I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16800

The Labour Market Returns to Graduation: Reconciling Administrative and Survey Data Estimates

Anna Adamecz
Matt Dickson
Nikki Shure

FEBRUARY 2024

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16800

The Labour Market Returns to Graduation: Reconciling Administrative and Survey Data Estimates

Anna Adamecz
KRTK KTI and UCL

Matt Dickson
University of Bath and IZA

Nikki Shure
University College London and IZA

FEBRUARY 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Labour Market Returns to Graduation: Reconciling Administrative and Survey Data Estimates

This paper contributes to the literature on the earnings returns to university graduation. Recent evidence using administrative earnings data from England suggests a zero return to graduation for men and positive returns to graduation for women in annual earnings at age 26. We show that once hours worked are taken into account – typically not available in administrative tax data – returns to graduation are zero for women too. Graduate women work more hours than comparable non-graduate women, explaining their annual earnings return, but in terms of hourly wages, average returns to graduation at this early career stage are around zero for both sexes. This highlights the importance of using both survey and administrative data sources when estimating the returns to university graduation.

JEL Classification: I23, I26

Keywords: returns to graduation, university, gender differences, survey data

Corresponding author:
Matt Dickson
Department of Social and Policy Sciences
University of Bath
Bath BA2 7AY
United Kingdom
E-mail: m.dickson@bath.ac.uk

1. Introduction

The labour market value of higher education is a topic of keen public and policy debate in many advanced economies. This is in part due to the nature of the public investment and there being both public and private returns which arise from it. A range of literature has highlighted the positive returns to the individual from completing a university degree – from higher earnings (Britton et al., 2022; Maurin & McNally, 2008; Webber, 2016), to better health (Herd et al., 2007; Raghupathi & Raghupathi, 2020), to positive assortative mating (Elsayed & Shirshikova, 2023; Hu & Qian, 2016) – see Oreopoulos & Petronijevic (2013) for a review. This has raised concerns about how higher education should be funded and the debate has been particularly salient in England where participation in higher education has increased markedly from around 15% in 1990 to just over 53% in 2019/20 (Office for National Statistics (ONS), 2021; Walker & Zhu, 2013). This expansion has been accompanied first by the introduction of tuition fees (1998) and then sharp increases in them (2006, 2012) to fund the extra supply of places, with fees now standing at £9,250 per year for a full-time undergraduate course (Wyness, 2010). For most students, these fees and maintenance costs are paid for up-front by government loans that are then repaid by graduates once they are in the labour market and earning above a certain salary threshold.

This rebalancing of higher education costs towards the individual has invigorated the research literature on the estimated value of university graduation to graduates**,** who are now expected to repay the cost of their tertiary education. The most recent research on the topic in England (Belfield et al., 2018) has exploited newly available earnings data from tax records, linked with administrative information on higher education participation, prior educational attainment and family background, to estimate the returns to higher education.[1] This type of rich administrative data is increasingly being made available to researchers in England and elsewhere. A limitation with this data, however, is that earnings information comes from tax returns, which are calculated annually, and therefore reflect both hourly earnings and annual hours worked. This issue is common to estimates of returns to higher education using administrative data in numerous other countries too (Anelli, 2020; Zimmerman, 2019; Kirkeboen et al., 2016; Hastings et al., 2013).

For women in particular, differences in annual hours between graduates and non-graduates can distort the apparent graduate premium. Across the world, women are more likely to work part-time than their male counterparts (European Commission, 2013), for example over all OECD countries in 2022, on average 24% of employed women are working part-time, but only 9.6% of employed men are part-time (OECD, 2022). This rises to 32.9% for employed women in the UK (11.2% for men), moreover non-graduate women are especially are more likely to work part-time compared to graduate women (Department for Education, 2023). As such, recent estimates on the returns to university graduation that do not account for hours worked may present an incomplete picture and lead to incorrect policy conclusions on the gender differences in the returns to graduation.

[1] This linked administrative data resource is known as the Longitudinal Education Outcomes (LEO) dataset.

In this paper, we use data from an English longitudinal cohort study linked with administrative schooling data and self-reported higher education and hourly wage information to estimate the hourly earnings returns to graduation. We find that, as per the recent literature, in annual earnings at age 25/26 there is a positive premium for women (6.9%) but zero for men. When we take hours worked into account, the return to graduation in hourly wages falls and is not significantly different to zero for either sex. Our findings highlight the continuing importance of using survey data to complement studies undertaken using administrative data.

2. Recent Literature

The positive association between higher education and a wide range of later life outcomes, and the extent of causation in these relationships, is the subject of an extensive literature in economics and the social sciences more generally, see Oreopoulos & Petronijevic (2013) and Hout (2012) for comprehensive reviews. The most populous sub-division within the literature on the returns to higher education focuses on estimating the impact of an undergraduate (bachelor's) degree on labour market earnings. This literature has in recent years been reinvigorated by the increasing availability of linked administrative datasets that provide accurate measures of background characteristics, prior educational attainment, university subject and institution and crucially earnings from national (or state) tax registers.

In the US this has seen several recent papers exploiting state-level administrative datasets to go beyond estimating the return to a higher education degree and look at returns to specific college majors and how they vary according to the quality of match between the student and the course. Andrews et al. (2022) exploit earnings data from Texas, linked with school and college information, to show the return to different majors and how the subject of major also affects earnings growth and variability. Similarly, Mountjoy & Hickman (2021) exploit administrative data from Texas to estimate the value-added of the state's public universities, and how this varies by college selectivity and student characteristics. This follows earlier work by Dale & Krueger (2014) exploring the relationship between college selectivity and earnings returns. They use the College and Beyond survey linked to Social Security Administrative earnings data and find that while there is a high return to college selectivity, once the student choice sets were controlled for, these selectivity returns fall to zero, albeit with some large returns remaining for Hispanic and black students. Jepsen et al. (2014) use linked administrative data from Kentucky to estimate the earnings returns to community college degrees and similarly to Andrews et al. (2022) demonstrate how returns to associate degrees, diplomas and certificates vary substantially according to subject. Likewise, Liu et al (2014) estimate returns to community college qualifications, up to and including bachelor's degrees, attained in the North Carolina Community College system, showing that while the returns to certificates and diplomas were low, there were strong returns for associate's and bachelor's degrees, with returns for females being higher than for males.

Outside the US, numerous studies have used administrative data to examine the returns to degrees and the importance of institutional selectivity and/or quality, subject of major, and individual characteristics in determining the return. Anelli (2020) estimates a substantial return to enrolment in a highly selective elite higher education institution in Italy, exploiting a unique dataset of applicants and attendees linked to their tax register earnings. Hastings et al. (2013)

find large positive effects of enrolment on selective degree programs and for particular subjects (Health, Sciences and Social Sciences) in Chile, with little variation in returns to selectivity by students' socio-economic status. Conversely, focusing more narrowly on elite business-focused degrees, Zimmerman (2019) finds that the large returns associated with these particular programs in Chile are completely driven by males from high-tuition, private secondary schools, with zero returns for females or males from other school types. For Norway, Kirkeboen et al. (2016) find that returns to selectivity are low relative to the variation related to subject of major, with Sciences, Technology, Business and Law consistently providing high returns. In many cases they find the return to a particular field is as large as the college premium itself suggesting that the choice of major is as important as the initial enrolment decision. This conclusion is echoed for England where Belfield et al. (2018) were the first to exploit the availability of linked administrative registers to estimate the return to an undergraduate degree, and how this varies by subject and institution. They find an overall earnings return at age 29 of 26% for women and 6% for men but substantial variation around this by both choice of subject and institution.

However, a common feature of this recent literature from around the world exploiting linked administrative datasets, is that earnings are recorded on an annual or quarterly basis and have no adjustment for hours worked. This limitation is likely to be particularly acute for women, given their greater likelihood of part-time work (Blau & Kahn, 2017), and presents an additional issue when looking at graduate premia given differences in hours worked between graduate and non-graduate women. In this paper, we overcome this widespread issue by using information on hours worked as well as annual earnings to compare the returns to a degree in annual and hourly earnings, highlighting the importance of this more detailed information for returns estimates and the policy implications that derive from them.

3. Data and methods

We use Next Steps, a longitudinal study which follows a cohort born in 1989/1990 and comprises eight waves of data up to the age of 25/26 (University College London, UCL Institute of Education, Centre for Longitudinal Studies, 2018). Next Steps has been linked with the National Pupil Database (NPD), which provides a census of pupils attending schools in England, allowing us to access their school exam results. This includes compulsory, high-stakes, end of secondary school (GCSE) exams, and the exams typically necessary for university entry (A-level exams).

Next Steps is the closest English cohort study in age that matches the administrative data used in Belfield et al. (2018). The young people in their analysis took their GCSEs between 2002-07, while the young people in Next Steps took their GCSEs in 2005-06. This means we should be able to broadly replicate their results with our sample.

The eighth wave of Next Steps covers 7,707 individuals, however, following Belfield et al. (2018), we restrict the sample to those who have at least 5 A*-C GCSEs (this is usually the minimum attainment threshold for progression to study university entry qualifications), and are in sustained employment i.e. had paid employment at the time of data collection, and worked for at least five of the previous six months. Our sample consists of 1,112 men and 1,593 women.

We look at three outcome variables: log annual wage, log hourly wage, and hours worked, all observed on average at age 26 (Figure A1 in the Appendix). This is slightly younger than the primary age examined by Belfield et al. (2018), who look at annual earnings at age 29; however, they also produce earlier age estimates for this cohort (see Table 12 in Belfield et al., 2018). Following their methods, we control for the following characteristics:

- demographic and family background: age in months, mother's and father's social class (NS-SEC), region, ethnicity;
- early and pre-university educational attainment: GCSE and A-level (age 18) raw scores, indicator variables for A-level subjects (Math, Sciences, Social Science, Humanities, Arts, Languages and Other), a binary variable for having vocational qualifications, a binary variable capturing whether the individual attended a private (fee paying) secondary school at age 13/14.

Descriptive statistics for the sample[2] are shown in Appendix Table A1. Just over half (56.5% of men, 55.6% of women) of our sample graduated from university by age 26 which gives us confidence that most of the individuals who attend higher education will have completed by this point. Table A1 breaks down the descriptive statistics by gender and by graduate or non-graduate status. We can see that both men and women select into higher education on the basis of pre-university characteristics. Male graduates in particular are much more likely to be from higher social class than non-graduates: for 54.2% of graduates their father is in the highest social class (NS-SEC groups 1-2) compared to 40.2% for non-graduates, with corresponding figures of 44.8% v 32.9% for mother's. Graduates are more likely than non-graduates to have been educated at an independent school (5.8% versus 3.7% for males, 5.1% versus 3.4% for females), and have higher attainment at age 16: average GCSE points are 509.5 (503.9) for male (female) graduates versus 479.5 (475.2) for non-graduates. Those who do not go on to attain a university degree are much more likely to study a vocational qualification at level 3, particularly amongst males (42.6% vs. 27.3%), and for those who do study A-levels, non-graduates are more likely to be in the lower quintiles of attainment at A-level and less likely to be in the highest quintiles. These differences highlight the importance of controlling for background characteristics and prior attainment when estimating the returns to graduation. The raw figures in Table A1 show that amongst men, graduates and non-graduates work approximately the same average weekly hours (41.82 vs. 41.02) with hourly wages slightly higher for graduates (£13.16 vs. £12.72). For females, hourly wages see a greater raw graduate premium (£12.11 vs. £11.26) but weekly hours are notably higher for female graduates than non-graduates (40.07 vs. 37.38), underlining the importance of taking work hours into account.

Our methods follow Belfield et al. (2018) since we are firstly aiming to replicate their estimates for annual earnings before going on to examine the impact on returns when we account for

[2] The descriptive statistics in Table A1 are for our preferred IPWRA estimation sample which is slightly smaller than the full sample but characteristics are very similar for both.

hours worked. We estimate standard Mincer-type wage models using ordinary least squares (OLS), separately by gender:

$$ln\, y_i = \alpha + X_i'\beta + \gamma Grad_i + \varepsilon_i \quad (1)$$

in which y_i is either annual earnings, hourly wage, or weekly hours, X_i is the vector of the control variables listed above, $Grad_i$ is an indicator for being a graduate, and ε_i is a well-behaved error term. Standard errors are clustered at the school level.

In Model 1, we look at the raw wage difference between graduates and non-graduates. In Model 2, we control for all variables listed above. Lastly, in model 3 we apply inverse probability weighting regression adjustment (IPWRA) (Wooldridge, 2007), which reweights the sample so that the first moments of the control variables do not differ between graduates and non-graduates. We operationalize the IPWRA approach using "teffects" in Stata (StataCorp, 2013)[3], and we do the estimation on the common support sample of graduates and non-graduates only. Hence, the sample size is somewhat smaller in our preferred approach of Model 3 than in Models 1 and 2. However, as a robustness check we restrict the sample for all models to be the same as for the IPWRA estimates and results remain the same (see Appendix Tables A2 and A3).

4. Results

Our results on the returns to graduation in terms of log annual earnings are similar to those of Belfield et al. (2018) using administrative data. Without controlling for any background characteristics or prior attainment, male graduates earn on average 8% (8.0 log points, Table 1) more than their non-graduate peers, whereas for female graduates the average premium is 24% (21.7 log points). This is similar to Belfield et al. (2018)'s pattern of findings at age 26 without any controls (Figure 2, Belfield et al., 2018, p.16). Once we control for background characteristics and apply IPWRA, the estimated coefficients reduce to an insignificant 0.2% for men but a still significant at 7% for women. This again is similar to Belfield et al. (2018), who find -3% returns to higher education for men and 14.9% for women (results in their Table 12, p.63) at this age.[4] It is worth noting that our sample comprises those aged 25 and 26, which makes our estimates comparable to something between the age 24 and age 26 estimates of Belfield et al (2018), e.g. for women this would be between 0.04 log points and 0.13 log points, so close to our estimate of 0.069 log points. This broad congruence is reassuring given our earnings data is self-reported whereas Belfield et al. (2018) have access to administrative tax records.

[3] We estimate the IPWRA weights separately by gender, using the following control variables: age, ethnicity, region, father's and mother's social class, boost sample indicator, private school, GCSE and A-level (age 18) raw scores, indicator variables for A-level subjects (Math, Sciences, Social Science, Humanities, Arts, Languages and Other), and having a vocational qualification.

[4] Belfield et al. (2018) estimate returns to HE attendance rather than graduation and so their estimates include those who drop-out as well as graduates; however, restricting to graduates only has little effect on their estimates, see Table 8, p. 38.

Table 1: Returns to graduation: log annual earnings and hourly wages

	(1)	(2)	(3)	(4)	(5)	(6)
		Men			Women	
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
			Outcome: log annual earnings			
Returns to graduation	0.080***	0.002	0.002	0.217***	0.073**	0.069**
	(0.029)	(0.034)	(0.032)	(0.028)	(0.030)	(0.029)
Constant	NR	9.351***	8.155***	NR	7.510***	7.091***
		(0.946)	(1.636)		(0.913)	(1.176)
No. of unweighted observations						
R-squared	1,342	1,342	1,112	1,872	1,872	1,593
			Outcome: log hourly wage			
Returns to graduation	0.073***	-0.014	-0.016	0.104***	0.026	0.026
	(0.027)	(0.030)	(0.028)	(0.022)	(0.023)	(0.022)
Constant	NR	2.254***	1.254	NR	-0.366	-1.259
		(0.826)	(1.150)	(0.016)	(0.693)	(1.085)
No. of unweighted observations						
R-squared	1,342	1,342	1,112	1,872	1,872	1,593
			Control variables and weighting			
Control variables		Yes	Yes		Yes	Yes
IPWRA weighting			Yes			Yes
Wave 8 sampling weights	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Sample of those having at least 5 A*-C GCSE examinations and in sustained employment. Model 1 and 2: linear regression models estimated by OLS, weighted using wave 8 weights. Model 3: IPWRA-weighted regressions. Robust standard errors clustered by sampling school are in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Control variables: Sample boost: whether the individual belongs to the sample boost added to the survey in Wave 4. Family background: age in months as a continuous variable, mother's and father's social class, region, ethnicity. Early and pre-university educational attainment: GCSE and A-level raw scores, indicator variables for A-level subjects as Math, Sciences, Social science, Humanities, Arts, Languages and Other, a binary variable for having vocational qualifications, a binary variable capturing whether the individual attended independent secondary school at age 13/14. Missing observations are controlled for using missing flags.

Source: University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2018). Next Steps: Sweeps 1-8, 2004-2016: Secure Access. DOI: 10.5255/UKDA-SN-7104-4

Looking at log hourly wages (lower panel Table 1), however, reveals the importance of hours worked in generating these premia, in particular for women. In the uncontrolled regressions (Model 1) returns to graduation for women are about half as large as they are for annual wages (11% compared to 24%). Once we control for individual characteristics (Model 2) and employ

IPWRA (Model 3), returns to graduation become small (2.6%) and insignificant for women, as they are for men.

Table 2: Returns to graduation: number of hours worked per week

	(1)	(2)	(3)	(4)	(5)	(6)
		Men			Women	
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
			Outcome: number of hours worked per week			
Graduation	0.202	0.737	0.779	3.706***	1.641***	1.515***
	(0.547)	(0.586)	(0.568)	(0.530)	(0.547)	(0.545)
Constant	NR	27.835	18.467	NR	47.683***	52.035**
		(18.592)	(27.384)		(16.795)	(24.265)
No. of unweighted observations						
R-squared	1,342	1,342	1,112	1,872	1,872	1,593
			Control variables and weighting			
Control variables		Yes	Yes		Yes	Yes
IPWRA weighting			Yes			Yes
Wave 8 sampling weights	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Sample of those having at least 5 A*-C GCSE examinations and in sustained employment. Model 1 and 2: linear regression models estimated by OLS, weighted using wave 8 weights. Model 3: IPWRA-weighted regressions. Robust standard errors clustered by sampling school are in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. Control variables: Sample boost: whether the individual belongs to the sample boost added to the survey in Wave 4. Family background: age in months as a continuous variable, mother's and father's social class, region, ethnicity. Early and pre-university educational attainment: GCSE and A-level raw scores, indicator variables for A-level subjects as Math, Sciences, Social science, Humanities, Arts, Languages and Other, a binary variable for having vocational qualifications, a binary variable capturing whether the individual attended independent secondary school at age 13/14. Missing observations are controlled for using missing flags.

Source: University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2018). Next Steps: Sweeps 1-8, 2004-2016: Secure Access. DOI: 10.5255/UKDA-SN-7104-4

Table 2 shows the same models, with hours worked per week as the outcome variable. Interestingly, it is only true for women that graduates work more hours than non-graduates. In the raw model, the difference is on average 3.7 hours per week[5]; this reduces to 1.5 hours after controlling for characteristics and applying the IPWRA estimator. This relatively small difference in weekly working hours between graduate and non-graduate women is enough to

[5] The raw difference in Table A1 of descriptive statistics (2.68) refers to the common support sample used for the IPWRA estimates (model 3) and is slightly smaller than the 3.7 difference for the full sample used for model 1 in Table 2. The estimates of model 1 for the smaller sample are in Table A3 where the estimated impact of graduation on hours is 2.681 for women as per the descriptive statistics.

reduce the magnitude of the graduation premium at this early career stage by almost two-thirds, and render it statistically insignificantly different from zero.[6]

5. Discussion

The availability of administrative tax records linked to individual education and background information in numerous countries has allowed estimates of the return to graduating from university to be estimated on large samples, providing new insights on graduate premia. However, a widespread limitation in these studies is the lack of information on hours worked, hence these premia reflect both the productivity enhancement from higher education (hourly wage premia) and the effect on hours worked. Our results for England, using a smaller sample but with richer labour market information than is available from administrative records, show the importance of adjusting for hours worked when estimating returns to graduation, particularly for women where patterns of employment and fertility differ between graduates and non-graduates.

The differences between female graduates and non-graduates in terms of employment have been explored in the UK Labour Force Survey. In 2015, the year of data collection for Next Steps, 31.2% of female graduates worked part-time as compared to 46.8% of female non-graduates (Department for Education, 2023). In contrast, the difference in part-time working between male graduates and non-graduates was less than five percentage points (8.1% and 12.7% respectively). Therefore, failing to take hours worked into account can result in the returns to graduation being heavily overestimated for women.

For both sexes the estimated hourly earnings premium for university graduation is small (negative for men) and insignificantly different from zero at age 26. However, this is a relatively early stage in the labour market career of graduates who would typically have four or five years of labour market experience at this point, as compared with around eight years for their comparable peers who entered the labour market at age 18. This difference in experience is relatively large and plays a part in limiting the graduate premium at this point in the career. As further experience is gained, graduates tend to have a steeper earnings profile, with premia increasing into mid-career, resulting in positive lifetime returns to higher education for both men and women (Britton et al., 2020).

One important takeaway from our results is the robustness of self-reported earnings. In terms of magnitude, our annual earnings results broadly replicate findings from administrative data, which indicates a high degree of accuracy in self-reporting, at least for this cohort study. Previous work has highlighted that certain types of income are more accurately reported (e.g. regular, monthly sources of income) than others (Alwin et al., 2014). Given the relatively young age of the individuals in our sample, their primary source of earnings will be labour

[6] The difference in working hours of 1.5 hours per week for graduate women compared to non-graduate women translates to 78 hours per year, which is 4% of the unconditional average annual hours of non-graduate women (1,955). Adjusting the annual return (1.069) for this 4% difference in hours gives an approximate expected hourly return of $1.069/1.040=1.028 \approx 2.8\%$, which is almost exactly the 2.6% return in hourly wages we estimate in our IPWRA estimates.

income, which is more likely to be regular and therefore easier for them to report accurately. This should assuage concerns around using self-reported earnings in survey data.

Our results further highlight the value of rich, longitudinal survey data. Despite smaller sample sizes and self-reported earnings data, we can replicate the returns to graduation for a cohort estimated using administrative data. The fact that administrative data lacks key variables (e.g. on hours worked or more nuanced measures of socioeconomic disadvantage) means that it cannot always provide more robust estimates. When used in combination with survey data, however, both have the potential to shed light on key policy debates.

References

Alwin, D. F., Zeiser, K., & Gensimore, D. (2014). Reliability of Self-reports of Financial Data in Surveys: Results From the Health and Retirement Study. *Sociological Methods & Research*, *43*(1), 98–136. https://doi.org/10.1177/0049124113507908

Andrews, R., Imberman, S. A., Lovenheim, M., & Stange, K. (2022). *The Returns to College Major Choice: Average and Distributional Effects, Career Trajectories, and Earnings Variability* (NBER Working Paper No. w30331).

Anelli, M. (2020). The Returns to Elite University Education: A Quasi-Experimental Analysis. *Journal of the European Economic Association*, *18*(6), 2824–2868. https://doi.org/10.1093/jeea/jvz070

Belfield, C., Britton, J., Buscha, F., Dearden, L., Dickson, M., Erve, L. van der, Sibieta, L., Vignoles, A., Walker, I., & Zhu, Y. (2018, November 27). *The impact of undergraduate degrees on early-career earnings*. https://www.ifs.org.uk/publications/13731

Blau, F. D., & Kahn, L. M. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, *55*(3), 789–865. https://doi.org/10.1257/jel.20160995

Britton, J., Dearden, L., Waltmann, B., & van der Erve, L. (2020). *The impact of undergraduate degrees on lifetime earnings* (Research Report DFE-RR974). Department for Education.

Britton, J., Van Der Erve, L., Belfield, C., Vignoles, A., Dickson, M., Zhu, Y., Walker, I., Dearden, L., Sibieta, L., & Buscha, F. (2022). How much does degree choice matter? *Labour Economics*, *79*, 102268. https://doi.org/10.1016/j.labeco.2022.102268

Dale, S. B., & Krueger, A. B. (2014). Estimating the Effects of College Characteristics over the Career Using Administrative Earnings Data. *The Journal of Human Resources*, *49*(2), 323–358.

Department for Education. (2023). *Graduate labour market statistics*. GOV.UK.

Elsayed, A., & Shirshikova, A. (2023). The women-empowering effect of higher education. *Journal of Development Economics*, *163*, 103101. https://doi.org/10.1016/j.jdeveco.2023.103101

European Commission. (2013). *Sex-Discrimination in Relation to Part-Time and Fixed-Term Work*. European Network of Legal Experts in the Field of Gender Equality.

Hastings, J. S., Neilson, C., & Zimmerman, S. D. (2013). *Are Some Degrees Worth More than Others? Evidence from College Admission Cutoffs in Chile*. *w19241*.

Herd, P., Goesling, B., & House, J. S. (2007). Socioeconomic Position and Health: The Differential Effects of Education versus Income on the Onset versus Progression of Health Problems. *Journal of Health and Social Behavior*, *48*(3), 223–238. https://doi.org/10.1177/002214650704800302

Hout, M. (2012). Social and Economic Returns to College Education in the United States. *Annual Review of Sociology*, *38*(1), 379–400. https://doi.org/10.1146/annurev.soc.012809.102503

Hu, A., & Qian, Z. (2016). Does Higher Education Expansion Promote Educational Homogamy? Evidence from Married Couples of the Post-80s Generation in Shanghai, China. *Social Science Research*, *60*, 148–162. https://doi.org/10.1016/j.ssresearch.2016.05.001

Jepsen, C., Troske, K., & Coomes, P. (2014). The Labor-Market Returns to Community College Degrees, Diplomas, and Certificates. *Journal of Labor Economics*, *32*(1), 95–121. https://doi.org/10.1086/671809

Kirkeboen, L. J., Leuven, E., & Mogstad, M. (2016). Field of Study, Earnings, and Self-Selection*. *The Quarterly Journal of Economics*, *131*(3), 1057–1111. https://doi.org/10.1093/qje/qjw019

Maurin, E., & McNally, S. (2008). Vive la Révolution! Long-Term Educational Returns of 1968 to the Angry Students. *Journal of Labor Economics*, *26*(1), 1–33. https://doi.org/10.1086/522071

Mountjoy, J., & Hickman, B. R. (2021). *The Returns to College(s): Relative Value-Added and Match Effects in Higher Education* (NBER Working Paper 29276). National Bureau of Economic Research. https://doi.org/10.3386/w29276

OECD. (2022). *Employment: Share of employed in part-time employment by sex and age-group*. https://stats.oecd.org/index.aspx?queryid=54746

Office for National Statistics (ONS). (2021). *Participation measures in higher education: 2006 to 2020*. https://www.gov.uk/government/statistics/participation-measures-in-higher-education-2006-to-2020

Oreopoulos, P., & Petronijevic, U. (2013). *Making College Worth It: A Review of Research on the Returns to Higher Education* (Working Paper 19053). National Bureau of Economic Research. https://doi.org/10.3386/w19053

Raghupathi, V., & Raghupathi, W. (2020). The influence of education on health: An empirical assessment of OECD countries for the period 1995–2015. *Archives of Public Health*, *78*(1), 20. https://doi.org/10.1186/s13690-020-00402-5

StataCorp. (2013). *Stata: Release 13. Statistical Software.* StataCorp LP.

University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2018). *Next Steps: Sweeps 1-8, 2004-2016: Secure Access [data collection]*. https://doi.org/10.5255/UKDA-SN-7104-4

Walker, I., & Zhu, Y. (2013). *The impact of university degrees on the lifecycle of earnings: See further analysis* (112; BIS Research Paper).

Webber, D. A. (2016). Are college costs worth it? How ability, major, and debt affect the returns to schooling. *Economics of Education Review*, *53*, 296–310. https://doi.org/10.1016/j.econedurev.2016.04.007

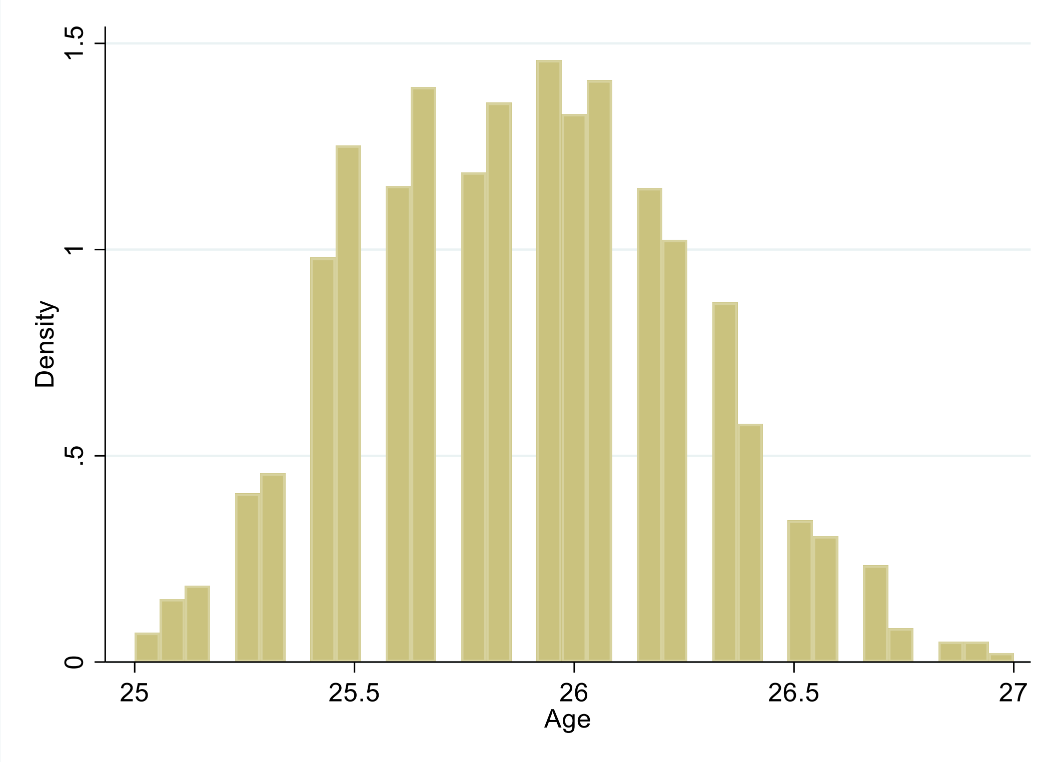
Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics*, *141*(2), 1281–1301. https://doi.org/10.1016/j.jeconom.2007.02.002

Wyness, G. (2010). *Policy changes in UK higher education funding, 1963-2009* (10–15; DoQSS Working Papers). Department of Quantitative Social Science - UCL Institute of Education, University College London. https://ideas.repec.org/p/qss/dqsswp/1015.html

Zimmerman, S. D. (2019). Elite Colleges and Upward Mobility to Top Jobs and Top Incomes. *American Economic Review*, *109*(1), 1–47. https://doi.org/10.1257/aer.20171019

6. Appendix

Figure A1: This distribution of age of observation in the sample



Notes: No. of unweighted observations: 3,214.

Source: University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2018). Next Steps: Sweeps 1-8, 2004-2016: Secure Access. DOI: 10.5255/UKDA-SN-7104-4

Table A1a: Descriptive Statistics - Women

	(1)	(2)	(3)	(4)	(5)	(6)
	Non-graduates			Graduates		
	N	Mean	SD	N	Mean	SD
White	707	0.873	0.334	886	0.833	0.373
North East	707	0.061	0.240	886	0.081	0.273
North West	707	0.112	0.315	886	0.128	0.334
Yorkshire and The Humber	707	0.103	0.304	886	0.113	0.316
East Midlands	707	0.100	0.300	886	0.105	0.306
West Midlands	707	0.130	0.336	886	0.114	0.318
East of England	707	0.103	0.304	886	0.112	0.315
London	707	0.120	0.326	886	0.151	0.358
South East	707	0.159	0.366	886	0.113	0.316
South West	707	0.113	0.317	886	0.084	0.278
Mother's NS-SEC: 1-2	707	0.358	0.480	886	0.393	0.489
Mother's NS-SEC: 3-5	707	0.284	0.451	886	0.322	0.468
Mother's NS-SEC: 6-7	707	0.278	0.448	886	0.202	0.401
Mother's NS-SEC: missing	707	0.080	0.272	886	0.083	0.276
Father's NS-SEC: 1-2	707	0.452	0.498	886	0.512	0.500

Father's NS-SEC: 3-5	707	0.299	0.458	886	0.316	0.465
Father's NS-SEC: 6-7	707	0.205	0.404	886	0.136	0.343
Father's NS-SEC: missing	707	0.045	0.207	886	0.036	0.185
Independent school	707	0.034	0.182	886	0.051	0.220
GCSE test scores	707	475.2	88.73	886	503.9	89.84
GCSE missing	707	0	0	886	0	0
Vocational qualification	707	0.386	0.487	886	0.246	0.431
A-levels quintile, lowest	707	0.165	0.372	886	0.096	0.295
A-levels quintile, lower-middle	707	0.137	0.345	886	0.123	0.329
A-levels quintile, middle	707	0.142	0.349	886	0.166	0.372
A-levels quintile, upper-middle	707	0.128	0.334	886	0.172	0.377
A-levels quintile, highest	707	0.090	0.286	886	0.195	0.396
A-level quintile missing	707	0.193	0.395	886	0.120	0.326
No A-levels	707	0.145	0.352	886	0.127	0.333
A-level in math	707	0.121	0.326	886	0.200	0.400
A-level in sciences	707	0.362	0.481	886	0.458	0.499
A-level in social sciences	707	0.193	0.395	886	0.201	0.401
A-level in art	707	0.282	0.450	886	0.246	0.431
A-level in humanities	707	0.494	0.500	886	0.545	0.498
A-level in languages	707	0.155	0.363	886	0.218	0.413
A-level in other	707	0.597	0.491	886	0.633	0.482
Graduation	707	0	0	886	1	0
Hours worked	707	37.38	9.387	886	40.07	9.247
Annual wage	707	21,209	17,126	886	25,129	25,982
Log annual wage	707	9.822	0.527	886	9.976	0.511
Hourly wage	707	11.26	12.03	886	12.11	11.87
Log hourly wage	707	2.293	0.418	886	2.367	0.421

No. of unweighted observations: 707+886=1,593. Sample of those having at least 5 A*-C GCSE examinations and in sustained employment.

Source: University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2018). Next Steps: Sweeps 1-8, 2004-2016: Secure Access. DOI: 10.5255/UKDA-SN-7104-4

Table A1b: Descriptive Statistics - Men

	(1)	(2)	(3)	(1)	(2)	(3)
	Non-graduates			Graduates		
	N	Mean	SD	N	Mean	SD
White	484	0.887	0.317	628	0.868	0.339
North East	484	0.083	0.276	628	0.028	0.165
North West	484	0.131	0.338	628	0.155	0.362
Yorkshire and The Humber	484	0.085	0.279	628	0.112	0.315
East Midlands	484	0.089	0.285	628	0.083	0.276
West Midlands	484	0.156	0.363	628	0.115	0.319
East of England	484	0.116	0.320	628	0.119	0.324
London	484	0.120	0.325	628	0.136	0.343
South East	484	0.111	0.315	628	0.169	0.375
South West	484	0.110	0.313	628	0.085	0.279
Mother's NS-SEC: 1-2	484	0.329	0.470	628	0.448	0.498

Mother's NS-SEC: 3-5	484	0.298	0.458	628	0.278	0.449
Mother's NS-SEC: 6-7	484	0.272	0.445	628	0.212	0.409
Mother's NS-SEC: missing	484	0.102	0.302	628	0.062	0.241
Father's NS-SEC: 1-2	484	0.402	0.491	628	0.542	0.499
Father's NS-SEC: 3-5	484	0.364	0.482	628	0.292	0.455
Father's NS-SEC: 6-7	484	0.189	0.392	628	0.146	0.354
Father's NS-SEC: missing	484	0.045	0.207	628	0.020	0.138
Independent school	484	0.037	0.188	628	0.058	0.233
GCSE test scores	484	479.5	92.24	628	509.5	96.76
GCSE missing	484	0	0	628	0	0
Vocational qualification	484	0.426	0.495	628	0.273	0.446
A-levels quintile, lowest	484	0.196	0.397	628	0.138	0.345
A-levels quintile, lower-middle	484	0.143	0.350	628	0.175	0.380
A-levels quintile, middle	484	0.106	0.308	628	0.140	0.348
A-levels quintile, upper-middle	484	0.098	0.298	628	0.155	0.362
A-levels quintile, highest	484	0.065	0.247	628	0.160	0.367
A-level quintile missing	484	0.208	0.407	628	0.107	0.309
No A-levels	484	0.183	0.387	628	0.125	0.331
A-level in math	484	0.228	0.420	628	0.376	0.485
A-level in sciences	484	0.395	0.489	628	0.529	0.500
A-level in social sciences	484	0.136	0.343	628	0.183	0.387
A-level in art	484	0.144	0.351	628	0.154	0.361
A-level in humanities	484	0.447	0.498	628	0.508	0.500
A-level in languages	484	0.088	0.284	628	0.158	0.365
A-level in other	484	0.542	0.499	628	0.627	0.484
Graduation	484	0	0	628	1	0
Hours worked	484	41.02	7.971	628	41.82	8.263
Annual wage	484	27,097	23,397	628	28,339	19,038
Log annual wage	484	10.07	0.467	628	10.13	0.470
Hourly wage	484	12.72	10.77	628	13.16	8.987
Log hourly wage	484	2.430	0.409	628	2.469	0.429

No. of unweighted observations: 484+628=1,112. Sample of those having at least 5 A*-C GCSE examinations and in sustained employment.

Source: University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2018). Next Steps: Sweeps 1-8, 2004-2016: Secure Access. DOI: 10.5255/UKDA-SN-7104-4

Table A2: Returns to graduation: log annual and hourly wages – robustness check on the IPWRA sample

	(1)	(2)	(3)	(4)	(5)	(6)
		Men			Women	
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
		Outcome: log annual wage				
Returns to graduation	0.058*	0.001	0.002	0.154***	0.071**	0.069**
	(0.033)	(0.034)	(0.032)	(0.030)	(0.030)	(0.029)
Constant	10.074***	8.563***	8.155***	9.822***	7.942***	7.091***
	(0.025)	(1.077)	(1.636)	(0.024)	(1.033)	(1.176)

No. of unweighted observations						
R-squared	1,112	1,112	1,112	1,593	1,593	1,593
Outcome: log hourly wage						
Returns to graduation	0.039	-0.015	-0.016	0.075***	0.028	0.026
	(0.030)	(0.030)	(0.028)	(0.024)	(0.023)	(0.022)
Constant	2.430***	1.770*	1.254	2.293***	0.382	-1.259
	(0.022)	(0.921)	(1.150)	(0.019)	(0.754)	(1.085)
No. of unweighted observations						
R-squared	1,112	1,112	1,112	1,593	1,593	1,593
Control variables and weighting						
Control variables		Yes	Yes		Yes	Yes
IPWRA weighting			Yes			Yes
Wave 8 sampling weights	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Sample of those having at least 5 A*-C GCSE examinations and in sustained employment. Model 1 and 2: linear regression models estimated by OLS, weighted using wave 8 weights. Model 3: IPWRA-weighted regressions. Robust standard errors clustered by sampling school are in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. Control variables: Sample boost: whether the individual belongs to the sample boost added to the survey in Wave 4. Family background: age in months as a continuous variable, mother's and father's social class, region, ethnicity. Early and pre-university educational attainment: GCSE and A-level raw scores, indicator variables for A-level subjects as Math, Sciences, Social science, Humanities, Arts, Languages and Other, a binary variable for having vocational qualifications, a binary variable capturing whether the individual attended independent secondary school at age 13/14. Missing observations are controlled for using missing flags.

Source: University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2018). Next Steps: Sweeps 1-8, 2004-2016: Secure Access. DOI: 10.5255/UKDA-SN-7104-4

Table A3: Returns to graduation: number of hours worked per week – robustness check on the IPWRA sample

	(1)	(2) Men	(3)	(4)	(5) Women	(6)
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Outcome: number of hours worked per week						
Graduation	0.802	0.698	0.779	2.681***	1.503***	1.515***
	(0.576)	(0.601)	(0.568)	(0.543)	(0.546)	(0.545)
Constant	41.020***	20.348	18.467	37.384***	40.310**	52.035**
	(0.417)	(21.373)	(27.384)	(0.400)	(18.094)	(24.265)
No. of unweighted observations						
R-squared	1,112	1,112	1,112	1,593	1,593	1,593
Control variables and weighting						

Control variables		Yes	Yes		Yes	Yes
IPWRA weighting			Yes			Yes
Wave 8 sampling weights	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Sample of those having at least 5 A*-C GCSE examinations and in sustained employment. Model 1 and 2: linear regression models estimated by OLS, weighted using wave 8 weights. Model 3: IPWRA-weighted regressions. Robust standard errors clustered by sampling school are in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. Control variables: Sample boost: whether the individual belongs to the sample boost added to the survey in Wave 4. Family background: age in months as a continuous variable, mother's and father's social class, region, ethnicity. Early and pre-university educational attainment: GCSE and A-level raw scores, indicator variables for A-level subjects as Math, Sciences, Social science, Humanities, Arts, Languages and Other, a binary variable for having vocational qualifications, a binary variable capturing whether the individual attended independent secondary school at age 13/14. Missing observations are controlled for using missing flags.

Source: University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2018). Next Steps: Sweeps 1-8, 2004-2016: Secure Access. DOI: 10.5255/UKDA-SN-7104-4