



Department
for Education

GCSE attainment and lifetime earnings

Research report

June 2021

**Louis Hodge, Allan Little and Matthew
Weldon: Department for Education**



Government
Social Research

Contents

List of Figures	3
List of Tables	4
Acknowledgments	5
Executive summary	6
Key findings	6
Interpretation of findings	7
Introduction	9
Literature	11
Survey-based estimates	11
Quasi-experimental evidence	12
Data	14
Longitudinal Educational Outcomes (LEO)	14
Labour Force Survey (LFS)	15
Methodology	18
Modelling lifetime earnings	19
Employment	20
Rank dependence	21
Real growth	22
Retirement	24
Simulation	24
Estimation of marginal returns	25
Nonlinear grade-effects model	27
Pre-treatment controls	27
Controlling for unobservables	28
Post-treatment controls	29
Results	30
Lifetime earnings and PV statistics	30
Returns to 'global' grade improvement	32
Marginal grade returns	33
Marginal returns by subject	34
Non-linear returns by grade boundary	37
Discussion	40
Causal effects, human capital and signalling	40

Multiple skills	41
Implications for recent and future pupil cohorts	42
Wellbeing	42
Macroeconomic and wider benefits	43
Extensions	43
Subject choice	43
Key Stages 1, 2 and 5	44
Exchequer returns	44
Appendices	45
A Sklar's Theorem, copula's and lifetime earnings	45
B t-copula persistence parameters	47
C Age-Period-Cohort problem	48
D State pension age	49
E Accuracy of earnings simulations	50
F Multiple imputation	52
G Variables in GCSE attainment model	53
H Subject mapping	54
I Additional non-linear results	55
J Additional analysis of non-linear grade effects	59
K Weekly vs. Hourly earnings	62
References	64

List of Figures

Figure 1:	LFS Earnings	17
Figure 2:	Schematic Overview	19
Figure 3:	Mean Earning Trajectories	31
Figure 4:	Global effect for 11.2 grade improvement: FSM status and gender	33
Figure 5:	Average standardised marginal effect for all subjects: number of GC-SEs, FSM status, KS2 attainment and gender	35
Figure 6:	Average marginal effects by subject	36
Figure 7:	Average marginal effects by subject and gender	37
Figure 8:	Non-linear returns: Maths, Double Science and English	38
Figure 9:	Non-linear returns: Geography, History and English Literature	39
Figure 10:	Unsmoothed Copula Persistence Parameters	47
Figure 11:	Smoothed Real Earnings Growth	48
Figure 12:	Actual vs. Simulated Distributions	50
Figure 13:	Comparing Rank Movement	51
Figure 14:	Non-linear returns: Business, IT and PE	55
Figure 15:	Non-linear returns: French, German and Spanish	56
Figure 16:	Non-linear returns: Drama, Art and Music	57
Figure 17:	Non-linear returns: Physics, Chemistry and Biology	58
Figure 18:	Scatter of grades against PV for French	60
Figure 19:	Scatter of grades against PV for Music	60
Figure 20:	Scatter of grades against PV for English Literature	61
Figure 21:	Scatter of grades against PV for Maths	61
Figure 22:	Hourly vs. Weekly LFS Earnings	63

List of Tables

Table 1:	LFS Summary Statistics	14
Table 2:	LFS Summary Statistics	16
Table 3:	Summary Statistics by GCSE Subject	31
Table 4:	Summary Statistics by demographics	32
Table 5:	Legislated Changes in State Pension Age	49
Table 6:	Variables in GCSE attainment model	53
Table 7:	Subject Mapping	54

Acknowledgments

This analysis was undertaken by economists and statisticians at the Department for Education (DfE). We are grateful to the members of our Academic Advisory Panel for their constructive advice on this project. The Panel's methodological input has been invaluable. Responsibility for the methodology and results remains with DfE. The Panel included:

Dr Claire Crawford, *University of Birmingham*

Emily Hunt, *Education Policy Institute*

Professor Lindsey Macmillan, *Centre for Education Policy and Equalising Opportunities - University College London*

Professor Anna Vignoles, *The Leverhulme Trust*

Ben Waltmann, *Institute for Fiscal Studies and University of Oxford*

Professor Yu Zhu, *University of Dundee*

We would also like to acknowledge extensive input in preparing the data and analysis from colleagues at DfE, notably Oliver Clifton-Moore, Amy Morgan and Moira Nelson.

Executive summary

The main contribution of this report is to provide estimates of the lifetime earnings return to small improvements in GCSE grades. The data, and the flexible methods we employ, reveal important differences in these estimates by GCSE subject and pupil subgroups. Within government, such estimates are used frequently in the economic appraisal of education policy. They can also inform debate about the economic importance of compulsory education.

We use the Longitudinal Educational Outcomes (LEO) data, an administrative dataset that links education records with income and employment records, for every child in the state school system in England. The advent of LEO transformed what is possible in the measurement of earnings linked to attainment. The report adds to a handful of studies using LEO that have estimated the returns to post-compulsory qualifications (for example, Britton et al. (2020) and Espinoza and Speckesser (2019)).

Previous estimates of the returns associated with GCSE performance have been limited largely to the measurement of the total number of GCSEs taken, or to threshold levels of attainment, such as attaining 5 A* to C grades. We leverage the detailed pupil records and large samples in LEO to estimate marginal grade improvements, across a number of GCSE subjects.

Although LEO provides detailed histories of education and earnings, it is limited by the fact linked records only exist for the 1985/86 birth cohort onwards; hence, we only observe individuals' annual earnings through their twenties. We use additional information from the UK Labour Force Survey (LFS) to complete each individual's earnings trajectory, up to retirement age.

Key findings

Our findings indicate that:

- An average student, that took their GCSE exams between 2001-2004, will go on to earn £1.3 million in undiscounted earnings during their life, or £515,000 in present values at age 16.¹
- A one-grade improvement² in overall GCSE attainment is associated with an average increase in the present value of lifetime earnings of £8,500. This implies

¹Using the standard HM Treasury discount rate of 3.5% over years 1 to 30, and 3% thereafter (HM Treasury (2020)).

²For those pupils that took their GCSE between 2001-2004. It is not necessarily directly applicable to a one grade improvement today.

that a one-standard deviation (11.2 grades) improvement in overall GCSE performance is associated with an increase in discounted lifetime earnings of approximately £96,000. This is nearly 20% of average discounted lifetime earnings, and 46% of the standard deviation of discounted lifetime earnings.

- The value of an additional grade in undiscounted earnings is £23,000, which represents about three quarters of the average full-time annual salary in the UK in 2019.³ The estimate for female (male) students is £20,000 (£24,000).
- There is wide variation in the marginal grade returns by individual GCSE subjects. A one-grade improvement in Maths is associated with a discounted return of £14,500, whereas in English the return is £7,300 and in Music it is £5,500.
- On average, men are found to have 18% larger marginal returns than women; those not eligible for Free School Meals have 9% larger marginal returns than those that are eligible.
- The marginal return varies with the grade boundary: the largest marginal returns are associated with moving from grade D to grade C, and from grade C to grade B, in most subjects. In some subjects, there are robust returns to top grades (A and A*) but in others the return is not measurably different from zero. The returns to top grades are generally larger for male students than female. The returns to achieving grades below C are typically smaller or not statistically different from zero.

Interpretation of findings

There are a few caveats, which must accompany these estimates:

1. The dataset cannot measure every variable that could possibly affect both grades and earnings. If there are important unseen traits that are correlated with both grades and earnings, this will add bias to the estimated returns.
2. We cannot distinguish between earnings returns caused by gains in useful skills and knowledge (human capital) and returns caused by possession of the qualification itself, which provides admission to further study and employment, irrespective of skills and knowledge (signalling).
3. Lifetime earnings represent a private return to education – they do not capture the effects of each person's education on the productivity, welfare, or labour market outcomes of others, or on the overall size of the economy.

³ONS (2019a) estimate that the gross weekly earnings of full-time employees was £584.90 in 2019.

4. Cohorts in the sample completed their GCSE's almost twenty years ago. Using old data is unavoidable if we wish to measure long-run outcomes. The implications of the twenty-year delay should be considered when making predictions of current policy impacts, based on these estimates.

Taken at face value, the large variation in grade returns by subject appear to imply that some subjects are 'worth' more than others. However, this interpretation would be misleading for several reasons. We estimate returns to marginal grade improvements, not the return to taking each subject. Students that take different subjects are not random samples from the population. The processes whereby students select different subjects, according to their own preferences, in conjunction with schools and parents, are complex and hamper straightforward comparisons between subjects. As an example, Triple Sciences has a relatively low marginal return per grade, but it is likely that this reflects diminishing marginal returns at the top of the earnings distribution, as students taking those subjects are among those with the highest grades and highest expected earnings.

Taking the above caveats into account, the findings imply that the largest and most robust difference in returns, when averaging across subjects, is between female and male students. In almost all splits of the data, female students have, not only smaller lifetime earnings overall (which is well documented⁴), but also a smaller average absolute return to grade improvements. We have not estimated proportional returns, so it is possible that this smaller absolute return masks female respondents achieving similar or even greater proportional returns, compared to women's lower expected earnings. It is nevertheless a discouraging finding.

Returns for FSM-eligible students are also lower on average, although not in every split of the data. Again, we cannot say if this lower absolute return reflects a smaller or larger relative return.

⁴For example, see figure 7 of ONS (2019b)

Introduction

This report presents estimates of the lifetime earnings value of improvements in GCSE attainment.⁵ Measurements of the economic value of educational attainment are used to facilitate cost-benefit analysis of policies and legislation. We are motivated in part by the availability of the Longitudinal Educational Outcomes (LEO) dataset, linking data on education performance to earnings data through the tax system. This report is the first to use LEO to estimate the associations between earnings and attainment at the end of Key Stage 4 (KS4), a key juncture for pupils later educational and labour market pathways.

Through its link to the National Pupil Database (NPD), LEO contains details on individual subject performance and a broader set of pupil and school characteristics, hitherto unavailable in studies of the earning returns to GCSEs. LEO also provides much larger sample sizes than representative surveys, used in previous studies; these samples allow us to disaggregate grade returns by pupil characteristics and (for the first time) by subject.

There have been innovations in the estimation of earnings returns to educational attainment as a result of LEO, notably through research on university degrees (Belfield et al. (2018b); (2018a); Britton et al. (2020)). These studies also sought to explore earnings variation in fine detail, including by subject, institution and student characteristics. Most recently, variation in these returns has been explored by ethnicity (Britton, Dearden and Waltmann (2021)). Collectively, this research deploys state-of-the-art methods and represents a step change in our understanding of the returns to undergraduate degrees.

Our report seeks similar advancements in the estimation of earnings returns associated with Key Stage 4 performance. We share the same inherent limitations faced by Belfield et al. (2018b), (2018a) and Britton et al. (2020). Whilst our models include an extensive set of controls, some causal factors inevitably remain unobserved in the LEO data. This ‘selection-on-observables’ assumption implies that, if there are any unseen traits that are positively correlated with grades and earnings, our estimates will be biased to some degree.

To calculate a measure of lifetime earnings we rely on data from the UK Labour Force Survey (LFS) to estimate later life earnings. This adds uncertainty. The labour market value of qualifications changes over time. Syllabuses and curriculums evolve, as does the value of specific skills in the labour market. Hence, the relationship between GCSEs and earnings for past cohorts is an imperfect guide to the prospective returns for more recent and future cohorts, whose labour market outcomes we are yet to observe.

By design, our report solely focuses on the earnings returns associated with improvements in GCSE grades, which we describe as an ‘intensive’ margin. Further

⁵GCSEs are a national qualification, typically taken by pupils at the age of 16 in between 7 and 10 subjects. Around 80% of pupils took 7-10 GCSEs in 2019 (Ofqual (2019)).

research could usefully consider ‘extensive’ margins, through two responses: a ‘switch effect’, changing one subject for another; and a ‘number effect’, taking additional GCSEs in a given subject.⁶ Investigation into the associations between GCSE subject choice and earnings would be a useful direction for further research.

Our focus on marginal grade improvements is a matter of priority, based on the typical use of earnings estimates in policy appraisal. Most policy interventions target improved attainment, evaluated by changes in test scores.⁷ A smaller number of policies seek to influence the total number of GCSE subjects taken, or subject choices at the beginning of KS4.⁸

Note that our report does not consider the costs, or cost-effectiveness of any given education policy intervention. Nor do we consider wider socio-economic outcomes associated with GCSE performance. Where the reader intends to use estimates from our report within social cost benefit analyses, we recommend the Department for Education’s companion piece, the Schools Policy Appraisal Handbook (Hodge, Little and Weldon (2021)).

The Handbook, published alongside this report, was motivated by the need to ensure that KS4 earnings estimates are used appropriately in appraisal. This informed the following selection of topics, which we cover briefly in the Discussion section of this report, but require more attention than we could afford in this report:

- the relationship between compulsory schooling and wellbeing;
- distributional effects, determining differences in the utility value of wages;
- impacts on macroeconomic growth and productivity, over and above those captured by individual earnings; and
- the relative importance of ‘human capital’ and ‘signalling’ effects in explaining causal relationships between KS4 qualifications and labour market outcomes.

All of these issues are of crucial importance, when using earnings estimates to appraise the societal value of KS4 outcomes.

The rest of this report discusses the existing returns to education literature, our methodology, outlines the results and provides further discussion.

⁶In addition, some interventions involve a combination of effects at both the intensive and extensive margins – for example, encouraging both uptake of, and improved attainment in, STEM subjects.

⁷Changes in grade performance may be summarised by measures of total GCSE points scores, standardised improvements (effect sizes) or by progress measures at pupil or school-level.

⁸These extensive margins do confer economic value. While estimation strategies would be feasible, the modelling approaches required to identify earnings by grade are substantively different to those used to estimate subject mix, recognising that pupils will face constrained choice sets, in part framed by schools.

Literature

Since Mincer (1958), there has been extensive use of regression specifications to estimate variation in earnings, associated with educational attainment. We focus selectively on the main forerunners to our report.

Survey-based estimates

Several studies have developed methods to estimate the earnings returns to attainment using the UK Labour Force Survey (LFS) as it provides representative UK data on both qualifications and labour market outcomes. Surveys are, however, undertaken with adults and so rely on accurate recall of historical information on qualifications. LFS and similar UK surveys do not include information on GCSE subjects, nor detailed test scores in these subjects. Studies using these surveys tend to observe earnings variation across threshold levels of attainment, based on the total number of GCSE passes.

For example, Greenwood et al. (2007) estimated that individuals whose highest qualifications are five or more 'good' GCSEs have marginal wage returns of ~25%, compared to those who hold no qualifications.⁹ Similarly, McIntosh (2006) found returns to 5+ good GCSEs of 25-30% for women and 28-31% for men. Hayward et al. (2014) allowed returns to qualifications to vary over the lifecycle, by interacting qualification level with a polynomial in age. They found lifetime productivity returns to 5+ good GCSEs, relative to those with 1-4 good GCSEs, was £105,000 for men and £100,000 for women, in present values. At the lower end of the distribution, attaining 1-2 good GCSE passes over no qualifications had a return of £171,000 for men and £110,000 for women.

The LFS includes information on key demographic characteristics; however, it lacks details on known confounders, such as prior attainment at the end of primary school, Free School Meal eligibility and Special Education Needs and Disabilities (SEND) status. More recently, administrative education records have been linked with employment, earnings and benefits data to create the Longitudinal Educational Outcomes (LEO) dataset. This provides more detailed records, on many more individuals, than the available UK surveys. The level of detail allows for much more granular subject and sub-group analyses. It also minimises issues with sample selection.

Britton, Shephard and Vignoles (2019) compare earnings data in the LFS with that from HMRC tax records, finding some differences. They find a similar share of zero earners but

⁹'Good' GCSEs are defined as those awarded at grade 'C' or above under the letter grading system, phased out in summer 2019. Five or more 'good' GCSEs included English and Maths under this definition.

a much higher share of low earners in the admin data¹⁰ and generally higher earnings throughout the distribution. The LEO data, therefore, may have the drawback of not capturing all realised earnings if the observed differences are caused by under-reporting of earnings for tax purposes.

The use of LEO has, thus far, focused on higher-level qualifications rather than compulsory schooling. The focus has largely been on “snapshots” of the economic returns at discrete ages, and on early career pathways. Espinoza and Speckesser (2019) estimated earnings returns to NVQ level 4, 5 and 6 qualifications at age 30. Belfield et al. (2018b), (2018a), use LEO to explore returns by university degree subject and institution, five years after graduation and at age 29.

Britton et al. (2020) is the first study to extend this analysis of LEO, to examine ‘lifetime’ earnings, i.e., up to retirement. The LEO data currently only extends to age 29 and so the LFS is used to complete estimates of future earnings paths. We follow a similar estimation strategy in this report. Britton et al. (2020) estimated a total net lifetime earnings premium of £130k for men and £100k for women from attending university, in present values. This represented a 20% gain in average net lifetime earnings. They report variation by subject and student characteristics, in detail.

Quasi-experimental evidence

The ideal way to measure the returns to education would be to compare the earnings of the same individual with multiple different educational outcomes. Each individual, however, only completes their schooling once and only one set of educational outcomes are observed.

There are two main approaches to manoeuvre around this problem. Belfield et al. (2018b); (2018a), Britton et al. (2020), Espinoza and Speckesser (2019), and the estimates in this report adopt a similar type of study design, broadly described as a selection-on-observed variable approach.

A selection-on-observed variables approach aims to compare outcomes for pupils who are observably similar, on the dimensions that can be measured within the data. In many of the studies mentioned above, including those using LEO, information on pupil, school and area characteristics are reasonably detailed, making this strategy a useful and informative approach. It is, however, not possible to test whether all relevant variables are sufficiently controlled for and there may be relevant but unobserved characteristics. The possibility of omitted variable and selection biases therefore remains.

¹⁰They find that of those in employment, 14% earn less than £8,000 in the admin data compared to only 5% in the LFS.

These biases motivate a smaller number of studies that employ quasi-experimental designs to estimate the impact of qualifications on labour market outcomes.

Quasi-experimental methods exploit effectively random variation in exposure to the treatment of interest, typically generated as a by-product of a policy or institutional feature. These studies can provide a compelling approach to identify the causal effects of educational attainment.

For example, Machin, McNally and Ruiz-Valenzuela (2020) assessed the effects of gaining a C grade in English GCSE on appeal, compared to a D grade.¹¹ The argument is that the test score differences between these pupils is essentially random, in a very small window close to the 'C' grade cut-off. Just missing out on a grade 'C' decreased the probability of enrolling on a Level 3 (A-Level or equivalent) course by 9 percentage points and increased the probability of dropping out of school before 18 without employment by 2 percentage points.

De Phillipis (2017) observed the staggered roll-out of a policy mandating the provision of 'triple-science'¹², for high ability pupils. The reform increased the share of pupils taking triple science, from 4% to 20% over 9 years, amounting to an additional 5 hours of science per week. It increased the probability of enrolling in a STEM degree at university by 1.2 percentage points (or 7%), and the probability of graduating from a STEM degree by 3.0 percentage points (or 17%).

Quasi-experimental evidence has, therefore, helped to corroborate the hypothesis that KS4 has causal impacts on later outcomes. While confidence in these causal inferences is higher in quasi-experimental designs, this does not imply that they are necessarily preferable to selection-on-observed variables. An important limitation of quasi-experimental designs is that the findings are context specific, limited to the particular setting from which they are generated. Taking Machin, McNally and Ruiz-Valenzuela (2020) as an example, the effect of passing the C/D grade boundary, in English, for one pupil cohort, is not generalisable for all grade boundaries, pupils or GCSE subjects.

Our choice of a selection-on-observed variables approach is motivated by the need to inform a broad range of KS4 policy appraisal. We cannot rely on the happenstance of where quasi-experiments present themselves. Rarely do policy interventions coincide with the availability of this narrower, causal evidence. The best approach is, in our view, to use evidence from a range of empirical strategies, in combination.

¹¹ A pass of at least a 'C' grade at GCSE has been viewed as a key threshold for progression in education and certain occupations – as such, many pupils appealed their score if they fell just below the required mark to gain a 'C' grade.

¹² Triple Science (also called 'advanced science') covers relatively complex science topics and entails more instruction hours and examination than double- or single-science.

Data

Longitudinal Educational Outcomes (LEO)

Since 2015, it has been possible to link the National Pupil Database (NPD) held by the Department of Education with income data (both PAYE and self-assessment) from HMRC and benefit data from DWP.¹³ This linked dataset forms the Longitudinal Educational Outcomes study (LEO). The NPD contains detailed records of pupils in the state education system.¹⁴ This includes detailed pupil characteristics such as gender, Free School Meal (FSM) eligibility, Special Educational Needs and Disability (SEND) status, ethnicity, and geographical location. It also contains records of attainment, both prior KS2 results and subject level GCSE grades.

Our analysis uses the earliest four GCSE cohorts available in LEO, who completed their GCSEs in 2001/02; 2002/03, 2003/04 and 2004/05.¹⁵ Through pooling over several cohorts, we create a larger sample that allows better identification of effects. Using earlier cohorts means we can observe longer and more stable earnings trajectories, whereas for later pupil cohorts we could only observe their early labour market histories. There is a trade-off between the number of years we observe in the labour market and the difference in the policy environment further in the past, and now.

Table 1: LFS Summary Statistics

Variable	Value	Statistic
Gender	Female	1,039,332 (48%)
Gender	Male	1,105,611 (52%)
FSM	FSM	361,991 (17%)
FSM	Not FSM	1,782,952 (83%)
N GCSEs	4 or fewer	175,068 (8.2%)
N GCSEs	5-6	214,613 (10%)
N GCSEs	7-8	581,268 (27%)
N GCSEs	9	553,949 (26%)
N GCSEs	10+	620,045 (29%)
Earnings (GBP)	Earnings (GBP)	1,329,405 (552,721)
PV (GBP)	PV (GBP)	514,571 (225,831)
Grade points	Grade points	37 (20)

Notes: Total Observations = 2,144,943. Statistics presented: n (%); mean (SD)

¹³This was made possible by the passing of The Small Business, Employment and Enterprise Act 2015 in conjunction with the Education and Skills Act 2008.

¹⁴Our analysis is conducted on only those who attended state-funded schools.

¹⁵Currently, the earliest linked cohort is the 1985/86 birth cohort who took their GCSE exams in summer 2002. LEO provides a rich panel of individuals for all subsequent cohorts.

Labour Force Survey (LFS)

We observe earnings up to age 29 in our LEO sample. To estimate the impact of attainment on 'lifetime' earnings, we simulate earnings over the rest of individuals' working lives, using the UK Labour Force Survey (LFS) between 2003 and 2018.

The LFS is a nationally representative rotating panel¹⁶, in which participants are surveyed for five consecutive quarters. We only observe labour market outcomes in the 1st and 5th quarters (one year apart). We use this information to estimate transition probabilities between employment states and changes in earnings at each age.

We divide our sample into 12 groups, each defined by an individual's gender and highest level of qualification achieved. All our quantities of interest are then calculated separately for each group at each pair of adjacent working ages.

Whilst the LFS does contain other socio-demographics by which the sample could be grouped (for example ethnicity or geographic location), we deem gender and education level to be most relevant to earning patterns. There is an explicit trade-off between the number of characteristics we stratify by and the number of observations within each group that can be used to estimate each parameter. A summary of the data used can be found in table 2.¹⁷

We use measures of weekly pay in our analysis, rather than hourly wages. This more accurately reflects the earnings returns we are trying to estimate, implicitly capturing variation in the number of hours worked. Appendix K discusses this in further detail and provides a comparison between hourly and weekly measures. Figure 1 shows median positive weekly earnings at each age, inflating earnings to 2019 prices using the Consumer Prices Index (CPI), for each sub-group defined by gender and highest qualification obtained.

A similar approach to simulating future earning trajectories would have been possible using the British Household Panel Survey (BHPS).¹⁸ The data available on qualifications and labour market outcome is very similar to that of the LFS. Due to the extended panel nature of the BHPS it contains information for more than only two adjacent years. This is potentially advantageous as it would allow more detailed understanding of trajectories over extended periods of time. However, the LFS has the benefit of containing a larger number of observations, especially when pooled over many years, than the BHPS.

¹⁶Groups of individuals are surveyed for five consecutive quarters before being replaced by a new set of survey participants on a rolling basis. For more details see ONS (2019c).

¹⁷There are slightly more men than women in our overall sample despite there being more women in the overall population. This is due to differences in pension age between genders over the period of the sample. There are more women than men in our sample below the age of 60.

¹⁸Started in 1991, the BHPS followed over 5,000 households with over 10,000 individuals for 18 waves until 2009. Subsequently, the majority of BHPS participants have been transferred to the Understanding Society (US) study.

Table 2: LFS Summary Statistics

Variable	Value	Statistic
Gender	Male	124,838 (52%)
Gender	Female	114,945 (48%)
Highest Qualification	Level 2 (GCSE grades A*-C or equivalent)	48,157 (20%)
Highest Qualification	Level 3 (GCE, A-level or equivalent)	52,034 (22%)
Highest Qualification	Level 4,5 (CertHE, DipHE or equivalent)	24,822 (10%)
Highest Qualification	Level 6 (Degree or equivalent)	60,050 (25%)
Highest Qualification	No qualification	31,267 (13%)
Highest Qualification	Other qualifications	23,453 (9.8%)
Employment Flows	Employed -> Employed	171,047 (71%)
Employment Flows	Employed -> Not Employed	9,006 (3.8%)
Employment Flows	Not Employed -> Employed	6,948 (2.9%)
Employment Flows	Not Employed -> Not Employed	52,782 (22%)
Age	25-34	37,841 (16%)
Age	35-44	62,272 (26%)
Age	45-54	69,778 (29%)
Age	55-64	61,048 (25%)
Age	65+	8,844 (3.7%)

Notes: Total Observations = 239,783. Statistics presented: n (%)

Other studies have compared using both these surveys, in similar contexts. Walker and Zhu (2013), using their two-stage quantile regression approach, found that “the LFS estimates of the lifecycle parameters are more precise than the BHPS panel estimates of the same parameters”. Using copula techniques similar to those this report uses, Dearden et al. (2008) find that “simulations of the lifetime earnings distributions from both data sources are very similar”.

Figure 1: LFS Earnings



Methodology

Our quantities of interest are the lifetime earnings returns to marginal changes in GCSE performance. We attempt to identify effects associated with an improvement in grades (the “intensive” margin), as opposed to obtaining more GCSEs (the “extensive” margin), which as discussed above has typically been the focus of prior studies.

We seek to estimate the total effect of GCSEs on earnings, including direct effects (the additional value of improved performance in school) and induced changes (e.g., resulting from subsequent educational and occupational choices).

To estimate such a fine-grained quantity requires highly detailed data on educational performance linked to labour market outcomes. LEO provides us with exactly that.

As noted above, individuals are observed in the LEO data only for the 1985/86 birth cohort onward and so we are unable to observe complete lifetime earnings pathways. To address this issue, we model later life earnings using data from the LFS, capturing labour market dynamics across periods. Whilst there is a degree of randomness in individual earning paths, there is also a degree of dependence on earnings realisations in previous periods. We model this persistence process using copula functions. In addition, we also extract employment status transition probabilities and their impact on lifecycle earnings growth. This model allows us to then simulate complete age-earning profiles for each individual in the LEO data, resolving the missing data problem.

This missing data problem is, as we have described, not unique to our study. Historically, this problem has been addressed by estimating polynomial or ARMA¹⁹ models to construct age-earnings profiles. There is precedent, however, for using copula techniques instead. In one of the earliest examples of using copulas to estimate lifetime age-earning distributions, Dearden et al. (2006) used the BHPS to test various functional forms and compare their results to traditional ARMA models. They show that t-copulas²⁰ “more closely approximate the structure of the data”.

In a more direct comparison, Britton et al. (2020) use copula methods as part of their strategy to estimate the earnings returns to undergraduate degrees. They complete lifetime age-earning profiles using parameters on sub-groups based on subject, institution type and gender. The copula method for estimating the simulation parameters is described in detail in a sub-section below.

We use the copula model to predict a lifetime earnings trajectory for each person in the LEO sample, from age 30 through to retirement.²¹ Each trajectory is then collapsed into a

¹⁹AutoRegressive-Moving-Average. These are typical time series models, used to describe stationary stochastic processes.

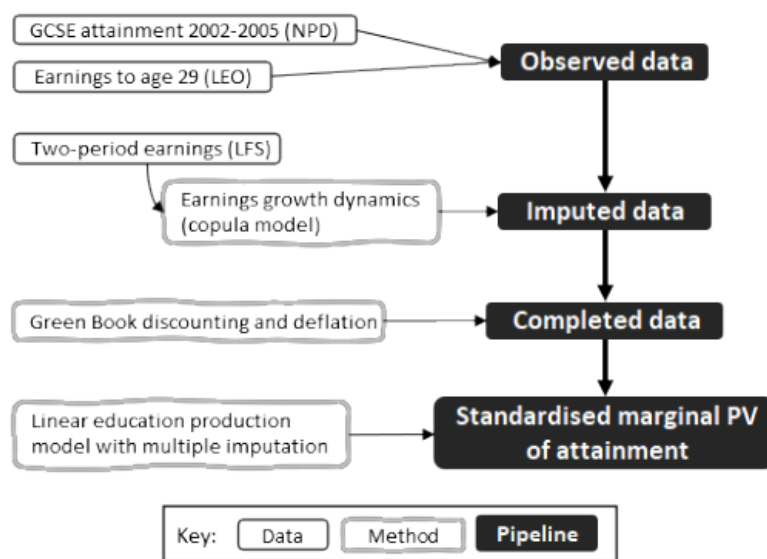
²⁰t-copulas model dependence using a multivariate Student’s t distribution.

²¹Retirement is modelled at the State Pension Age, more details are provided in a sub-section.

discounted present value (i.e. a weighted sum of lifetime earnings)²², which is the intended outcome variable for our analysis. These estimates are the dependent variable in our set of regression equations.

As these predictions include an element of randomness, we must account for the additional uncertainty this brings to the estimates. We do this by simulating each person’s earnings trajectory ten times, to generate ten potentially different outcome (present value lifetime earnings) variables for each person. The regressions are estimated separately for each simulated outcome variable, and the ten sets of results are pooled using Rubin’s (1987) multiple imputation method. These regression equations are described in detail in a sub-section below. The multiple imputation method is described in appendix F.

Figure 2: Schematic Overview



The diagram in figure 2 provides a visual overview of the methods used to estimate the returns to GCSE grades, outlined above.

Modelling lifetime earnings

As discussed, we use data taken from five-quarter LFS panels to simulate complete age-earning profiles for the individuals in our LEO data. We estimate a series of processes which, when combined, can describe the evolution of an individual’s earnings over time. We estimate the parameters of these processes separately at each adjacent age pair and for each of 12 characteristic based groups (by gender and highest qualification obtained). To guard against small bins and outliers all estimates are subsequently smoothed over age using Nadaraya-Watson kernel regression.²³ These smoothed parameter estimates are then combined to simulate individual lifetime earnings trajectories.

²²Discounted present values are calculated using the method described in the Green Book (HM Treasury (2020); appendix A6).

²³Nadaraya-Watson kernel regression is a nonparametric technique that estimates a local weighted average.

Employment

Firstly, we estimate the likelihood of individuals being employed in the current period given their employment status in previous periods. The typical definition of employment divides working age individuals into three groups; employed, searching for employment (meeting the ILO definition of unemployment²⁴) and economically inactive. For the purposes of our model, we simply define a binary employment variable that takes the value 1 if a person is employed and 0 otherwise. The LFS panels include employment status in quarter 1 ($a - 1$) and quarter 5 (a), one year apart. They also include information on individuals longer term employment history which we use to assign an employment indicator one year prior ($a - 2$). We assume if non-employed individuals left their last job within the last year they were employed at $t - 2$, else we assume they were non-employed. We therefore estimate three transition probabilities:

$$P(E_{i,g,a} = 1 | E_{i,g,a-1} = 1) \quad (1)$$

$$P(E_{i,g,a} = 1 | E_{i,g,a-1} = 0 \& E_{i,g,a-2} = 1) \quad (2)$$

$$P(E_{i,g,a} = 1 | E_{i,g,a-1} = 0 \& E_{i,g,a-2} = 0) \quad (3)$$

We would expect each of these probabilities to vary over the lifetime, by age (a) and our 12 characteristic groups (g).

The probability of being employed in the current period ($E_{i,g,a}$) given employment in the previous period ($E_{i,g,a-1}$) is estimated as a probit model²⁵ using individuals rank in the earnings distribution, in the previous period ($r_{i,g,a-1}$):

$$P(E_{i,g,a} = 1 | E_{i,g,a-1} = 1) = \Phi(\alpha_{0,g,a} + \alpha_{1,g,a} r_{i,g,a-1}) \quad (4)$$

Where Φ is the CDF of the standard normal distribution.

In the case that an individual is not employed in the previous period ($a - 1$) we have no information on individual's earnings in any prior periods. Thus, we only make use of the binary employment indicators. We estimate the probability of entering employment as the within group, age sample means.

²⁴The ILO (International Labour Organization) define unemployed persons as those; without work, currently available for work (in the next two weeks) and seeking work (having actively sought work in the last four weeks).

²⁵A probit model describes a binary response variable with two outcomes (in this case employed or not employed), where the probability of a given outcome is modelled using the standard normal distribution.

For those that have only been unemployed for one period we calculate:

$$\hat{P}(E_{i,g,a} = 1 | E_{i,g,a-1} = 0 \& E_{i,g,a-2} = 1) = \frac{1}{N^g} \sum_{i=1}^{N^g} E_{i,g,a} 1(E_{i,g,a-1} = 0 \& E_{i,g,a-2} = 1) \quad (5)$$

And for those that have not been employed for at least two periods we calculate:

$$\hat{P}(E_{i,g,a} = 1 | E_{i,g,a-1} = 0 \& E_{i,g,a-2} = 0) = \frac{1}{N^g} \sum_{i=1}^{N^g} E_{i,g,a} 1(E_{i,g,a-1} = 0 \& E_{i,g,a-2} = 0) \quad (6)$$

Rank dependence

We then estimate the rank dependence of earnings between adjacent time periods. Specifically, we use a series of bivariate copula functions to describe the dependency structure. This approach allows us to estimate earnings dynamics over a whole lifetime, despite the lack of longitudinal data to directly capture changes in earnings.

Copulas allow us to capture the fact that, for example, a person whose income is in the bottom 10% of incomes at age 30, is not likely to be in the top 10% of incomes at age 31, although it allows for this with a very small probability. More detail on copula functions is provided in appendix A. We estimate series of bivariate copulas at each adjacent age, allowing us to model path dependence without making strong assumptions about the shape of the earnings distribution. This is purely determined by the empirical distribution of earnings in the data. Most importantly for our purpose, once we have estimated the model on one dataset, it is straightforward to simulate a set of earnings trajectories on a new dataset, which then has the temporal correlation structure of the first dataset, but the empirical earnings distribution of the new data.

Whilst we do not need to place restrictions on the marginal distributions of earnings, we do need to parameterise the copula function itself. We use Student's t-distribution to estimate a t-copula. Dearden et al. (2008) find that the t-copula best fits earnings data in the LFS. Additionally, our own comparison of 40 different bivariate copula families using Akaike Information Criterion (AIC)²⁶ found that the t-copula best fits the LFS data.

²⁶AIC provides a useful tool for model selection. AIC is a calculated quantity that gives a measure of model goodness-of-fit whilst guarding against over-fitting. Mathematically:

$$AIC = 2k - 2\ln(\mathcal{L})$$

Following from equation (16) in appendix A we can define the 2-dimensional t-copula:

$$C_t(u_1, u_2; \rho, \nu) = t_{\rho, \nu}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2)) \quad (7)$$

where ν is the degrees of freedom parameter, t_ν^{-1} is the inverse of the univariate standard Student's t-distribution function, and $t_{\rho, \nu}$ is the multivariate standard Student-t distribution parametrised by the correlation matrix ρ and degrees of freedom ν .

Therefore this requires us to estimate just two parameters; a degrees of freedom (ν) and a persistence parameter (ρ) that describes the correlation between the marginal distributions. Using the rank distribution of earnings ($\sim U[0, 1]$) we estimate a t-copula, using maximum likelihood, for each adjacent age, within each of the 12 characteristic groups. The estimated persistence parameters can be found in appendix B. Earnings are found to be most persistent in middle age across groups, and persistence is generally higher for men and those who have obtained higher qualification levels. These estimates are smoothed using Nadaraya-Watson kernel regression before being used in our simulation model.

For those individuals who are unemployed in the first period but find employment in the second period we cannot estimate their re-entry rank as a function of previous rank. Therefore, we draw their earnings rank randomly from the unit interval, however, we subsequently apply an 'earnings penalty' if an individual is returning from a spell of unemployment. This is estimated as the ratio between the earning rank distribution at age a given employment in the previous period, $(a - 1)$. This is again done separately for each characteristic grouping:

$$\delta_{i,g,a} = \frac{\frac{1}{N^g} \sum_{i=1}^{N^g} r_{i,g,a} 1(E_{i,g,a-1} = 0)}{\frac{1}{N^g} \sum_{i=1}^{N^g} r_{i,g,a} 1(E_{i,g,a-1} = 1)} \quad (8)$$

Therefore, in effect we model individuals that return to employment as facing only a (lower) subset of the population rank distribution of earnings.

Real growth

We identify the differential real earnings growth rate with age for each of our 12 groups. This is non-trivial due to what is known as the 'age-period-cohort (APC) problem'. Wages evolve over the lifecycle with age (e.g., with experience) but we would also expect them to

k = number of estimated parameters in a model \mathcal{L} = maximum value of the likelihood function for a model
Given a collection of models for the data, the model with the lowest AIC is said to be relatively better than the other models.

vary by time period (e.g., with macroeconomic conditions) and by birth cohort (e.g., with education regime during school age). The APC problem arises from the simple fact that each of the three variables of interest can be expressed as a linear function of the other two:

$$Cohort = Period - Age$$

Thus, perfect collinearity exists. Typical methods for dealing with collinearity in samples are not applicable as this collinearity is present in the underlying data generating process itself.

The simplest way to deal with the APC problem is to assume that one of the three effects does not exist. Given we are interested in the evolution of wages over lifetimes, this leaves us with two simple approaches.²⁷ Take a period view, in which we assume cohort effects to be zero, or take a cohort view, in which we assume period effects to be zero.

To estimate the growth rates, we run a fixed effects specification for each gender-qualification group, accounting for either period or cohort effects. The growth rate over age is extracted by taking differences between each adjacent age ($growth_a = \delta_a - \delta_{a-1}$).

In the *period* view:

$$\log(y_{at}) = \beta_0 + \delta_a + \gamma_t + \epsilon_{at} \quad (9)$$

In the *cohort* view:

$$\log(y_{ac}) = \beta_0 + \delta_a + \mu_c + \epsilon_{ac} \quad (10)$$

There is no theoretical reason why one ‘view’ is better than the other, however, we prefer the period view as we believe macroeconomic conditions have had the largest impact on earnings growth in our sample, in part owing to the recession between 2007-2009. To ensure our decision does not impact the model results, we compare the results from the two approaches as a robustness check. (See appendix E)

²⁷Several studies have attempted to produce more complicated models to resolve the APC problem (see, for example, Deaton and Paxson (1994) and Chamon and Prasad (2010)) but for ease of use and to be consistent with previous literature we stick to simple models.

Retirement

As a simplifying assumption we assume retirement at the State Pension age (SPa) for all individuals. This has not been stable over our sample though and will increase in the future.²⁸ It is also plausible that further changes may be made to the SPa before cohorts observed in our sample reach retirement age. We, however, make the assumption that the retirement age of those in our sample will be 68, the planned increase in the SPa. Along similar lines to Britton, van der Erve and Shephard (2019) and Britton et al. (2020), to account for this increase in retirement age, we add in extra years to the age-earning profile at the point when lifecycle earnings growth is zero. We hold all estimated model parameters constant for these three inserted years.

It is unclear how all women reacted to the staggered increases in the retirement age during our sample period. It has been argued that many women were unaware of the change.²⁹ This potentially distorted women's labour market behaviours towards the end of their working life. There is therefore greater uncertainty in our estimation of lifecycle parameters for women in their 60's.

Simulation

The earnings growth dynamics model described above provides the following set of parameter estimates, that can be used to generate a sequence of annual earnings for any individual in the LEO sample:

- Employment probabilities for those unemployed in year $t - 1$
- The parameters of a model to determine employment given income rank, for those employed in year $t - 1$
- Earnings growth in year t
- A correlation parameter to simulate earnings rank dependence in year t
- A penalty to apply to earnings rank on return from unemployment

LEO provides a series of observed earnings for each individual up to age 29. We use the model described by the parameters listed above, to predict the continuation of the earnings sequence to age 68. Each imputed earnings trajectory is a stochastic process, meaning with the same parameter values it will be different each time it is estimated. To account for the uncertainty induced by the randomness of the process, we run each

²⁸The Pensions Act 1995 legislated a path for the women's SPa to increase from 60 to 65 in a staggered fashion from 2010 such that by 2020 the SPa would be 65 for both men and women. Subsequent a series of acts of parliament have legislated for increases to the SPa for both men and women. A table of legislated SPa changes can be found in appendix D.

²⁹See the Women Against State Pension Inequality (WASPI) Campaign.

individual's trajectory ten times, and store the aggregate present value (PV) of each. These ten PVs for each person are then used in a multiple imputation pooling described in appendix F.

Earnings are specified in real terms, so there is no need to deflate future earnings. However, to account for predicted whole-economy growth in future years, we apply the OBR rate of 2% growth in each year. This figure also ensures consistency with the Green Book discount rate.

Estimation of marginal returns

The completed estimation dataset contains data on GCSE attainment, discounted lifetime earnings, and pupil characteristics. To estimate the marginal lifetime earnings return of an improvement in grades, we can impose a production function of the form:

$$PV_i = \alpha + \sum_{s=1}^S \beta_s grade_{is} + \epsilon_i \quad (11)$$

where s denotes subjects. The outcome variable on the left-hand side of the equation is the present value of lifetime earnings.

This formula implies constant returns to scale for each GCSE and overall, and perfect substitutability between GCSEs.³⁰ It assumes that the return to not obtaining a GCSE in a particular subject is equivalent, *ceteris paribus*, to obtaining a fail grade in that GCSE. In practice, this assumption will not hold, because of non-random selection into taking a given GCSE.

However, we do not estimate the marginal effects using the regression equation above. As well as the non-random selection into taking each subject, the equation above also suffers from extreme correlation between grades in different subjects, which would make estimation of each grade effect unstable. Instead, we isolate the marginal effect of a *ceteris paribus* improvement in each subject by estimating a series of univariate regressions, one for each subject, controlling for total attainment in the other subjects.

Each regression is estimated using only the sub-group of pupils who take that subject at GCSE. We therefore estimate the equations:

³⁰An alternative model that does not assume constant returns to scale or perfect substitutes would be a Cobb-Douglas model, which can be estimated by specifying the equation in log-log form. However, the Cobb-Douglas model makes equally strong assumptions about the functional form of returns to scale and complementarities and is not as easy to use to predict policy outcomes.

$$PV_i = \pi_{g[i]} + \beta x_{si} + \delta \left(\sum_{\substack{r=1 \\ r \neq s}}^S x_{ri} \right) + \gamma' \mathbf{z}_i + \epsilon_i \quad (12)$$

where π_g is a fixed effect for group g , x_s is the grade in subject s and \mathbf{z} is a vector of control variables. A full list of variables can be found in appendix G. To construct the variable x , the letter grades $\{U, G, F, E, D, C, B, A, A^*\}$ are mapped to numbers, $\{0, 1, 2, 3, 4, 5, 6, 7, 8\}$. This imposes the assumption that the marginal effect on lifetime earnings at each grade boundary is equal. The letter grades thus mapped are not equivalent to the new regime of numeric grades that runs from 0 to 9.

We also investigate how these returns vary for subgroups of the student population. Therefore, we estimate the marginal grade effects (β_s) separately for 48 subgroups, breaking down the pupil population by; male and female, FSM and non-FSM pupils, high, middle and low terciles of KS2 attainment, and for pupils taking different numbers of GCSE's (1-6, 7-8, 9, 10+).

Although the full dataset is very large, the process of repeated subdividing means that some sub-groups will be very small, with as few as one or two members. Consequently, without any adjustment the estimates from small groups would be noisy, and in some cases, it would be impossible to obtain any kind of estimate. In this context, we use Empirical Bayes estimates to partially pool, or 'shrink', estimates from similar sub-groups towards a common mean, in inverse proportion to the amount of information in the sample.

In a moderately sized dataset, it would be feasible to obtain partially pooled estimates by fitting a multi-level model with varying slope coefficients. In datasets with a very large number of observations and/or sub-groups, this becomes prohibitively computationally expensive. We therefore estimate the model using two-step Empirical Bayes (Robbins (1956); Efron and Morris (1972)). In the Empirical Bayes procedure, a multi-level model is fitted to a sub-sample of the full data, and then the group-level estimates are obtained by ordinary least squares (OLS), with shrinkage applied using the estimated top-level parameters (fixed coefficients and variance components). This is more computationally tractable, as each individual regression has a computational cost equivalent to OLS.

Since the above model is a linear model, the marginal grade effect of interest is the estimate of parameter β . We also present results of the standardised marginal effect, representing the effect of increasing total grades by one standard deviation. The standard deviation of total grades, 11.2, is calculated within groups defined by the number of GCSE's taken, by subtracting the mean grades within each subgroup. The reason for this is that we are interested in the 'intensive' margin where grades improve for a fixed GCSE bundle. The standard deviation of total grades, without holding the number of GCSE's fixed, is about 50% larger. We call the standardised effect thus calculated the 'global'

effect.

Nonlinear grade-effects model

The model specified above assumes that the marginal effect of a grade improvement does not depend on the grade boundary – for example, the additional lifetime earnings associated with moving from a *D* grade to a *C* grade are assumed to be the same as those associated with moving from a *B* grade to an *A* grade. This is a strong assumption. To relax this assumption, we additionally fit a variation of the model in which the marginal effect is allowed to vary with the grade boundary.

To allow the marginal grade effect to vary by grade boundary, we define a model where the single grade variable x_{si} is decomposed into a vector \mathbf{x}_{si} of piecewise effects. The simplest way to define this would be to specify \mathbf{x} as a vector with $K = 8$ elements (that being the number of nonzero grades), and setting each element k of \mathbf{x} as a binary dummy taking a value of 1 if $x_{si} = k$, zero otherwise. The updated equation is

$$PV_i = \pi_{g[i]} + \mathbf{x}'_{si}\beta + \delta_q \left(\sum_{\substack{r=1 \\ r \neq s}}^S x_{ir} \right) + \gamma' \mathbf{z}_i + \epsilon_i \quad (13)$$

The estimated components of β give the levels of each earnings effect compared to the reference level (zero, representing grade *U*). These cumulative effects can be converted to marginal effects by differencing adjacent components. Rather than using binary dummies, we implement the model by designing the vector \mathbf{x}_{si} as a set of piecewise linear splines, which allows more flexibility in the model specification. With the piecewise linear spline it is straightforward to allow some marginal effects to be constrained to be equal, if necessitated by not having enough data to estimate the full model.

Pre-treatment controls

As well as controlling for attainment in other subjects, we also include a wider set of pre-treatment controls, guided by theory and empirical evidence. There are well-documented ‘pre-treatment’ effects of primary school, pre-school and the home learning environment on both GCSE outcomes and wages (Sammons et al. (2014)). As such, we include controls for pupil’s Key Stage 2 attainment, averaged over English and Maths scores and standardised within each year. Key Stage 2 performance, however, can only provide partial control for these pre-treatment effects. The inclusion of prior

attainment mitigates, but does not remove, the possibility of selection effects and unobservable pupil heterogeneity.

Levels of attainment also vary by pupil subgroup. We include controls for gender and disadvantage (location-based, through IDACI, and pupil-level, through FSM eligibility³¹) Girls do better than boys in all headline measures at KS4 and in the vast majority of individual subjects taken (DfE (2018b)). Gender differences do though vary by subject. For example, boys marginally outperformed girls in Physics, whereas girls significantly outperform boys in English Literature.³² Attainment is lower for economically disadvantaged pupils compared to all other pupils, and across all headline KS4 attainment measures. In 2017, White pupils who are eligible for FSM are the lowest attaining major ethnic group, with an average Attainment 8 score 14.0 points below the national average; the gap increases to 17.1 points when only White FSM boys are considered (DfE (2018a)). A full list of controls can be found in appendix G.

Controlling for unobservables

We can never be certain that all the relevant control variables are included in the regression model. Unobservable factors may correlate with the dependent and independent variables, leading to omitted variable bias. In our model we are particularly worried that there is unobservable variation across schools. If school is assumed to be time invariant, then including school level fixed effects will eliminate any worry of omitted variable bias. Fixed effects in principle allow a different intercept for each group, in our case each school. In practice, rather than estimating each intercept we simply demean all variables, using school level means.

Including school fixed effects in this way controls for average differences between schools. Controlling between school variation allows us to estimate effects within schools and so the results are generalisable and hold irrelevant of the school a pupil attended. This inclusion of school fixed effects mitigates against a likely large source of unobservable variation. This is in addition to controlling for an extensive set of observable variables. This does not, however, rule out the possibility that some un-observable variation still exists. For instance, there may be unobserved variation between classes or teachers within schools.

³¹Free School Meals (FSM) eligibility is determined by the receipt of certain income-related benefits. For a detailed list see page 5 of DfE (2018a). DfE (2021) report that 17.3% of school age pupils were eligible for free school meals.

³²Based on 2017/18 provisional data, showing the percentage achieving a grade A*-C or 9-4 under the new points grade: 91% of boys taking Physics achieved these higher grades, compared to 90% of girls; whereas 65% of boys taking English Literature achieved these grades, compared to 80% of girls (DfE (2018b)).

Post-treatment controls

One additional option would be to include post-treatment effects, noting that information on later attainment at Key Stage 5 (A-levels and equivalents), further and higher educational outcomes are all available in LEO. In contrast to pre-treatment effects (e.g. KS2 performance), the inclusion of post-treatment controls is problematic, statistically speaking. This would create a version of traditional selection bias (see Angrist and Pischke (2009)). The composition of the groups taking different pathways is altered by changes in GCSE attainment and so the estimated effect would contain bias. It would be implicitly comparing different groups of people. Simulation studies show that this form of bias can be severe, even with only small amounts of confounding.

The value of policies to improve GCSE attainment include any educational and labour market pathways that this opens up. This report aims to obtain the total earnings return of improved GCSE performance, through both direct effects (the inherent value of GCSEs in the labour market) and indirect effects (mediated through later education). Hence investigating later educational pathways is not essential for our purposes. Further research on the causal pathways, after compulsory education, would though be of interest in its own right. For example, Chowdry et al. (2013) found that GCSE and A-level attainment could explain a substantial proportion of the socioeconomic differences in Higher Education (HE) participation. They control for GCSE performance using the standard threshold measure of achieving 5 A*-C grades (including Maths and English). Again, LEO would allow researchers to explore the role of prior attainment in finer detail, across more GCSE grade boundaries, by subject, and linked to similarly detailed information on A-level results.

Results

This section presents the main results of both the linear and non-linear grade returns models, split down by demographics and subject. As well as setting an overall scale for the returns to attainment, we find evidence supporting the existence of several empirical regularities. In summary, the main empirical regularities we find are the following:

1. Returns to grade improvements are greater for male students, especially for those with high prior attainment.
2. Returns are also greater for students not eligible for Free School Meals (FSM), especially for those with high prior attainment.
3. Marginal returns to KS4 attainment increase with increasing prior attainment at KS2 for males, but slightly decrease for female students.
4. Returns to marginal grade improvements decrease in the number of GCSE's taken.
5. There is large subject variation in grade returns, with the highest returns almost three times as large as the smallest.
6. There is grade boundary variation: the largest returns are observed at D-C and C-B boundaries. Returns are sharply decreasing at B-A and A-A* boundaries for female students, and to a lesser extent for males.

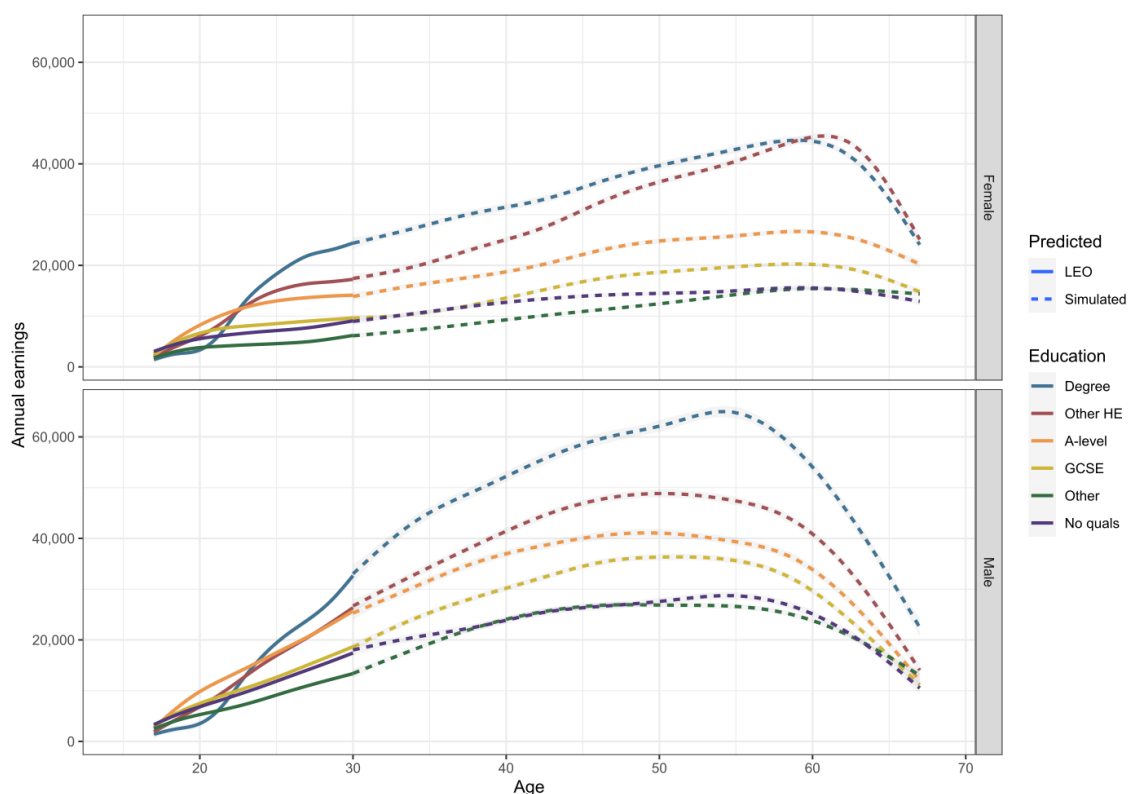
Before presenting the regression results, the next section presents summaries of the simulated earnings trajectories, and the calculated lifetime earnings (both discounted and undiscounted) broken down by demographics and subject.

Lifetime earnings and PV statistics

Figure 3 shows the conditional means of earnings by age, gender and highest level of education. This is a combination of observed earnings to age 29, and simulated predictions, thereafter. The difference in peak earnings for male and female respondents with degree-level education is almost £20,000 per year. Earnings for respondents with at least first degree-level education drop sharply from the late fifties, due to a combination of changes in working patterns as respondents approach retirement, and inflation eroding salaries.³³

³³Working patterns would appear to be the dominant force, see the comparison of LFS hourly and weekly earnings in appendix K.

Figure 3: Mean Earning Trajectories



As a robustness check we also start our simulation model earlier, at age 26 so we can compare how the simulation performs compared to actual observed outcomes in LEO at ages 27, 28 and 29. We compare key statistics from the earnings distributions generated by the simulation to those observed. Additionally, we compare differences in rank earnings after 1, 2 and 3 years to ensure we are adequately capturing individuals' earnings persistence. We find that the simulations appear to perform well. Further detail and graphics can be found in appendix E.

Table 3: Summary Statistics by GCSE Subject

Subject	N. Observations	Undiscounted Earnings	Discounted Earnings	FSM	Male	KS2 Maths	KS2 English	N. GCSEs	GCSE pts
English	2,144,943	£1,330,615	£514,957	16.88%	51.55%	0.00	0.00	8.14	37.19
Maths	2,144,943	£1,330,615	£514,957	16.88%	51.55%	0.00	0.00	8.14	37.19
Eng Literature	1,818,032	£1,383,419	£536,478	14.41%	49.46%	0.12	0.14	8.70	41.57
Double Science	1,683,149	£1,360,430	£527,833	14.72%	50.78%	0.06	0.07	8.72	40.30
French	978,097	£1,413,684	£548,219	11.87%	47.38%	0.19	0.22	9.04	44.42
Geography	677,789	£1,466,134	£570,711	11.45%	57.06%	0.20	0.15	9.04	44.40
History	675,949	£1,476,738	£571,178	11.12%	50.9%	0.25	0.32	9.14	46.59
Art	642,040	£1,297,617	£499,637	15.85%	43.34%	-0.02	0.03	8.71	40.43
PE	446,575	£1,427,888	£561,931	13.28%	65.11%	0.11	0.00	8.85	40.28
German	406,864	£1,476,126	£573,156	8.63%	48.91%	0.35	0.37	9.22	47.73
RS	406,233	£1,420,934	£547,879	14.75%	43.11%	0.22	0.29	8.92	45.10
DT Resistant mat.	391,259	£1,441,481	£564,833	16.01%	80.03%	-0.04	-0.21	8.61	37.00
DT graphics	376,613	£1,448,418	£560,107	12.96%	57.44%	0.18	0.17	9.01	43.41
DT Food	373,163	£1,207,913	£467,880	15.17%	28.73%	-0.12	0.00	8.69	39.35
Drama	326,018	£1,293,103	£498,827	15.55%	37.61%	-0.02	0.10	8.84	41.46
Business	303,880	£1,469,181	£574,448	11.98%	58.26%	0.23	0.17	9.14	43.71
IT	299,786	£1,467,405	£571,183	13.19%	59.8%	0.19	0.13	9.20	44.09
DT textiles	181,463	£1,135,718	£435,294	16.17%	3.39%	-0.02	0.17	8.83	42.93
Music	167,167	£1,469,926	£563,056	11.2%	48.59%	0.33	0.38	9.18	47.67
Spanish	162,860	£1,429,522	£553,127	12.68%	41.83%	0.31	0.37	9.12	46.82
Triple Science	104,305	£1,809,577	£698,412	5.17%	59.37%	1.01	0.90	10.15	62.91
DT Electrical	33,887	£1,673,815	£654,800	12.79%	93.13%	0.34	0.11	8.97	43.49
DT Systems	23,681	£1,720,674	£672,653	10.47%	92.09%	0.46	0.23	9.10	46.23

Table 3 shows sample sizes taking each subject, alongside the average aggregate lifetime earnings (undiscounted and discounted), proportion FSM eligible, proportion male,

average standardised KS2 Mathematics score, average standardised KS2 English score, and average number of GCSE's taken. Everyone in the sample takes English and Mathematics, and so the averages in the top two rows are the overall population averages for each measure.

The subject for which the earnings of those taking it are largest is Triple Sciences, with an average undiscounted lifetime earnings of £1.8 million. Students taking Triple Sciences also have the highest KS2 scores, the largest average number of GCSEs, and the lowest proportion FSM eligible. 59% of those taking Triple Sciences are male.

The subject with the lowest average earnings is DT Textiles, with an average undiscounted lifetime earnings of £1.1 million. It does not have the highest average number of FSM students or the lowest KS2, but it has the lowest proportion male, with only 3.4% male students. Note that all subjects apart from English and Maths are additional subjects, which explains why almost all subjects have slightly more advantaged, higher attaining samples than the population.

Table 4: Summary Statistics by demographics

Gender	FSM	KS2 Tercile	N. Observations	Undiscounted Earnings	Discounted Earnings	N. GCSEs	GCSE pts
Female	Not FSM	Bottom	231,973	£880,213	£335,948	7.54	26.79
Female	Not FSM	Middle	283,867	£1,090,915	£420,178	8.41	39.58
Female	Not FSM	Top	344,523	£1,348,779	£521,862	9.25	55.33
Female	FSM	Bottom	86,736	£745,972	£273,136	6.53	19.00
Female	FSM	Middle	62,692	£889,388	£330,890	7.22	27.76
Female	FSM	Top	29,541	£1,120,795	£427,430	8.39	43.40
Male	Not FSM	Bottom	294,443	£1,350,119	£530,264	7.28	23.44
Male	Not FSM	Middle	308,483	£1,559,818	£610,507	8.24	35.64
Male	Not FSM	Top	319,663	£1,889,477	£734,372	9.21	52.01
Male	FSM	Bottom	92,787	£1,207,229	£458,811	6.23	16.10
Male	FSM	Middle	63,867	£1,336,998	£509,549	6.95	23.74
Male	FSM	Top	26,368	£1,587,776	£612,973	8.34	40.11

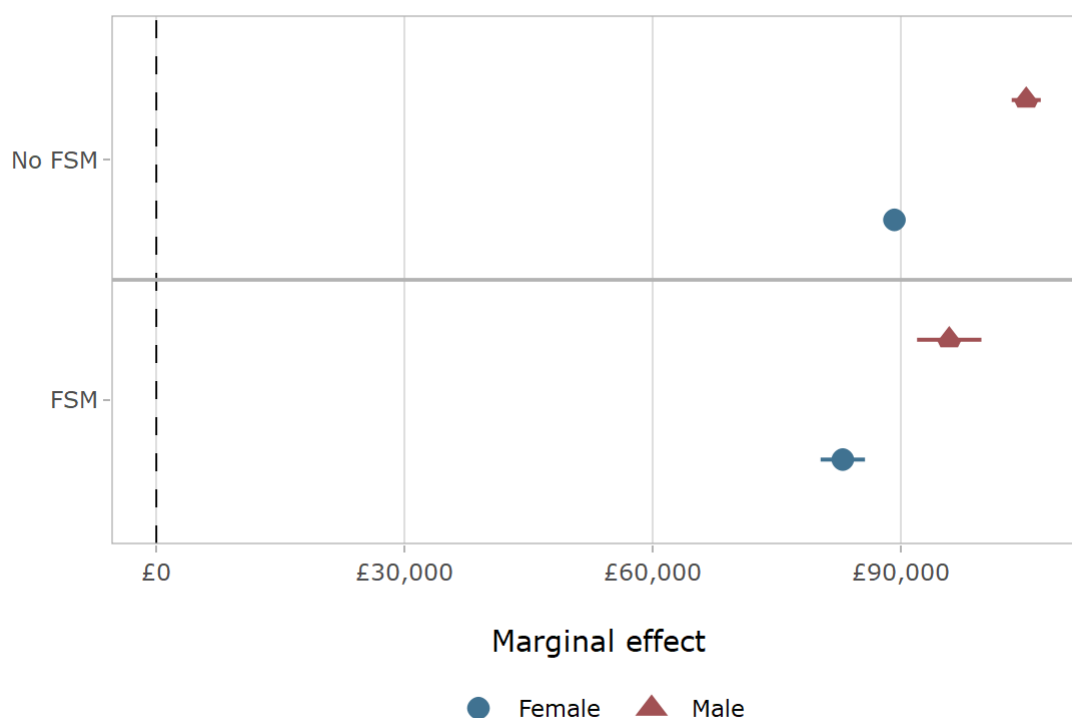
Table 4 presents sample sizes, earnings and other descriptive statistics for each demographic group. The lowest earning group are female, FSM eligible students in the bottom third of the KS2 attainment distribution. The predicted undiscounted earnings for this group is less than £746,000. In contrast, the earnings for male FSM eligible, low KS2 attaining students is £1.2 million, even though the female students have better GCSE grades on average (19 vs 16 points). The group with the largest predicted earnings are male, non-FSM eligible, top-3rd KS2 attainers, whose average earnings are £1.9 million.

Returns to 'global' grade improvement

We first show results across the full distribution of GCSE points scores. Figure 4 shows the marginal return to an improvement of 11.2 grades across all GCSEs taken. This is the standard deviation (SD) of total grade points in our sample, when conditioning on the number of GCSEs taken. The average return is £96,111 (£95,154—97,069).

The return for male students is larger than the return for female students (these are scaled versions of the unit-grade effects) and the return for non-FSM greater than those for FSM. The average return for a one-SD improvement in grades represents 19% of the mean discounted lifetime earnings. Dividing this amount by the SD of discounted lifetime earnings £211,288³⁴ provides a standardised effect size of 0.455 (0.450-0.459). This means that a one-SD change in grades corresponds to a 0.455 SD change in discounted earnings.

Figure 4: Global effect for 11.2 grade improvement: FSM status and gender



Marginal grade returns

In this section we explore the marginal grade return, setting aside subject. The reported marginal effects are based on a letter (A*-G) grade scale, used to assess the cohorts we observe in LEO. The average return to a one-grade improvement is estimated to be £8,566 (£8,481—8,651).

We interact the marginal returns by gender, free school meal (FSM) status, number of GCSEs taken, and tercile of prior attainment at Key Stage 2. Figure 5 presents the fully interacted estimates. We consistently find higher marginal returns for male (£9,091—9,377) than for female students (£7,766—7,945). The difference between male

³⁴Standard deviations in this section are calculated conditional on a fixed number of GCSE's taken. This is for consistency with estimation of the coefficients, which was undertaken holding the number of GCSE's fixed.

and female marginal returns appears to be largest for those with high prior attainment, who sit a larger number of GCSEs. The return to grade improvements with additional GCSEs appears to be more sharply decreasing for female students than male students. There is also a smaller return per grade for FSM eligible pupils (£7,763—8,186) than ineligible pupils (£8,593—8,779).

The relationship between marginal returns and KS2 attainment are mediated by gender, but this is only clear for non-FSM students; the marginal returns appear to be decreasing with increasing prior attainment for females, but increasing with increasing prior attainment for males, conditional on the number of GCSEs sat. Both the gender disparity, and the decreasing marginal return to increasing number of GCSEs are most stark in the bottom-right pane of figure 5, representing top-tercile KS2, non-FSM students.

We make no attempt to model changes in GCSE subject choices. We do, however, look at how the marginal returns to grade improvements vary by how many GCSE subjects' pupils take. In our sample, over 80% of pupils sat seven or more GCSEs, but around 8% took four or less. In 2019, these figures were not dissimilar.³⁵ We find that the return to a grade improvement decreases slightly for those taking a greater number of GCSEs.

Marginal returns by subject

We estimate the marginal earning returns to an improvement of one (A*-G) grade for each of the most popular GCSE courses.³⁶ These estimated effects by subject are reported in figure 6, and by gender in figure 7. A one-grade improvement in Maths is estimated to be associated with an increase of £14,579 in present value of lifetime earnings, whereas the figure is around a third for German: £5,704.

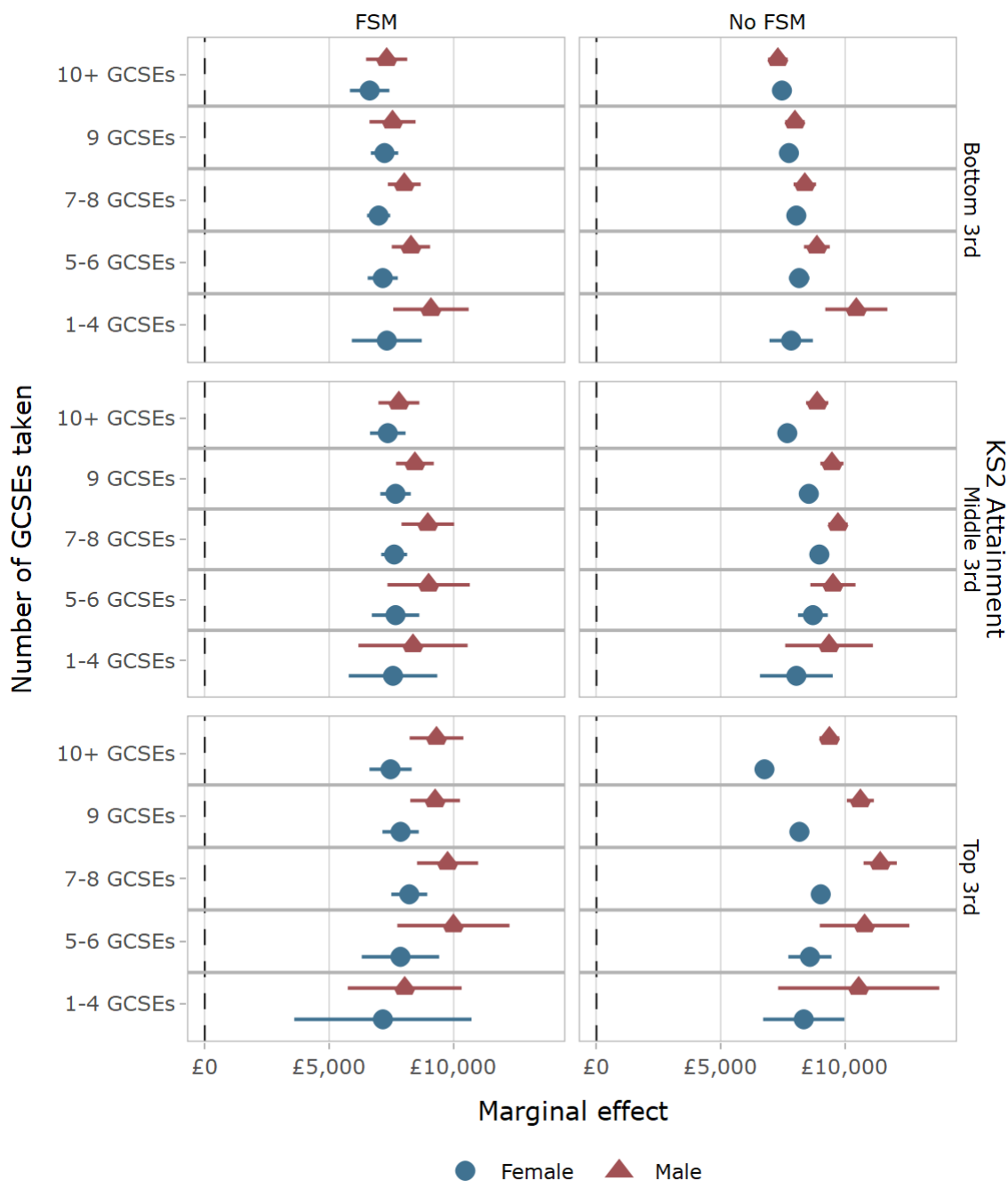
For most subjects, we find that male students receive a higher earnings return from an equivalent improvement in performance than female students. Improvements in Drama and Music, however, are on average associated with larger gains for women. For example, a one-grade improvement in Music is worth £6,052 for women and £4,820 for men. We also find the relative difference in return by gender is largest for those subjects that have the highest returns. For the five subjects that have the highest returns to improved performance (Maths, Business, IT, Geography and PE), a one-grade change is worth between £2,106-£3,980 more in present value terms for men than it is for women.

The number of entrants for each subject differs. Some GCSEs such as English and Mathematics are compulsory; other subjects are not and have only a fraction of the number of entries. As discussed above, our estimation strategy can only account for

³⁵See Ofqual (2019).

³⁶Those with more than 20,000 entries in summer 2019, which roughly equates to 4% of all pupils.

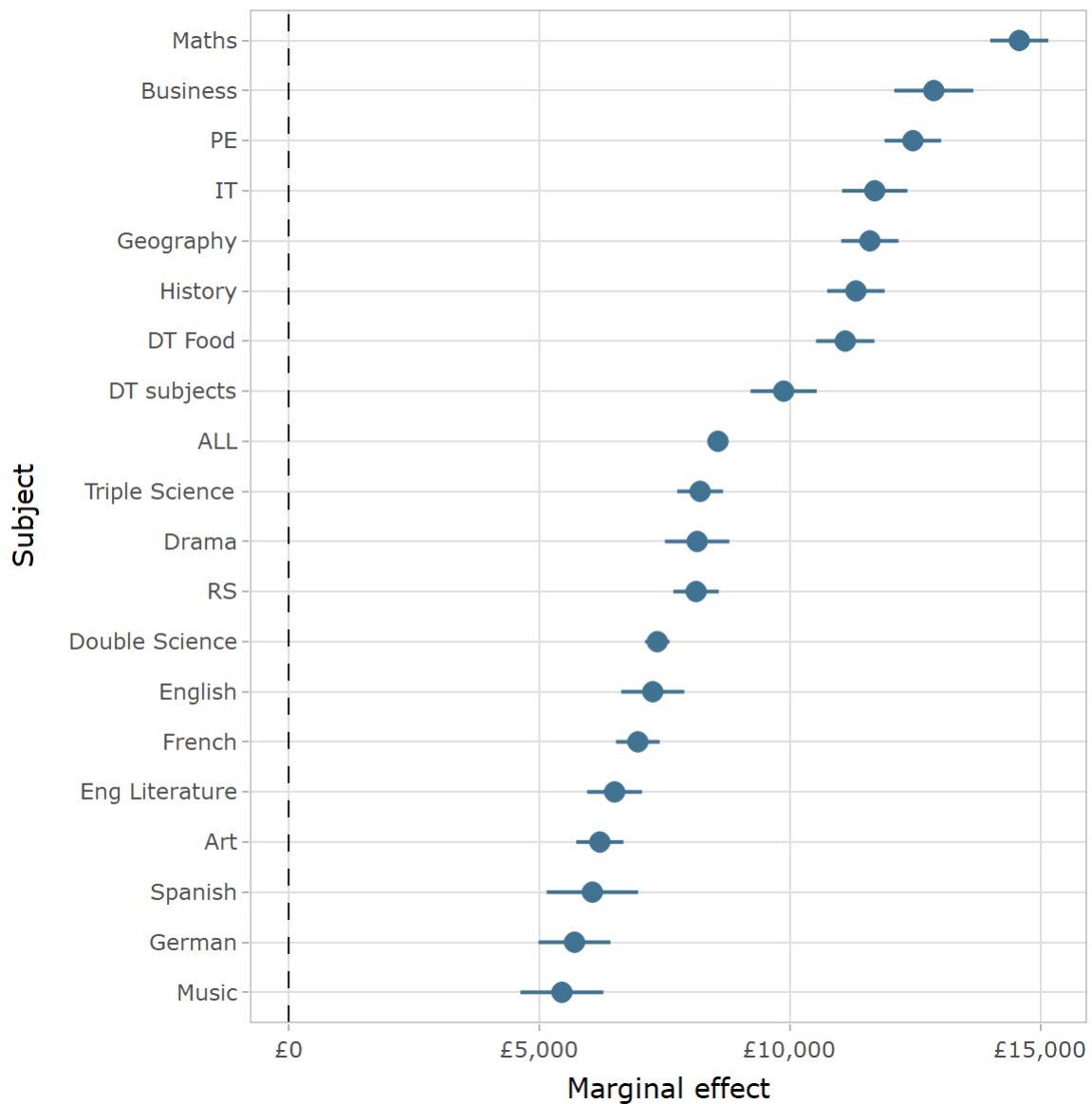
Figure 5: Average standardised marginal effect for all subjects: number of GCSEs, FSM status, KS2 attainment and gender



observable differences between pupils. Therefore, some selection into taking different subjects based on unobservable characteristics cannot be ruled out.

There is little risk of selection effects in English and Mathematics, given they are both taken by almost every pupil. However, we find differential returns between these two subjects; the marginal return from a one-grade improvement in Mathematics is £14,579 in present value of lifetime earnings, compared to £7,266 for English. This would lend itself to the argument that the subset of human capital required to perform well in GCSE Mathematics is more valuable in the labour market than the set needed to achieve highly

Figure 6: Average marginal effects by subject

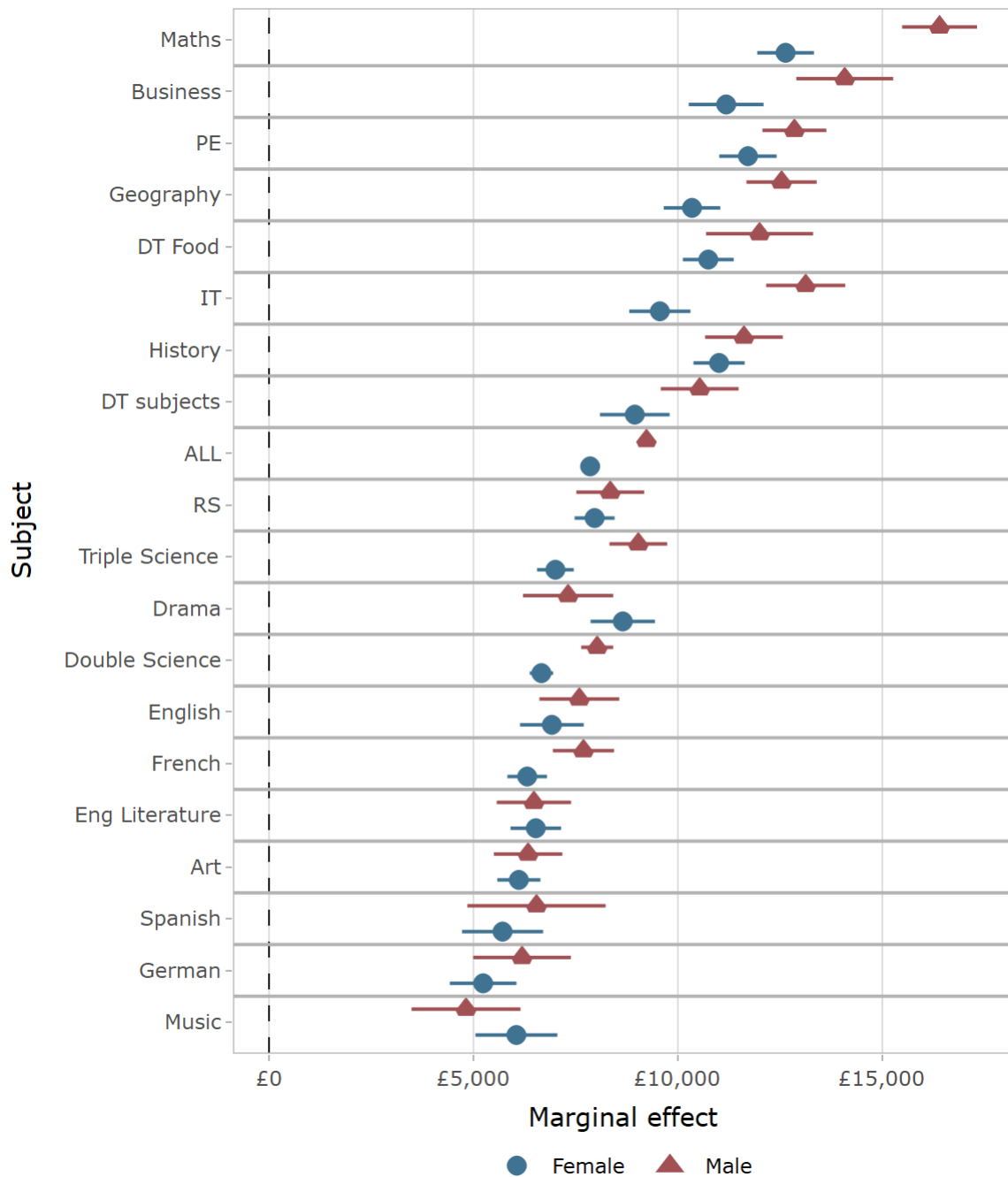


in GCSE English. Equally, we cannot rule out that signalling effects are also stronger for improved grades in Mathematics, potentially through screening and sorting mechanisms used by employers and education institutions (see section on human capital and signalling).

The number of entrants for each subject has also changed over time. This has potential implications for the interpretation of the results in a present-day context. For example, there has been a substantial rise in the number of pupils taking Triple Science. Between 2007 and 2012 the percentage of pupils in state-funded schools taking Triple Science rose from 6% to 23%.³⁷ It is not clear or obvious exactly what impact this is likely to have but it is possible the unobserved characteristics and motivations of those taking different types of GCSE science qualifications has changed, as a result.

³⁷DfE written memorandum to the Science and Technology Select Committee. "Educating tomorrow's engineers: the impact of Government reforms on 14-19 education" (DfE (2013)).

Figure 7: Average marginal effects by subject and gender

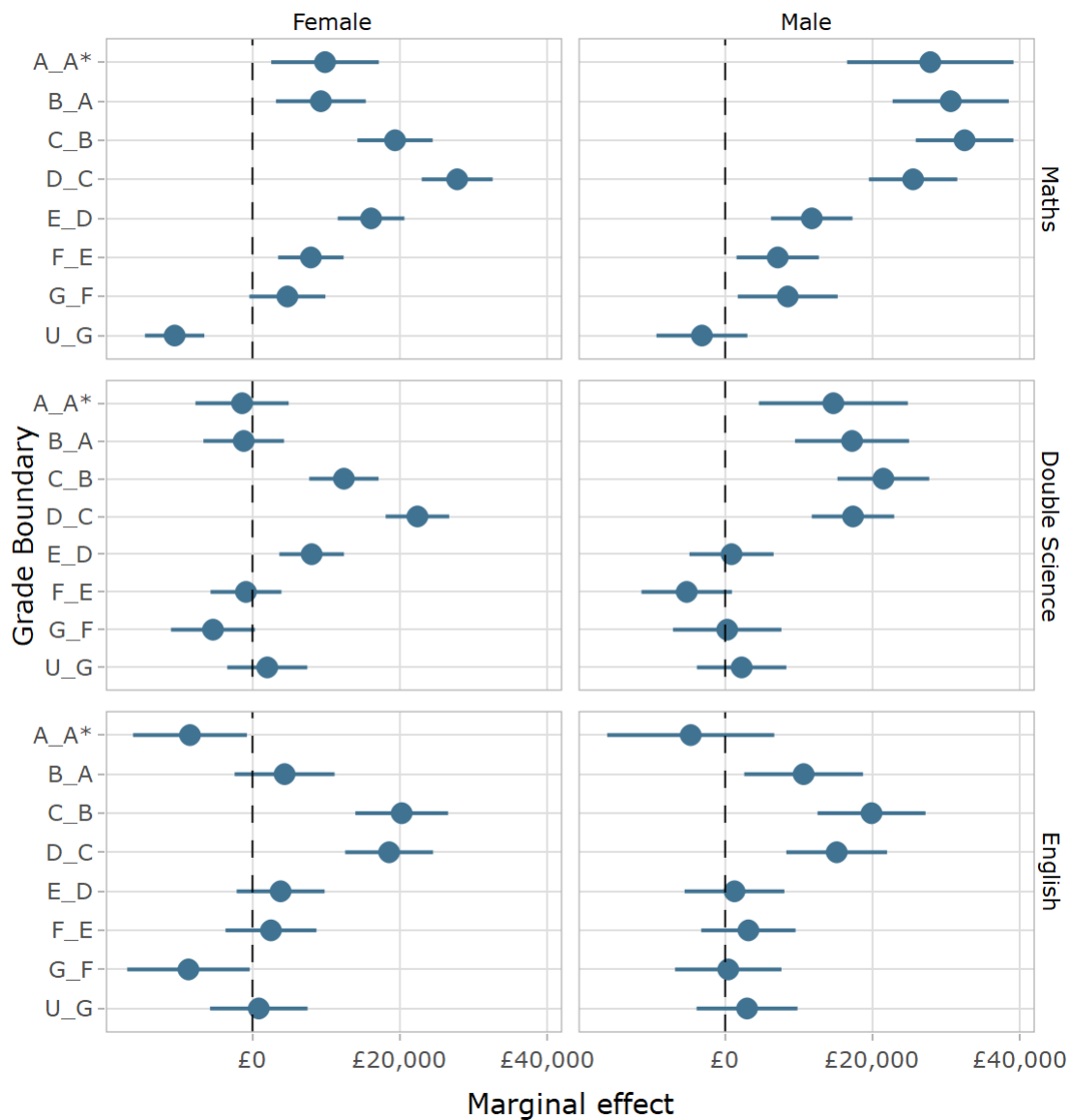


Non-linear returns by grade boundary

Figures 8 and 9 present the results of the model when the marginal returns are allowed to vary by grade boundary. To compensate for the additional flexibility of the model it has only been interacted with gender. The estimates for six key subjects are presented in these figures, the estimates for the remaining subjects can be found in appendix I.

These results, again, indicate the earnings return to a marginal improvement in grades. For example, the change in present value of lifetime earnings associated with achieving

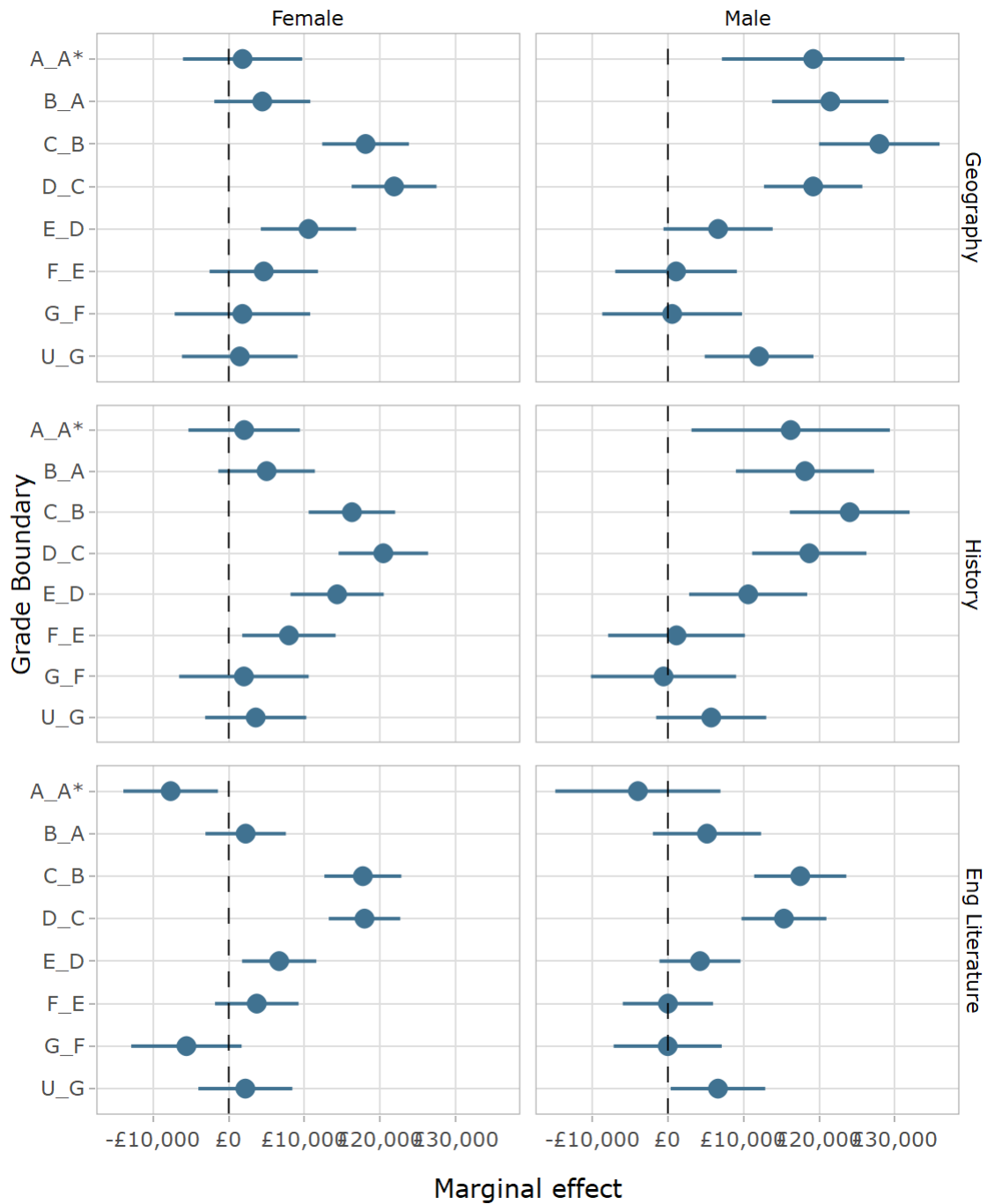
Figure 8: Non-linear returns: Maths, Double Science and English



an A instead of a B in English is ~£5,000 for female pupils. This is not the total return of achieving an A in GCSE English. The general pattern we observe is that marginal returns below the D-C boundary are smaller, and in many cases not measurably different from zero. Whilst the marginal returns at the D-C boundary and C-B boundaries are typically the largest in magnitude.

The pattern of marginal returns at the B-A and A-A* boundaries depend on both gender and the subject. A striking regularity is that returns to top grades are usually much smaller for female students than male students. For example, in Maths the marginal returns to A* and A grades for male students are only slightly smaller than the marginal return to a B grade. For female students though, the additional returns to A and A* grades drop off sharply. In other subjects the returns to top grades are small, non-existent, or even negative for both genders. For example, in English and English Literature, the return to an A* grade is not measurably different to zero for male students and is negative for female

Figure 9: Non-linear returns: Geography, History and English Literature



students. The figures in appendix I contain other examples of negative returns to top grades; in French, German, and Music.

The negative returns raise the question of whether these represent mainly people who have only done well in one or two subjects (including the subject in question), or whether they are driven by people who do well in all subjects, or both. Appendix J uses scatter plots to analyse the returns to grades as a function of total grades in other subjects. We find evidence that small and negative marginal returns are driven by decreasing marginal returns to additional grades for those who do well in all subjects. It should be noted that

these scatters do not represent exactly the same model as the main estimates, as they do not adjust for any fixed effects or other controls.

Discussion

The objective of our research was to obtain robust estimates of changes in lifetime earnings, associated with marginal improvements in GCSE grades. This report provides the first published earnings returns estimates at KS4, which exploit the LEO administrative data. We provide detailed breakdowns by subject, by grade, and by pupil characteristics. This level of detail was hitherto unavailable to researchers using UK survey data. Indeed, the measurement of GCSE performance in these surveys was limited to the total number of GCSEs obtained and crossing broad threshold-levels, such as achieving five or more A*-C grades.

Estimates of observed variation in earnings should not be used in isolation, to guide targeted education policy. For example, whilst we report relatively high marginal grade returns for pupils who are male and non-FSM, it does not follow that policy ought to target these groups. There are (normative and positive) value judgements to consider, alongside variation in the costs of intervention by sub-group. Longitudinal estimates of earnings differentials inevitably reflect existing social and economic disadvantages in education and labour markets, which policy intervention itself may intend to overcome, rather than perpetuate.

Therefore, this section briefly highlights the need for circumspect interpretation and application of our estimates in economic appraisal. For practitioners, we provide more detailed recommendations in our companion piece, the Schools Policy Appraisal Handbook (Hodge, Little and Weldon (2021)).

Causal effects, human capital and signalling

Our strategy for identification of returns relies upon an assumption that observable variables (school, demographics, prior attainment, other subjects studied) capture all the relevant variation between individuals, that may affect outcomes. If this assumption is violated by the presence of unseen variables that are correlated with both GCSE grade improvements and lifetime earnings, the estimates will be biased. Indeed, whilst the rich longitudinal data available in LEO allows us to observe a substantial number of relevant variables, it is extremely likely that we have not been able to capture all the characteristics that jointly determine grades and earnings. If the correlation between grades and earnings, induced by such unseen traits, is positive, then the implied bias in our estimates is positive; that is, they are larger than they should be.

Causal mechanisms are also uncertain. Consider, for example, the relatively high marginal returns associated with attainment in Maths, particularly at the C/D grade boundary. Following human capital theory, higher earnings could result from the productivity-enhancing effects associated with the study of KS4 Maths, to a good level. On the other hand, test scores could be a signal a learner's innate skills, motivation and abilities, irrespective of whether achieving a C grade augmented their productivity.

Gambin et al. (2014) argues that it “makes little difference” to the individual learner whether higher earnings arise from the effects of human capital or signalling. Yet, the distinction is important to policy makers. Human capital theory provides a more compelling case for public investment to education to grow the economy as a whole. In the presence of signalling, lifetime earnings could overestimate the return on human capital investment.

As the empirical predictions of both human capital and signalling theory are so similar it is hard to disentangle them from one another. Several studies have made compelling attempts to measure the value of skills on labour market outcomes. In a recent review of the human capital and signalling literature, Wyness, Macmillian and Anders (2021) concluded that:

“there is a clear picture of substantial importance for human capital theory as a dominant part of the link between education and later earnings”

Even so, causal interpretations remain elusive in our estimates and so we cannot disentangle human capital and signalling effects. Their relative importance may well explain some observed variation in earnings, by grade boundary, by subject and even by demographics.

Multiple skills

Achieving high grades in different GCSE subjects often requires different skills. Until now we have used the term ‘human capital’ as a catch all description of the knowledge and skills individuals hold. Human capital can more accurately be broken down into multiple component skills. There is, for instance, an ever-growing literature that attempts to disentangle the relative effects of cognitive and non-cognitive skills.³⁸ In the context of GCSEs it is important to recognise that different subjects will reward the possession of different skills. For example, the set of skills required to get an A grade in Mathematics were not the same as those needed to achieve an A in Art, although these may overlap. Different skill sets are likely to then have different earnings returns in the labour market, the quantity of interest in this report.

³⁸ Almlund et al. (2011) provide a helpful review of non-cognitive skills and personality traits in education economics.

This should not, however, detract from the usefulness and value of less-easily monetised skills to society. Care should be taken to avoid skewing policy intervention toward more measurable skills (GCSE results), or measurable components of productivity (earnings). Our estimates are instructive in policy appraisal, only where they are appropriately balanced with evidence of decisive, less-easily monetised impacts.

Implications for recent and future pupil cohorts

As noted in the introduction, forecasting the impact of educational attainment on labour market outcomes is inherently difficult. The premium that employers attach to different types and levels of skill changes, unpredictably. The type and scarcity of foundational skills acquired during KS4 also changes. With regard to signalling effects, the distribution of GCSE test scores differs from one cohort to the next, again with uncertain effects on labour market returns.

In the current context, these challenges have been brought into focus by the Coronavirus pandemic, both through its disruption to GCSE exams and its negative effects on the labour market. In addition there are potential longer-term structural shifts in employment caused, for instance, by automation.

We do not broach these forms of uncertainty in this report. When interpreting or using these estimates, it ought to be acknowledged that an observed relationship between GCSEs and earnings is a highly imperfect guide to the future.

Wellbeing

Traditional economic indicators, such as wages, cannot fully capture the effects of schooling on people's lives. Giving regard to direct measures of personal wellbeing can improve decision-making and wellbeing appraisal continues to evolve in the UK.

There are data constraints in the analysis of education policy on wellbeing. We could not, for instance, emulate our method to predict wage impacts of Key Stage 4 performance, to make similar predictions about adult life satisfaction. Again, we recognise that these constraints relate to what we can measure, not what matters. Ultimately, the economic objectives of school policy ought to be grounded in broader conceptions of social welfare.

In the context of this report, it is particularly important to note that income generally improves personal wellbeing and utility at a decreasing rate (Layard et al. (2018)). That is, the value of an extra pound is worth more to people on lower incomes than to those on higher incomes. HM Treasury (2020) proposes the use of 'welfare weights' to better represent the utility value of income, in cost-benefit analysis. Our Schools Policy

Appraisal Handbook (Hodge, Little and Weldon (2021)) describes how these weights could be applied to the lifetime earnings estimates in this report.

Macroeconomic and wider benefits

We use a microeconomic strategy, using pupil-level data to estimate the first-order, private returns to GCSE performance. By contrast, macroeconomic estimates can take account of external benefits to the whole economy, which often motivate public investment in education. A more extensive discussion of how best to reconcile these approaches is provided in the Schools Policy Appraisal Handbook (Hodge, Little and Weldon (2021)) and in Crawford and Cattan (2013). In sum, macroeconomic estimates are considered an ‘upper bound’ on the total benefit of GCSE attainment. Microeconomic estimates retain the advantage of finer measurement by subject and grade, with greater control for pupil- and school-level characteristics.

Extensions

Further limitations relate to the scope of analysis, some of which could be overcome through further research. Here, we highlight three viable extensions that may be of value to policy makers.

Subject choice

Our modelling approach was designed to estimate earnings associated with marginal improvements in attainment. That is, we explore attainment in each subject at the end of KS4, not the effect of taking subject choices at the beginning of KS4.

There is scope for research, using LEO, to explore labour market outcomes associated with subject choice. This would require a different strategy. GCSEs cannot be viewed as standalone qualifications as, typically, pupils take between 7 and 10 GCSEs in different subjects. The composition of this subject ‘bundle’ is likely dependent on a combination of pupils, parents and schools’ preferences.³⁹ One would need to consider at least two behaviour responses: a ‘switch effect’, changing one subject for another; and a ‘number effect’, taking additional GCSEs in a given subject.⁴⁰

³⁹There is evidence that schools constrain the ‘bundle’ choice. Jin et al. (2011) report that one in five pupils could not study a subject they would like to and so have their choice constrained in some way. In about a third of cases this was because the school simply did not offer the course. Anders et al. (2018) show schools explain about a 25% of the variation in the academic selectivity of the subjects even after controlling for school demographics.

⁴⁰In addition, some interventions involve a combination of effects at both the intensive and extensive

Key Stages 1, 2 and 5

This report focuses on GCSEs, proceeding recent earnings estimates at degree-level (Belfield et al. (2018b), (2018a); Britton et al. (2020)) and for higher vocational qualifications (Espinoza and Speckesser (2019)). Estimates associated with attainment at KS1, KS2 and KS5 are yet to be explored, at similarly detailed levels of analyses using LEO. These are feasible extensions and likely to be of significant interest to policy makers.

Exchequer returns

We estimate changes in individuals' gross earnings. We do not estimate benefits to the taxpayer, who typically bear the cost of investment in public education. Modelling the fiscal returns to improved GCSE performance is complex, requiring the simulation of multiple direct and indirect effects on the welfare system. There is precedent in estimating lifetime earnings returns associated with degree-level qualifications, Britton et al. (2020) also estimated the returns accruing to the Exchequer.⁴¹ This would be a useful extension in the context of policy appraisal.

margins, for example, to encourage both uptake of, and improved attainment in, STEM subjects.

⁴¹Returns to the exchequer were limited to largely first-order effects income tax, national insurance contributions and student loan repayments. Britton et al. (2020) did not attempt to estimate savings through the benefits system, such as support for low or no earners through Universal Credit. The exchequer will also benefit from indirect benefits of education, for instance where improved health and lower crime reduces public expenditure.

Appendices

A Sklar's Theorem, copula's and lifetime earnings

Copulas allow us to capture rich correlation structures, unlike linear correlation coefficients (e.g., Pearson) which by definition only capture linear dependence among variables. Copulas provide a flexible statistical tool which allow the separation of the dependency structure from the marginal distributions in a multivariate distribution. In essence, a copula is a function which simply “couples” together a series of one-dimensional marginal distribution functions.

Copula methods are a direct result of Sklar's Theorem (Sklar (1959)).

Theorem A.1 (Sklar's Theorem) For any multivariate cumulative distribution function F there always exists a copula C that links the marginal distribution functions F_1, F_2, \dots, F_N :

$$F(x_1, x_2, \dots, x_N) = C[F_1(x_1), F_2(x_2), \dots, F_N(x_N)]$$

Given that, regardless of its distribution, any variable X when passed through its own cumulative distribution function (CDF) will always be transformed to a standard uniform distribution:

$$U = F_X(X) \sim U[0, 1] \tag{14}$$

Equivalently:

$$X = F_X^{-1}(U) \tag{15}$$

A copula function C can be defined as:

$$C(u_1, u_2, \dots, u_N) = F(F^{-1}(u_1), F^{-1}(u_2), \dots, F^{-1}(u_N)) \tag{16}$$

A copula function therefore describes a N-dimensional probability distribution function on the hypercube $[0, 1]^N$. It links a series of marginal standard uniform distributions together.

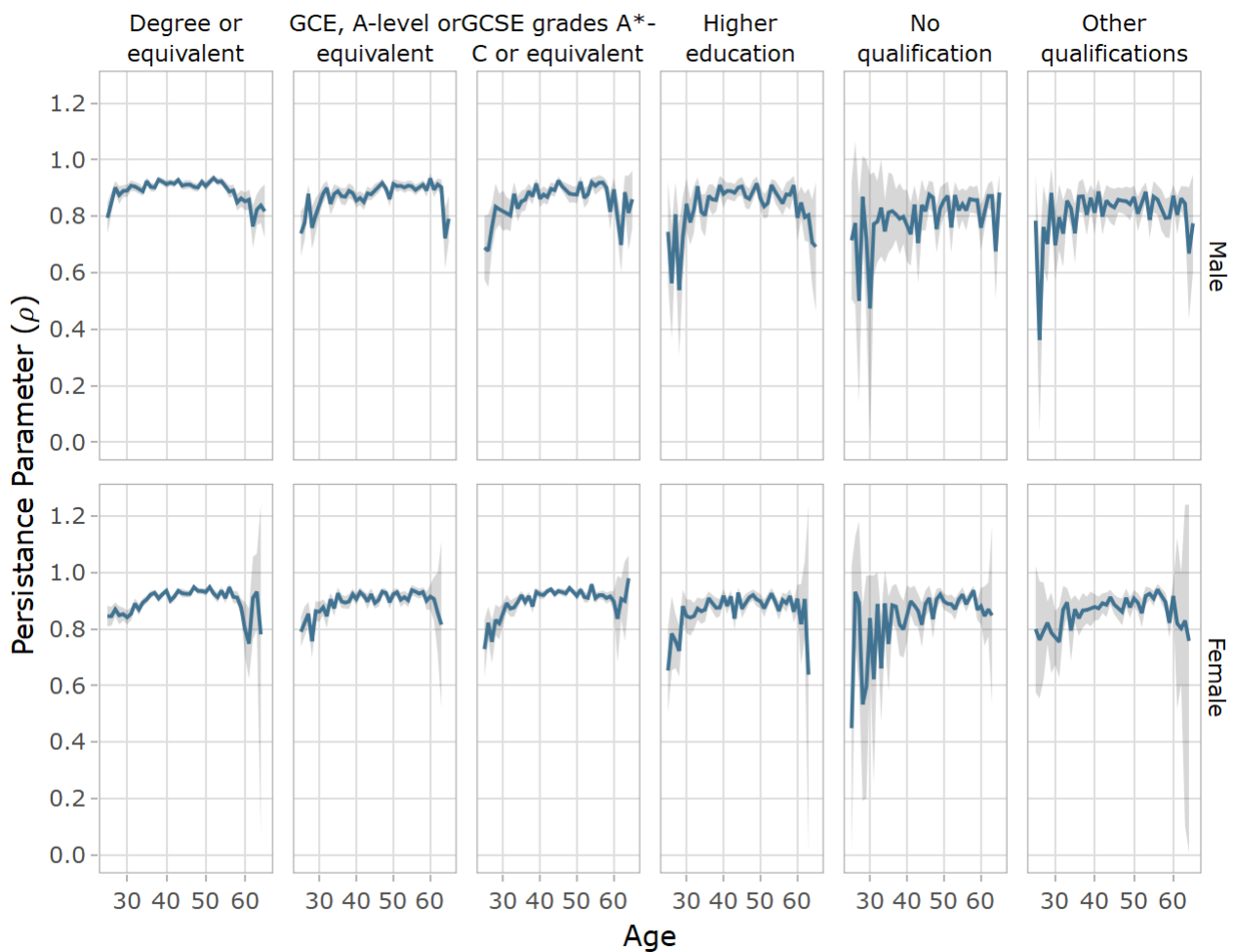
Applying this to lifetime earnings, the full joint distribution of lifetime earnings can be defined by the marginal distributions of earnings at each age and a series of bivariate

copula functions. The copula C , therefore, captures the rank dependence between adjacent ages, in effect “coupling” together the marginal distributions. This is a helpful strategy as it means we do not need to place any restrictions on the distribution of earnings taken from the LFS, we need only parameterise the copula function. This approach allows us to estimate earnings dynamics over a whole lifetime, despite the lack of longitudinal data to directly capture individuals changes in earnings over an extended period.

B t-copula persistence parameters

Figure 10 illustrates the raw, unsmoothed estimates of the persistence parameter (ρ) at each age for each of our 12 characteristic groups. The shaded area represents the 95% confidence interval surrounding each estimate. Estimates tend to be less precise at the start and end of individuals working lives. Furthermore, point estimates for those with lower levels of education have both more variability in age and are less precisely estimated, attracting larger uncertainty.

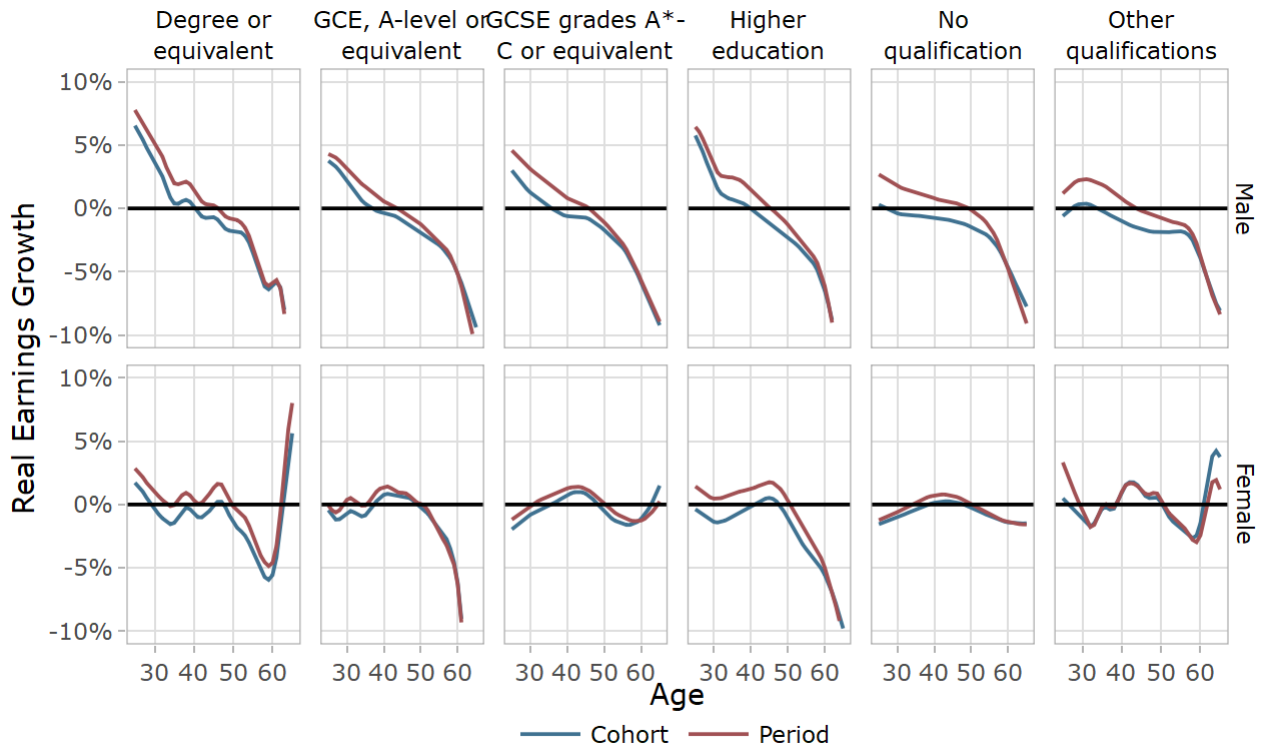
Figure 10: Unsmoothed Copula Persistence Parameters



C Age-Period-Cohort problem

As discussed in the main text, we implement the simple period view to manoeuvre around the age-period-cohort problem. Figure 11 illustrates the smoothed earnings growth by group produced by this model and compares them with a model that instead takes a simple cohort view. The period view generally leads to larger real earnings growth; however, this is not true for all groups or all ages.

Figure 11: Smoothed Real Earnings Growth



D State pension age

Table 5: Legislated Changes in State Pension Age

Year	Female	Male
2010-2018	60-65	–
2018-2020	65-66	65-66
2026-2028	66-67	66-67
2044-2046	67-68	67-68

Source:

Pensions Act 2011, c. 19.

Pensions Act 2014, c. 19.

Pensions Act 2007, c. 22.

E Accuracy of earnings simulations

To assess the performance of our simulation model, we start the simulation at age 26 rather than 29. This allows us to compare actual realised outcomes with those from the simulation model.

Figure 12: Actual vs. Simulated Distributions

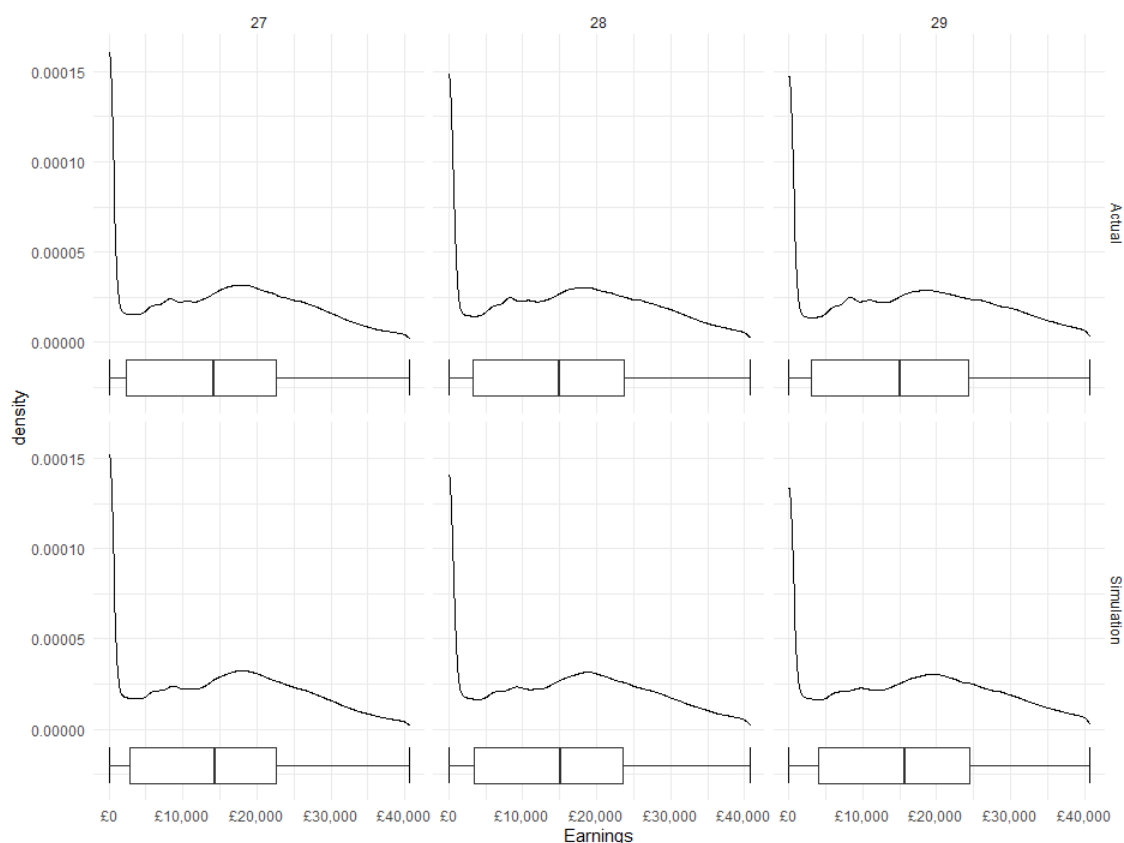
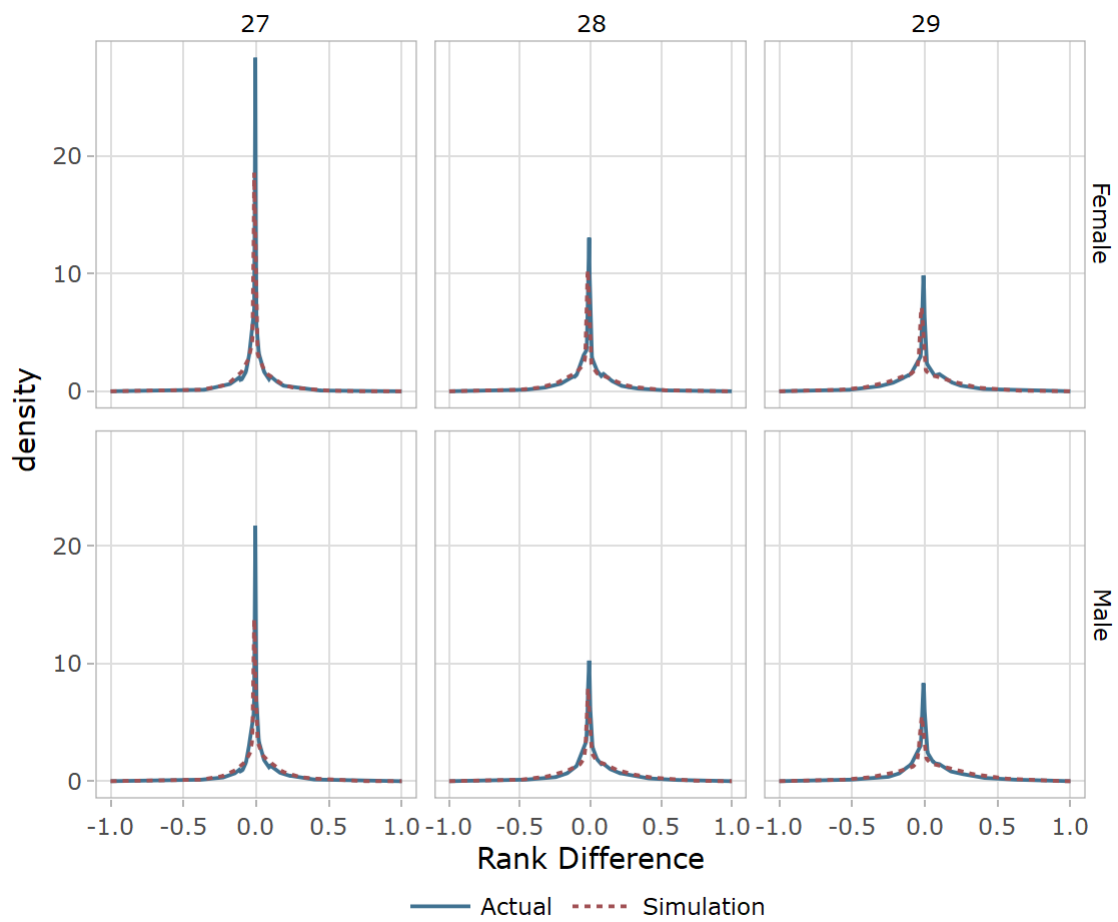


Figure 12 compares the actual distributions of earnings at ages 27, 28 and 29 observed in LEO, with our simulated earnings. The shape of the simulated earnings distributions at each age shows a good likeness to the actual data and key distributional statistics⁴², shown in the box plots, are found to be similar. As are the mean earnings at each age (£15,548, £16,582, and £17,604 in the simulated data and £15,613, £16,864, and £17,782 in the actual data).

As a further check of our simulation model, we compute the difference in rank between earnings at age 26 and the ages of 27, 28 and 29. This helps verify that we are accurately capturing individual age-earning pathways, not simply the overall distribution at each age. Figure 13 compares these differences, for males and females, between the simulated and the observed data. As expected, there is a larger difference in rank the more intervening years that have passed, leading to a more dispersion in the distribution of differences. This is true in both the observed and simulated data, although the simulation model generates

⁴²25th, 50th and 75th percentiles.

Figure 13: Comparing Rank Movement



slightly more dispersion than the actual data, at all time points. Finally, there does not seem to be any difference in the performance of the simulation between men and women.

F Multiple imputation

The method used to complete lifetime earnings and employment sequences, which are only observed up to age 29, produces a stochastic prediction of remaining lifetime earnings and employment spells for each individual. Each prediction is deflated to current prices, discounted to age 16, and then summed to give a single aggregate PV of lifetime earnings for each individual in the main dataset. This PV is the dependent variable in the estimation of the education production function. The random nature of the process adds additional uncertainty to estimates; to take account of this additional uncertainty we undertake multiple imputation as is outlined here.

First, we obtain K independent predictions of lifetime earnings for each student. Then, the education production model is estimated K times: once using each of the K imputed outcome variables. This produces K separate estimates of each parameter, with K standard error estimates. The multiple imputation method of Rubin (1987) constructs the pooled estimate as

$$\beta_{MI} = \frac{1}{K} \sum_{k=1}^K \beta_k \quad (17)$$

The pooled standard error combines two components, the ‘within’ variance:

$$V_w = \frac{1}{K} \sum_{k=1}^K se(\beta_k)^2 \quad (18)$$

and the ‘between’ variance

$$V_b = \frac{1}{K-1} \sum_{k=1}^K (\beta_k - \beta_{MI})^2 \quad (19)$$

using the formula

$$se_{MI}(\beta) = \sqrt{V_w + \left[1 + \frac{1}{K}\right] V_b} \quad (20)$$

In the analysis we use $K = 10$ imputations of the discounted lifetime productivity.

G Variables in GCSE attainment model

Table 6 contains all the variables used in the GCSE attainment model. Variables are included in the model as either:

- Fixed effects: the variable is included in an implicit fixed effect, that is not estimated but is differenced out of the regression equation before estimation.
- Interactions: the model coefficients are estimated separately for each subgroup of the cross-classification of the interaction variables.
- Controls: variable or dummies of variable are included in the right-hand-side of the regression equation. For categorical controls, the reference level is indicated by '(ref)'.

Table 6: Variables in GCSE attainment model

Variable	Inclusion in model	Notes
School	Fixed effect	
Gender	Fixed effect, interaction	Female, Male
FSM status	Fixed effect, interaction	Was the student FSM eligible at any time in last six years? No, Yes
Tercile of aggregate KS2 score	Fixed effect, interaction	Top, middle and bottom 3rd on combined English and Maths
Number of GCSEs taken	Fixed effect	NB the variable included in the FE is the ungrouped number of GCSEs. The grouped number is nested within it.
Grouped number of GCSEs taken	Interaction	1-4, 5-6, 7-8, 9, 10+
IDAC1 rank	Control	Standardised
KS2 maths score	Control	Standardised
KS2 english score	Control	Standardised
Major ethnic group	Control	White (ref), Asian, Black, Chinese, Mixed, Any other Ethnic group, Unknown
English as Additional Language	Control	No (ref), Yes
Special Educational needs	Control	No (ref), Yes – no statement, Yes – with statement
Month of birth (grouped)	Control	Sep-Dec (ref), Jan-Apr, May-Aug
Academic year that KS4 was completed	Control	2001/02, 2002/03, 2003/04, 2004/05 (ref) This is not to adjust for prices, which are adjusted separately. This is to control for cohort effects in labour market.
Children in Need	Not included	Not available in data for early 2000s
Region	Not included	Cannot control for variable that is collinear with school ID
Post-GCSE qualifications	Not included	Post-treatment effects problematic, statistically speaking.

H Subject mapping

There have been several overhauls of the GCSE curriculum between the early 2000's, when we observe pupils taking GCSE exams in our LEO sample, and the present day. This has involved the renaming of some subjects, for example, 'Information Technology' has now become 'Computing'. Table 7 describes all these changes.

Table 7: Subject Mapping

Subject in 2002-2005	Subject in 2019
Maths	Maths
English	English
English Literature	English Literature
History	History
Geography	Geography
Religious Studies	Religious Studies
Double Sciences	Double Sciences
Triple Sciences	Triple Sciences
Physical Education	Physical Education/Sports Science
Business	Business
Information Technology	Computing
French	French
Spanish	Spanish
German	German
Art	Art and Design
Music	Music
Drama	Drama
DT Food	Food Science
DT Graphic Design	Design Technology
DT Textiles	Design Technology
DT Resistant Materials	Design Technology
DT Electrical Eng.	Design Technology
Other	Economics, Social Sciences, Citizenship, other MFL, Classics

I Additional non-linear results

Our estimated non-linear marginal earnings returns in Maths, English, English Literature, Double Science, History and Geography can be found in the main text. The results for the remaining subjects are shown here.

Figure 14: Non-linear returns: Business, IT and PE

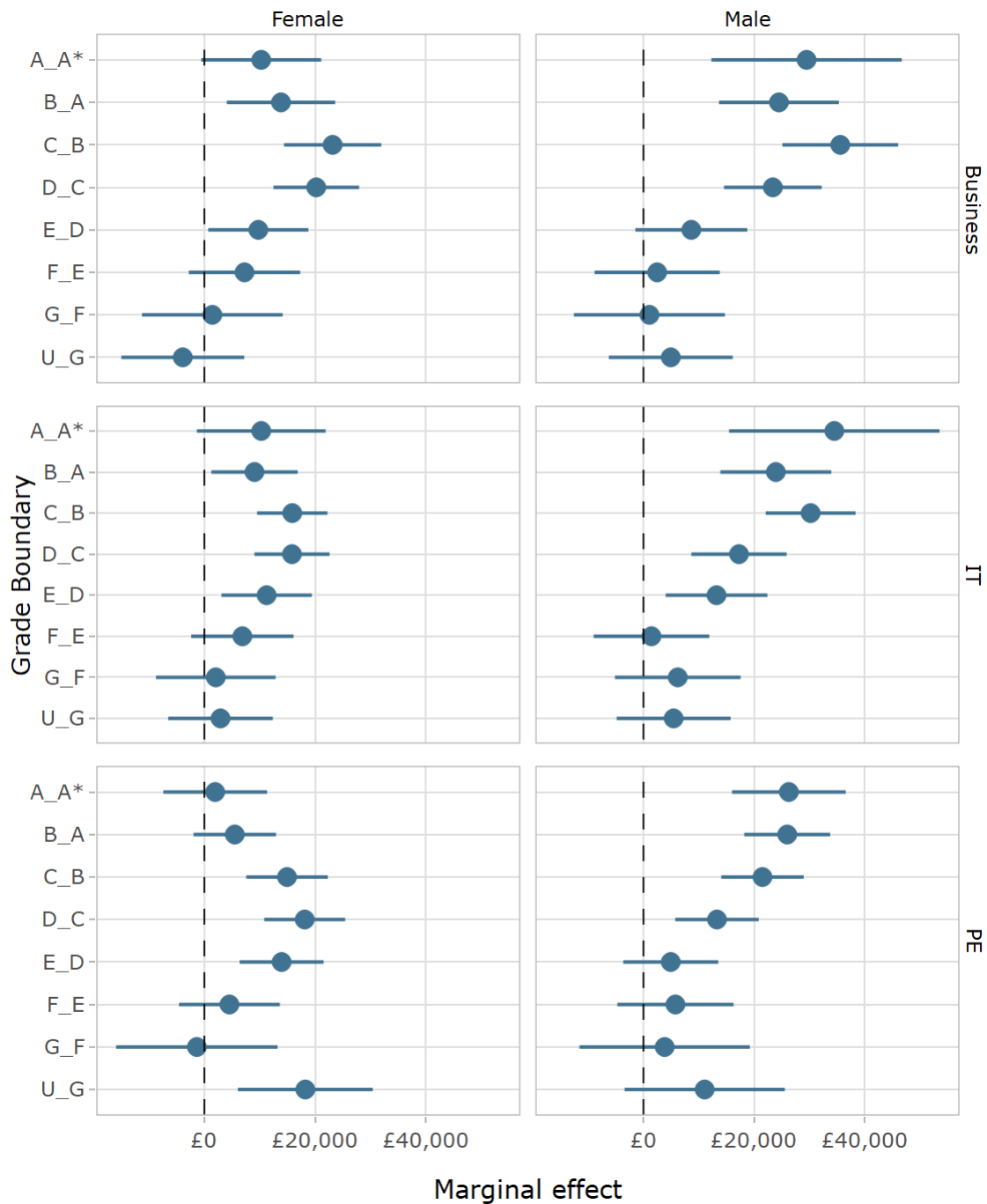


Figure 15: Non-linear returns: French, German and Spanish

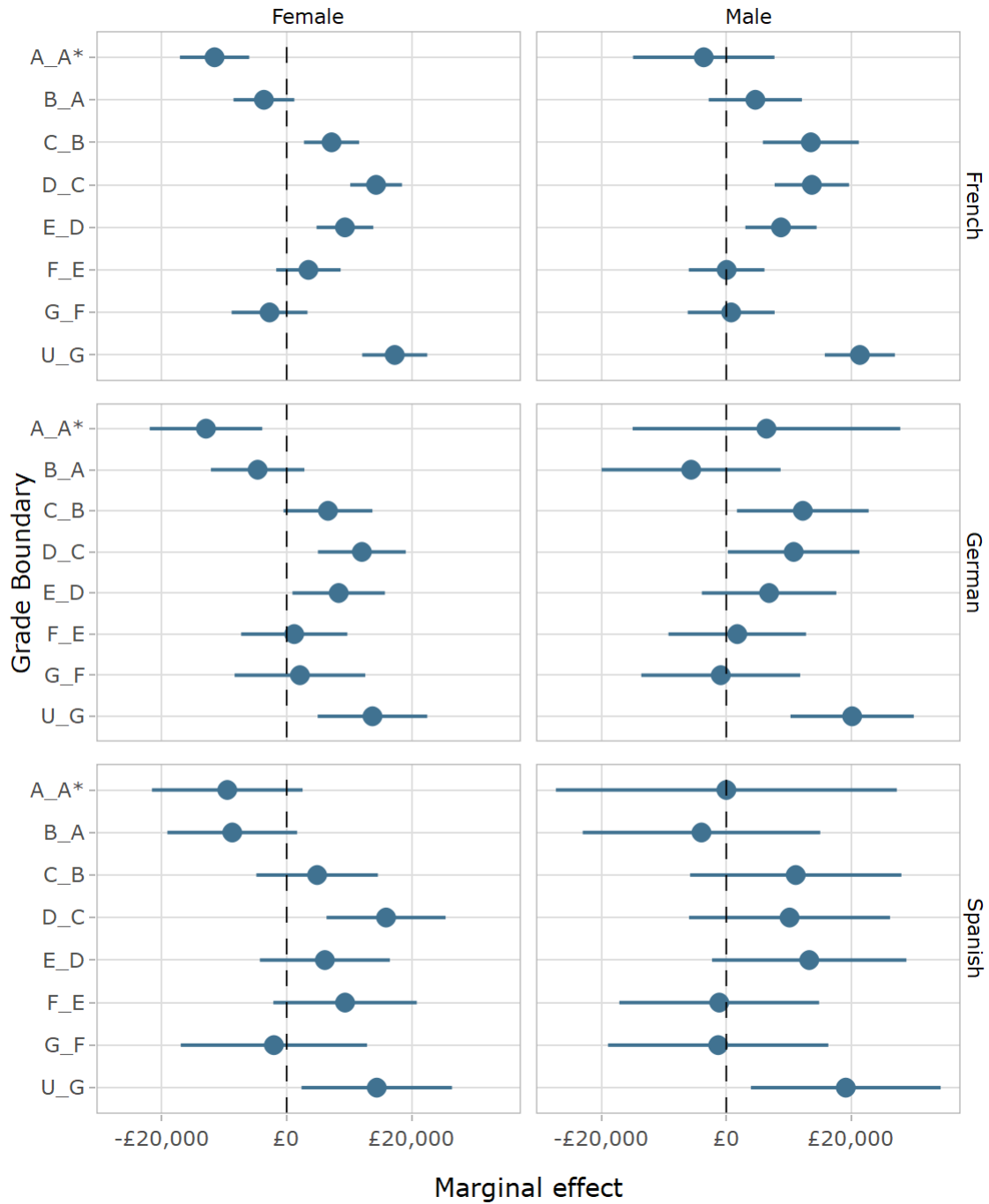


Figure 16: Non-linear returns: Drama, Art and Music

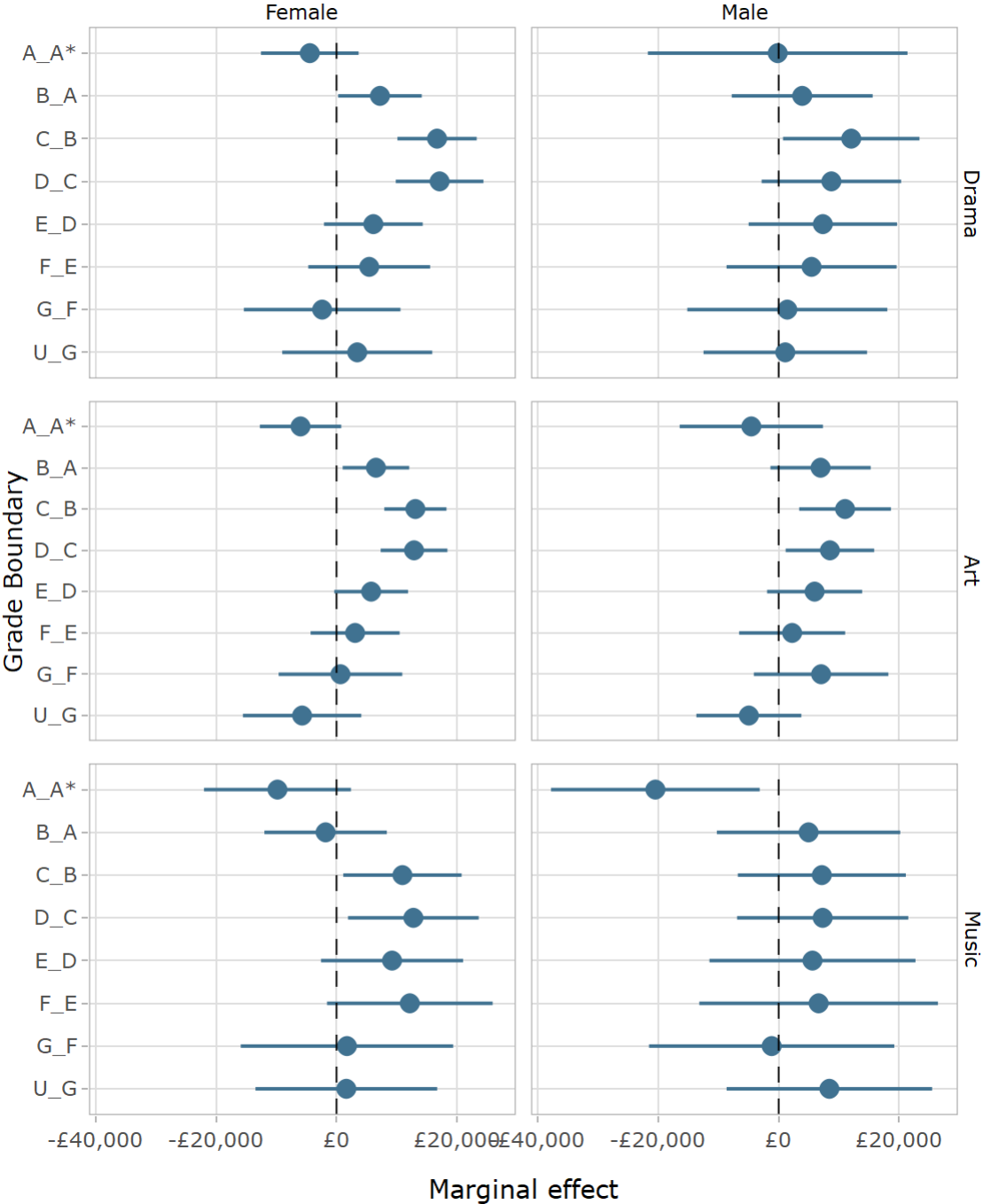
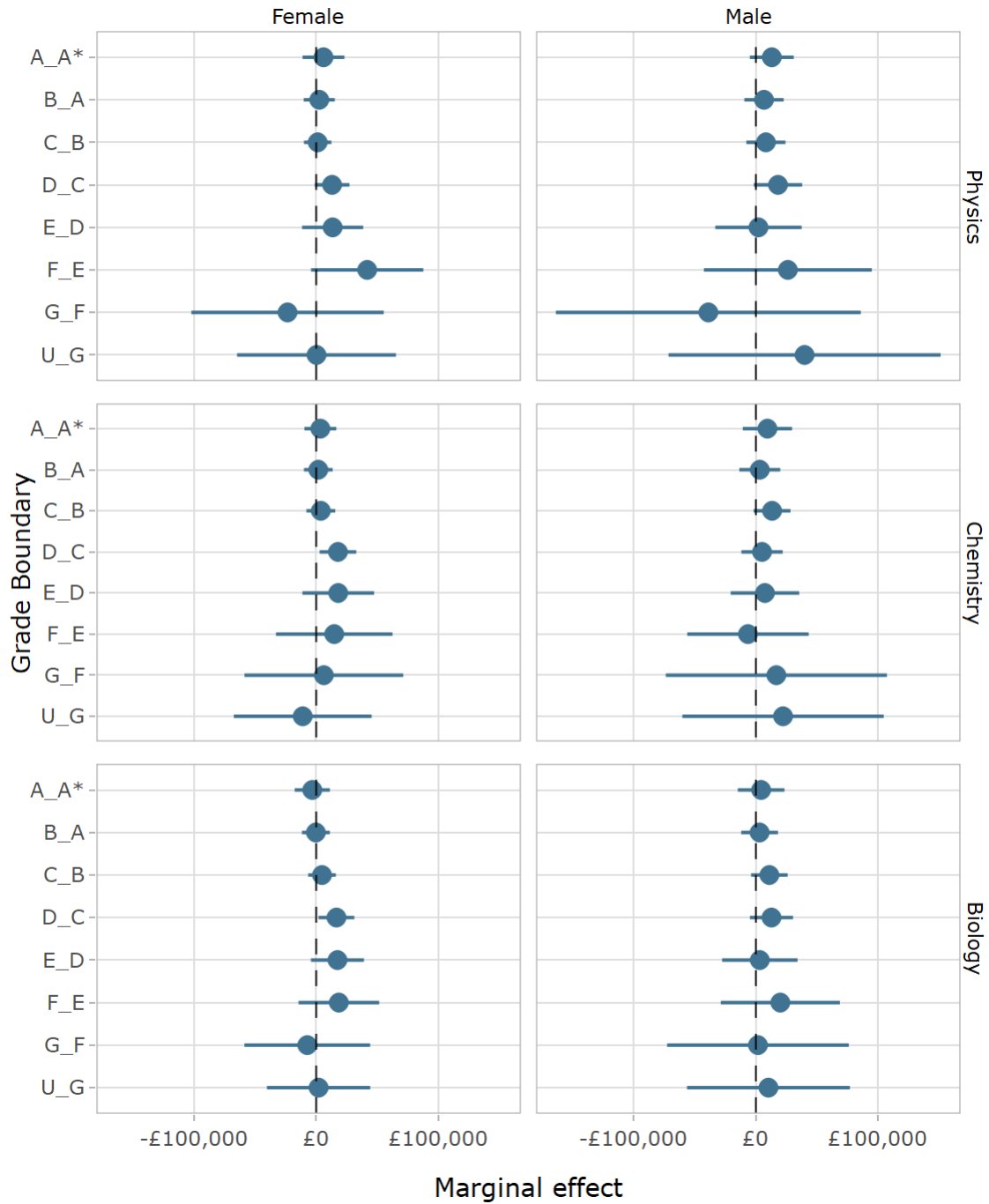


Figure 17: Non-linear returns: Physics, Chemistry and Biology



J Additional analysis of non-linear grade effects

The evidence of negative marginal effects at the top grades in some subjects raises the question: are the negative effects driven by people who do well in all subjects, or by people who only do well in the subject in question?

To shed light on this, the regression lines in figures 18, 19, 20 and 21 plot the present value of lifetime earnings against total grades, for students with different grades in Music, English Literature and, for comparison, Maths. For male students studying Music, there is evidence that the return to a top grade in Music is diminishing along the distribution of total grades: getting an A* is not worth as much additionally for a high-scoring student as it is for a low-scoring student. The lines cross at the top end, suggesting that for some students with higher grades overall, a top grade in Music is associated with a negative earnings return. Again, note that these effects are marginal and we cannot overlay causal interpretations.

Likewise, the same pattern is evident in English Literature scores, this time for female students. The pattern of diminishing returns at the top of the grade distribution suggests that it is the top-scoring female students (on overall grades) who have the least to gain from a top grade in English Literature and may indeed lose out from gaining the top grade. For comparison, the pattern in Maths shows strongly increasing returns in Maths grade. The lines do cross, slightly, at the high end, but this is more likely to be an artefact of the linear regression extrapolation. However, there is some evidence of a diminishing return to top grades for high-scoring students (especially female students).

Figure 18: Scatter of grades against PV for French

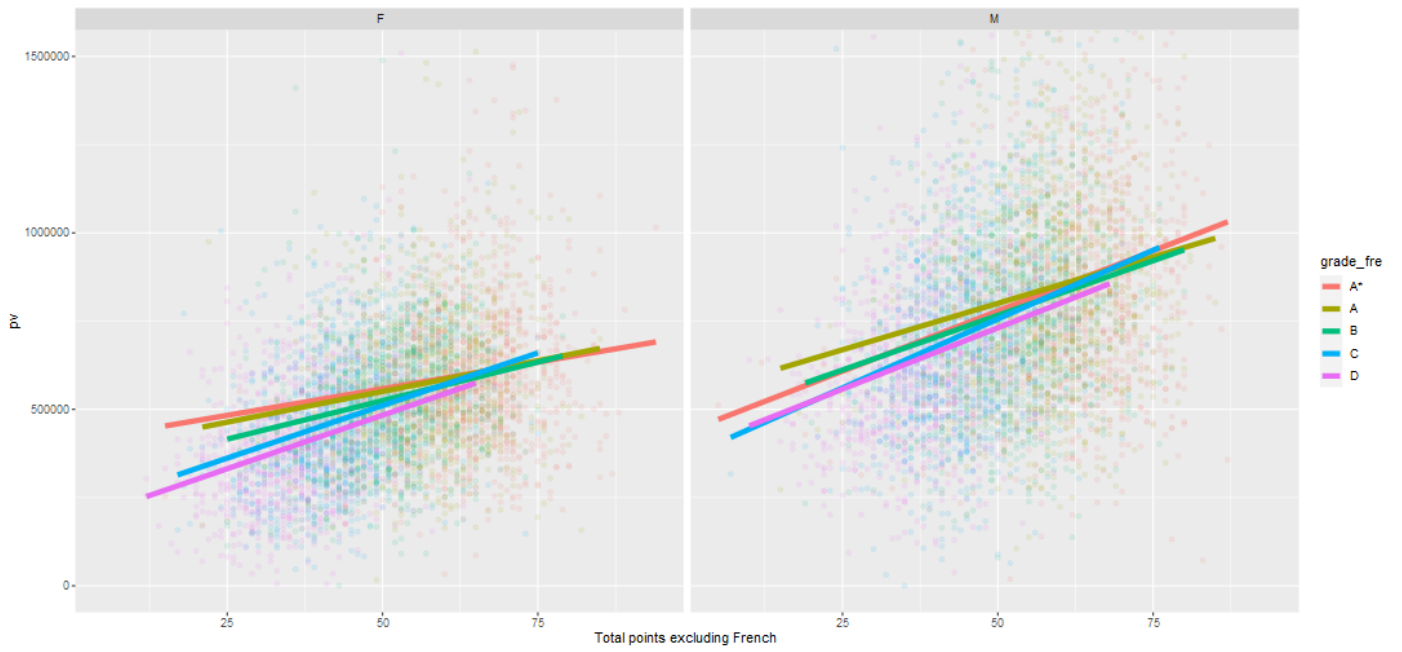


Figure 19: Scatter of grades against PV for Music

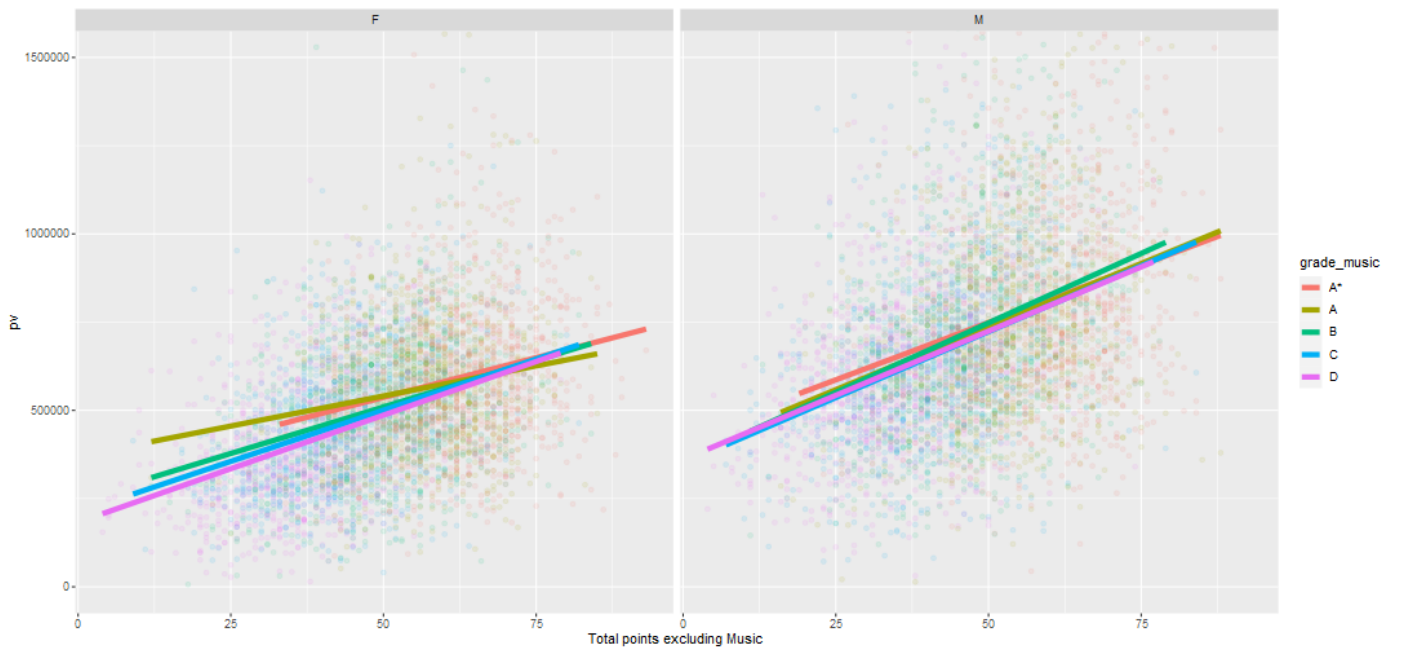


Figure 20: Scatter of grades against PV for English Literature

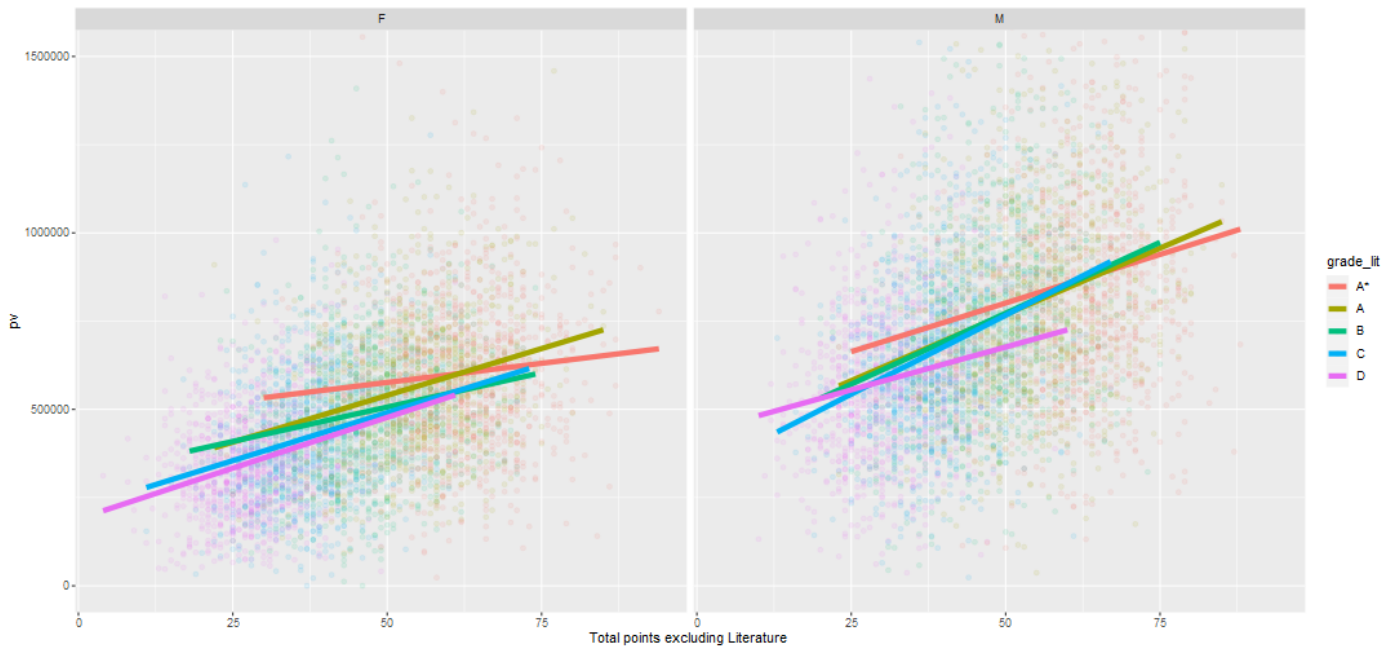
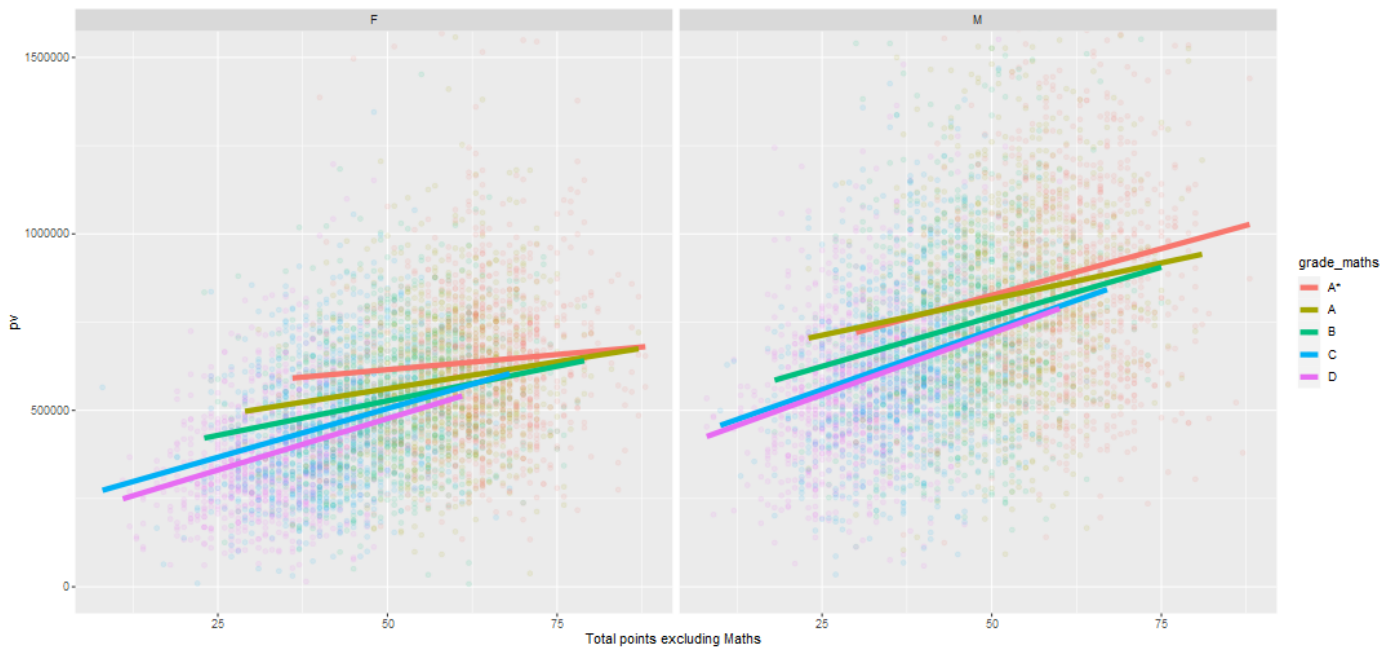


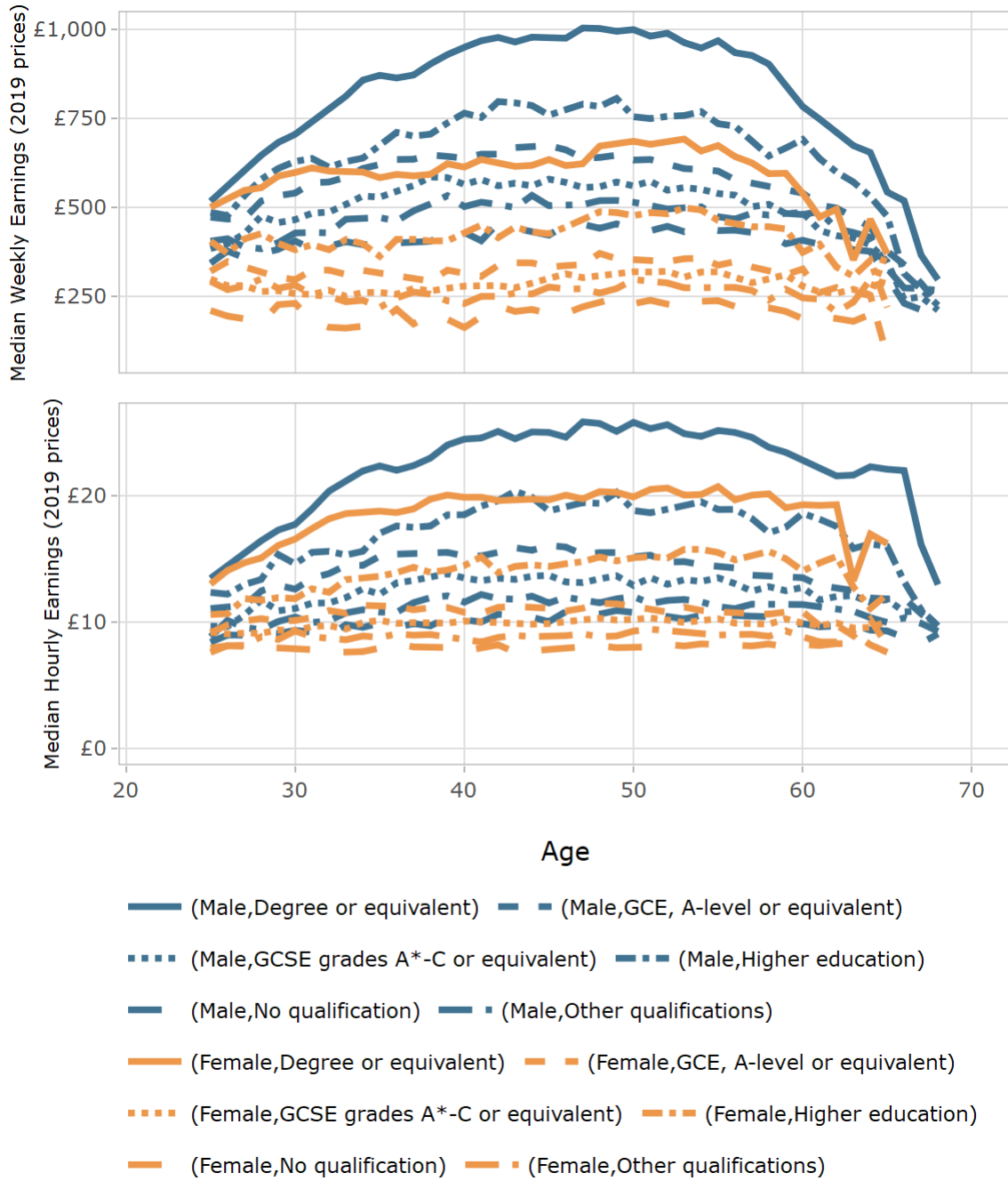
Figure 21: Scatter of grades against PV for Maths



K Weekly vs. Hourly earnings

Our approach uses gross weekly earnings as reported in the LFS. We prefer this to using hourly pay as this weekly measure accounts for the number of hours worked. This is important as we seek to estimate the association between total earnings and GCSE attainment, however this does mean we say nothing about the decisions and constraints that dictate the number of hours worked. Figure 22 compares the median average weekly and hourly earnings in the LFS for our gender, highest qualification groups. This shows that women have relatively higher earnings when hourly measures are used, implying that on average they work fewer hours a week.

Figure 22: Hourly vs. Weekly LFS Earnings



References

- Almlund, M., Duckworth, A. L., Heckman, J. and Kautz, T. (2011) 'Personality Psychology and Economics', in Hanushek, E., Machin, S., and Woessmann, L. (eds) *Handbook of the Economics of Education*. Elsevier, pp. 1–181.
- Anders, J., Henderson, M., Moulton, V. and Sullivan, A. (2018) 'The role of schools in explaining individuals' subject choices at age 14', *Oxford Review of Education*, 44(1), pp. 75–93.
- Angrist, J. D. and Pischke, J.-S. (2009) *Mostly Harmless Econometrics*. Princeton University Press.
- Belfield, C., Britton, J., Buscha, F., Dearden, L., Dickson, M., Erve, L. van der, Sibieta, L., Vignoles, A., Walker, I. and Zhu, Y. (2018a) *The impact of undergraduate degrees on early-career earnings*. Research Report. Department for Education.
- Belfield, C., Britton, J., Buscha, F., Dearden, L., Dickson, M., Erve, L. van der, Sibieta, L., Vignoles, A., Walker, I. and Zhu, Y. (2018b) *The relative labour market returns to different degrees*. Research Report. Department for Education.
- Britton, J., Dearden, L., Erve, L. van der and Waltmann, B. (2020) *The impact of undergraduate degrees on lifetime earnings*. Research Report. Department for Education.
- Britton, J., Dearden, L. and Waltmann, B. (2021) *The returns to undergraduate degrees by socio-economic group and ethnicity*. Research Report. Department for Education.
- Britton, J., Erve, L. van der and Shephard, N. (2019) *Econometrics of valuing income contingent student loans using administrative data: Groups of English students*. IFS Working Papers W19/04. Institute for Fiscal Studies.
- Britton, J., Shephard, N. and Vignoles, A. (2019) 'A comparison of sample survey measures of earnings of English graduates with administrative data', *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(3), pp. 719–754.
- Chamon, M. and Prasad, E. (2010) 'Why Are Saving Rates of Urban Households in China Rising?', *American Economic Journal: Macroeconomics*, 2(1), pp. 93–130.
- Chowdry, H., Crawford, C., Dearden, L., Goodman, A. and Vignoles, A. (2013) 'Widening participation in higher education: Analysis using linked administrative data', *Journal of*

the Royal Statistical Society: Series A (Statistics in Society), 176(2), pp. 431–457.

Crawford, C. and Cattan, S. (2013) *Assessing the economic benefits of education: Reconciling microeconomic and macroeconomic approaches*. CAYT Report No.4. Centre for Analysis of Youth Transitions.

Dearden, L., Fitzsimons, E., Goodman, A. and Kaplan, G. (2006) *Estimating Lifetime Earnings Distributions Using Copulas*. Unpublished Working Paper.

Dearden, L., Fitzsimons, E., Goodman, A. and Kaplan, G. (2008) 'Higher Education Funding Reforms in England: The Distributional Effects and the Shifting Balance of Costs', *The Economics Journal*, 118(526), pp. F100–F125.

Deaton, A. and Paxson, C. (1994) 'Saving, Growth, and Aging in Taiwan', in *Studies in the Economics of Aging*. University of Chicago Press, pp. 331–362.

DfE (2013) 'Science and Technology: Supplementary written evidence submitted by the Department for Education'. House of Commons.

DfE (2018a) 'Free school meals Guidance for local authorities, maintained schools, academies and free schools', *GOV.UK*.

DfE (2018b) 'GCSE and equivalent results: 2017 to 2018 (provisional)', *GOV.UK*.

DfE (2021) 'Schools, pupils and their characteristics, Academic Year 2019/20', *GOV.UK*.

Efron, B. and Morris, C. (1972) 'Empirical Bayes on Vector Observations: An Extension of Stein's Method.', *Biometrika*, 59(2), pp. 335–347.

Espinoza, H. and Speckesser, S. (2019) *A comparison of earnings related to higher level vocational/technical and academic education*. CVER Research Discussion Paper 019. Centre for Vocational Education Research.

Gambin, L., Beaven, R., Hogarth, T., May-Gillings, M. and Long, K. (2014) *Methodological Issues in Estimating the Value Added of Further Education, Higher Education and Skills: A review of relevant literature*. BIS Research Paper No. 166. Department for Business Innovation & Skills.

Greenwood, C., Jenkins, A. and Vignoles, A. (2007) *The Returns to Qualifications in England: Updating the Evidence Base on Level 2 and Level 3 Vocational Qualifications*. CEE Discussion Papers 0089. Centre for the Economics of Education, LSE.

- Hayward, H., Hunt, E. and Lord, A. (2014) *The economic value of key intermediate qualifications: Estimating the returns and lifetime productivity gains to GCSEs, A levels and apprenticeships*. Research Report. Department for Education.
- HM Treasury (2020) *The Green Book: Central government guidance on appraisal and evaluation*.
- Hodge, L., Little, A. and Weldon, M. (2021) *Schools Policy Appraisal Handbook*. Department for Education.
- Jin, W., Muriel, A. and Sibieta, L. (2011) *Subject and course choices at ages 14 and 16 amongst young people in England: Insights from behavioural economics*. Research Report. Department for Education.
- Layard, R., Clark, A., Flèche, S., Ward, G. and Powdthavee, N. (2018) *The origins of happiness: The science of wellbeing over the life course*. Princeton University Press.
- Machin, S., McNally, S. and Ruiz-Valenzuela, J. (2020) 'Entry through the narrow door: The costs of just failing high stakes exams', *Journal of Public Economics*, 190, pp. 104–224.
- Mcintosh, S. (2006) 'Further Analysis of the Returns to Academic and Vocational Qualifications', *Oxford Bulletin of Economics and Statistics*, 68(2), pp. 225–251.
- Mincer, J. (1958) 'Investment in Human Capital and Personal Income Distribution', *Journal of Political Economy*, 66(4), pp. 281–302.
- Ofqual (2019) 'Guide to GCSE results for England, 2019', *GOV.UK*.
- ONS (2019a) 'Employee earnings in the UK: 2019', *GOV.UK*.
- ONS (2019b) 'Human capital estimates in the UK: 2004 to 2018', *GOV.UK*.
- ONS (2019c) *LFS User Guide, Volume 11 – Longitudinal Datasets*.
- Philippis, M. de (2017) *STEM Graduates and Secondary School Curriculum: Does Early Exposure to Science Matter?* Temi di Discussione 1107. Bank of Italy.
- Robbins, H. (1956) 'An Empirical Bayes Approach to Statistics', in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. Berkeley: University of California Press, pp. 157–163.

Rubin, D. (1987) *Multiple imputation for nonresponse in surveys*. Wiley.

Sammons, P., Sylva, K., Melhuish, E. C., Siraj, I., Taggart, B., Smees, R. and Toth, K. (2014) *Influences on students' social-behavioural development at age 16: Effective pre-school, primary & secondary education project (EPPSE)*. Research Brief. Department of Education.

Sklar, A. (1959) 'Fonctions de répartition à n dimensions et leurs marges', *Publications de l'Institut Statistique de l'Université de Paris*, 8, pp. 229–231.

Walker, I. and Zhu, Y. (2013) *The impact of university degrees on the lifecycle of earnings: Some further analysis*. BIS Research Paper No. 112. Department for Business Innovation & Skills.

Wyness, G., Macmillan, L. and Anders, J. (2021) *Does education raise people's productivity or does it just signal their existing ability?* CEPEO Briefing Note No 12. Centre for Education Policy; Equalising Opportunities, UCL Institute of Education.



Department
for Education

© Department for Education 2021

Reference: DFE-RR1131

ISBN: 978-1-83870-264-9

For any enquiries regarding this publication, contact us at:

Allan.LITTLE@education.gov.uk or www.education.gov.uk/contactus

This document is available for download at www.gov.uk/government/publications