTRACT

Gender Wage Gap among Young Adults: A Comparison across British Cohorts*

We study the evolution of the gender wage gap among young adults in Britain between 1972 and 2015 using data from four British cohorts born in 1946, 1958, 1970 and 1989/90 on early life factors, human capital, family formation and job characteristics. We account for non-random selection of men and women into the labour market and compare the gender wage gap among graduates and non-graduates. The raw and covariate adjusted gender wage gaps at the mean decline over the period among nongraduates, but they rise among young graduates. The gender wage gap across the wage distribution narrows over time for lower wages. Adjusting for positive selection into employment increases the size of the gender wage gap in earlier cohorts, but selection is not apparent in the two most recent cohorts. Thus the rate of convergence in the wages of young men and women is understated when estimates do not adjust for positive selection in earlier cohorts. Differences in traditional human capital variables explain only a very small component of the gender wage gaps among young people in all four cohorts, but occupational gender segregation plays an important role in the later cohorts.

Corresponding author:
Francesca Foliano
University College London
55-59 Gordon Sq.
WC1H 0NU
London
Great Britain
E-mail: f.foliano@ucl.ac.uk

1 Introduction

In the UK gender inequalities in wages appear early in working life and gradually widen over the life course (Manning and Swaffield, 2008; Costa Dias *et al.*, 2020; Bryson *et al.*, 2020; Benny *et al.*, 2021), especially with the arrival of children (Costa Dias *et al.*, 2020; Kleven *et al.*, 2019). For most workers employment in early adulthood precedes the acquisition of family responsibilities, therefore a study of the gender wage gap at the beginning of the working life can help understand some of the mechanisms through which the gender wage gap first appears and creates path dependence over the subsequent ages.

In recent cohorts, women have surpassed men in terms of their academic attainment while they delayed the birth of the first child and increased their participation in the labour market (Albanesi *et al.*, 2022). At the same time a series of policies, the Equal Pay Act of 1970 implemented in 1976 with other Equal Opportunities legislation, the Equal Pay Amendment Act of 1980, the National Minimum Wage Act 1998 and the Equality Act 2010, aimed to reduce discrimination and should have created a more favourable environment for gender pay equity for later cohorts of workers. As a consequence, studying the evolution of the gender wage gap and its main sources across successive cohorts of young adults can shed light on how demographic and economic shifts, including changing non-random selection into the labour market, may have affected the gender disparity in earnings over time while legislation likely contributed to reduce discrimination.

The gender wage gap for all workers in the UK, as for most developed countries, has substantially decreased over time (Kunze, 2018). While we might anticipate a similar decline among young adults, convergence in qualifications and experience is unlikely to have affected the gender wage gap among young people because previous studies find that human capital variables explain only a small proportion of the earnings gap among young adults (Fortin, 2008; Combet and Oesch, 2019). Indeed, some research suggests gender convergence in earnings may be heterogeneous across the skill and wage distribution. Goldin (2021) shows that in the United States graduate women's earnings stopped catching up with graduate men's earnings after 1990. In Europe, research shows that the gender wage gap is larger at the top of the earnings distribution than at the bottom, but that, in some cases, it has widened among low earners, where women are concentrated (Albrecht *et al.*, 2003; Arulampalam *et al.*, 2007).

A complicating factor in the study of the gender convergence in earnings is that the increasing labour market participation of young women may have changed the profile of their selection into employment. It is important to consider any changing sample selection when comparing the observed wages of men and women over time. Non-random selection into employment may differ at different stages of life and across cohorts. If women are more positively selected into employment than men, as earlier research for Britain suggested (Neuburger *et al.*, 2011), failure to account for this bias leads to an underestimation of the gender wage gap. Furthermore, if the positive selection into employment has diminished over time, as Bryson *et al.* (2020) show for workers born in 1958 and 1970, failure to account for such a change will underestimate convergence in the pay of young men and women.

In this paper we examine the gender wage gap among young adults across four British cohorts, over five decades. We use four nationally representative British birth cohort studies whose members were born in 1946, 1958, 1970 and 1989-90, who were interviewed at age 26 (1972), 23 (1981), 26 (1996) and 25 (2015) respectively. We ask four questions: first, how has the gender wage gap evolved among young adults in Britain, over a period when policies were being introduced to promote equal pay? Secondly, were graduates more or less likely than non-graduates to have benefited from improved opportunities? Third, how does selection into the labour market affect these trends? Fourth, which human capital, family and job characteristics help explain the size of the gender wage gap?

We study the evolution of the raw gender wage gap at the (logarithmic) mean and across the distribution, separately, for graduates and non-graduates. In doing so we compare raw (observed) wage gaps with those adjusted for covariates on personal characteristics. We then present an additional set of estimates of the gender wage gap that accounts for selection into employment. We use standard decomposition methods to explore the role of early life factors, human capital, marital and parental status in explaining the gender wage gap. Similarly we estimate the gender wage gap across the wage distribution to investigate whether wage inequalities among young adults vary for low and high earners over time and after accounting for selection into employment; we then decompose this gap into a component explained by the distribution of characteristics and an unexplained component. For the two most recent cohorts we have detailed information about the occupations held by respondents: we therefore include observed job characteristics in the covariates to assess how they contribute to the observed gender wage gap at the mean and across the wage distribution.

The paper contributes to the literature on the gender wage gap in four ways. First, it extends earlier studies examining the gender wage gap among young adults using nationally representative samples (see, for example, Neuburger, 2010; Neuburger *et al.*, 2011; Fortin, 2008; Manning and Swaffield, 2008; Combet and Oesch, 2019), by incorporating a cohort of individuals born as early as 1946 and a much more recent birth cohort - born in 1989/90 - to examine trends over 43 years. The 1946 cohort study, although it has limited sample size, includes unique microdata evidence from the pre-equal pay and opportunities legislation era. We find suggestive evidence that the equal pay policies had an impact in reducing the gap between the 1970s and 1980s, and that the NMW policy reduced the gap for the lowest wages in the later cohort.

The second contribution is to assess and compare the trends in the gender wage gap among graduates and non-graduates in the population at large. Other studies on the gender wage gap among young people focus on homogeneous samples of high-skilled young adults (Dolton *et al.*, 1996; Chevalier, 2007; Goldin and Katz, 2008; Bertrand *et al.*, 2010; Azmat and Ferrer, 2017). We find that the gender wage gap among non-graduates declines over time while wages stop converging after the mid-1990s among young graduates.

Our third contribution is to study gender inequalities in earnings across cohorts by accounting for differential selection into employment over time, given the changing participation of young men and women in the labour market and their shifts in educational attainment and fertility. We find that adjusting for positive selection into employment increases the size of the gender wage gap in earlier cohorts, but selection is not apparent in the two most recent cohorts. Thus, the rate of convergence in the wages of young men and women is understated when estimates do not adjust for positive selection in earlier cohorts.

Our fourth contribution is to investigate the role of specific job characteristics in explaining the gender inequality in this particular age group as well as human capital and family. As women started to look more like men by catching up with them in terms of their main human capital traits, researchers turned to differences in job characteristics to explain the gender wage gap (Cortes and Pan, 2018). Groshen (1991) reports that occupational segregation explains a sizeable part of the gap in five US industries. Bayard *et al.* (2003) use administrative data for all sectors and occupations in the US to study the contribution

of occupational gender segregation to the gender wage gap. More recently Blau and Kahn (2017) show that differences over time in occupations and sectors explain a rising proportion of the gender wage gap for all women in the US. We find that occupational gender segregation plays an increasingly important role, accounting for much of the observed gap in 2015.

The rest of the paper is structured as follow: section 2 describes the data; section 3 details the empirical strategy to estimate the gap; section 4 presents the results of our decomposition analysis; section 5 concludes.

2 Data

We examine gender wage gaps among young people in their twenties from four nationally representative British cohort studies (see Ferri *et al.*, 2003; University College London, b,a,c, various years, for documentation). The first is the National Study of Health and Development (NSHD), a cohort born in 1946 and interviewed as young adults, at 26 in 1972. The second is the National Child Development Study (NCDS) whose members were born in 1958 with an interview in 1981 at age 23. The third is the British Cohort Study (BCS), born in 1970, whose members responded to a postal survey in 1996 at age 26. The fourth is Next Steps[1], born in 1989-90, in England only, whose members were interviewed in 2015 at age 25.

All four cohort studies include a rich set of demographic and longitudinal socio-economic characteristics of the cohort members. NSHD, NCDS and BCS follow the cohort members since birth. Next Steps started following cohort members from the age of 14 which means that early individual and family characteristics are not available for this cohort. All four studies include a set of cognitive measures recorded at pre-employment ages which would not be available in cross-sectional sources.[2] Detailed information about the highest qualifications achieved by the cohort members at the time of data collection allows us to map these qualifications into levels of education across cohorts, by taking into account the significant changes in the structure of education that happened over time. This allows us to study the evolution of the gender wage gap separately among graduates and non-graduates. Graduates are defined consistently across the four cohort studies as those respondents who

---

[1]Originally known as Longitudinal Survey of Young People in England.

[2]Next Steps is linked to administrative records of educational attainment. Therefore it includes scores in Reading and Maths from the tests taken by the cohort members at the end Key Stage 2, aged 11-12.

achieved post-secondary professional qualifications and higher education diplomas, university degrees or higher degrees.[3] Gross hourly wages are derived from information collected in every cohort about gross weekly wages and hours worked per week; these are then deflated by the RPI to January 2000 prices.

The studies were affected by sample attrition over time, an issue we tackle with attrition weights as described in Section 3. The initial samples used in our study include only respondents who participated in each survey in the year of interest (1972 for NSHD, 1981 for NCDS, 1996 for BCS and 2015 for Next Steps). The full number of observations for men and women in each sample is the following: 1901 and 1851 in the NSHD, 6043 and 6033 in the NCDS, 3959 and 4756 in the BCS and 3371 and 4198 in Next Steps.

The evolution of female and male employment rates is crucial to understanding any change in selection affecting the comparison of cohorts. The proportion of young men and women in employment across the four cohorts is reported in the left plot of Figure 1. For young men this dropped sharply between 1972 and 1981 (from 94% to 76%) and remained constant after that. The employment rate of young women increased between 1972 and 1981 by 14 percentage points (from 48% to 62%), coinciding with the Equal Pay and opportunities legislation that came into force in 1976, and a sharp drop in childbearing by women in their early twenties. It continued to increase between 1981 and 1996 by 9 percentage points. The increase was only marginal in the last cohort (from 73% to 77%). The steep convergence in employment rates between 1981 and 1996 suggests the types of men and women entering the labour market possibly changed too, reducing the positive selection of women and increasing that of men.

The comparison of gender inequalities in wages by education levels over time requires the consideration of possible differential selection into higher education. If young women become less positively selected into higher education over time compared to men, the convergence of wages will be underestimated for both graduates and non-graduates. Young women were

---

[3]Graduates are defined as individuals who achieved level 4 or 5 of National Vocational Qualifications (NVQ), which include: BTEC Higher Certificate/Diploma, HNC/HND, Professional degree level qualifications, Nursing qualifications, RSA Higher diploma, Non-Vocational Qualification level 4 and 5, Degree, HE diploma, PGCE, other teaching qualifications, higher degree.

less likely than young men to have a degree in 1972, as shown by the right plot in Figure 1.[4] After that, the proportion of young men and women with higher education was similar and it increased over time until 2015, when the young women were more likely than men to be graduates. The low proportion of graduate women compared to men in the earliest cohort suggests more positive selection of women compared to men into higher education in that cohort. This is confirmed by descriptive statistics of cognitive test scores recorded pre-labour market entry presented in tables A1 to A4 in Appendix A. A comparison between graduates and non-graduates shows that across all cohorts there is positive selection into higher education, i.e. those with a degree have higher pre-market cognitive scores in school. However, the advantage has declined over time particularly for women. Graduate women had higher scores than men in reading and maths cognitive tests in 1972. This difference changed in favour of men in 1981, and then it appeared to settle in later cohorts where women performed better in reading and men performed better in maths. We do not observe gender differences in test scores among non-graduates. If young women have been less positively selected into higher education over time, compared to men, this might have contributed to a growing divergence in the earnings of graduate men and women over time, and increased the rate of convergence in earnings between non-graduate men and women.

The variables used in the covariate adjustment of the gender wage gap can be grouped into four main sets.

- Early years factors: mother's school leaving age, father's school living age and father's social class.

- Human capital variables: maths and reading standardised test scores, indicators for the highest qualification achieved, indicators for subjects at higher education (science, technology, engineering, or mathematics, law, economics or management, others), months of full-time and part-time work experience, squared months of full-time and part-time work experience, the number of different spells of work experience.

- Family formation variables: indicators for whether the young adult is married/cohabiting,

---

[4]The apparent dip in the proportion of male graduates between 1972 and 1981 should be qualified by the consideration that the NCDS data comes from the lowest age, 23. The series disguises an upward trend across all of these cohorts in the proportions having graduate qualifications by their thirties, see Ferri *et al.* (2003) Table 2.1a. By that age then, NCDS had overtaken the 1946 cohort in the proportion graduate through qualifications gained post 23.

Figure 1: Employment and higher education rates among young Men and Women in the British cohort studies
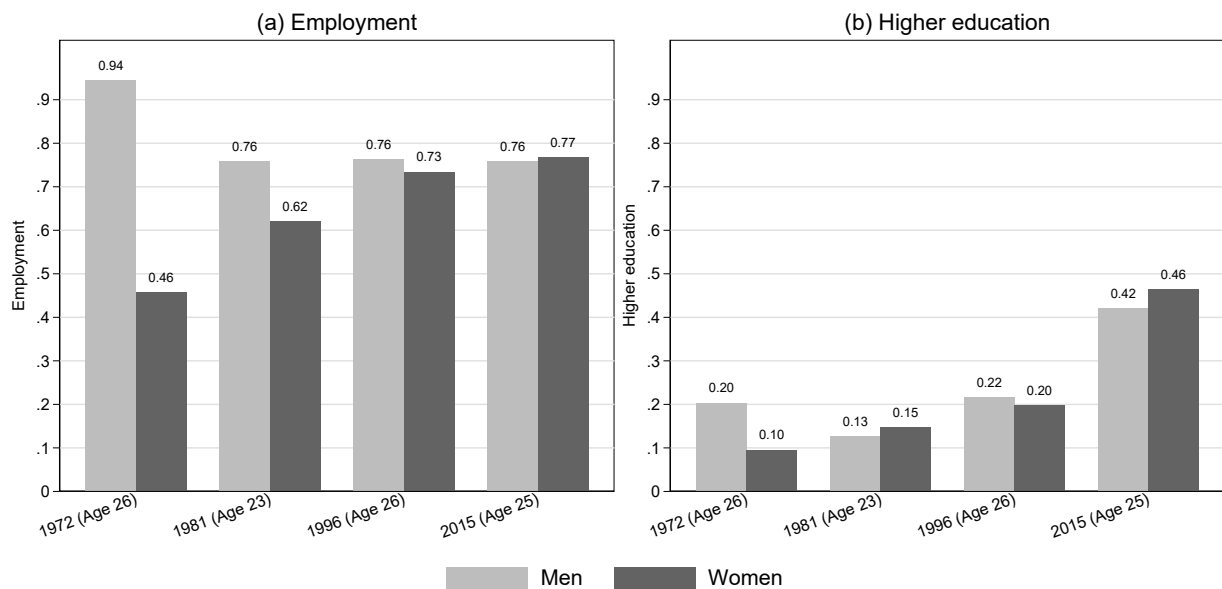
Note: Authors' calculation based on NSHD data in 1972, NCDS data in 1981, BCS data in 1996 and Next Steps data for England in 2015. Weights are used to account for attrition in all four cohorts and for stratification in NSHD.

  for the presence of at least one dependent child and for the presence of more than one dependent child.

- Job characteristics: sets of indicators for weekly hours worked (30 hrs or less, 30 to 45 hrs and more than 45hrs); and for the major group of the Standard Occupational Classification for the job held by the cohort member; the proportion of women by occupation defined at four digit level of the Standard Occupational Classification. This last variable is defined only for BCS and Next Steps by using the contemporary Labour Force Surveys.

Descriptive statistics for these variables in the estimation samples are reported by higher education status and cohort in tables A1 to A4 in Appendix A.

3 Empirical Methodology

We compute and compare three variants of the gender wage gap: the raw gender wage gap; a covariate adjusted gap; and a covariate adjusted gap that also accounts for selection into employment. We define the raw gender wage gap as the difference in mean log hourly wages between men and women. The adjusted gender wage gap is the estimated coefficient γ of the female indicator variable from the following model:

$$Ln\left(HW_i\right) = \gamma Female_i + X_i\beta + \varepsilon_i \tag{1}$$

Where HW_i is the real hourly wage derived from reported weekly wage and hours worked; X_i include all the variables for early years factors, human capital and family formation presented in tables A1 to A4 in Appendix A.

To account for the effect of differential selection into employment on the estimates of the gender wage gap we adopt an imputation-based approach similar to Bryson *et al.* (2020) and Neuburger *et al.* (2011).[5] Missing wages for individuals who are in employment without a wage observation (NSHD, NCDS, BCS and Next Steps), unemployed or economically inactive (NSHD, NCDS, BCS and Next Steps); self-employed (NCDS and BCS) or students (Next Steps) are imputed using 'potential wages' obtained from individuals with an observed wage who are closer in terms of their propensity to be in waged employment. First we estimate, by gender and cohort, the propensity for waged employment using probits for the (0,1) outcome of having an observed wage at the time of the interview. Tables A5 to A7 in Appendix A reports the list of variables used in the estimation of the probit models.[6] We use propensity score matching to find for each individual without an observed wage, the nearest neighbour in terms of propensity for waged employment within the same sex and cohort. The common support requirement is implemented by dropping respondents whose propensity for

---

[5]Other approaches that deal with non-random selection include the Heckman two-step selection model (Mulligan and Rubinstein, 2008) and estimating bounds for the gender wage gap based on quantiles of the wage distribution (Blundell *et al.*, 2007). However, both these methodologies require a credible exclusion restriction which was not available to us.

[6]Our assumption is that selection into employment over time is based on observed characteristics. Heckman *et al.* (2006) show that pre-market non-cognitive traits, mostly unobserved, are important drivers for labour market outcomes. Unfortunately, while British cohort studies include some non-cognitive measures, these are not consistent across cohorts and therefore they could not be used for a cross-cohorts comparison.

waged employment is smaller than the lowest probability in the waged employee sample in each cohort.[7] The wage of the nearest neighbour is then used as 'potential' wage to impute the missing wage.

The selection adjusted gender wage gap is the estimated coefficient of the female indicator from model 1 where employees' log hourly wages are supplemented with the imputed log hourly wages for individuals for whom we do not observe earnings in the sweep.

We adjust for sample attrition over time by weighting each individual observation by the inverse probability of responding to each sweep. The inverse probability is obtained as the inverse of the predicted probability estimated with a logit for the probability of responding in the sweep of interest given a valid response in the sweep at age 2 (the first follow up) for NSHD, the sweep at age 10/11 for BCS and NCDS; and in the sweep at age 14 for Next Steps.[8] Tables A8 to A11 in Appendix A report the list of variables used to estimate the probability of responding at the sweep of interest. We deal with item non-response in individual characteristics by creating a binary indicator for the missing observations if the variable of interest is categorical, while if the variable is continuous we impute the missing value with the gender-specific mean of the variable obtained by using non-missing only observations and we include an indicator for the missing observations.

To understand the contribution of each set of variables (early life factors, human capital and family formation) to the gender wage gap we carry out a Kitagawa-Oaxaca-Blinder (Kitagawa, 1955; Oaxaca, 1973; Blinder, 1973, KOB hereafter;) decomposition of the selection adjusted gap. To this end log hourly wages for each gender (g) are estimated using the following equation:

$$Ln\left(HW_{ig}\right) = \ \gamma Female_i + X_{ig}\beta_g + \varepsilon_{ig} \tag{2}$$

---

[7]After implementing the common support restriction the number of observations for men and women in each sample adjusted for selection is the following: 1879 and 1831 in the NSHD, 5911 and 5810 in the NCDS, 3721 and 4341 in the BCS and 3322 and 4085 in Next Steps.

[8]The NSHD consists of a socially stratified target sample of individuals born in one week in March 2016. This longitudinal survey only followed up one in four urban working-class children (whose fathers were in urban, manual occupations at birth) and therefore stratification weights are used for descriptive statistics, sample sizes and for the analysis based on these data.

The models for females and males can be subtracted from each other to decompose the mean gender wage gap into mean differences in observed characteristics and differences in returns to these characteristics. To define the differences in returns, one needs to choose the reference parameters. Studies suggests results can be sensitive to different choices. We follow the approach suggested by Neumark (2004) and Jann (2008) by selecting as reference parameters the coefficients from regressions pooled over males and females with the addition of a gender indicator. The decomposition is thus:

$$\overline{Ln\left(HW_{iM}\right)} - \overline{Ln\left(HW_{iW}\right)} = \left(\overline{X}_M - \overline{X}_W\right)\beta^* + \overline{X}_M\left(\beta_M - \beta^*\right) + \left(\beta^* - \beta_W\right)\overline{X}_W \qquad (3)$$

Where β^* is the reference coefficient estimated in Equation 1.[9] The first term on the right-hand side of Equation 3 represents the explained component of the gender wage gap. The second and third terms together represent the unexplained component: the gap in parameters weighted by the mean of characteristics in the male sample plus the deviations of female parameters from the reference set, weighted by mean female characteristics. This unexplained component reflects any unequal treatment/discrimination in the labour market, the impact of gendered social norms and preferences, and of course, of any other factors differentiating the sexes which may not be captured by the characteristics included in the model. In an additional set of analyses we explore the role of job characteristics to explain the gender wage gap among young adults. These decompositions of the gender gap are based on observed wages only because we cannot impute wages where we do not have information on occupations.

Finally, we study the relative importance of differences in observable characteristics and their returns in explaining wage inequalities across the wage distribution, without and with adjustment for selection. To study the unconditional gender gap across the wage distribution and its decomposition between the effect of characteristics and coefficients we rely on the methodology proposed by Chernozhukov *et al.* (2013). This method decomposes unconditional gender gaps at a given percentile of the distribution of log wages into a component due to the distribution of characteristics (the explained part of the gap) and a component due to different wage functions conditional on characteristics (the unexplained part of the gap). The empirical implementation of this method first requires the estimation of conditional

---

[9]We deal with the sensitivity to choice of the reference categories for categorical variables by using the option "normalize" from the Stata command oaxaca (Jann, 2008).

wage distributions by using quantile regressions. Then, unconditional wage distributions are estimated through integration over the distribution of the covariates.

4 Results

4.1 Selection adjustment and wages

After imputing potential wages for individuals without observed earnings, thus adjusting the sample of working individuals for non-random selection into the labour market, we compare the samples with and without selection adjustment. The aim is to show how the main characteristics of young women and men in and out of the labour market have changed over time, bearing in mind that the exact age at observation varies by cohort.

Tables A1-A4 in Appendix A report descriptive statistics, by gender and higher education status, for the early life factors, human capital and family formation variables that will later be included in the estimation of the covariate adjusted gap. The first four columns present descriptive statistics for the sample that includes observed and potential wages while the last four columns report the same for the sample that includes only observed wages. By comparing some of the main individual characteristics, such as cognitive test scores, full-time experience and parental status, between these two samples, by gender and higher education status, we can approximate whether selection into the labour market evolved over time differently between men and women.

Relative to men, graduate women in the first cohort and non-graduate women in the first two cohorts, were positively selected into work in terms of cognitive test scores and months of work experience. In all cohorts working women are much less likely to have children than women in the broader sample while in none of the cohorts did the proportion of men with children differ by whether they are selected into work or not.

By considering only respondents who have an observed wage across cohorts in 1996 and 2015 it is possible to observe how differences in job characteristics for men and women have changed over time. Young women are more likely than men to work in part-time jobs and less likely to work long hours. However, between 1996 and 2015, women narrow the gap in hours, particularly in long hours worked. We do not observe noticeable differences in the

proportion of men and women in managerial or professional jobs while the proportion of all young people in these occupations has decreased across the two cohorts.[10] On average both graduate and non-graduate women hold occupations that are more female-intensive than their male counterparts. But the gap between men and women in terms of female concentration in the occupation has narrowed between 1996 and 2015, as more women joined more integrated and male dominated jobs.

Figure 2 plots, for each cohort study and by higher education status, the distribution of men's and women's log hourly wages before and after imputing potential earnings. A visual inspection of these distributions shows that observed wages of women in the labour market in 1972 and, marginally, in 1981 were higher than the potential wages of women who were not in employment, confirming the hypothesis that young women in NSHD and NCDS positively selected into employment. Comparison of the distribution of wages before and after imputation for BCS and Next Steps shows that the stronger positive selection of young women reduced over time.

4.2   Trends in gender wage gap

A comparison of the raw gender wage gap and the covariate adjusted gap for waged employees in the four cohorts is presented in Figure 3. The gender wage gaps reported here are the estimates of the coefficients for the female indicator in the Mincer wage equation 1. Results are reported for the whole sample (black) and for graduates (dark grey) and non-graduates (light grey) respectively. For each sample the figure reports the estimates of the raw gap (dot marker) and of the conditional gap (triangular marker) from a model that include early life characteristics, qualifications, experience, marital and parental status, and the relative 95% confidence intervals. The estimates are negative as they represent the female mean log hourly wage minus the male mean log hourly wage. The full sets of estimated coefficients for these regressions are reported in tables B1 to B4 in Appendix B.

Measured in points below the zero mark, the raw gap for the whole sample shrinks from 41 log points in 1972 to 10 log points in 1996 and 8 log points in 2015. Adjusting for early life factors, human capital and family formation characteristics only reduces the overall gap

---

[10]Managerial and professional occupations are defined as Major Groups 1 (managers, directors and senior officials) and 2 (professional occupations) in the Standard Occupational Classification 2000 and 2010.

Figure 2: Distribution of log hourly wages for graduates and non-graduates by cohort

(a) Without selection adjustment

(b) With selection adjustment

Men   Women
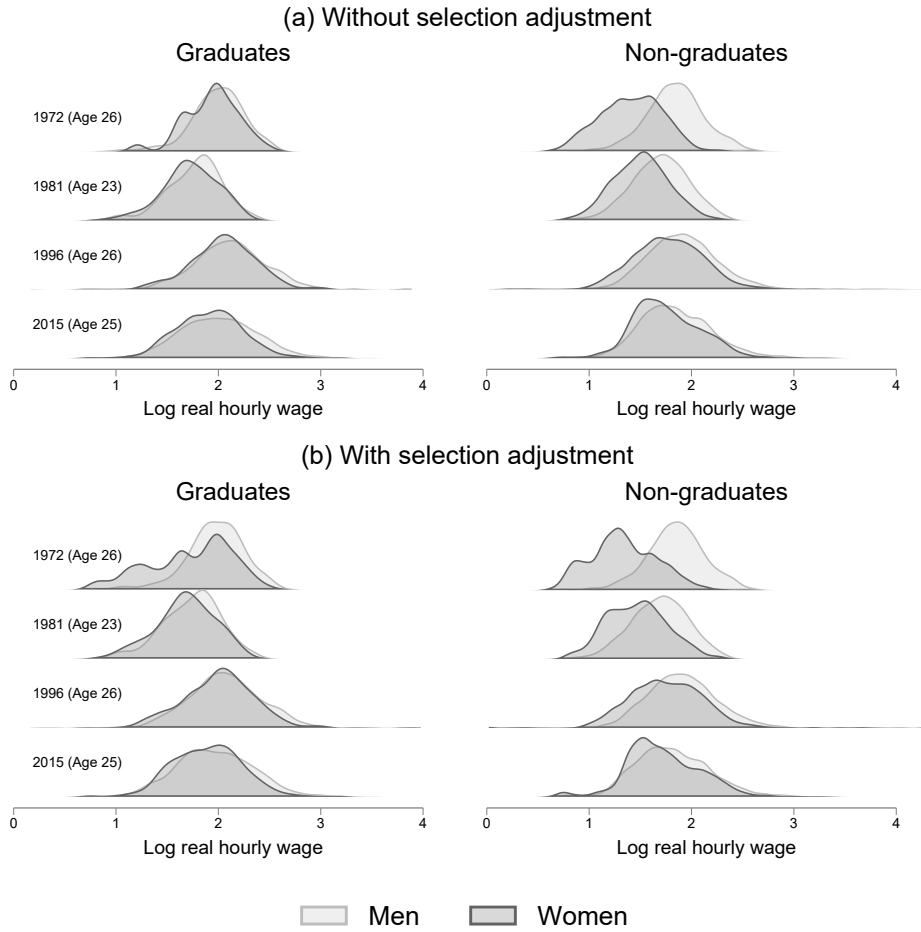
Note: These plots present the kernel density of observed log hourly wages (top panel) and observed as well as potential log hourly wages (bottom panel) by gender, higher education status and cohort. Potential log hourly wages are obtained from nearest neighbour 'donors', defined as those in waged employment in the same cohort and from the same gender who are closer in their propensity to waged employment to the respondents for whom a wage is not observed.

Figure 3: Raw and covariate-adjusted gender wage gap (with 95% CI)
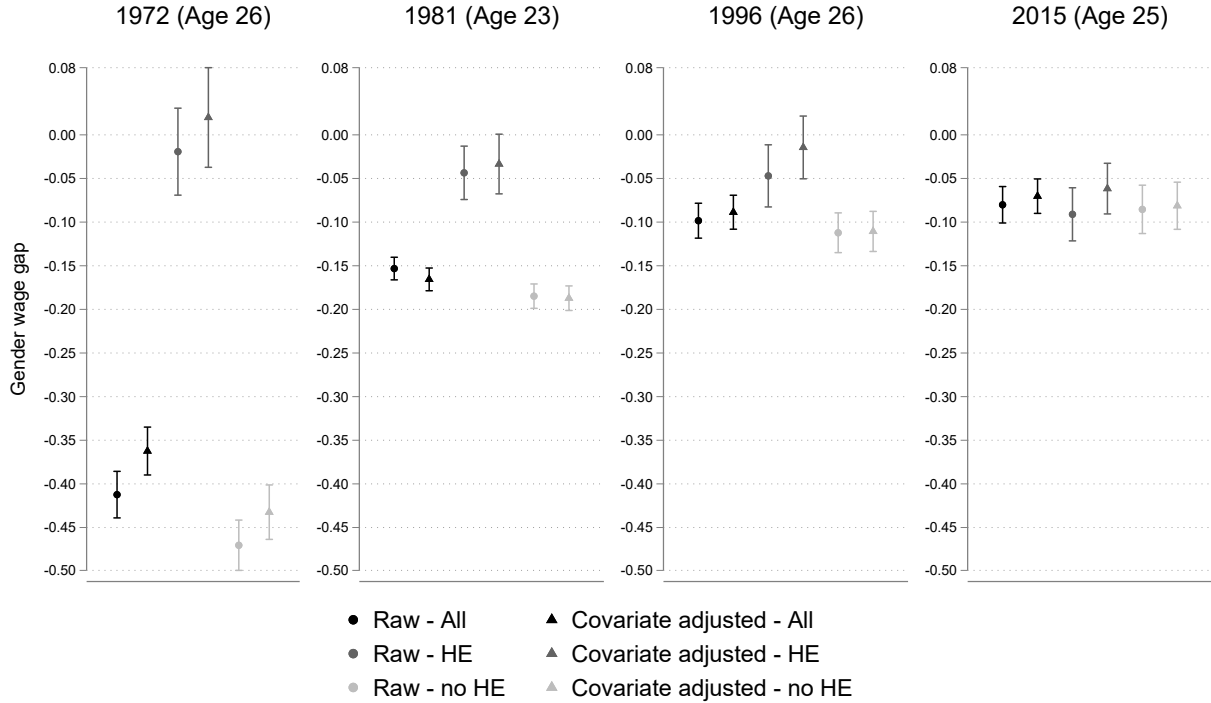
Note: These plots report, by cohort, the estimated coefficients for the female indicator from Equation 1 and its 95% confidence interval for different samples and models. Estimates marked in black refer to the whole sample; estimates in dark grey refer to graduates; estimates in light grey refer to non-graduates. For each sample the dot marker reports the estimate of the coefficient for the female indicator in a wage model without additional controls, while the triangular marker reports the estimate of the coefficient for the female indicator in a wage model that includes early life characteristics, qualifications, experience, marital and parental status. Table A1 to A4 in Appendix A report the full list of variables included in the model for each cohort. Full set of estimates are reported in Appendix B, tables B1 to B4.

significantly in the earliest cohort.

The raw gender wage gap among graduates, represented by the dark grey dots, increases between 1972 and 2015 (from 2 log points to 9.1 log points). The regression adjusted gender wage gap among graduates is small and statistically non-significant in the first three birth cohorts, but is sizeable (6.3 log points), and statistically significant, in the Next Steps cohort.

The two light grey estimates in Figure 3 present the same analysis, for the non-graduate subsample. Among non-graduates the raw gender wage gap declined steadily between 1972

15

and 2015: from 42 to 9 log points. As with the overall sample, adjusting for observed characteristics reduces the gap only in the earliest cohort but does not in any of the subsequent cohorts.

Figure 4: Covariate-adjusted and covariate and selection adjusted gender wage gap (with 95% CI)
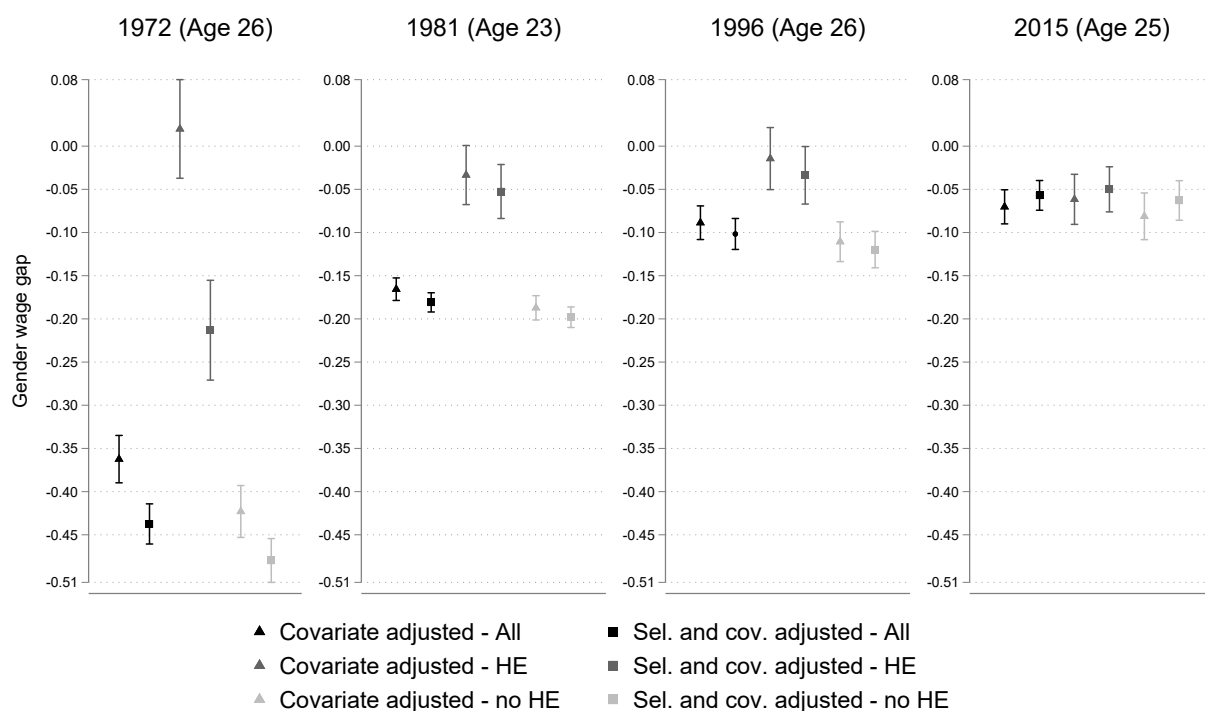
Note: These plots report, by cohort, the estimated coefficients of the female indicator from Equation 1 and its 95% confidence interval for different samples and models. Estimates marked in black refer to the whole sample; estimates in dark grey refer to graduates; estimates in light grey refer to non-graduates. For each sample the triangular markers report estimates based on observed wages while the square markers report estimates based on observed and potential wages. All models include the full set of controls for personal characteristics (early life characteristics, qualifications, experience, marital and parental status). Table A1 to A4 in Appendix A report the full list of variables included in the model for each cohort. Full set of estimates are reported in Appendix B, tables B1 to B4.

Figure 4 compares the covariate adjusted gender wage gap (triangular marker) to the selection and covariate adjusted gender wage gap (square marker). The full sets of estimated coefficients for these regressions are also reported in tables B1 to B4 in Appendix B.

16

Adjusting for employment selection as well as individual characteristics increases the gender wage gap for the whole sample in a statistically significant way among the NSHD cohort members (by 8 log points, from 36 to 44) and the NCDS cohort members (by 1.4 log points from 17 to 19). The increase in the gap estimate in BCS of 1.1 log points (from 8.9 to 10) is not statistically significant, while the gender wage gap among Next Steps cohort members decreases (from 7.1 to 6.3 log points), but this change is not statistically significant.

Accounting for non-random selection into employment increases the gender wage gap among young graduates in 1972 by 23 log points. Furthermore, it doubles the size of the estimated graduate gender wage gap in 1981 (a statistically significant rise from 3 to 6 log points) while the change in the estimate of the gap in 1996 and 2015 in not statistically significant.

Similarly adjusting for selection increases the gender wage gap for non-graduates in the NSHD (from 42 log points to 48) and marginally for the NCDS cohort (from 19 to 20 log points and the difference is statistically significant), while it does not vary the estimates of the gap among non-graduates in the BCS and Next Steps.

In summary, whilst the overall gender wage gap decreases over time this is driven by the decrease in the gap among non-graduates. Gender wage inequality among graduates worsened over time. Adjusting for observed characteristics explains some of the gap for the whole sample in 1972, and only among graduates in 1996 and 2015. Taking into account selection into employment widens the gap among young adults in the earliest cohort, mostly driven by positive selection among graduates, but it has little or no bearing on estimates of the gaps for the most recent cohorts. The contrast between the wage gaps in 1972 and 1981, which are adjusted for the differences in education and family formation, can be assumed to reflect, at least in part, the impact of intervening equal opportunities legislation on the otherwise unexplained female pay penalty.

4.3 Decomposition of the mean gender wage gap (Kitagawa-Oaxaca-Blinder decomposition)

Figure 5 presents the KOB decomposition of the gender wage gap for graduate (on the left) and non-graduate (on the right) cohort members. This figure reports how the main sets of characteristics account for the gender wage gap. The full decomposition estimates with standard errors for each cohort and by higher education status is reported in tables B5 and B6 in Appendix B. The top two plots in Figure 5 report the decomposition without selection adjustments in the sample. The bottom two plots present the decomposition of the gender wage gap that accounts for non-random selection into employment. Negative numbers represent differences in characteristics and returns that work in favour of women. Differences in observables characteristics explain a very small component of the gender wage gap across all cohorts.

Results of the decompositions without selection adjustment show that in 1972 the small gap among graduates was entirely explained by differences in early life characteristics, qualifications (that include cognitive test scores and subjects studied), and parental status. Differences in experience explain a small portion of the gap in 1981 while differences in qualifications explain some of the gap among graduates in 1996 and 2015.

Among non-graduates differences in experience explain a small proportion of the gender wage gap across the four cohorts while qualifications, that for this group include indicators for secondary school levels as well as cognitive test scores, represent consistently a small negative component of the gap, reflecting the higher attainment of women in these sub-samples. Accounting for non-random selection increases the unexplained component of the gender wage gap in the earlier two cohorts for both graduates and non-graduates. The unexplained component of the gender wage gap decreases over time among non-graduates. Among graduates instead the unexplained component is greater in the most recent cohort compared to the previous two.[11]

---

[11]When the samples are pooled across those with and without higher education (Tables B7 and B8 in Appendix B) the contribution of qualifications to explaining the gender gap is apparent in the 1946 cohort. The small contribution of human capital to explaining the gender wage gap reaches 7.5 log points in this cohort (in the selection adjusted model) with the higher qualifications of men accounting for 2.3 point of the gap. In the later cohorts the explained gap remains negligible, as does the estimate for qualifications, reflecting the convergence of educational attainments. The contribution of qualifications to the overall pay

Figure 5: Decomposition of the mean gender wage gap for graduates and non-graduates by cohort: early life factors, human capital, marital and parental status

(a) KOB decomposition without selection adjustment

Graduates Non-graduates

(b) KOB decomposition with selection adjustement

Graduates Non-graduates

■ Unexplained ■ Early life ■ Qualifications ■ Experience ■ Family
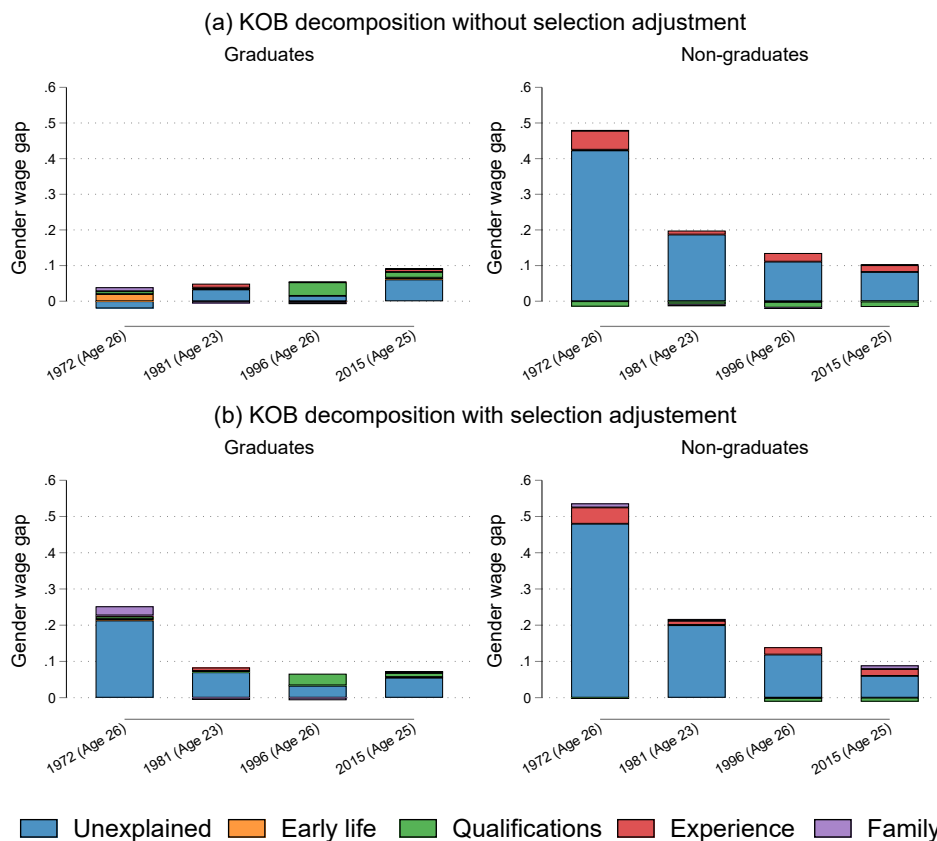
Note: The figure reports the Kitagawa-Oaxaca-Blinder decomposition of the log wage gap among graduates and non-graduates without selection adjustment (top panel) and with selection adjustment (bottom panel). The full decomposition with confidence intervals for each cohort is reported in tables B5 and B6 in Appendix B. The variables used in the decomposition are those reported in tables A1 to A4 in Appendix A.

4.4 Decomposition across the wage distribution

Previous research has highlighted barriers in the labour market that affect women at different points of the wage distribution: women may be prevented from escaping low-paid jobs or progressing to the highest paid positions. These barriers are conventionally referred to as sticky floor and glass ceiling and they may change differentially over time as shown by Blau and Kahn (2017) for the US. Our study of the gender wage gap at the mean among

───────────────────────

gap is sometimes negative, depending on the cohort and model estimated, but it was not always so.

young adults has shown that the convergence of wages for non-graduates has been consistently faster than the one for graduates. In order to understand whether the gender wage gap evolved differently among individuals with lower and higher wages we now investigate the gender wage gap across the wage distribution.

Figure 6 reports the decomposition across the wage distribution for the sample of observed wages (black lines) and for that of imputed as well as observed wages (light grey lines) using the method proposed by Chernozhukov *et al.* (2013) . The variables used in the decomposition are those in the KOB decomposition reported in Figure 5. The plot presents the results of the decomposition between the 1st and 9th decile of the unconditional distribution of wages for each cohort study.

The solid lines show the total difference between the unconditional distributions of log hourly wages of men and women at every decile. The y-axis shows the wage differential in favor of men expressed in log points. The long- dashed line represents the portion of the total gap explained by observed characteristics of the individuals in the sample. The short-dotted line shows the residual portion of the gap that is not explained by differences in observed characteristics: it captures the additional log hourly wages women would receive if their attributes were rewarded in the same way as men would be for those same attributes.

As observed in the decomposition of the mean gender wage gap, male/female differences in observed characteristics explain very little at every decile for the three later cohorts. For the NSHD cohort, described in the first plot on the top left, the raw gender wage gap declines as we move up the wage distribution (as indicated by the downward sloping black line), that is it is highest for the low paid at the 10th percentile. The decline is mostly accounted for by gender differences in the returns while differences in characteristics explain a small component of the gap for low and high wages. Allowing for non-random selection increases the gap for every decile.

For the NCDS cohort, the plot at the top right shows the raw gender wage gap declining only marginally after the 8th decile. The gender wage gap is wholly accounted for by differences in returns to characteristics. Selection adjustment increases the gap among lower wages however does not increase the explained component.

Figure 6: Decomposition across the wage distribution by cohort: early life factors, human capital, marital and parental status
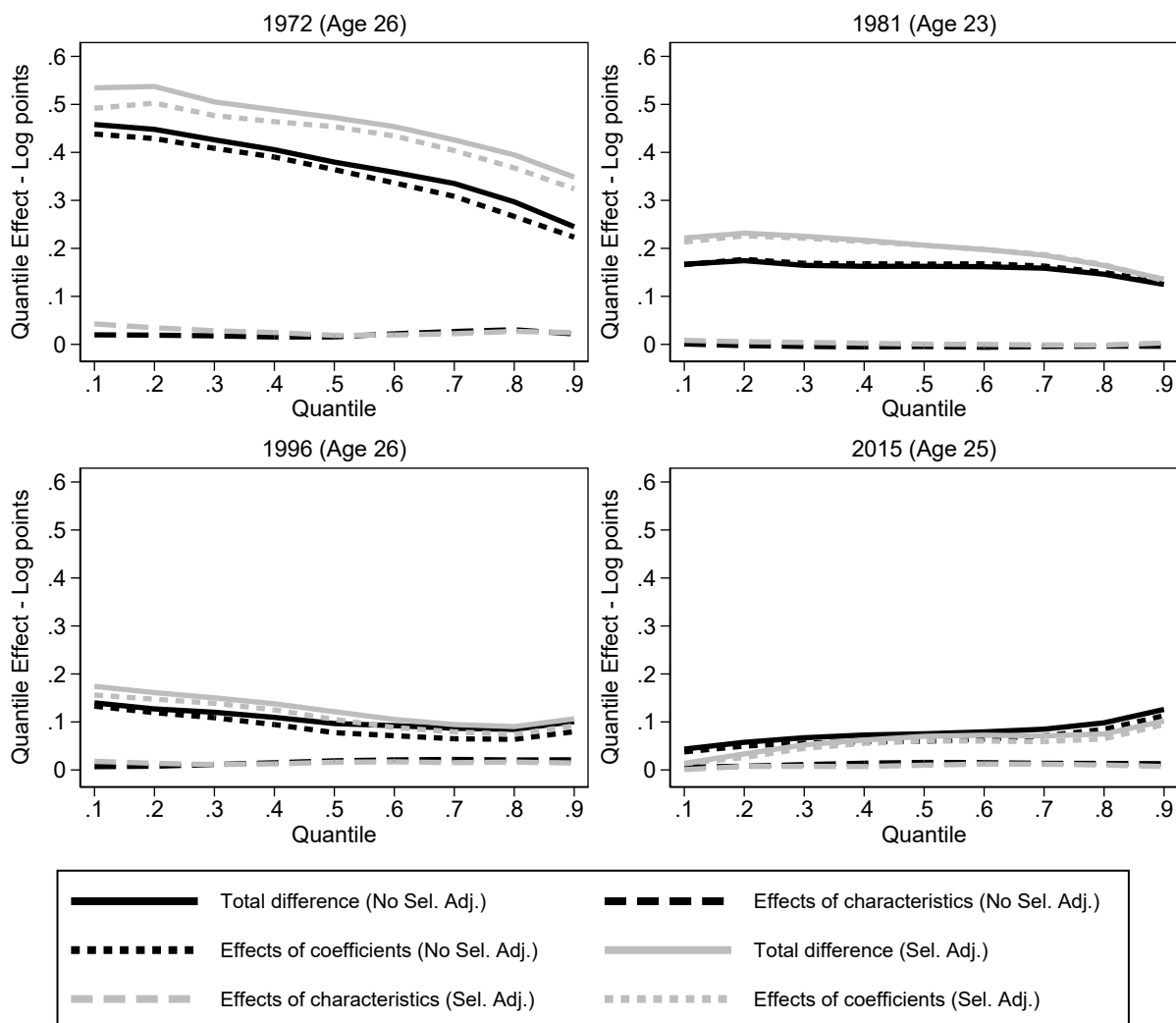
Note: The four plots report the decomposition between the 1st and 9th decile of the unconditional difference between the male and the female log wage distribution for the sample of observed wages (black lines) and for that of imputed and observed wages (grey lines) for each cohort study. The variables used in the decomposition are those included in the KOB decomposition reported in Figure 5. The solid lines show the total difference between the unconditional distributions of log hourly wages of men and women at every decile. The y-axis shows the wage differential in favor of men expressed in log points. The long-dashed line represents the portion of the total gap explained by observed characteristics of the individuals in the sample. The short-dotted line shows the residual portion of the gap that is not explained by differences in observed characteristics.

The decomposition of the gender wage gap for the BCS cohort is shown in the bottom left plot. The raw gender wage gap is smaller than in the NCDS, across the whole of the wage distribution. As in the NCDS cohort, the gender wage gap declines as we move up the wage distribution, although it turns up at around the 8th decile. Differences in human capital characteristics contribute only to a very small portion of the gap. Once again, nearly all of the gap is unexplained by the differences in the attributes of men and women. Instead, it is the gendered returns to those attributes that account for the gap. Accounting for selection does little to the gap across the distribution of wages.

The plot at the bottom right reports the results for Next Steps. The covariate and selection-adjusted gender wage gap is lower than for the BCS across the wage distribution but, in contrast to NSHD, NCDS and BCS, the gap rises as one moves up the wage distribution, in line with our findings about graduates, who are likely to be among the highest earners. The closing of the gender gap at low wages is plausibly attributable to the introduction of the National Minimum Wage in 1999, from which low paid women were the beneficiaries (Stewart and Swaffield, 2002; Connolly and Gregory, 2002). Again, most of the gap is unaccounted for by differences in men's and women's characteristics. Accounting for non-random selection does not increase the gap at any point of the wage distribution for this most recent cohort.

In summary, the analysis of the gender inequalities across the wage distribution has some important findings. First, as already observed in the decomposition at the mean, differences in characteristics account for little if any of the gender wage gap across the wage distribution in all cohorts. Second, accounting for selection into the labour market increases the total gender gap across the wage distribution in the first two cohorts. Third, gender inequalities in wages for lower paid jobs decrease over time. Finally, the gap at the top of the wage distribution decreases dramatically only between 1972 and 1981 while in the most recent cohort it is higher than for low paid jobs.

4.5 Decomposition of the gender wage gap with job characteristics

So far our results show that little of the wage gap is explained by differences in human capital and other personal characteristics/family formation. However, as discussed earlier, a growing literature points to the importance of job characteristics in explaining the gap. Detailed

occupations are available and comparable for the BCS and Next Steps cohort members. Results of the decompositions that include occupational characteristics for these two cohorts are reported in Figure 7. To reiterate, these decompositions do not account for selection into employment as we do not have information about occupations for individuals who are not in the labour market.

Estimates of the wage equation that includes occupational characteristics are reported in table B9 in Appendix B while results for the KOB decomposition with standard errors are reported in table B10.

Figure 7: Decomposition of the mean gender wage gap: early life factors, human capital, family formation and job characteristics
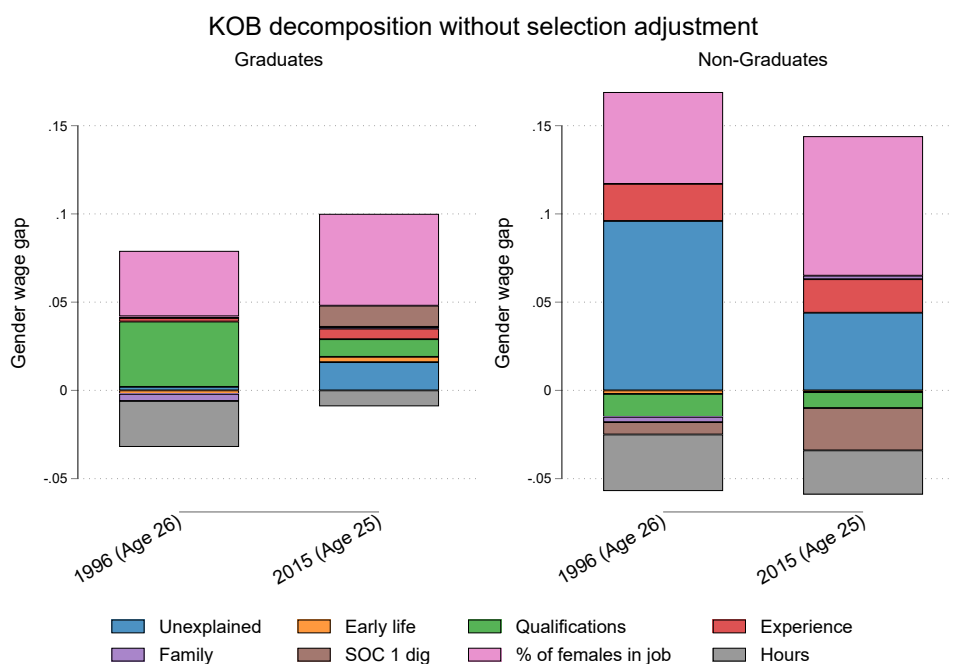
Note: The stacked bar charts report, by cohort, the KOB decomposition of the log wage gap among graduates (left panel) and non-graduates (right panel) based on observed wages. Estimates of the coefficients for job characteristics are reported in table B9 in Appendix B.

The decomposition shows gender differences in hours worked reduce the gender wage gap in the BCS cohort among graduates, and in both cohorts among non-graduates. Table B9 in Appendix B shows that this result, ceteris paribus, stems from the negative estimated

returns to hours worked, conditional on type of occupation and occupational gender segregation.[12] Returns to longer hours become less negative among graduates over time while they become more negative among non-graduates. At the same time the proportion of young women working long hours increases, particularly among graduates. Negative returns to hours worked are not surprising: Denning *et al.* (2022) find that wage returns to mean hours worked across occupations are positive while wage returns are very small or negative at individual level within occupations. Among non-graduates we observe that type of occupation and qualifications represent further negative components of the gap across both cohorts. This implies that women are in broad types of occupations and have qualifications with higher returns compared to men.

Differences in the proportion of females in the job nationally, ceteris paribus, explains a sizeable and growing component of the gender wage gap across both cohorts. This result implies that women (and men) employed in more female-dominated occupations are worse paid than other workers. Descriptive statistics in tables A3 to A4 in Appendix A and estimates of the wage equation in table B9 in Appendix B show that young women tend more than young men to work in female-dominated occupations while the returns to the proportion of women in the occupation becomes more negative over time.

One main takeaway of this analysis is that differences in the nature of jobs undertaken by men and women play an important role in explaining the otherwise unexplained part of the gender wage gap, but different aspects of those jobs have offsetting effects. Over time however, occupational segregation accounts for a rising proportion of the total gap for both graduates and non-graduates.

Turning to the full distribution of wages, the top two plots in Figure 8 report the overall raw gap and the explained and unexplained component by deciles when type of occupation and proportion of females in the occupation at national level are added to the initial set of individual characteristics in the decomposition proposed by Chernozhukov *et al.* (2013). The additional two plots at the bottom of Figure 8 include additionally hours worked. By comparing Figure 8 and Figure 6 we can recover what component of the gap across the wage distribution is attributable to differences in job characteristics. The comparison shows that over time more of the gap across the wage distribution is explained by occupational

---

[12]The excluded category for hours worked in the regression is 'less than 30 hours'.

24

Figure 8: Decomposition across the wage distribution - No selection adjustment

(a) Early life factors, human capital characteristics, marital and parental status, indicators for major group of occupations, proportion of female in the occupation
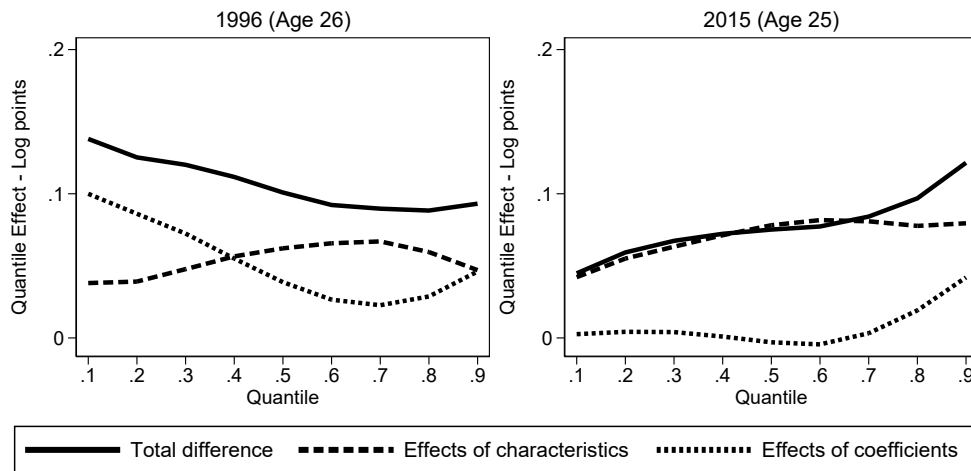
(b) Early life factors, human capital characteristics, marital and parental status, indicators for major group of occupations, proportion of female in the occupation and weekly hours worked
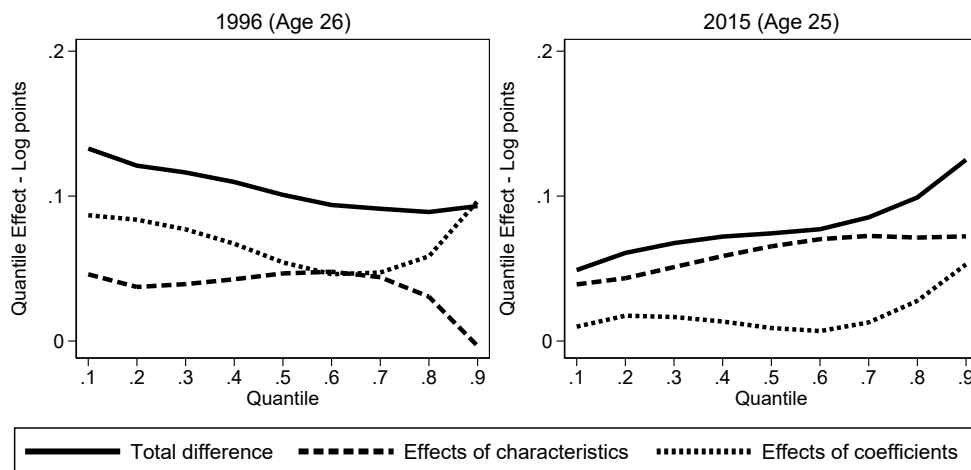
Note: The two plots in the top panel report the decomposition between the 1st and 9th decile of the unconditional difference between the male and the female log wage distribution for the sample of observed wages in 1996 and 2015. The variables used in the decomposition include early years characteristics, qualifications, experience, parental and marital status, indicators for broad types of occupations (major groups in Standard Occupational Classification 2000 and 2010) and proportion of women at national level in detailed occupations (4 digits Standard Occupational Classification 2000 and 2010). The two figures in the bottom panel report the same decomposition where the controls additionally include hours worked.

segregation, while hours worked reduces the explained component of the gap in both cohorts.

A visual inspection of the decomposition in the 1970 cohort (left plots) shows that differences in job characteristics account for less than half of the gender wage gap until the 7th decile. Between the 7th and 9th decile their role in explaining the gap is reduced and coefficients for individual and job attributes explain most of the gap. The decomposition in Next Steps is reported in the right plot: here differences in job characteristics explain most of the gap up to the 6th decile. After the 6th decile of the distribution of wages differences in coefficients of job characteristics begin to account for a small but rising component of the gap while the greatest part of the gender wage gap is still explained by differences in job characteristics, in particular by occupational segregation. These results are in line with findings from Levanon *et al.* (2009) and Blau and Kahn (2017) showing that female-dominated occupations pay less than male occupations for similar workers with the consequence that, over time, occupational segregation has become the main factor accounting for gender disparities in earnings as women caught up with men in terms of qualifications and experience.

5   Discussion and conclusion

We argue that to better understand the convergence of men's and women's earnings it is important to study the gender wage gap at the beginning of the working life and how it evolved across cohorts through socio-demographic changes and the implementation of new policies.

This paper documents the gender wage gap among young adults across four cohorts born between 1946 and 1990. It compares the gender wage gap among graduates and non-graduates over time, and across the distribution of wages. It explores the role of changing women's selection into employment over time in explaining the convergence of women's and men's earnings. Lastly it presents the relative contribution of early life factors, human capital, family formation and job characteristics in accounting for the gender wage gap among young adults over two decades.

The main results show that the overall raw and covariate adjusted gap, as expected, narrows over time. What was not well known for Britain is that this shift is driven by convergence in non-graduate men's and women's wages. The gender wage gap among graduates

increases between 1972 and 2015. This first evidence seems to confirm that graduate young women starting when their participation in higher education was particularly low and positive selection into employment was high, fared increasingly worse than young graduate men over time. Unlike the graduates, non-graduates benefitted considerably the equal opportunity measures of the 1970s and then from the National Minimum Wage legislation. In the last cohort we observe this reduced gender wage inequalities for very low paid jobs, as also shown by Amadxarif *et al.* (2020).

We find that non-random selection into the labour market affects relatively more graduate women than men in the earliest cohort, where they were least numerous. It also affects marginally the estimates of the gender wage gap for the cohort born in 1958 while it does not significantly affect the estimates in other cohorts.

In line with previous literature, the findings from the decomposition of the gender wage gap show that differences in traditional human capital characteristics and marital and parental status do not explain any significant component of the gap among young workers or potential workers- except perhaps workers born in 1946. It is perhaps too early in their life course for these factors to make much difference. The wage gap decreases over historical time and is mainly driven by a decrease in the unexplained component.

By decomposing the gap across the wage distribution, we observe that gender inequalities are greater for those with the lowest wages in the first two cohorts. The opposite happens in the latest cohort where differences in wages are smallest among the lowest paid while the gap widens for higher wages. This last result is consistent with the trend in the mean log gender wage gap among graduates, who are likely the highest earners in the samples, and non-graduates.

The proportion of women in the occupation accounts for a sizeable part of the gender wage gap among young adults in the two most recent cohorts, and it accounts for most of the gap for graduates in the later cohort. This result suggests that women, and in particular high-skilled women, concentrate in occupations that are likely to have lower pay than more mixed occupations or occupations mainly employing men.

Overall, these findings show that young non-graduate females' and males' wages have

been converging at a faster rate than the pay of graduate men and women. Changing selection of women into higher education may have contributed to these different rates of convergence. Occupational segregation is the main contributor to the lingering wage gap among young adults in recent years, as the unexplained gap has fallen faster in the less segregated occupations. This result is in line with Blau and Kahn (2017) who show that occupational segregation explains an increasing proportion of the gender wage gap over time in the US as women catch up with men in terms of human capital. Occupational segregation could be driven by women who perceive the greater risk of sexual harassment in mixed or predominantly male workplaces and therefore choose female segregated occupations (Folke and Rickne, 2022; Batut *et al.*, 2021). As such, harassment should be tackled by employers with more determined efforts to change workplace culture. Occupational segregation could also be the result of women making pre-market choices that have lower returns: they may choose jobs that are more flexible and with lower wages, even before forming a family, with the expectation of needing less rigid working patterns (see for instance Chevalier, 2007; Wiswall and Zafar, 2018). In addition, the post-COVID-19 world is likely to introduce changes towards flexibility and teleworking for a broader set of occupations. Future research will have to establish whether these changes are likely to favour women and increase their number in jobs where they are less represented.

References

ALBANESI, S., OLIVETTI, C. and PETRONGOLO, B. (2022). *Families, Labor Markets, and Policy*. Tech. rep., National Bureau of Economic Research.

ALBRECHT, J., BJÖRKLUND, A. and VROMAN, S. (2003). Is there a glass ceiling in sweden? *Journal of Labor economics*, **21** (1), 145–177.

AMADXARIF, Z., ANGELI, M., HALDANE, A. and ZEMAITYTE, G. (2020). Understanding pay gaps.

ARULAMPALAM, W., BOOTH, A. L. and BRYAN, M. L. (2007). Is there a glass ceiling over europe? exploring the gender pay gap across the wage distribution. *Ilr Review*, **60** (2), 163–186.

AZMAT, G. and FERRER, R. (2017). Gender gaps in performance: Evidence from young lawyers. *Journal of Political Economy*, **125** (5), 1306–1355.

BATUT, C., COLY, C., SCHNEIDER-STRAWCZYNSKI, S. *et al.* (2021). *It's a man's world: culture of abuse,# MeToo and worker flows*. Università Bocconi, The Dondena Centre.

BAYARD, K., HELLERSTEIN, J., NEUMARK, D. and TROSKE, K. (2003). New evidence on sex segregation and sex differences in wages from matched employee-employer data. *Journal of labor Economics*, **21** (4), 887–922.

BENNY, L., BHALOTRA, S. R. and FERNÁNDEZ, M. (2021). *Occupation flexibility and the graduate gender wage gap in the UK*. Tech. rep., ISER Working Paper Series.

BERTRAND, M., GOLDIN, C. and KATZ, L. F. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American economic journal: applied economics*, **2** (3), 228–255.

BLAU, F. D. and 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*, pp. 436–455.

BLUNDELL, R., GOSLING, A., ICHIMURA, H. and MEGHIR, C. (2007). Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica*, **75** (2), 323–363.

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*, **36** (4), 836–854.

CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and MELLY, B. (2013). Inference on counterfactual distributions. *Econometrica*, **81** (6), 2205–2268.

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.

COMBET, B. and OESCH, D. (2019). The gender wage gap opens long before motherhood. panel evidence on early careers in switzerland. *European sociological review*, **35** (3), 332–345.

CONNOLLY, S. and GREGORY, M. (2002). The national minimum wage and hours of work: implications for low paid women. *Oxford Bulletin of Economics and Statistics*, **64**, 607–631.

CORTES, P. and PAN, J. (2018). Occupation and gender. *The Oxford handbook of women and the economy*, pp. 425–452.

COSTA DIAS, M., JOYCE, R. and PARODI, F. (2020). The gender pay gap in the uk: children and experience in work. *Oxford Review of Economic Policy*, **36** (4), 855–881.

DENNING, J. T., JACOB, B. A., LEFGREN, L. J. and VOM LEHN, C. (2022). The return to hours worked within and across occupations: Implications for the gender wage gap. *ILR Review*, **75** (5), 1321–1347.

DOLTON, P., O'NEILL, D. and 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*, pp. 549–565.

FERRI, E., BYNNER, J. and WADSWORTH, M. (2003). *Changing Britain, changing lives*. Institute of Education Press.

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. (2008). The gender wage gap among young adults in the united states the importance of money versus people. *Journal of Human Resources*, **43** (4), 884–918.

GOLDIN, C. (2021). Career and family. In *Career and Family*, Princeton University Press.

— and KATZ, L. F. (2008). Transitions: Career and family life cycles of the educational elite. *American Economic Review*, **98** (2), 363–369.

GROSHEN, E. L. (1991). The structure of the female/male wage differential: Is it who you are, what you do, or where you work? *Journal of Human Resources*, pp. 457–472.

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.

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 ZWEIMÜLLER, J. (2019). Child penalties across countries: Evidence and explanations. In *AEA Papers and Proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, vol. 109, pp. 122–126.

KUNZE, A. (2018). The gender wage gap in developed countries. *The Oxford handbook of women and the economy*, pp. 369–394.

LEVANON, A., ENGLAND, P. and ALLISON, P. (2009). Occupational feminization and pay: Assessing causal dynamics using 1950–2000 us census data. *Social forces*, **88** (2), 865–891.

MANNING, A. and SWAFFIELD, J. (2008). The gender gap in early-career wage growth. *The Economic Journal*, **118** (530), 983–1024.

MULLIGAN, C. B. and RUBINSTEIN, Y. (2008). Selection, investment, and women's relative wages over time. *The Quarterly Journal of Economics*, **123** (3), 1061–1110.

NEUBURGER, J. (2010). *Trends in the unequal pay of women and men across three British generations*. Ph.D. thesis, Institute of Education, University of London.

—, KUH, D. and JOSHI, H. (2011). Cross-cohort changes in gender pay differences in britain from 1972 to 2004: accounting for selection into employment using wage imputation. *Longitudinal and Life Course Studies*, **2** (3), 260–285.

NEUMARK, D. (2004). Employers' discriminatory behavior and the estimation of wage discrimination. In *Sex Differences in Labor Markets*, Routledge, pp. 163–177.

OAXACA, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, pp. 693–709.

STEWART, M. B. and SWAFFIELD, J. K. (2002). Using the bhps wave 9 additional questions to evaluate the impact of the national minimum wage. *Oxford Bulletin of Economics and Statistics*, **64**, 633–652.

UNIVERSITY COLLEGE LONDON (a). 1970 british cohort study [data collection]. UK Data Service, SN: 2000001 https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=200001.

UNIVERSITY COLLEGE LONDON (b). National child development study [data collection]. UK Data Service, SN: 2000032, https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=2000032.

UNIVERSITY COLLEGE LONDON (c). Next steps (previously the longitudinal study of young people in england (lsype1)). UK Data Service, SN: 2000030 https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=2000030.

WISWALL, M. and ZAFAR, B. (2018). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics*, **133** (1), 457–507.

A Descriptive statistics and list of variables

Table A1: Descriptive statistics by higher education status of variables used in the decomposition (sample of cohort members with observed wages and cohort members with potential wages) - NSHD

| | With selection adjustment | | | | Without selection adjustment | | | |
| | Graduates | | Non-Graduates | | Graduates | | Non-graduates | |
	Males	Females	Males	Females	Males	Females	Males	Females
Log real hourly earnings	1.96	1.71	1.87	1.34	1.99	1.97	1.86	1.40
	(0.28)	(0.41)	(0.28)	(0.33)	(0.26)	(0.26)	(0.28)	(0.31)
Father education Primary	0.57	0.45	0.78	0.81	0.57	0.39	0.79	0.81
	(0.50)	(0.50)	(0.41)	(0.39)	(0.50)	(0.49)	(0.41)	(0.39)
Father education Primary and Tech/comm. Diploma	0.05	0.07	0.03	0.03	0.06	0.07	0.03	0.03
	(0.23)	(0.26)	(0.16)	(0.16)	(0.24)	(0.26)	(0.16)	(0.17)
Father education Secondary	0.20	0.24	0.10	0.10	0.20	0.29	0.10	0.10
	(0.40)	(0.43)	(0.30)	(0.31)	(0.40)	(0.45)	(0.30)	(0.29)
Father education Secondary and tech/Comm Diploma	0.12	0.19	0.03	0.03	0.11	0.20	0.03	0.03
	(0.32)	(0.40)	(0.17)	(0.16)	(0.31)	(0.40)	(0.17)	(0.18)
Father education Missing	0.06	0.05	0.06	0.03	0.06	0.05	0.05	0.03
	(0.25)	(0.21)	(0.23)	(0.18)	(0.23)	(0.22)	(0.22)	(0.17)
Mother education Primary only	0.62	0.56	0.83	0.86	0.63	0.52	0.84	0.84
	(0.49)	(0.50)	(0.38)	(0.34)	(0.48)	(0.50)	(0.37)	(0.36)
Mother education Primary and Tech/comm. Diploma	0.03	0.07	0.03	0.03	0.03	0.07	0.02	0.04
	(0.18)	(0.26)	(0.16)	(0.16)	(0.17)	(0.26)	(0.15)	(0.19)
Mother education Secondary only	0.21	0.18	0.08	0.06	0.20	0.21	0.08	0.06
	(0.41)	(0.39)	(0.27)	(0.23)	(0.40)	(0.41)	(0.27)	(0.25)
Mother education Secondary and tech/Comm Diploma	0.08	0.14	0.02	0.02	0.09	0.15	0.02	0.02
	(0.27)	(0.34)	(0.13)	(0.14)	(0.28)	(0.36)	(0.13)	(0.14)
Mother education Missing	0.05	0.04	0.05	0.03	0.05	0.04	0.05	0.03
	(0.23)	(0.21)	(0.22)	(0.18)	(0.21)	(0.21)	(0.21)	(0.18)
Father soc. Class Professional	0.10	0.12	0.03	0.02	0.10	0.17	0.03	0.02
	(0.30)	(0.33)	(0.17)	(0.14)	(0.30)	(0.37)	(0.17)	(0.14)
Father soc. Class Intermedite	0.27	0.34	0.13	0.14	0.28	0.40	0.11	0.12
	(0.44)	(0.47)	(0.34)	(0.34)	(0.45)	(0.49)	(0.32)	(0.33)
Father soc. Class Skilled (non-Manual)	0.14	0.19	0.07	0.09	0.14	0.15	0.07	0.10
	(0.35)	(0.40)	(0.26)	(0.29)	(0.35)	(0.36)	(0.26)	(0.29)
Father soc. Class Skilled (Manual)	0.36	0.26	0.42	0.48	0.34	0.20	0.42	0.50
	(0.48)	(0.44)	(0.49)	(0.50)	(0.48)	(0.40)	(0.49)	(0.50)
Father soc. Class 5 Partly skilled	0.10	0.08	0.24	0.18	0.11	0.07	0.25	0.16
	(0.30)	(0.28)	(0.43)	(0.38)	(0.31)	(0.25)	(0.43)	(0.37)
Father soc. Class Unskilled	0.02		0.10	0.09	0.02		0.10	0.09
	(0.15)		(0.30)	(0.28)	(0.15)		(0.30)	(0.28)
								Continued on next page

	With selection adjustment				Without selection adjustment			
	Graduates		Non Graduates		Graduates		Non graduates	
	Males	Females	Males	Females	Males	Females	Males	Females
Father soc. Class Missing	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	(0.10)	(0.08)	(0.11)	(0.09)	(0.10)	(0.10)	(0.10)	(0.09)
Std. maths score	0.56	0.78	-0.20	-0.10	0.56	0.72	-0.20	-0.01
	(0.69)	(0.58)	(1.16)	(1.05)	(0.65)	(0.69)	(1.16)	(1.00)
Std. reading score	0.78	1.16	-0.17	-0.08	0.80	1.11	-0.16	0.03
	(0.93)	(0.86)	(1.07)	(1.01)	(0.96)	(0.84)	(1.06)	(0.99)
Std. maths score missing	0.08	0.05	0.09	0.08	0.09	0.06	0.08	0.08
	(0.27)	(0.21)	(0.28)	(0.27)	(0.28)	(0.25)	(0.28)	(0.27)
Std. reading score missing	0.06	0.03	0.05	0.04	0.07	0.04	0.04	0.04
	(0.25)	(0.17)	(0.21)	(0.21)	(0.25)	(0.19)	(0.20)	(0.19)
No qualifications			0.52	0.47			0.52	0.38
			(0.50)	(0.50)			(0.50)	(0.49)
Vocational qualification			0.07	0.12			0.08	0.12
			(0.26)	(0.32)			(0.27)	(0.33)
GCE'O' Level or Burnham C			0.36	0.39			0.37	0.47
			(0.48)	(0.49)			(0.48)	(0.50)
Burnham A2	0.52	0.66			0.52	0.67		
	(0.50)	(0.47)			(0.50)	(0.47)		
First degree	0.48	0.34			0.48	0.33		
	(0.50)	(0.47)			(0.50)	(0.47)		
Qualification missing			0.04	0.03			0.03	0.03
			(0.20)	(0.16)			(0.18)	(0.17)
Full-time/part-time experience	51.69	36.14	80.26	59.94	51.43	36.60	81.20	73.27
	(28.55)	(17.22)	(10.03)	(51.90)	(27.49)	(16.23)	(7.93)	(57.87)
N. of different spells of work	6.33	4.68	8.08	9.94	5.07	3.18	7.90	10.94
	(16.62)	(14.22)	(15.86)	(49.72)	(13.30)	(8.97)	(16.16)	(60.10)
Any child	0.24	0.36	0.51	0.68	0.21	0.05	0.50	0.33
	(0.43)	(0.48)	(0.50)	(0.47)	(0.41)	(0.22)	(0.50)	(0.47)
More than one child	0.08	0.09	0.26	0.43	0.07	0.02	0.26	0.18
	(0.27)	(0.29)	(0.44)	(0.49)	(0.25)	(0.13)	(0.44)	(0.39)
N	**480**	**252**	**1399**	**1579**	**386**	**142**	**1114**	**601**
N weighted by stratification weights	**891**	**408**	**3451**	**3880**	**713**	**211**	**2767**	**1477**

Note: The table reports mean and standard deviation (in parentheses) of the variables used in the decomposition of the gender wage gap. Descriptive statistics are weighted to account for attrition and sample stratification.

Table A2: Descriptive statistics by higher education status of variables used in the decomposition (sample of cohort members with observed wages and cohort members with potential wages) - NCDS

	With selection adjustment				Without selection adjustment			
	Graduates		Non-Graduates		Graduates		Non-graduates	
	Males	Females	Males	Females	Males	Females	Males	Females
Log real hourly earnings	1.75	1.67	1.69	1.47	1.77	1.72	1.69	1.51

	With selection adjustment				Without selection adjustment			
	Graduates		Non Graduates		Graduates		Non graduates	
	Males	Females	Males	Females	Males	Females	Males	Females
	(0.29)	(0.30)	(0.29)	(0.30)	(0.28)	(0.28)	(0.28)	(0.28)
Father education (less than GCSE)	0.40	0.41	0.57	0.57	0.41	0.41	0.59	0.56
	(0.49)	(0.49)	(0.49)	(0.50)	(0.49)	(0.49)	(0.49)	(0.50)
Father education (GCSE or A-levels)	0.18	0.17	0.08	0.08	0.19	0.17	0.08	0.09
	(0.38)	(0.38)	(0.27)	(0.27)	(0.39)	(0.38)	(0.27)	(0.29)
Father education (Higher education)	0.15	0.14	0.04	0.04	0.15	0.13	0.04	0.05
	(0.35)	(0.35)	(0.20)	(0.20)	(0.35)	(0.34)	(0.18)	(0.21)
Father education (Missing)	0.27	0.28	0.31	0.31	0.26	0.29	0.30	0.30
	(0.44)	(0.45)	(0.46)	(0.46)	(0.44)	(0.45)	(0.46)	(0.46)
Mother education (less than GCSE)	0.44	0.43	0.58	0.58	0.45	0.44	0.60	0.57
	(0.50)	(0.50)	(0.49)	(0.49)	(0.50)	(0.50)	(0.49)	(0.50)
Mother education (GCSE or A-levels)	0.21	0.19	0.09	0.10	0.23	0.18	0.09	0.11
	(0.41)	(0.39)	(0.29)	(0.30)	(0.42)	(0.39)	(0.29)	(0.32)
Mother education (Higher education)	0.09	0.11	0.03	0.03	0.08	0.11	0.03	0.03
	(0.29)	(0.32)	(0.17)	(0.17)	(0.27)	(0.31)	(0.17)	(0.18)
Mother education (Missing)	0.26	0.26	0.29	0.29	0.24	0.27	0.28	0.28
	(0.44)	(0.44)	(0.46)	(0.45)	(0.43)	(0.44)	(0.45)	(0.45)
Father soc. Class 1	0.10	0.10	0.03	0.03	0.10	0.09	0.03	0.03
	(0.30)	(0.30)	(0.16)	(0.17)	(0.30)	(0.28)	(0.16)	(0.18)
Father soc. Class 2	0.22	0.25	0.10	0.10	0.21	0.26	0.09	0.11
	(0.42)	(0.43)	(0.30)	(0.30)	(0.41)	(0.44)	(0.29)	(0.32)
Father soc. Class 3	0.12	0.12	0.08	0.08	0.12	0.11	0.09	0.10
	(0.32)	(0.32)	(0.28)	(0.28)	(0.33)	(0.31)	(0.28)	(0.30)
Father soc. Class 4	0.37	0.38	0.47	0.47	0.38	0.39	0.49	0.46
	(0.48)	(0.48)	(0.50)	(0.50)	(0.49)	(0.49)	(0.50)	(0.50)
Father soc. Class 5	0.06	0.05	0.12	0.13	0.06	0.04	0.12	0.12
	(0.24)	(0.21)	(0.32)	(0.33)	(0.25)	(0.21)	(0.32)	(0.33)
Father soc. Class 6	0.03	0.03	0.10	0.09	0.03	0.04	0.09	0.07
	(0.16)	(0.18)	(0.30)	(0.28)	(0.16)	(0.19)	(0.29)	(0.26)
Father soc. Class missing	0.10	0.08	0.11	0.11	0.10	0.08	0.10	0.10
	(0.30)	(0.27)	(0.31)	(0.31)	(0.30)	(0.27)	(0.30)	(0.30)
GCSEs			0.23	0.31			0.25	0.34
			(0.42)	(0.46)			(0.43)	(0.48)
Alevels			0.18	0.14			0.20	0.18
			(0.39)	(0.35)			(0.40)	(0.39)
First degree	1.00	1.00			1.00	1.00		
Higher degree								
Other qualification			0.29	0.31			0.30	0.26
			(0.45)	(0.46)			(0.46)	(0.44)
Qualification missing			0.29	0.24			0.26	0.21
			(0.46)	(0.43)			(0.44)	(0.41)
STEM	0.33	0.11			0.32	0.12		
	(0.47)	(0.31)			(0.47)	(0.32)		
LEM	0.16	0.06			0.18	0.07		

	With selection adjustment				Without selection adjustment			
	Graduates		Non Graduates		Graduates		Non graduates	
	Males	Females	Males	Females	Males	Females	Males	Females
	(0.37)	(0.24)			(0.38)	(0.25)		
Other subjects	0.39	0.80			0.37	0.80		
	(0.49)	(0.40)			(0.48)	(0.40)		
Subject missing	0.12	0.03			0.13	0.02		
	(0.32)	(0.16)			(0.34)	(0.14)		
Std. maths score	0.77	0.61	-0.15	-0.14	0.74	0.61	-0.13	-0.01
	(0.84)	(0.84)	(0.90)	(0.87)	(0.85)	(0.83)	(0.88)	(0.86)
Std. reading score	0.69	0.60	-0.17	-0.12	0.67	0.58	-0.13	-0.00
	(0.81)	(0.78)	(0.94)	(0.86)	(0.82)	(0.76)	(0.92)	(0.82)
Std. maths score missing	0.13	0.13	0.13	0.13	0.12	0.13	0.13	0.12
	(0.33)	(0.33)	(0.34)	(0.34)	(0.33)	(0.34)	(0.33)	(0.33)
Std. reading score missing	0.13	0.13	0.13	0.13	0.12	0.13	0.13	0.12
	(0.33)	(0.33)	(0.34)	(0.34)	(0.33)	(0.34)	(0.33)	(0.33)
Full-time experience	39.92	35.32	65.59	56.13	43.69	36.17	70.52	64.56
	(30.82)	(26.26)	(26.08)	(27.57)	(30.07)	(25.76)	(22.37)	(24.58)
Part-time experience	0.69	2.88	0.73	3.54	0.59	2.56	0.78	3.44
	(4.54)	(10.27)	(5.73)	(11.79)	(4.15)	(9.93)	(5.93)	(12.07)
N. of spells of work	1.77	2.15	2.44	2.41	1.81	2.10	2.40	2.45
	(1.16)	(1.38)	(1.52)	(1.48)	(1.07)	(1.32)	(1.48)	(1.51)
Married/cohabiting	0.31	0.45	0.42	0.63	0.33	0.43	0.44	0.54
	(0.46)	(0.50)	(0.49)	(0.48)	(0.47)	(0.50)	(0.50)	(0.50)
Any child	0.06	0.09	0.19	0.35	0.06	0.03	0.18	0.10
	(0.23)	(0.28)	(0.39)	(0.48)	(0.23)	(0.18)	(0.39)	(0.31)
More than one child	0.01	0.01	0.06	0.14	0.01		0.06	0.03
	(0.11)	(0.11)	(0.24)	(0.34)	(0.11)	(0.05)	(0.23)	(0.17)
N	**798**	**901**	**5113**	**4909**	**597**	**694**	**3556**	**2802**

Note: The table reports mean and standard deviation (in parentheses) of the variables used in the decomposition of the gender wage gap. Descriptive statistics are weighted to account for attrition.

Table A3: Descriptive statistics by higher education status of variables used in the decomposition (sample of cohort members with observed wages and cohort members with potential wages) - BCS

	With selection adjustment				Without selection adjustment			
	Graduates		Non-Graduates		Graduates		Non-graduates	
	Males	Females	Males	Females	Males	Females	Males	Females
Log real hourly earnings	2.06	2.00	1.89	1.76	2.10	2.06	1.88	1.77
	(0.37)	(0.37)	(0.36)	(0.44)	(0.37)	(0.35)	(0.37)	(0.40)
Father education								
(less than GCSE)	0.32	0.31	0.65	0.63	0.32	0.30	0.65	0.62
	(0.47)	(0.46)	(0.48)	(0.48)	(0.47)	(0.46)	(0.48)	(0.49)
Father education								
(GCSE or A-levels)	0.23	0.27	0.16	0.17	0.25	0.28	0.17	0.19

	With selection adjustment				Without selection adjustment			
	Graduates		Non Graduates		Graduates		Non graduates	
	Males	Females	Males	Females	Males	Females	Males	Females
	(0.42)	(0.44)	(0.37)	(0.38)	(0.43)	(0.45)	(0.38)	(0.39)
Father education								
(Higher education)	0.31	0.30	0.07	0.08	0.30	0.30	0.07	0.08
	(0.46)	(0.46)	(0.26)	(0.27)	(0.46)	(0.46)	(0.25)	(0.28)
Father education (Missing)	0.14	0.12	0.11	0.12	0.14	0.12	0.11	0.11
	(0.35)	(0.32)	(0.32)	(0.32)	(0.34)	(0.32)	(0.31)	(0.31)
Mother education								
(less than GCSE)	0.31	0.34	0.68	0.67	0.32	0.34	0.67	0.65
	(0.46)	(0.47)	(0.47)	(0.47)	(0.47)	(0.47)	(0.47)	(0.48)
Mother education								
(GCSE or A-levels)	0.32	0.31	0.19	0.19	0.32	0.31	0.20	0.21
	(0.47)	(0.46)	(0.40)	(0.39)	(0.47)	(0.46)	(0.40)	(0.40)
Mother education								
(Higher education)	0.25	0.26	0.05	0.06	0.24	0.26	0.05	0.06
	(0.43)	(0.44)	(0.22)	(0.23)	(0.43)	(0.44)	(0.22)	(0.24)
Mother education (Missing)	0.12	0.09	0.08	0.08	0.12	0.09	0.08	0.08
	(0.33)	(0.29)	(0.27)	(0.27)	(0.33)	(0.29)	(0.27)	(0.27)
Father soc. Class 1	0.14	0.14	0.02	0.02	0.14	0.15	0.02	0.02
	(0.35)	(0.35)	(0.14)	(0.13)	(0.35)	(0.35)	(0.15)	(0.14)
Father soc. Class 2	0.18	0.21	0.08	0.08	0.18	0.21	0.08	0.08
	(0.38)	(0.41)	(0.28)	(0.28)	(0.39)	(0.41)	(0.27)	(0.28)
Father soc. Class 3	0.15	0.16	0.10	0.10	0.16	0.17	0.11	0.11
	(0.36)	(0.36)	(0.30)	(0.30)	(0.36)	(0.38)	(0.31)	(0.31)
Father soc. Class 4	0.30	0.28	0.44	0.46	0.30	0.28	0.44	0.47
	(0.46)	(0.45)	(0.50)	(0.50)	(0.46)	(0.45)	(0.50)	(0.50)
Father soc. Class 5	0.06	0.07	0.15	0.14	0.06	0.06	0.16	0.13
	(0.24)	(0.25)	(0.36)	(0.34)	(0.23)	(0.24)	(0.37)	(0.34)
Father soc. Class 6	0.02	0.01	0.07	0.06	0.01		0.06	0.05
	(0.12)	(0.09)	(0.25)	(0.23)	(0.12)	(0.07)	(0.23)	(0.21)
Father soc. Class missing	0.15	0.14	0.14	0.14	0.15	0.13	0.14	0.14
	(0.36)	(0.35)	(0.34)	(0.35)	(0.36)	(0.34)	(0.34)	(0.35)
GCSEs			0.47	0.53			0.49	0.56
			(0.50)	(0.50)			(0.50)	(0.50)
Alevels			0.11	0.13			0.12	0.15
			(0.31)	(0.34)			(0.33)	(0.36)
First degree	0.80	0.80			0.82	0.80		
	(0.40)	(0.40)			(0.39)	(0.40)		
Higher degree	0.20	0.20			0.18	0.20		
	(0.40)	(0.40)			(0.39)	(0.40)		
Other qualification			0.33	0.27			0.32	0.24
			(0.47)	(0.44)			(0.46)	(0.42)
Qualification missing			0.09	0.07			0.08	0.06
			(0.29)	(0.26)			(0.26)	(0.23)
STEM	0.29	0.14			0.31	0.15		
	(0.45)	(0.35)			(0.46)	(0.36)		
LEM	0.13	0.12			0.13	0.13		

	With selection adjustment				Without selection adjustment			
	Graduates		Non Graduates		Graduates		Non graduates	
	Males	Females	Males	Females	Males	Females	Males	Females
	(0.33)	(0.32)			(0.34)	(0.34)		
Other subject	0.34	0.57			0.31	0.55		
	(0.47)	(0.50)			(0.46)	(0.50)		
Subject missing	0.24	0.17			0.25	0.17		
	(0.43)	(0.38)			(0.43)	(0.37)		
Std. maths score	0.51	0.38	-0.21	-0.24	0.54	0.41	-0.14	-0.19
	(0.81)	(0.74)	(0.90)	(0.81)	(0.79)	(0.73)	(0.87)	(0.80)
Std. reading score	0.37	0.47	-0.31	-0.16	0.41	0.48	-0.26	-0.10
	(0.78)	(0.74)	(0.93)	(0.83)	(0.76)	(0.73)	(0.91)	(0.81)
Std. maths score missing	0.28	0.28	0.23	0.24	0.29	0.28	0.23	0.23
	(0.45)	(0.45)	(0.42)	(0.43)	(0.45)	(0.45)	(0.42)	(0.42)
Std. reading score missing	0.28	0.27	0.23	0.24	0.29	0.28	0.23	0.23
	(0.45)	(0.45)	(0.42)	(0.42)	(0.45)	(0.45)	(0.42)	(0.42)
Full-time experience	47.08	45.28	94.02	79.18	49.95	47.36	98.58	87.17
	(29.37)	(27.06)	(35.26)	(39.40)	(28.35)	(25.82)	(30.77)	(36.52)
Part-time experience	2.11	4.44	1.46	8.65	1.93	4.00	1.28	8.06
	(9.20)	(13.34)	(8.08)	(21.20)	(9.37)	(13.08)	(7.55)	(21.36)
N. of spells of work	2.05	2.20	2.36	2.39	2.11	2.22	2.42	2.47
	(1.30)	(1.36)	(1.54)	(1.55)	(1.27)	(1.33)	(1.54)	(1.55)
Married/cohabiting	0.37	0.50	0.54	0.65	0.39	0.52	0.57	0.65
	(0.48)	(0.50)	(0.50)	(0.48)	(0.49)	(0.50)	(0.49)	(0.48)
Any child	0.05	0.06	0.23	0.34	0.05	0.04	0.23	0.21
	(0.21)	(0.24)	(0.42)	(0.47)	(0.22)	(0.21)	(0.42)	(0.41)
More than one child	0.01	0.01	0.10	0.14	0.01	0.01	0.09	0.08
	(0.09)	(0.12)	(0.30)	(0.35)	(0.10)	(0.08)	(0.29)	(0.27)
Hours<30					0.03	0.10	0.02	0.19
					(0.16)	(0.30)	(0.13)	(0.39)
Hours>=30 & <45					0.71	0.76	0.62	0.74
					(0.45)	(0.43)	(0.49)	(0.44)
Hours>=45					0.26	0.14	0.36	0.08
					(0.44)	(0.35)	(0.48)	(0.27)
Proportion of women in the occupation					0.32	0.58	0.26	0.68
					(0.24)	(0.26)	(0.26)	(0.25)
Major group occupation 1 (Managers and senior officials)					0.20	0.17	0.14	0.15
					(0.40)	(0.38)	(0.34)	(0.35)
Major group occupation 2 (Professional occupations)					0.30	0.33	0.04	0.02
					(0.46)	(0.47)	(0.19)	(0.14)
Major group occupation 3 (Associate professional)					0.21	0.19	0.08	0.08
					(0.41)	(0.40)	(0.27)	(0.27)
Major group occupation 4 (Administrative)					0.09	0.14	0.10	0.34
					(0.28)	(0.35)	(0.30)	(0.47)

	With selection adjustment				Without selection adjustment			
	Graduates		Non Graduates		Graduates		Non graduates	
	Males	Females	Males	Females	Males	Females	Males	Females
Major group occupation 5 (Skilled trades)					0.04	0.01	0.21	0.04
					(0.19)	(0.09)	(0.41)	(0.19)
Major group occupation 6 (Personal service)					0.02	0.05	0.11	0.14
					(0.16)	(0.22)	(0.32)	(0.35)
Major group occupation 7 (Sales and customer service)					0.03	0.03	0.04	0.09
					(0.18)	(0.17)	(0.20)	(0.29)
Major group occupation 8 (Process, plant and machine operatives)					0.02		0.16	0.05
					(0.13)	(0.06)	(0.37)	(0.21)
Major group occupation 9 (Elementary occupations)					0.01		0.06	0.03
					(0.11)	(0.07)	(0.23)	(0.18)
Major group occupation missing					0.08	0.06	0.06	0.06
					(0.27)	(0.24)	(0.24)	(0.24)
N	963	1021	2758	3320	754	814	1963	2329

Note: The table reports mean and standard deviation (in parentheses) of the variables used in the decomposition of the gender wage gap. Descriptive statistics are weighted to account for attrition.

Table A4: Descriptive statistics by higher education status of variables used in the decomposition (sample of cohort members with observed wages and cohort members with potential wages) - Next Steps

	With selection adjustment				Without selection adjustment			
	Graduates		Non-Graduates		Graduates		Non-graduates	
	Males	Females	Males	Females	Males	Females	Males	Females
Log real hourly earnings	1.97	1.88	1.81	1.74	2.00	1.91	1.85	1.76
	(0.39)	(0.36)	(0.38)	(0.37)	(0.38)	(0.35)	(0.37)	(0.35)
Std. maths score	0.45	0.21	-0.14	-0.27	0.50	0.28	-0.06	-0.15
	(0.92)	(0.93)	(1.03)	(0.97)	(0.86)	(0.90)	(1.03)	(0.94)
Std. reading score	0.23	0.42	-0.40	-0.08	0.26	0.45	-0.32	0.04
	(0.87)	(0.87)	(1.03)	(1.00)	(0.83)	(0.83)	(1.01)	(0.96)
Std. maths score missing	0.09	0.08	0.08	0.08	0.08	0.07	0.06	0.07
	(0.28)	(0.27)	(0.28)	(0.28)	(0.28)	(0.25)	(0.24)	(0.25)
Std. reading score missing	0.09	0.07	0.10	0.09	0.08	0.07	0.07	0.07
	(0.28)	(0.26)	(0.30)	(0.28)	(0.28)	(0.25)	(0.26)	(0.25)
Full-time experience (months)	34.64	33.10	56.09	42.65	38.44	36.39	63.18	51.03
	(25.40)	(23.13)	(35.87)	(34.51)	(24.36)	(22.08)	(32.62)	(33.64)
Part-time experience (months)	7.15	10.12	8.38	15.31	7.55	10.50	8.56	16.83

	With selection adjustment				Without selection adjustment			
	Graduates		Non Graduates		Graduates		Non graduates	
	Males	Females	Males	Females	Males	Females	Males	Females
	(17.39)	(20.76)	(19.85)	(26.57)	(18.30)	(21.51)	(20.20)	(28.46)
N. of spells of work	1.92	2.11	2.07	2.03	2.10	2.29	2.25	2.34
	(1.23)	(1.30)	(1.36)	(1.39)	(1.21)	(1.25)	(1.35)	(1.36)
Father soc. Class 1 and 2	0.21	0.16	0.09	0.10	0.20	0.17	0.09	0.11
	(0.41)	(0.36)	(0.28)	(0.30)	(0.40)	(0.37)	(0.29)	(0.31)
Father soc. Class 3 and 3	0.21	0.22	0.16	0.15	0.23	0.22	0.17	0.17
	(0.41)	(0.41)	(0.37)	(0.36)	(0.42)	(0.42)	(0.38)	(0.38)
Father soc. Class 5	0.03	0.04	0.03	0.03	0.04	0.04	0.04	0.03
	(0.18)	(0.19)	(0.17)	(0.17)	(0.19)	(0.19)	(0.19)	(0.17)
Father soc. Class 6 and 7	0.14	0.12	0.14	0.12	0.13	0.12	0.14	0.13
	(0.35)	(0.32)	(0.35)	(0.33)	(0.34)	(0.33)	(0.34)	(0.34)
Father soc. Class 5	0.09	0.09	0.13	0.11	0.09	0.10	0.14	0.11
	(0.28)	(0.29)	(0.34)	(0.32)	(0.29)	(0.30)	(0.35)	(0.32)
Mother: higher education	0.33	0.29	0.17	0.16	0.33	0.29	0.18	0.18
	(0.47)	(0.45)	(0.38)	(0.37)	(0.47)	(0.45)	(0.39)	(0.38)
Mother: A level/GCSE	0.34	0.37	0.39	0.40	0.35	0.39	0.41	0.42
	(0.47)	(0.48)	(0.49)	(0.49)	(0.48)	(0.49)	(0.49)	(0.49)
Mother: other qual.	0.25	0.26	0.34	0.36	0.23	0.23	0.32	0.32
	(0.43)	(0.44)	(0.48)	(0.48)	(0.42)	(0.42)	(0.47)	(0.47)
Mother: missing education	0.09	0.09	0.09	0.08	0.09	0.09	0.09	0.08
	(0.28)	(0.28)	(0.29)	(0.27)	(0.28)	(0.28)	(0.28)	(0.27)
Father: higher education	0.28	0.24	0.13	0.14	0.29	0.24	0.14	0.15
	(0.45)	(0.43)	(0.34)	(0.35)	(0.45)	(0.43)	(0.35)	(0.36)
Father: A level/GCSE	0.28	0.28	0.31	0.29	0.29	0.29	0.32	0.31
	(0.45)	(0.45)	(0.46)	(0.45)	(0.45)	(0.45)	(0.47)	(0.46)
Father: other qual.	0.17	0.18	0.25	0.22	0.17	0.18	0.24	0.21
	(0.38)	(0.38)	(0.43)	(0.42)	(0.38)	(0.38)	(0.43)	(0.41)
Father: missing education	0.26	0.31	0.31	0.35	0.25	0.29	0.29	0.32
	(0.44)	(0.46)	(0.46)	(0.48)	(0.43)	(0.46)	(0.46)	(0.47)
GCSEs			0.39	0.38			0.41	0.39
			(0.49)	(0.49)			(0.49)	(0.49)
Alevels			0.31	0.37			0.35	0.43
			(0.46)	(0.48)			(0.48)	(0.50)
Other qualification			0.30	0.24			0.25	0.18
			(0.46)	(0.43)			(0.43)	(0.38)
STEM	0.45	0.36			0.47	0.38		
	(0.50)	(0.48)			(0.50)	(0.49)		
LEM	0.19	0.19			0.20	0.20		
	(0.39)	(0.39)			(0.40)	(0.40)		
Other subject	0.16	0.25			0.15	0.25		
	(0.37)	(0.43)			(0.36)	(0.43)		
Subject missing	0.20	0.21			0.18	0.18		
	(0.40)	(0.40)			(0.39)	(0.38)		
Married/cohabiting	0.31	0.39	0.33	0.43	0.35	0.39	0.36	0.43
	(0.46)	(0.49)	(0.47)	(0.50)	(0.48)	(0.49)	(0.48)	(0.50)
Any child	0.04	0.12	0.15	0.33	0.05	0.08	0.17	0.24

| | With selection adjustment | | | | Without selection adjustment | | | |
| | Graduates | | Non Graduates | | Graduates | | Non graduates | |
	Males	Females	Males	Females	Males	Females	Males	Females
	(0.20)	(0.32)	(0.36)	(0.47)	(0.21)	(0.28)	(0.38)	(0.43)
Proportion of women								
in the occupation					0.51	0.69	0.46	0.70
					(0.23)	(0.16)	(0.27)	(0.15)
Hours<30					0.06	0.09	0.07	0.23
					(0.24)	(0.29)	(0.25)	(0.42)
Hours>=30 & <45					0.63	0.66	0.65	0.63
					(0.48)	(0.47)	(0.48)	(0.48)
Hours>=45					0.31	0.25	0.28	0.14
					(0.46)	(0.43)	(0.45)	(0.35)
Major group occupation 1 & 2					0.16	0.16	0.06	0.05
					(0.37)	(0.36)	(0.25)	(0.22)
Major group occupation 3					0.13	0.10	0.07	0.05
					(0.33)	(0.30)	(0.26)	(0.22)
Major group occupation 4					0.08	0.11	0.07	0.12
					(0.27)	(0.32)	(0.25)	(0.32)
Major group occupation 5 & 6					0.08	0.13	0.14	0.14
					(0.27)	(0.33)	(0.35)	(0.34)
Major group occupation 7					0.19	0.23	0.21	0.28
					(0.39)	(0.42)	(0.41)	(0.45)
Major group occupation 8 & 9					0.14	0.10	0.23	0.18
					(0.35)	(0.30)	(0.42)	(0.39)
Major group occupation								
missing					0.23	0.17	0.20	0.18
					(0.42)	(0.37)	(0.40)	(0.39)
N	**1315**	**1830**	**2007**	**2255**	**966**	**1358**	**1294**	**1429**

Note: The table reports mean and standard deviation (in parentheses) of the variables used in the decomposition of the gender wage gap. Descriptive statistics are weighted to account for attrition.

Table A5: Variables used in the estimation of the propensity score - NSHD

Variable	Description
In waged employment	Dummy = 1 if cohort member in employment with valid hourly wage
Work experience	Months in paid employment since age 16
Work experience squared	Months in self defined full-time paid employment since age 16 - squared
Jobs	Number of jobs since 16
Qualifications	
Level 1	Dummy = 1 if none attempted
Level 2	Dummy = 1 if Vocational course or Sub GCE
Level 3	Dummy = 1 if GCE 'O'
Level 4	Dummy = 1 if GCE 'A'
Level 5	Dummy = 1 if Burnham A2 or first or postgraduate degree
Missing	Dummy = 1 if qualification missing
Test scores Maths score	Standardised maths test score taken at age 11

Variable	Description
Missing maths score	Dummy = 1 if maths score missing
Reading score	Standardised reading test score taken at age 11
Missing reading score	Dummy = 1 if reading score missing
Fathers social class	
I	Dummy = 1 if at birth father's job in RG class I
II	Dummy = 1 if at birth father's job in RG class II
III NM	Dummy = 1 if at birth father's job in RG class III Non-Manual
III M	Dummy = 1 if at birth father's job in RG class III Manual
IV	Dummy = 1 if at birth father's job in RG class IV
V	Dummy = 1 if at birth father's job in RG class V
Missing	Dummy = 1 if information at birth on father's job is missing
Mother's age	Mother's age last birthday in years at birth sweep
Mother's age missing	Dummy = 1 if information missing
Mother's qualification	
Level 1	Dummy = 1 if Primary only
Level 2	Dummy = 1 if Primary and Tech/Comm Diploma
Level 3	Dummy = 1 if Secondary only
Level 4	Dummy = 1 if Secondary and tech./comm. diploma
Missing	Dummy = 1 if qualification missing
Father's qualification	
Level 1	Dummy = 1 if Primary only
Level 2	Dummy = 1 if Primary and Tech/Comm Diploma
Level 3	Dummy = 1 if Secondary only
Level 4	Dummy = 1 if Secondary and tech./comm. diploma
Missing	Dummy = 1 if qualification missing
Young child	Dummy = 1 if a child aged under 5 in the household at the time of the survey
More than 1 child	Dummy = 1 if more than one child in the household at the time of the survey

Table A6: Variables used in the estimation of the propensity score - NCDS and BCS

Variable	Description
In waged employment	Dummy = 1 if cohort member in employment with valid hourly wage
Jobs	Number of jobs since 16
Part time worker	Dummy = 1 if self-defined part-time worker (<30 hours per week)
Work experience	
Full-time experience	Months in self defined full-time paid employment since age 16
Full-time experience squared	Months in self defined full-time paid employment since age 16 - squared
Part-time experience	Months in self defined part-time paid employment since age 16
Part-time experience squared	Months in self defined part-time paid employment since age 16 - squared
Highest qualification	
NVQ Level 1	Dummy = 1 if highest qualification is NVQ level 1 or equivalent
NVQ Level 2	Dummy = 1 if highest qualification is NVQ level 2 or equivalent
NVQ Level 3	Dummy = 1 if highest qualification is NVQ level 3 or equivalent
NVQ Level 4	Dummy = 1 if highest qualification is NVQ level 4 or equivalent

Variable	Description
NVQ Level 5	Dummy = 1 if highest qualification is NVQ level 5 or equivalent
Missing	Dummy = 1 if information on highest qualification is missing
Test Scores	
Maths score	Standardised maths test score taken at age 10 (1970 cohort) or 11 (1958 cohort)
Missing maths score	Dummy = 1 if maths score missing
Reading score	Standardised reading test score taken at age 10 (1970 cohort) or 11 (1958 cohort)
Missing reading score	Dummy = 1 if reading score missing
Region	
London or SE	Dummy = 1 if living in London or the South East at time of survey
Presence of Children	
Children in household	Dummy = 1 if a child in the household by the time of the survey
Young child	Dummy = 1 if a child aged under 5 in the household at the time of the survey
More than 1 child	Dummy = 1 if more than one child in the household at the time of the survey
Social class of first job	
I	Dummy = 1 if first job in RG class I
II	Dummy = 1 if first job in RG class II
III NM	Dummy = 1 if first job in RG class III Non-Manual
III M	Dummy = 1 if first job in RG class III Manual
IV	Dummy = 1 if first job in RG class IV
V	Dummy = 1 if first job in RG class V
Missing	Dummy = 1 if information on occupation of first job is missing
Fathers social class	
I	Dummy = 1 if at birth father's job in RG class I
II	Dummy = 1 if at birth father's job in RG class II
III NM	Dummy = 1 if at birth father's job in RG class III Non-Manual
III M	Dummy = 1 if at birth father's job in RG class III Manual
IV	Dummy = 1 if at birth father's job in RG class IV
V	Dummy = 1 if at birth father's job in RG class V
Missing	Dummy = 1 if information at birth on father's job is missing
Age of parents	
Mother's age	Mother's age last birthday in years at birth sweep
Mother's age missing	Dummy = 1 if information missing
Father's age	Father's/ husband's age at birth sweep
Father's age missing	Dummy = 1 if information missing
Age mother left education	
Left before 16	Dummy = 1 if age left was less than 16
Left aged 16 or 17	Dummy = 1 if age left was 16 or 17
Left at 18 or more	Dummy = 1 if age left was 18 or more
Missing	Dummy = 1 if information missing
Age father left education	
Left before 16	Dummy = 1 if age left was less than 16
Left aged 16 or 17	Dummy = 1 if age left was 16 or 17
Left at 18 or more	Dummy = 1 if age left was 18 or more
Missing	Dummy = 1 if information missing
Number of siblings at age 16	
Only child	Dummy = 1 if had no siblings at age 16
One sibling	Dummy = 1 if had one sibling at age 16

Variable	Description
Two or three siblings	Dummy = 1 if had two or three siblings at age 16
Four or more siblings	Dummy = 1 if had four or more sibling at age 16

Table A7: Variables used in the estimation of the propensity score - Next Steps

Variable	Description
In waged employment	Dummy = 1 if cohort member in employment with valid hourly wage
Jobs	Number of jobs since 16
Full-time experience	Months in self defined full-time paid employment since age 16
Full-time experience squared	Months in self defined full-time paid employment since age 16 - squared
Part-time experience	Months in self defined part-time paid employment since age 16
Part-time experience squared	Months in self defined part-time paid employment since age 16 - squared
Qualifications	
NVQ Level 1	Dummy = 1 if highest qualification is NVQ level 1 or equivalent
NVQ Level 2	Dummy = 1 if highest qualification is NVQ level 2 or equivalent
NVQ Level 3	Dummy = 1 if highest qualification is NVQ level 3 or equivalent
NVQ Level 4	Dummy = 1 if highest qualification is NVQ level 4 or equivalent
NVQ Level 5	Dummy = 1 if highest qualification is NVQ level 5 or equivalent
Missing	Dummy = 1 if information on highest qualification is missing
Test scores	
Maths score	Standardised maths test score taken at age 12 (Key Stage 2)
Maths score missing	Dummy = 1 if maths score missing
Reading score	Standardised reading test score taken at age 12 (Key Stage 2)
Reading score missing	Dummy = 1 if reading score missing
Children in household	Dummy = 1 if a child in the household by the time of the survey
Young child	Dummy = 1 if a child aged under 5 in the household at the time of the survey
More than 1 child	Dummy = 1 if more than one child in the household at the time of the survey
Fathers social class	
I	Dummy = 1 if at birth father's job in RG class I
II	Dummy = 1 if at birth father's job in RG class II
III NM	Dummy = 1 if at birth father's job in RG class III Non-Manual
III M	Dummy = 1 if at birth father's job in RG class III Manual
IV	Dummy = 1 if at birth father's job in RG class IV
V	Dummy = 1 if at birth father's job in RG class V
Missing	Dummy = 1 if information at birth on father's job is missing
Mother's qualification	
Level 1	Dummy = 1 if age left was 16 or 17
Level 2	Dummy = 1 if age left was 18 or more
Missing	Dummy = 1 if information missing
Father's qualification	
Level 1	Dummy = 1 if age left was 16 or 17
Level 2	Dummy = 1 if age left was 18 or more
Missing	Dummy = 1 if information missing
Only child	Dummy = 1 if had no siblings at age 16

Variable	Description
One sibling	Dummy = 1 if had one sibling at age 16
Two or three siblings	Dummy = 1 if had two or three siblings at age 16
Four or more siblings	Dummy = 1 if had four or more sibling at age 16

Table A8: Variables used in NSHD models adjusting for sample attrition

Variable	Description
Test Scores	
Maths score	Standardised maths test score taken at age 11
Missing maths score	Dummy = 1 if maths score missing
Reading score	Standardised reading test score taken at age 11
Missing reading score	Dummy = 1 if reading score missing
Fathers social class	
I	Dummy = 1 if at birth father's job in RG class I
II	Dummy = 1 if at birth father's job in RG class II
III NM	Dummy = 1 if at birth father's job in RG class III Non-Manual
III M	Dummy = 1 if at birth father's job in RG class III Manual
IV	Dummy = 1 if at birth father's job in RG class IV
V	Dummy = 1 if at birth father's job in RG class V
Missing	Dummy = 1 if information at birth on father's job is missing
Mother's age	Mother's age last birthday in years at birth sweep
Mother's age missing	Dummy = 1 if information missing
Mother's qualification	
Level 1	Dummy = 1 if Primary only
Level 2	Dummy = 1 if Primary and Tech/Comm Diploma
Level 3	Dummy = 1 if Secondary only
Level 4	Dummy = 1 if Secondary and tech./comm. diploma
Missing	Dummy = 1 if qualification missing
Father's qualification	
Level 1	Dummy = 1 if Primary only
Level 2	Dummy = 1 if Primary and Tech/Comm Diploma
Level 3	Dummy = 1 if Secondary only
Level 4	Dummy = 1 if Secondary and tech./comm. diploma
Missing	Dummy = 1 if qualification missing

Table A9: Variables used in NCDS models adjusting for sample attrition

Variable	Description
Present in birth sweep	Dummy = 1 if cohort member was present
Birthweight	Weight in ounces

Variable	Description
Missing birthweight	Dummy = 1 if birthweight missing
Older siblings	Number of older siblings at age 16
Missing older sibling	Dummy = 1 if information on older siblings is missing
Younger siblings	Number of younger siblings at age 16
Missing younger sibling	Dummy = 1 if information on younger siblings is missing
Test Scores	
Maths score	Standardised maths test score taken at age 11
Missing maths score	Dummy = 1 if maths score missing
Reading score	Standardised reading test score taken at age 11
Missing reading score	Dummy = 1 if reading score missing
Rutter score	
Normal	Dummy = 1 if between 1 and 9
High	Dummy = 1 if 10 or more
Smoking	
Non-smoker	Dummy = 1 if non-smoker by age 16
Smoker	Dummy = 1 if ever smoked up to age 16
Drinking alcohol	
None	Dummy = 1 if had not drunk alcohol by age 16
Past week	Dummy = 1 if had drunk alcohol in past week at age 16
Past month	Dummy = 1 if had drunk alcohol in past month at age 16
Past year	Dummy = 1 if had drunk alcohol in past year at age 16
Mother's characteristics	
Mother's age	Mother's age last birthday in years at birth sweep
Mother's age missing	Dummy = 1 if information missing
Mother's marital status	
Married	Dummy = 1 if married or stable union at birth
Not married	Dummy = 1 if not married at birth
Mother's smoking	
Non-smoker	Dummy = 1 if non-smoker during pregnancy
Stopped smoking	Dummy = 1 if stopped smoking during pregnancy
Smoker	Dummy = 1 if smoker during pregnancy
Breastfeeding	
None	Dummy=1 if did not breastfeed
Up to 1 month	Dummy=1 if breastfed for up to 1 month
More than 1 month	Dummy=1 if breastfed for more than 1 month
Fathers social class	
I	Dummy = 1 if at birth father's job in RG class I
II	Dummy = 1 if at birth father's job in RG class II
III NM	Dummy = 1 if at birth father's job in RG class III Non-Manual
III M	Dummy = 1 if at birth father's job in RG class III Manual
IV	Dummy = 1 if at birth father's job in RG class IV
V	Dummy = 1 if at birth father's job in RG class V
Missing	Dummy = 1 if information at birth on father's job is missing
Housing tenure at birth	
Owner occupied	Dummy = 1 if owner occupied
Council rented	Dummy = 1 if council rented
Private rented	Dummy = 1 if private rented

Variable	Description
Rent free	Dummy = 1 if rent free
Other	Dummy = 1 if other
Persons per room - age 7	
1 or fewer	Dummy = 1 if 1 or fewer
1 to 1.5	Dummy = 1 if more than 1 up to 1.5
1.5 to 2	Dummy = 1 if more than 1.5 up to 2
More than 2	Dummy = 1 if more than 2
Region	Dummies = 1 if region living in at birth was: North;
	North West; East and West Riding; North Midlands; Midlands;
	East; South East; South; South West; Wales; Scotland.

Table A10: Variables used in BCS models adjusting for sample attrition

Variable	Description
Present in birth sweep	Dummy = 1 if cohort member was present
Birthweight	Weight in grams
Missing birthweight	Dummy = 1 if birthweight missing
Older siblings	Number of older siblings at age 16
Missing older sibling	Dummy = 1 if information on older siblings is missing
Younger siblings	Number of younger siblings at age 16
Missing younger sibling	Dummy = 1 if information on younger siblings is missing
Test Scores	
Maths score	Standardised maths test score taken at age 10
Missing maths score	Dummy = 1 if maths score missing
Reading score	Standardised reading test score taken at age 10
Missing reading score	Dummy = 1 if reading score missing
Rutter score	
Normal	Dummy = 1 if between 1 and 9
High	Dummy = 1 if 10 or more
Smoking	
Never smoker	Dummy = 1 if never smoked by age 16
Ex-smoker	Dummy = 1 if previously smoked, but non-smoker at age 16
Smoker	Dummy = 1 if smoker at age 16
Drinking alcohol	
None in last week	Dummy = 1 if had not drunk alcohol in the last week (at age 16)
Once in last week	Dummy = 1 if had drunk alcohol once in the last week (at age 16)
2 -3 times in last week	Dummy = 1 if had drunk alcohol 2-3 times in the last week (at age 16)
Most days	Dummy = 1 if had drunk alcohol most days in the last week (at age 16)
Mother's characteristics	
Mother's age	Mother's age last birthday in years at birth sweep
Mother's age missing	Dummy = 1 if information missing
Mother's marital status	
Married	Dummy = 1 if married at birth
Not married	Dummy = 1 if single, divorced, widowed, separated at birth

Variable	Description
Mother's smoking	
Non-smoker	Dummy = 1 if non-smoker during pregnancy
Stopped smoking	Dummy = 1 if stopped smoking during pregnancy
Smoker	Dummy = 1 if smoker during pregnancy
Breastfeeding	
None	Dummy=1 if did not breastfeed
Up to 1 month	Dummy=1 if breastfed for up to 1 month
1-3 months	Dummy=1 if breastfed for between 1 and 3 months
More than 3 months	Dummy=1 if breastfed for more than 3 months
Fathers social class	
I	Dummy = 1 if at birth father's job in RG class I
II	Dummy = 1 if at birth father's job in RG class II
III NM	Dummy = 1 if at birth father's job in RG class III Non-Manual
III M	Dummy = 1 if at birth father's job in RG class III Manual
IV	Dummy = 1 if at birth father's job in RG class IV
V	Dummy = 1 if at birth father's job in RG class V
Missing	Dummy = 1 if information at birth on father's job is missing
Housing tenure at age 5	
Owner occupied	Dummy = 1 if owner occupied
Being bought	Dummy = 1 if housing being bought
Council rented	Dummy = 1 if council rented
Private rent unfurnished	Dummy = 1 if private rented unfurnished
Private rent furnished	Dummy = 1 if private rented furnished
Rent free	Dummy = 1 if rent free
Tied to occupation	Dummy = 1 if tied to occupation
Other	Dummy = 1 if other
Persons per room - age 5	Ratio of people in household to number of rooms.
Missing person per room	Dummy = 1 if information missing
Region	Dummies = 1 if region living in at birth was: North;
	North West; East and West Riding; North Midlands; Midlands;
	East; South East; South; South West; Wales; Scotland.

Table A11: Variables used in Next Steps models adjusting for sample attrition

Variable	Description
Present in age 14 sweep	Dummy = 1 if cohort member was present
Birthweight	Weight in grams
Missing birthweight	Dummy = 1 if birthweight missing
Older siblings	Number of older siblings at age 16
Missing older sibling	Dummy = 1 if information on older siblings is missing
Younger siblings	Number of younger siblings at age 16
Missing younger sibling	Dummy = 1 if information on younger siblings is missing
Test score	
Maths score	Standardised maths test score taken at age 10

Variable	Description
Missing maths score	Dummy = 1 if maths score missing
Reading score	Standardised reading test score taken at age 10
Missing reading score	Dummy = 1 if reading score missing
Smoking	Dummy = 1 if smoker at age 14
Smoking missing	Dummy = 1 if information missing
Drinking alcohol	Dummy = 1 if had drunk alcohol in the last week (at age 14)
Drinking alcohol missing	Dummy = 1 if information missing
Parents know when cohort member is out	
Always	Dummy = 1 if always
Usually	Dummy = 1 if usually
Never	Dummy = 1 if never
Other	Dummy = 1 if other
Playing a music instrument	Dummy = 1 if cohort member plays music instrument
Playing a music instrument missing	Dummy = 1 if other missing
Name calling	Dummy = 1 if endured namecalling
Name calling missing	Dummy = 1 if information about name calling is missing
Mother's characteristics	
Mother's age	Mother's age last birthday in years at birth sweep
Mother's age missing	Dummy = 1 if information missing
Mother's marital status	
Married	Dummy = 1 if married at birth
Not married	Dummy = 1 if single, divorced, widowed, separated at birth
Fathers social class	
I	Dummy = 1 if at birth father's job in RG class I
II	Dummy = 1 if at birth father's job in RG class II
III NM	Dummy = 1 if at birth father's job in RG class III Non-Manual
III M	Dummy = 1 if at birth father's job in RG class III Manual
IV	Dummy = 1 if at birth father's job in RG class IV
V	Dummy = 1 if at birth father's job in RG class V
Missing	Dummy = 1 if information at birth on father's job is missing
Housing tenure at age 14	
Owner	Dummy = 1 if housing being bought
Council rented	Dummy = 1 if council rented
Private rent	Dummy = 1 if private rented
Rent free	Dummy = 1 if rent free
Other	Dummy = 1 if other

B Estimates of OLS regressions and Kitagawa-Oaxaca-Blinder decompositions

Table B1: OLS estimates of Gender Wage Gap by higher education status in NSHD by higher education status - Raw, covariate adjusted, selection and covariate adjusted

	Raw	*Covariate Adjusted*	*Covariate + Selection Adjusted*	*Raw*	*Covariate Adjusted*	*Covariate + Selection Adjusted*
		Graduate			Non-Graduate	
Female	-0.019	0.020	-0.213***	-0.465***	-0.423***	-0.480***
	(0.025)	(0.029)	(0.029)	(0.014)	(0.015)	(0.013)
Std. maths score		-0.058***	-0.066***		0.015*	-0.006
		(0.020)	(0.022)		(0.009)	(0.007)
Std. reading score		0.050***	0.054***		0.039***	0.036***
		(0.015)	(0.016)		(0.009)	(0.008)
Vocational qualification					-0.012	-0.001
					(0.025)	(0.021)
GCE'O' Level or Burnham C					0.055***	0.043***
					(0.018)	(0.015)
Burnham A2		-0.123***				
		(0.030)				
First degree			0.123***			
			(0.032)			
Qualification missing					0.042	0.025
					(0.038)	(0.031)
Full-time/ part-time experience		0.005***	0.004**		0.005***	0.002***
		(0.002)	(0.002)		(0.001)	(0.000)
Full-time/ part-time experience sq.		-0.000**	-0.000		-0.000***	-0.000***
		(0.000)	(0.000)		(0.000)	(0.000)
N. of different spells of work		-0.003***	-0.002**		-0.000	-0.000
		(0.001)	(0.001)		(0.000)	(0.000)
Any child		0.047	-0.193***		0.004	-0.074***
		(0.037)	(0.032)		(0.018)	(0.015)
More than one child		0.034	-0.002		0.016	-0.001
		(0.060)	(0.050)		(0.022)	(0.016)
Father soc. Class Intermediate		-0.042	-0.063*		-0.046	-0.030
		(0.037)	(0.038)		(0.036)	(0.030)
Father soc. Class Skilled (non-Manual)		-0.045	-0.075*		-0.030	-0.026
		(0.041)	(0.044)		(0.038)	(0.031)
Father soc. Class Skilled (Manual)		-0.044	-0.102**		-0.032	-0.017
		(0.047)	(0.050)		(0.037)	(0.031)

Continued on next page

Table B1 – continued from previous page

	Raw	Covariate Adjusted	Covariate + Selection Adjusted	Raw	Covariate Adjusted	Covariate + Selection Adjusted
		Graduate			Non-Graduate	
Father soc. Class 5						
Partly skilled		-0.049	-0.044		-0.043	-0.041
		(0.057)	(0.061)		(0.038)	(0.032)
Father soc. Class						
Unskilled		0.166	0.074		-0.028	-0.012
		(0.114)	(0.136)		(0.044)	(0.037)
Father soc. Class						
Missing		-0.053	-0.121		-0.111	-0.046
		(0.123)	(0.138)		(0.077)	(0.064)
Father education						
Primary and Tech/comm. Diploma		0.070	0.088*		-0.030	-0.017
		(0.047)	(0.051)		(0.038)	(0.031)
Father education Secondary only		-0.021	0.007		0.063***	0.030*
		(0.033)	(0.035)		(0.022)	(0.018)
Father education Secondary						
and tech/Comm Diploma		-0.026	-0.026		0.016	-0.025
		(0.041)	(0.042)		(0.035)	(0.030)
Father education Missing		0.037	0.139		0.053	0.001
		(0.097)	(0.103)		(0.053)	(0.047)
Mother education Primary						
and Tech/comm. Diploma		-0.041	-0.020		0.028	0.010
		(0.055)	(0.057)		(0.038)	(0.031)
Mother education Secondary only		-0.063**	-0.066**		0.026	0.029
		(0.031)	(0.033)		(0.025)	(0.021)
Mother education Secondary						
and tech/Comm Diploma		-0.068*	-0.036		-0.014	-0.004
		(0.038)	(0.040)		(0.039)	(0.035)
Mother education Missing		-0.048	-0.135		0.023	0.014
		(0.098)	(0.102)		(0.054)	(0.048)
Constant	1.993***	1.792***	1.659***	1.848***	1.466***	1.766***
	(0.014)	(0.225)	(0.205)	(0.009)	(0.075)	(0.060)
N	528	528	732	1,715	1,715	2,978
R^2	0.005	0.156	0.245	0.325	0.420	0.43

Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B2: OLS estimates of Gender Wage Gap by higher education status in NCDS by higher education status - Raw, covariate adjusted, selection and covariate adjusted

	Raw	Covariate Adjusted	Covariate + Selection Adjusted	Raw	Covariate Adjusted	Covariate + Selection Adjusted
		Graduate			Non-Graduate	
Female	-0.044***	-0.033*	-0.070***	-0.185***	-0.187***	-0.200***
	(0.016)	(0.017)	(0.016)	(0.007)	(0.007)	(0.006)
Father education (GCSE or A-levels)		-0.033	-0.017		0.022*	0.020*
		(0.024)	(0.022)		(0.013)	(0.012)
Father education (Higher education)		-0.016	-0.020		0.012	0.008
		(0.030)	(0.027)		(0.020)	(0.017)
Father education (Missing)		-0.002	-0.000		-0.005	-0.000
		(0.053)	(0.052)		(0.021)	(0.017)
Mother education (GCSE or A-levels)		0.010	-0.002		0.015	0.014
		(0.022)	(0.021)		(0.012)	(0.010)
Mother education (Higher education)		0.011	-0.002		-0.046**	-0.019
		(0.032)	(0.029)		(0.022)	(0.019)
Mother education (Missing)		-0.020	-0.024		0.010	0.000
		(0.054)	(0.053)		(0.021)	(0.017)
Father soc. Class 2		-0.010	0.037		-0.019	-0.003
		(0.030)	(0.027)		(0.023)	(0.019)
Father soc. Class 3		0.036	0.060*		0.009	0.009
		(0.035)	(0.031)		(0.023)	(0.020)
Father soc. Class 4		0.046	0.067**		-0.011	-0.001
		(0.031)	(0.027)		(0.021)	(0.018)
Father soc. Class 5		0.029	0.054		-0.040*	-0.016
		(0.044)	(0.040)		(0.023)	(0.020)
Father soc. Class 6		0.031	0.047		-0.023	-0.004
		(0.051)	(0.048)		(0.024)	(0.020)
Father soc. Class missing		0.047	0.057*		-0.016	-0.004
		(0.036)	(0.033)		(0.023)	(0.020)
GCSEs					-0.049***	-0.045***
					(0.010)	(0.009)
Other qualification					-0.101***	-0.073***
					(0.011)	(0.010)
Qualification missing					-0.045***	-0.047***
					(0.011)	(0.009)
STEM		0.042**	0.016			
		(0.021)	(0.019)			
LEM		-0.113***	-0.080***			
		(0.025)	(0.024)			
Subject missing		-0.001	0.002			
		(0.034)	(0.031)			
Std. maths score		0.045***	0.043***		0.031***	0.027***
		(0.013)	(0.012)		(0.006)	(0.005)
Std.reading score		0.013	0.003		0.027***	0.015***

Continued on next page

	Raw	*Covariate Adjusted*	*Covariate + Selection Adjusted*	*Raw*	*Covariate Adjusted*	*Covariate + Selection Adjusted*
		Graduate			Non-Graduate	
		(0.013)	(0.012)		(0.006)	(0.005)
Std. maths score missing		0.060**	0.049**		0.210	0.140
		(0.025)	(0.023)		(0.199)	(0.161)
Std. reading score missing					-0.222	-0.140
					(0.199)	(0.162)
Full-time experience		0.003**	0.004***		0.002***	0.003***
		(0.001)	(0.001)		(0.001)	(0.000)
Full-time experience		-0.000*	-0.000***		-0.000*	-0.000***
		(0.000)	(0.000)		(0.000)	(0.000)
Part-time experience		-0.002	-0.003		-0.005***	-0.002**
		(0.003)	(0.002)		(0.001)	(0.001)
Part-time experience		0.000	0.000		0.000***	0.000**
		(0.000)	(0.000)		(0.000)	(0.000)
N. of spells of work		-0.036***	-0.028***		-0.019***	-0.011***
		(0.007)	(0.006)		(0.002)	(0.002)
Married/cohabiting		0.033	-0.114***		-0.007	-0.047***
		(0.042)	(0.031)		(0.012)	(0.009)
Any child		-0.071	0.062		-0.014	-0.051***
		(0.098)	(0.071)		(0.019)	(0.012)
More than one child		0.055***	0.059***		0.040***	0.031***
		(0.016)	(0.016)		(0.007)	(0.007)
Constant	1.768***	1.707***	1.639***	1.690***	1.689***	1.661***
	(0.011)	(0.042)	(0.037)	(0.005)	(0.028)	(0.022)
N	1,291	1,291	1,699	6,358	6,358	10,022
R^2	0.006	0.088	0.076	0.096	0.177	0.180

Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B3: OLS estimates of Gender Wage Gap by higher education status in BCS by higher education status - Raw, covariate adjusted, selection and covariate adjusted

	Raw	*Covariate Adjusted*	*Covariate + Selection Adjusted*	*Raw*	*Covariate Adjusted*	*Covariate + Selection Adjusted*
		Graduate			Non-Graduate	
Female dummy	-0.047***	-0.014	-0.033**	-0.112***	-0.111***	-0.119***
	(0.018)	(0.018)	(0.017)	(0.012)	(0.012)	(0.011)
GCSEs					-0.101***	-0.076***
					(0.018)	(0.017)

	Raw	Covariate Adjusted	Covariate + Selection Adjusted	Raw	Covariate Adjusted	Covariate + Selection Adjusted
		Graduate			Non-Graduate	
Other qualification					-0.198***	-0.162***
					(0.021)	(0.019)
Qualification missing					-0.177***	-0.103***
					(0.028)	(0.024)
STEM		0.135***	0.109***			
		(0.023)	(0.022)			
LEM		0.134***	0.123***			
		(0.028)	(0.026)			
Subject missing		0.161***	0.131***			
		(0.024)	(0.022)			
Std maths score		0.040**	0.032**		0.026***	0.028***
		(0.017)	(0.015)		(0.009)	(0.008)
Std reading score		-0.007	-0.012		0.033***	0.016**
		(0.017)	(0.016)		(0.009)	(0.008)
Std maths score missing		0.142	0.070		-0.127	0.011
		(0.119)	(0.114)		(0.092)	(0.079)
Std reading score missing		-0.125	-0.055		0.125	-0.008
		(0.119)	(0.114)		(0.092)	(0.079)
Full-time experience		0.009***	0.008***		0.001*	0.001**
		(0.001)	(0.001)		(0.001)	(0.001)
Full-time experience sq.		-0.000***	-0.000***		0.000	-0.000
		(0.000)	(0.000)		(0.000)	(0.000)
Part-time experience		0.002	0.001		-0.003***	-0.001
		(0.002)	(0.002)		(0.001)	(0.001)
Part-time experience sq.		-0.000	-0.000		0.000***	0.000
		(0.000)	(0.000)		(0.000)	(0.000)
N. of spells of work		-0.039***	-0.030***		-0.012***	-0.011***
		(0.007)	(0.007)		(0.004)	(0.004)
Married/cohabiting		-0.050	-0.081**		-0.028	-0.029**
		(0.045)	(0.041)		(0.017)	(0.015)
Any child		0.397***	0.146*		0.036	0.005
		(0.108)	(0.086)		(0.024)	(0.019)
More than one child		0.048***	0.055***		0.045***	0.030***
		(0.018)	(0.017)		(0.012)	(0.011)
Father education (GCSE or A-levels)		0.004	0.000		0.035**	0.013
		(0.025)	(0.024)		(0.016)	(0.015)
Father education (Higher education)		0.021	0.032		0.040	0.028
		(0.029)	(0.027)		(0.025)	(0.022)
Father education (Missing)		0.047	-0.025		0.028	0.018
		(0.063)	(0.055)		(0.034)	(0.030)

	Raw	Covariate Adjusted	Covariate + Selection Adjusted	Raw	Covariate Adjusted	Covariate + Selection Adjusted
		Graduate			Non-Graduate	
Mother education (GCSE or A-levels)		0.024	0.032		0.009	0.019
		(0.023)	(0.021)		(0.015)	(0.014)
Mother education (Higher education)		0.068**	0.068***		0.034	0.010
		(0.028)	(0.026)		(0.027)	(0.025)
Mother education (Missing)		0.024	0.136**		0.016	-0.008
		(0.071)	(0.063)		(0.039)	(0.035)
Father soc. Class 2		-0.018	0.011		-0.031	-0.044
		(0.031)	(0.028)		(0.044)	(0.041)
Father soc. Class 3		0.026	0.055*		-0.026	-0.023
		(0.034)	(0.031)		(0.043)	(0.041)
Father soc. Class 4		-0.033	-0.019		-0.037	-0.050
		(0.033)	(0.030)		(0.042)	(0.039)
Father soc. Class 5		-0.046	-0.040		-0.050	-0.063
		(0.047)	(0.043)		(0.043)	(0.041)
Father soc. Class 6		-0.166*	-0.224***		-0.139***	-0.092**
		(0.095)	(0.079)		(0.048)	(0.044)
Father soc. Class missing		-0.025	-0.049		-0.075	-0.067
		(0.048)	(0.044)		(0.046)	(0.043)
Constant	2.104***	1.776***	1.748***	1.883***	1.920***	1.936***
	(0.012)	(0.049)	(0.043)	(0.008)	(0.051)	(0.046)
N	1,568	1,568	1,984	4,292	4,292	6,078
$?R^2$	0.004	0.126	0.112	0.021	0.124	0.084

Note: Standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table B4: OLS estimates of Gender Wage Gap by higher education status in Next Steps by higher education status - Raw, covariate adjusted, selection and covariate adjusted

	Raw	Covariate Adjusted	Covariate + Selection Adjusted	Raw	Covariate Adjusted	Covariate + Selection Adjusted
		Graduate			Non-Graduate	
Female	-0.091***	-0.062***	-0.053***	-0.085***	-0.081***	-0.058***
	(0.016)	(0.015)	(0.013)	(0.014)	(0.014)	(0.011)
Other qualification					-0.193***	-0.172***

	Raw	Covariate Adjusted	Covariate + Selection Adjusted	Raw	Covariate Adjusted	Covariate + Selection Adjusted
		Graduate			Non-Graduate	
					(0.020)	(0.016)
GCSEs					-0.105***	-0.113***
					(0.016)	(0.013)
STEM		0.146***	0.101***			
		(0.019)	(0.017)			
LEM		0.118***	0.079***			
		(0.022)	(0.020)			
Subject missing		0.029	0.020			
		(0.024)	(0.021)			
Full-time experience		0.007***	0.007***		0.003***	0.004***
		(0.001)	(0.001)		(0.001)	(0.001)
Part-time experience		-0.006***	-0.004***		-0.002**	-0.001
		(0.001)	(0.001)		(0.001)	(0.001)
Full-time experience sq		-0.000***	-0.000***		-0.000***	-0.000***
		(0.000)	(0.000)		(0.000)	(0.000)
Part-time experience sq		0.000***	0.000***		0.000	0.000
		(0.000)	(0.000)		(0.000)	(0.000)
N. of spells of work		-0.023***	-0.020***		-0.012**	-0.001
		(0.006)	(0.006)		(0.005)	(0.005)
Married/cohabiting		-0.041	-0.042		-0.081***	-0.060***
		(0.033)	(0.028)		(0.022)	(0.017)
Any child		-0.105	-0.009		0.009	-0.029
		(0.070)	(0.050)		(0.031)	(0.023)
More than one child		0.014	0.017		0.046***	0.020*
		(0.015)	(0.014)		(0.015)	(0.012)
Father soc. Class 2		-0.026	-0.003		-0.037	-0.015
		(0.021)	(0.019)		(0.024)	(0.021)
Father soc. Class 3		-0.079**	-0.036		-0.119***	-0.045
		(0.038)	(0.036)		(0.040)	(0.034)
Father soc. Class 4		-0.056**	-0.028		-0.048*	-0.022
		(0.027)	(0.024)		(0.028)	(0.023)
Father soc. Class 5		-0.093***	-0.038		-0.023	-0.002
		(0.029)	(0.027)		(0.029)	(0.024)
Father soc. Class 6		-0.088**	-0.042		-0.057*	-0.034
		(0.036)	(0.032)		(0.032)	(0.026)
Father soc. Class 7		-0.097***	-0.049		-0.066**	-0.039
		(0.033)	(0.030)		(0.030)	(0.025)
Father soc. Class missing		-0.059*	-0.026		-0.095***	-0.065**
		(0.030)	(0.027)		(0.031)	(0.025)
Mother: missing education					-0.026	-0.034
					(0.029)	(0.023)
Mother: higher education		0.045	0.044*			
		(0.028)	(0.026)			

Table B4 – continued from previous page

	Raw	Covariate Adjusted	Covariate + Selection Adjusted	Raw	Covariate Adjusted	Covariate + Selection Adjusted
		Graduate			Non-Graduate	
Mother: A level/GCSE		0.024	0.036		-0.021	-0.016
		(0.027)	(0.025)		(0.018)	(0.016)
Mother: other qual.		0.018	0.013		-0.038*	-0.040**
		(0.030)	(0.027)		(0.021)	(0.017)
Father: missing education			0.021		-0.003	0.031
			(0.026)		(0.027)	(0.020)
Father: higher education		0.020	0.035			0.028
		(0.026)	(0.023)			(0.021)
Father: A level/GCSE		0.005	0.018		-0.036*	0.002
		(0.026)	(0.021)		(0.021)	(0.016)
Father: other qual.		0.002			-0.039	
		(0.029)			(0.025)	
Constant	2.012***	1.845***	1.780***	1.862***	1.966***	1.829***
	(0.012)	(0.044)	(0.040)	(0.011)	(0.034)	(0.030)
N	2,324	2,324	3,143	2,723	2,723	4,299
R^2	0.015	0.150	0.096	0.013	0.132	0.125

Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B5: Kitagawa-Oaxaca-Blinder decomposition without selection adjustment by higher education status

	NSHD		NCDS		BCS		Next steps	
	Characteristics	Coefficients	Characteristics	Coefficients	Characteristics	Coefficients	Characteristics	Coefficients
Graduates								
Difference	0.019		0.044		0.047		0.091	
	(0.030)		(0.016)		(0.019)		(0.015)	
Unexplained	-0.020		0.033		0.015		0.061	
	(0.034)		(0.018)		(0.019)		(0.014)	
Early life	0.020	-0.039	0.002	0.031	-0.003	-0.007	0.005	0.012
	(0.010)	(0.047)	(0.002)	(0.017)	(0.003)	(0.028)	(0.002)	(0.013)
Qualifications	0.008	-0.065	0.003	-0.086	0.038	0.017	0.016	0.031
	(0.012)	(0.056)	(0.008)	(0.025)	(0.006)	(0.025)	(0.003)	(0.010)
Experience	0.000	0.086	0.011	0.119	0.001	0.066	0.008	0.032
	(0.013)	(0.158)	(0.005)	(0.046)	(0.005)	(0.077)	(0.004)	(0.203)
Marital/parental	0.011	-0.007	-0.006	-0.011	-0.005	-0.218	0.001	-0.017
	(0.005)	(0.058)	(0.003)	(0.098)	(0.003)	(0.221)	(0.002)	(0.064)
Constant		0.005		-0.020		0.158		0.004
		(0.170)		(0.115)		(0.253)		(0.215)
N	528	528	1,291	1,291	1,568	1,568	2,324	2,324
Non graduates								
Difference	0.465		0.185		0.112		0.085	
	(0.019)		(0.007)		(0.012)		(0.014)	
Unexplained	0.423		0.187		0.111		0.082	
	(0.020)		(0.007)		(0.012)		(0.014)	
Early life	0.002	-0.007	-0.001	0.037	-0.002	0.027	-0.001	0.020
	(0.003)	(0.044)	(0.001)	(0.012)	(0.001)	(0.020)	(0.002)	(0.013)
Qualifications	-0.015	0.007	-0.008	-0.002	-0.016	0.014	-0.015	0.031
	(0.004)	(0.016)	(0.002)	(0.008)	(0.004)	(0.011)	(0.003)	(0.014)
Experience	0.053	-0.108	0.011	0.091	0.024	0.073	0.019	0.186
	(0.008)	(0.131)	(0.002)	(0.037)	(0.004)	(0.063)	(0.003)	(0.098)
Marital	0.002	-0.019	-0.005	-0.078	-0.003	-0.029	0.001	-0.014
	(0.003)	(0.016)	(0.001)	(0.019)	(0.001)	(0.022)	(0.002)	(0.020)
Constant		0.550		0.140		0.025		-0.140
		(0.132)		(0.042)		(0.075)		(0.099)
N	1,715	1,715	6,358	6,358	4,292	4,292	2,723	2,723

Note: Standard errors in parentheses.

Table B6: Kitagawa-Oaxaca-Blinder decomposition with selection adjustment by higher education status

	NSHD		NCDS		BCS		Next steps	
	Characteristics	*Coefficients*	*Characteristics*	*Coefficients*	*Characteristics*	*Coefficients*	*Characteristics*	*Coefficients*
				Graduates				
Difference	0.251		0.077		0.059		0.072	
	(0.035)		(0.014)		(0.017)		(0.013)	
Unexplained	0.213		0.070		0.033		0.055	
	(0.037)		(0.016)		(0.017)		(0.013)	
Early life	0.004	-0.116	0.000	0.013	0.001	-0.006	0.002	-0.001
	(0.008)	(0.054)	(0.002)	(0.016)	(0.003)	(0.025)	(0.002)	(0.009)
Qualifications	0.007	-0.057	0.004	-0.078	0.032	0.035	0.011	-0.039
	(0.012)	(0.069)	(0.007)	(0.024)	(0.005)	(0.031)	(0.002)	(0.011)
Experience	0.003	0.286	0.009	0.074	-0.000	0.027	0.003	0.285
	(0.014)	(0.142)	(0.004)	(0.037)	(0.004)	(0.054)	(0.004)	(0.093)
Marital/parental	0.025	-0.115	-0.006	-0.069	-0.007	-0.270	0.001	-0.060
	(0.009)	(0.047)	(0.003)	(0.069)	(0.003)	(0.185)	(0.002)	(0.056)
Constant		0.215		0.128		0.247		-0.130
		(0.158)		(0.086)		(0.207)		(0.112)
N	732	732	1,699	1,699	1,984	1,984	3,143	3,143
				Non graduates				
Difference	0.533		0.217		0.128		0.079	
	(0.013)		(0.006)		(0.011)		(0.012)	
Unexplained	0.480		0.200		0.119		0.060	
	(0.015)		(0.006)		(0.011)		(0.011)	
Early life	0.000	0.024	-0.000	0.030	-0.001	0.018	0.000	0.001
	(0.002)	(0.036)	(0.000)	(0.010)	(0.001)	(0.018)	(0.001)	(0.007)
Qualifications	-0.003	-0.001	0.001	-0.008	-0.010	0.007	-0.011	0.018
	(0.002)	(0.011)	(0.001)	(0.006)	(0.003)	(0.009)	(0.002)	(0.009)
Experience	0.045	-0.117	0.011	-0.001	0.020	0.031	0.019	0.085
	(0.009)	(0.074)	(0.002)	(0.022)	(0.004)	(0.043)	(0.003)	(0.048)
Marital	0.011	0.018	0.005	-0.064	-0.000	-0.055	0.010	-0.039
	(0.003)	(0.008)	(0.002)	(0.010)	(0.002)	(0.016)	(0.002)	(0.014)
Constant		0.556		0.244		0.118		-0.005
		(0.080)		(0.026)		(0.053)		(0.049)
N	2,978	2,978	10,022	10,022	6,078	6,078	4,299	4,299

Note: Standard errors in parentheses.

Table B7: Kitagawa-Oaxaca-Blinder decomposition by cohort for the whole sample and without selection adjustment

	NSHD		NCDS		BCS		Next steps	
	Characteristics	*Coefficients*	*Characteristics*	*Coefficients*	*Characteristics*	*Coefficients*	*Characteristics*	*Coefficients*
Difference	*0.412*		*0.153*		*0.098*		*0.080*	
	(0.018)		*(0.007)*		*(0.011)*		*(0.011)*	
Unexplained	*0.362*		*0.166*		*0.089*		*0.070*	
	(0.018)		*(0.007)*		*(0.010)*		*(0.010)*	
Early life	0.002	-0.026	0.000	0.036	-0.002	0.019	0.001	0.015
	(0.002)	(0.038)	(0.001)	(0.009)	(0.001)	(0.014)	(0.001)	(0.009)
Qualifications	0.017	0.077	-0.020	0.073	-0.002	0.189	-0.007	-0.059
	(0.007)	(0.022)	(0.003)	(0.019)	(0.004)	(0.025)	(0.003)	(0.013)
Experience	0.031	0.040	0.012	0.112	0.018	0.080	0.016	0.155
	(0.006)	(0.097)	(0.002)	(0.029)	(0.003)	(0.050)	(0.003)	(0.099)
Marital	0.000	-0.032	-0.005	-0.086	-0.004	-0.036	0.000	-0.019
	(0.002)	(0.018)	(0.001)	(0.019)	(0.001)	(0.023)	(0.001)	(0.021)
Constant		0.303		0.030		-0.163		-0.022
		(0.093)		(0.040)		(0.064)		(0.104)
N	2,243	2,243	7,649	7,649	5,860	5,860	5,047	5,047

Note: Standard errors in parentheses.

Table B8: Kitagawa-Oaxaca-Blinder decomposition by cohort for the whole sample and with selection adjustment

	NSHD		NCDS		BCS		Next steps	
	Characteristics	*Coefficients*	*Characteristics*	*Coefficients*	*Characteristics*	*Coefficients*	*Characteristics*	*Coefficients*
Difference	*0.512*		*0.197*		*0.116*		*0.080*	
	(0.013)		*(0.006)*		*(0.009)*		*(0.009)*	
Unexplained	*0.437*		*0.186*		*0.101*		*0.065*	
	(0.014)		*(0.006)*		*(0.009)*		*(0.008)*	
Early life	0.001	0.014	-0.000	0.023	-0.001	0.012	0.001	0.004
	(0.002)	(0.032)	(0.000)	(0.008)	(0.001)	(0.012)	(0.001)	(0.005)
Qualifications	0.023	0.035	-0.005	-0.017	0.002	0.070	-0.005	-0.009
	(0.004)	(0.016)	(0.002)	(0.021)	(0.003)	(0.026)	(0.002)	(0.010)
Experience	0.034	0.109	0.012	0.042	0.015	0.041	0.013	0.119
	(0.007)	(0.060)	(0.002)	(0.019)	(0.003)	(0.035)	(0.002)	(0.043)
Marital	0.017	0.013	0.004	-0.063	-0.001	-0.061	0.006	-0.030
	(0.003)	(0.008)	(0.001)	(0.010)	(0.001)	(0.017)	(0.002)	(0.015)
Constant		0.268		0.200		0.040		-0.019
		(0.065)		(0.032)		(0.045)		(0.049)
N	3,710	3,710	11,746	11,746	8,061	8,061	7,429	7,429

Note: Standard errors in parentheses.

Table B9: OLS estimates of occupational characteristics by cohort and higher education status

	BCS Graduates	BCS Non Graduates	Next Steps Graduates	Next Steps Non Graduates
Female	-0.001	-0.096***	-0.016	-0.044***
	(0.020)	(0.015)	(0.015)	(0.015)
Major group occupation 2	0.030	0.086**	0.134***	0.195***
	(0.026)	(0.035)	(0.031)	(0.030)
Major group occupation 3	0.036	0.109***	0.089***	0.109***
	(0.028)	(0.024)	(0.032)	(0.029)
Major group occupation 4	-0.091**	-0.042**	-0.002	0.043
	(0.036)	(0.020)	(0.036)	(0.032)
Major group occupation 5	-0.052	-0.086***	-0.045	-0.084**
	(0.058)	(0.022)	(0.058)	(0.035)
Major group occupation 6	-0.127***	-0.146***	-0.189***	-0.104***
	(0.048)	(0.021)	(0.040)	(0.034)
Major group occupation 7	-0.124**	-0.199***	-0.122***	-0.095***
	(0.052)	(0.026)	(0.038)	(0.031)
Major group occupation 8	-0.240***	-0.117***	-0.237***	-0.148***
	(0.082)	(0.023)	(0.077)	(0.040)
Major group occupation 9	-0.319***	-0.217***	-0.299***	-0.185***
	(0.096)	(0.029)	(0.052)	(0.033)
Major group occupation missing	0.026	-0.010	0.379**	0.095
	(0.038)	(0.026)	(0.164)	(0.108)
Hours>=30 & <45	-0.181***	-0.039*	-0.026	-0.096***
	(0.037)	(0.023)	(0.029)	(0.024)
Hours>=45	-0.273***	-0.128***	-0.161***	-0.181***
	(0.041)	(0.026)	(0.031)	(0.028)
Proportion of women in the occupation	-0.144***	-0.121***	-0.256***	-0.275***
	(0.040)	(0.026)	(0.034)	(0.033)
Constant	2.261***	2.005***	2.015***	1.946***
	(0.086)	(0.070)	(0.065)	(0.064)
N	1,568	4,292	2,324	2,723
R^2	0.192	0.183	0.300	0.266

Note: Regressions also include early life characteristics, qualifications, experience, marital and parental status. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B10: Kitagawa-Oaxaca-Blinder decomposition without selection adjustment and with job characteristics by higher education status

	BCS		NS	
	Characteristics	*Coefficients*	*Characteristics*	*Coefficients*
		Graduates		
Difference	*0.047*		*0.091*	
	(0.019)		*(0.015)*	
Unexplained	*0.002*		*0.016*	
	(0.020)		*(0.015)*	
Early life	-0.002	0.009	0.003	0.023
	(0.003)	(0.024)	(0.002)	(0.012)
Qualifications	0.037	-0.001	0.010	0.008
	(0.006)	(0.024)	(0.002)	(0.009)
Experience	0.002	0.040	0.006	0.001
	(0.005)	(0.071)	(0.004)	(0.178)
Marital	-0.004	-0.251	0.001	0.030
	(0.003)	(0.200)	(0.001)	(0.066)
Major group occupations	0.001	-0.122	0.012	0.053
	(0.005)	(0.041)	(0.006)	(0.021)
Female proportion in job	0.037	-0.004	0.052	-0.049
	(0.011)	(0.037)	(0.007)	(0.035)
Hours	-0.026	-0.010	-0.009	-0.023
	(0.006)	(0.045)	(0.003)	(0.021)
N	1,568	1,568	2,324	2,324
		Non-graduates		
Difference	*0.112*		*0.085*	
	(0.012)		*(0.014)*	
Unexplained	*0.096*		*0.044*	
	(0.015)		*(0.015)*	
Early life	-0.002	0.030	-0.001	0.016
	(0.001)	(0.019)	(0.001)	(0.012)
Qualifications	-0.013	0.013	-0.009	0.030
	(0.003)	(0.010)	(0.002)	(0.014)
Experience	0.021	0.069	0.019	0.071
	(0.004)	(0.060)	(0.004)	(0.085)
Marital	-0.003	-0.016	0.002	-0.012
	(0.001)	(0.022)	(0.002)	(0.019)
Major group occupations	-0.007	-0.009	-0.024	-0.056
	(0.007)	(0.007)	(0.009)	(0.029)
Female proportion in job	0.052	0.004	0.079	-0.033
	(0.012)	(0.028)	(0.011)	(0.041)
Hours	-0.032	-0.015	-0.025	-0.002
	(0.007)	(0.040)	(0.005)	(0.015)
N	4,292	4,292	2,723	2,723

Note: Standard errors in parentheses.