The Gender Wage Gap Over the Life-course: Does the Genetic Predisposition For Educational Attainment Matter?

Alex Bryson, Tim Morris, David Bann, Dave Wilkinson

UCL's Social Research Institute

CLS Lunchtime Talk

October 15th 2024

ESRC Grant No. ES/S012583/1

Motivation (1)

1. Education, together with work experience, is part of ‘human capital’ (Becker, 1964) which determines the worth of an individual to employers, and thus their wage
   - Reflected in the Mincerian wage equation (Mincer, 1958) which dominates economics
2. In the absence of full information on individuals’ productivity qualifications signal worker quality (Arrow, 1973; Spence, 1973)
   - Again, higher qualifications will raise earnings
3. Polygenic scores (PGSs) derived from Genome-wide Association Studies (GWASs) account for 12-16% of variance in educational attainment (Okbay et al, 2022)
4. By virtue of random genetic segregation, there should be no mean differences in the PGS by sex, yet for much of the post-war period men had higher educational attainment than women likely due to societal expectations of gendered roles (Becker, 1985)
   - Efficient division of household labour affects incentives to invest

Motivation (2)

1. Don't expect this genetic predisposition to vary by gender but women may:
    - Have fewer opportunities to convert this genetic predisposition into educational qualifications e.g. due to gendered norms
    - Face reduced returns to educational 'ability' e.g. discrimination (Becker, 1957; Phelps, 1972)
    - Choose to invest less in their human capital (Adda et al., 2017)

2. Research questions:
    - How much of the variance in men's and women's employment and earnings is accounted for by the genetic predisposition for educational attainment?
    - How does this association with employment and earnings vary at different points in the life-cycle?

Hypotheses

1. PGS for education -> +ve for probability of employment and time in employment
    - Larger effects for women due to +ve selection into employment – especially FT employment
    - More pronounced when women are of childrearing age
2. PGS for education -> +ve for earnings
    - attenuated in the case of women e.g. if they face barriers converting their 'ability' into higher earnings, e.g. due to discrimination

Overview of what we do

1. For a cohort born in 1958 we create polygenic scores for educational attainment from their genetic data and plug it into employment and wage equations
2. Recover the partial correlation between the PGS and (a) % time in employment (b) employment at point in time between ages 20 and 55 (c) log hourly earnings at interview
3. Models are parsimonious containing not much else.
4. Check sensitivity to parental education
5. Deal with some technical challenges, including accounting for selection into employment, survey sample attrition

Results Overview

1. PGS do not vary by sex at any point in achieved sample to age 55
2. PGS positively associated with % time in employment and % time in FT employment. Effects are larger for women
   - For FT employment PGS association is confined to women.
3. PGS accounts for a substantial percentage of the variance in earnings for men and women – 4-5%
4. At ages 33, 42, 50 and 55 1 s.d. increase in PRS is associated with a 6-10 log point increase in log hourly earnings
   - Persistent earnings advantage
   - Does not differ significantly across men and women
5. At age 23 PGS is not significant for male earnings but 1 s.d. increase raises female earnings by 5 log points.
   - Also true at age 33 without imputed earnings
6. Lower parental education (leaves at compulsory school leaving age) reduces cohort members' earnings but has little impact on the PGS coefficient in earnings equations

Literature

1. PGS for educational attainment positively associated with educational attainment (Okbay et al 2022)
2. PGS for educational attainment is positively associated with earnings, solely through its impact on educational attainment (Hogan-Hennessy 2024)
3. Those with higher genetic endowments for educational attainment benefit more from being first born than first borns with lower genetic endowments suggesting nature and nurture interact (Muslimova et al., 2024)
4. Conditioning on parental education helps tackle possibility that indirect genetic effects may inflate PGS associations (Morris et al., 2020)
   - because genes are shared between the cohort member and their parents, indirect genetic effects can arise if the shared genes have an impact on their parents' education, which in turn impacts the cohort member's home environment, for example through parenting
5. PGS for educational attainment has not featured in the gender wage gap literature

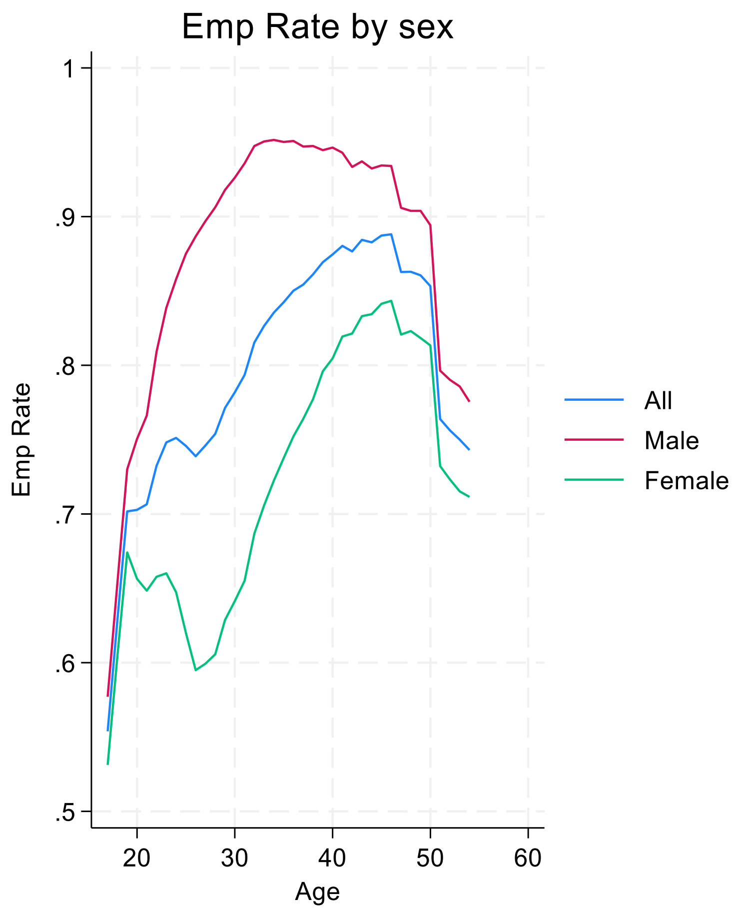
Data

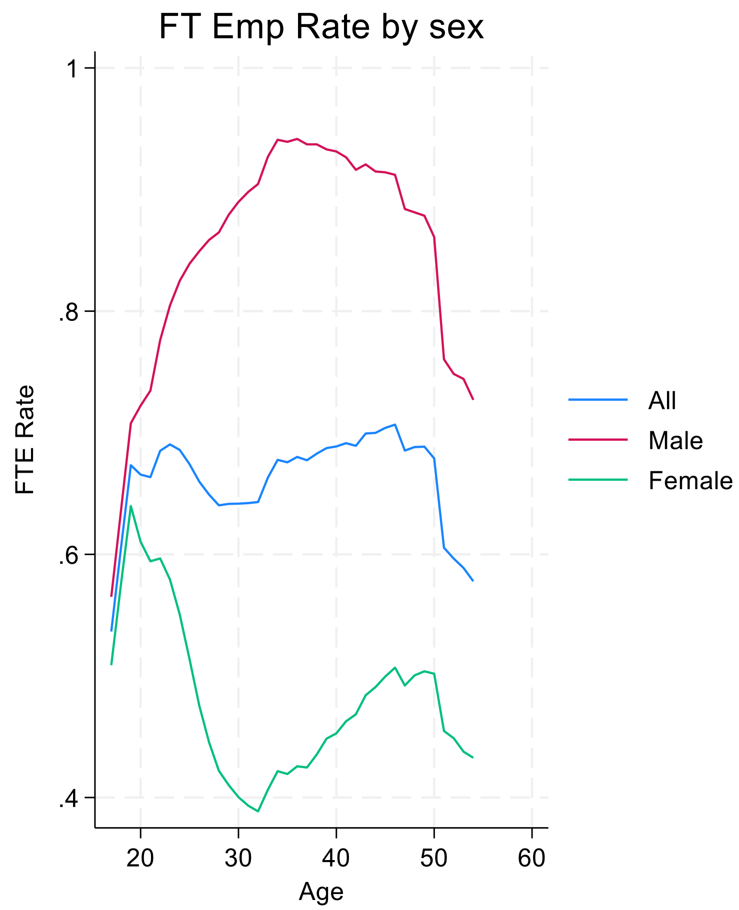
1. National Child Development Study (NCDS)
    - Cohort born in March 1958; age 55 in 2013
2. Dependent variables
    - % time in employment between age 20-55
    - % time in full-time employment between age 20-55
    - Employment and FT employment status at ages 23, 33, 42, 50, 55
    - Log hourly wages at ages 23, 33, 42, 50, 55
        - Gross wage, deflated to January 2000 prices
        - Dividing gross earnings including paid overtime by last or usual hours
        - With and without wage imputation
        - Drop outliers at top and bottom 1% of hourly earnings
3. Polygenic score (PGS) for educational attainment
    - SNPs from genetic code that predict educational attainment in genome-wide association studies (GWASs)
    - Collected at age 44 (2002) in Biomedical Survey
    - Two variants

Polygenic Scores for Educational Attainment

1. PGSs measure genetic predispositions towards traits using counts on individuals' alleles at specific genomes on the chromosome
2. PGSs are population-weighted estimates of the association between that allele and the outcome
3. By virtue of random genetic segregation, there should exist no mean differences in PGS between women and men as alleles on the autosomes are inherited randomly irrespective of sex.
4. Therefore, differential PGS associations with outcomes among women and men may be taken as indicative of sex-specific social or environmental factors (barriers).
5. Genotyped 13,738 samples for 6,431 individuals in NCDS age 44 of which 6,312 had useable data after quality control
    - Small Ns dropped due to data missingness, individuals related to one another, non-European ancestry
6. PGS constructed based on largest GWAS to date (Okbay et al., 2022)
7. Also derived a second more restrictive PGS

Employment and Full-time Employment Over the Life Course, by Sex





Estimation

1. Regression analyses containing standardized PRS (mean of zero, s.d of 1)
   - Interacted with sex
   - Separate regressions by sex
2. In earnings estimates we account for employment probability
   - Estimates with and w/out imputed earnings based on propensity score matching
   - Includes dummy variable identifying those with imputation
3. Time in employment estimation samples
   - 5,908 (2,913 men, 2,995 women)
   - Of these, 78% of men and 82% women had responded to 8 or 9 of the 9 sweeps between 7 and 55 years
   - Control for N years missing employment data (1.4 years for men, 1.8 for women)
4. Employment at each sweep models account for attrition with weights (1/prob of response each wave by sex)
5. All employment and earnings models include 20 principal components of inferred genetic structure to minimise likelihood of residual population stratification bias

Results

% Time in Employment Between Age 20-55

	Men	**Men**	**Women**	**Women**
Panel A: All Employment				
Standardized PGS	0.005+	**0.006***	**0.023***	**0.023***
	(1.84)	(2.07)	(5.21)	(5.09)
Parental Education	No	Yes	No	Yes
Adj R-sq	0.019	0.019	0.011	0.012
Panel B: Full-time Employment				
Standardized PGS	0.004	**0.005+**	**0.033***	**0.030***
	(1.29)	(1.65)	(5.73)	(5.20)
Parental Education	No	Yes	No	Yes
Adj R-sq	0.013	0.014	0.020	0.023
Unweighted N	2913	2913	2995	2995
Notes: (1) T-stats in parentheses (2) Significance: + p<0.10 * p<0.05 (3) All models contain 20 principal components of the PGS and a count variable for the number of missing years of work history.				

Employment and FT Employment at Survey Sweeps

	23	**33**	**42**	**50**	**55**
Panel A: All Employment, Men					
Standardized PGS	0.003	0.003	0.006	0.008	0.002
	(0.40)	(0.33)	(0.69)	(0.87)	(0.17)
Adj R-sq	0.008	0.005	0.010	0.008	0.010
Panel B: Full-time Employment, Men					
Standardized PGS	0.001	0.002	0.005	0.004	-0.011
	(0.17)	(0.20)	(0.60)	(0.46)	(1.10)
Adj R-sq	0.007	0.004	0.009	0.006	0.010
Panel C: All Employment, Women					
Standardized PGS	**0.070***	**0.028***	0.012	0.013+	-0.000
	(7.89)	(3.04)	(1.42)	(1.61)	(0.00)
Adj R-sq	0.028	0.012	0.005	0.008	0.007
Panel D: Full-time Employment, Women					
Standardized PGS	**0.075***	**0.055***	**0.022***	0.016+	0.007
	(8.21)	(6.28)	(2.51)	(1.70)	(0.71)
Adj R-sq	0.029	0.025	0.011	0.007	0.008
Unweighted N, Men	2,691	2,765	3,037	2,733	2,556
Unweighted N, Women	2,797	2,897	3,080	2,822	2,687
Notes: (1) T-stats in parentheses (2) Significance: + p<0.10 * p<0.05 (3) All models contain 20 principal components of the PGS and are weighted to account for attrition.					

Raw Gender Gap in Log Hourly Earnings at Survey Sweeps

	23	**33**	**42**	**50**	**55**
Panel A: Employees With a Valid Hourly Wage					
Female	-0.157*	-0.336*	-0.396*	-0.329*	-0.304*
	(22.93)	(28.51)	(27.47)	(24.56)	(19.76)
R-sq	0.07	0.11	0.10	0.10	0.08
Panel B: Including Imputed Earnings					
Female	-0.198*	-0.345*	-0.389*	-0.345*	-0.307*
	(34.87)	(37.50)	(33.83)	(32.41)	(25.95)
R-sq	0.13	0.14	0.11	0.11	0.08
Unweighted N, Panel A	7,528	6,588	7,001	5,901	4,819
Unweighted N, Panel B	11,693	10,839	10,981	9,386	8,501
Notes: (1) T-stats in parentheses (2) Significance: + p<0.10 * p<0.05 (3) Regressions are weighted with attrition weights. (4) Panel B models incorporate a dummy variable identifying those with imputed earnings.					

Log Hourly Earnings at Survey Sweeps

	23	**33**	**42**	**50**	**55**
Panel A: Men					
Standardized PGS	0.011	**0.089**	**0.130**	**0.110**	**0.128**
	(1.62)	(8.27)*	(9.81)*	(8.36)*	(9.03)*
Adj R-sq	0.015	0.052	0.053	0.056	0.77
Panel B: Women					
Standardized PGS	**0.065**	**0.137**	**0.121**	**0.119**	**0.098**
	(9.48)*	(12.70)*	(9.63)*	(10.90)*	(8.20)*
Adj R-sq	0.062	0.088	0.054	0.054	0.058
Panel C: Pooled					
Female	-0.165	-0.347	-0.411	0.352	-0.311
	(17.09)*	(22.74)*	(21.90)*	(21.19)*	(16.88)*
Standardized PGS	0.011	0.088	0.132	0.110	0.128
	(1.65)+	(8.23)*	(9.95)*	(8.38)*	(9.02)*
Female*Standardized PGS	**0.054**	**0.050**	-0.010	0.009	-0.030
	(5.68)*	(3.31)*	(0.54)	(0.53)	(1.59)
Adj R-sq	0.095	0.181	0.150	0.158	0.140
Unweighted N, Men	1,941	1,932	2,007	1,681	1,418
Unweighted N, Women	1,691	1,588	1,938	1,861	1,547
Unweighted N, Pooled	3,632	3,520	3,945	3,542	2,965

Notes: (1) T-stats in parentheses (2) Significance: + p<0.10 * p<0.05 (3) All models contain 20 principal components of the PGS and are weighted to account for attrition.

Log Hourly Earnings at Survey Sweeps Including those with Imputed Earnings

	23	**33**	**42**	**50**	**55**
Panel A: Men					
Standardized PGS	0.008	**0.066**	**0.097**	**0.085**	**0.075**
	(1.45)	(7.49)*	(8.97)*	(8.11)*	(6.54)*
Adj R-sq	0.011	0.039	0.046	0.037	0.034
Panel B: Women					
Standardized PGS	**0.045**	**0.083**	**0.076**	**0.096**	**0.062**
	(7.91)*	(9.60)*	(7.46)*	(10.25)*	(6.51)*
Adj R-sq	0.028	0.045	0.035	0.050	0.027
Panel C: Pooled					
Female	-0.196	-0.350	-0.392	-0.364	-0.327
	(23.94)*	(27.98)*	(25.66)*	(26.66)*	(22.10)*
Standardized PGS	0.009	0.065	0.096	0.084	0.076
	(1.62)	(7.38)*	(8.92)*	(8.07)*	(6.55)*
Female*Standardized PGS	**0.038**	0.018	-0.019	0.013	-0.012
	(4.74)*	(1.43)	(1.29)	(0.93)	(0.83)
Adj R-sq	0.141	0.168	0.133	0.148	0.113
Unweighted N, Men	2,641	2,707	2,974	2,642	2,474
Unweighted N, Women	2,746	2,814	2,996	2,760	2,542
Unweighted N, Pooled	5,387	5,521	5,970	5,402	5,016

Notes: (1) T-stats in parentheses (2) Significance: + p<0.10 * p<0.05 (3) All models contain 20 principal components of the PGS, a dummy variable identifying those with imputed earnings, and are weighted to account for attrition.

Sensitivity Tests

1. Inclusion of parental education (whether stayed on after compulsory education)
   - Associated with cohort member's employment and earnings but does not affect coefficient on PGS
2. Second PGS for education
   - Similar results throughout
3. Those with lower earning potential die more quickly (Fluharty et al 2021) and attrit (Hawkes and Plewis, 2006). Implies those with higher (lower) earnings – including PGS – are more (less) likely to be surveyed over time
   - Higher PGS for educational attainment predicts responding to more sweeps
   - But not significantly different by sex
   - At no sweep did PGS association with probability of response differ by sex
   - At no sweep was there any sex difference in mean PGS

Summary

1. Those with higher PGS for education spend more time in employment and FT employment and, when in employment, earn higher hourly wages
2. The employment associations are more pronounced for women than men
   - 4-times greater (1 sd change in PGS -> 0.5pp in male employment, 2.3pp for women)
   - Association with FT employment is confined to women
   - Associations vary over life-course as anticipated
3. Conditional on employment, the PGS wage associations are sizeable, persistent and similar for men and women between ages 33 and 55
   - A 1 sd increase in the PGS is associated with a 6-10 log point increase in hourly earnings when 33-55 years
   - However, at age 23, whereas a 1 sd increase in PGS is associated with a 5 log point rise in women's earnings it is not associated with earnings for men
   - So, no evidence that women's wage returns to PRS for education are lower than for men (cf Psacharopoulos and Patrinos, 2018 who find the same for educational qualifications)
4. These associations are robust to non-random selection into employment and controls for parental education

Additional Work

1. Association between PGS and educational attainment by sex
    - Same for men and women?
2. Effect of PGS on educational attainment coefficients in wage equations
    - Are both statistically significant?
    - Does one impact the other?
3. Moving beyond the mean to associations across the distribution
4. PGS for educational attainment as an instrumental variable?
    - Controversy regarding the exclusion restriction
5. Use Millenium Cohort Survey (MCS) with both parents' genetic data to distinguish between nature and nurture?
6. PGSs for other traits to examine gender gaps e.g. to investigate gender wellbeing gap and effects of reproductive health among women

The Paper

The Gender Wage Gap across Life:

Effects of Genetic Predisposition towards Higher Education

IZA Discussion Paper No. 17255

Bonus Slide

Imputing Earnings to Account for Non-Random Selection into Employment

1. Impute hourly wage for 4 types of respondent
   - in work, no valid wage; self-employed; unemployed; economically inactive
2. Imputed wages come from nearest neighbour wage ‘donors’ defined as those in the waged employment group at the same sweep from the same sex who are nearest in their propensity for waged employment to the non-waged individual
3. Propensity scores derived from (0,1) probits for having valid wage by sex and sweep where estimation sample is those with valid wage plus one of 4 types of non-wage respondents
4. Matching equation contains information on cohort members from childhood (e.g. test scores), birth (e.g. N siblings) and parental information (e.g. social class)
5. Adjust for sweep non-response with attrition weights
6. Drop those whose predicted employment probability is below the lowest probability for the waged employee sample at that sweep
7. Except for the self-employed, whose imputed earnings are like those of employees, imputed earnings are substantially below those of employees posting an hourly wage in the case of both men and women