# The impact of undergraduate degrees on early-career earnings

## Research report

## November 2018

**Chris Belfield, Jack Britton, Franz Buscha, Lorraine Dearden, Matt Dickson, Laura van der Erve, Luke Sibieta, Anna Vignoles, Ian Walker and Yu Zhu**

**Institute for Fiscal Studies**

# Contents

# List of Figures

# List of Tables

# Executive Summary

This report uses the new Longitudinal Educational Outcomes (LEO) administrative dataset to provide the latest estimates of the impact of Higher Education (HE) on individuals' early-career earnings after accounting for individuals' pre-university characteristics. This will provide vital evidence for prospective students choosing whether, where and what to study at university.

We estimate the overall average impact of attending HE on earnings at age 29, and show how this varies for individuals studying different degree subjects or at different Higher Education Institutions (HEIs), based on those who went to HE in the mid- to late-2000s. We also investigate how these returns might differ for students with different prior attainment, based on their GCSE results and whether or not they studied a maths or science ('STEM') A-level. We focus on those who start HE, rather than just on HE graduates, as this is the relevant decision facing prospective students.

This is the second in a series of reports by the Institute for Fiscal Studies, commissioned by the Department for Education (DfE), that makes use of the LEO dataset to improve information on the value of HE degrees. The dataset, developed in collaboration with the DfE, tracks English students through school, college, university and into the labour market. This report, for the first time, uses the dataset to compare individuals who went to HE to those with similar background characteristics who did not.

All of our estimates (unless otherwise stated) report the effect of attending HE at age 18 on annual gross earnings at age 29, conditional on being in "sustained employment". We compare our HE students with those who did not go to HE but had at least five A*-C GCSEs, controlling for differences in prior attainment, Key Stage 5 subject choices and family background.

This pioneering dataset enables us to account for many of the differences between those who do and do not attend HE, but there are many other factors which may affect this decision that are not accounted for, such as passion or preferences. Also these estimates focus on the monetary returns to university, which may not fully reflect the wider society benefits of these degrees. As such, some caution should be executed when interpreting these findings.

Our main findings are as follows:

- Those who attend HE earn a lot more on average than those who do not. At age 29 the average man who attended HE earns around 25% more than the average man (with five A*-C GCSEs) who did not. For women the gap is more than 50%.

- A large portion of this difference can be explained by differences in pre-university characteristics: a typical HE student has higher prior attainment and is more likely to have come from a richer family than someone who does not attend. They would therefore be expected to earn more, even had they not gone to university.

- Once we account for differences in pre-university characteristics, we estimate the average impact of attending HE on earnings at age 29 to be 26% for women and 6% for men. If we focus on the impact of graduating, these returns rise to 28% and 8% respectively.

- The higher returns for women may be driven by the fact that women who attend HE typically work longer hours than those who do not. This impact on working hours may well be causal, but it may also be larger at age 29 than at later ages if graduates delay having children.

- Importantly, age 29 is relatively early in an individuals' career. There is strong evidence that the earnings of men who attend HE continue to grow faster than their non-HE counterparts after age 30. This is likely to result in the returns to HE for men being larger at later ages than we are able to estimate here. For women, the divergent trends in earnings by education type after age 30 are less clear.

- Not all degrees are the same, and subject choice appears to be a very important determinant of returns. For men, studying creative arts, English or philosophy actually result in lower earnings on average at age 29 than people with similar background characteristics who did not go to HE at all. By contrast, studying medicine or economics appears to increase earnings by more than 20%. For women, there are no subjects that have negative average returns, and studying economics or medicine increases their age 29 earnings by around 60%.

- Institution choice also appears to be highly important, as there is considerable variation in returns. For men, there are 12 institutions (accounting for 4% of male students) for which we estimate statistically significantly negative returns at age 29 on average, while there are 18 universities with average returns of more than 20%. For women, despite high returns on average, there are still two institutions (0.4% of female students) which have statistically significantly negative returns at age 29, while there are 66 institutions with returns of more than 20%.

- We estimate that 67% of men and 99% of women (and hence 85% of students) attended universities that have significantly positive returns on average by age 29.

- For both men and women, there is wide variation in returns within every subject and every institution. For example, studying at Cambridge yields positive returns of around 30% on average for both men and women, but some subject choices - for example creative arts - actually appear to result in lower earnings at age 29 than not going to university at all.

- The returns to HE also differ considerably for different types of students (see table). Attending HE only increases the age 29 earnings of lower prior attainment men (based on GCSE grades) without a STEM A-level by 4%. This compares to 20% for their peers who also do not have a STEM A-level but have high GCSE grades. The return is low because students with lower prior attainment are more likely to take low-returning subjects like creative arts, communications and sport science, and are more likely to attend lower-returning universities. However, this is not the only explanation: even when they study the same subject or at the same type of university as their peers who have higher prior attainment, they experience lower returns.

6

Table E1: Returns overall and by prior attainment subgroup

|  | | No STEM A-level | | | STEM A-level | | |
| | **All** | Lower | Middle | Higher | Lower | Middle | Higher |
|---|---|---|---|---|---|---|---|
| Men | **6%** | 4% | 8% | 20% | 11% | 9% | 5% |
| Women | **26%** | 23% | 25% | 31% | 22% | 16% | 23% |

Note: These are percentage estimates based on Tables 9 and 10 of report

- This is a particularly important when considering the impacts of expansion in the HE system: in our period of study, 70% of all students with five A*-C GCSEs that did not attend university fell within this lower prior attainment, without STEM A-level group.

- Men with higher prior attainment and a STEM A-level have an estimated return of 5%, which might be lower than expected. This is hugely varied: studying law, medicine or economics increases their earnings by around 20%, and the return to attending a Russell Group for this group is around 10%. On the other hand, studying arts English, communications, psychology, languages and history, or attending Post-1992 or Other universities actually appears to result in lower earnings for this group than they would have achieved had they not gone to university (of course, these individuals may be making these choices for reasons other than to try to maximise their earnings). These particular estimates should be treated with caution, as overall only 5% of individuals in this group do not go to HE, and they are likely to be quite unusual - indeed, they have very high average earnings of around £40,000 per year by age 29.

- Among women, the overall returns to HE are high for all groups, though some similar patterns emerge. Higher prior attainment women without a STEM A-level have higher returns than their lower attainment peers. Unlike for men, there is little evidence of lower prior attainment women without a STEM A-level experiencing lower returns when studying the same subject as their higher attaining peers. Instead, the lower returns for this group appear to be driven by a higher propensity to study lower (although still significantly positive) returning subjects such as social care, sociology or education, and because they are more likely to attend lower-returning universities.

# 1 Introduction

The decision of whether or not to go to Higher Education (HE) is a very important one. Students decide whether or not to take three to five years out of the labour market at a crucial time early in their careers, forgoing possible earnings during this time. In many countries around the world, they also face tuition fee costs, such as in England where tuition fees are up to £9,250 per year.[1] At the same time as deciding whether to go to university at all, young people must also decide which subject they would like to study and which institution to attend. These choices will likely have a substantial bearing on their employment and earning opportunities throughout their adult life.

In this report, we estimate the impact on earnings of attending HE compared with not going. We detail how this varies by subject and institution of study, as well as how these returns vary by gender, prior educational attainment and the sorts of subjects individuals have studied up to age 18. To do so, we make use of the new Longitudinal Education Outcomes (LEO) dataset, which links together tax, benefit, higher education and school records to provide a rich description of individuals' trajectories through the education system and into the labour market. In so doing, we build on our previous report for the Department for Education (Belfield et al., 2018), which sought to estimate how the returns to different subjects and institutions vary compared with the returns to the average graduate. In other words, having detailed the relative returns to different subjects and institutions (how returns to different courses compare with those of the average graduate), we now estimate the absolute returns (what the returns to different courses are compared with not going to university).

There is much existing empirical evidence estimating the returns to attending HE across different countries and how these have changed over time (Card, 1999; Harmon et al., 2003; Psacharopoulos and Patrinos, 2004; Heckman et al., 2006). The evidence strongly suggests high positive causal returns, i.e. education directly increases individuals' stock of human capital, productivity and earning potential. There is also a broad consensus that over the last half century these returns have risen even as the number of graduates has increased, though more recently returns may be stabilising or starting to fall (Blundell et al., 2005; Walker and Zhu, 2011, 2013; Blundell et al., 2016). A key potential reason for increased returns to education is 'skill-biased technical change' - the labour market has become increasingly complex because of technological advances and this has led to a rise in the demand for workers with the complementary skills to manage and exploit such changes, at the expense of workers whose skills substitute for such changes (Goldin and Katz, 2009; Goos et al., 2009; Michaels et al., 2014).

Looking specifically at evidence from the UK, Blundell et al. (2005) find that attending HE generates average earnings returns of 27% at age 33 compared with not going for those entering HE in the late 1970s. Walker and Zhu (2011) find large returns of around 25-30% for women aged 26-28 who entered HE during the late 1980s and 1990s, and 15-25% for men. Blundell et al. (2016) further show that the gap in earnings between graduates and non-graduates has been largely stable through to the present day, despite the very large increase in the number attending HE in the UK.

---

[1]In England individuals can take out a government provided loan that covers the full amount of the fees and only has to be repaid after graduation, with repayments depending on the graduates' income.

While the UK evidence base confirms high causal returns to attending HE, there are limitations to this body of work. The UK evidence largely comes from survey data, which in many cases lacks detailed controls for individuals' prior education and attainment. Both of these are likely to affect individuals' earnings and to differ significantly between those who do and do not attend HE, and hence not controlling for them will bias the obtained results.

The second main limitation from using survey data is low sample sizes, which limits one's ability to look at the heterogeneity of returns by subject, institution and individual characteristics. Work from other countries, particularly Norway, has used rich data on all individuals linking their education and tax records combined with a robust research design to examine the rates of return to different university subjects and how these vary across individual characteristics (Kirkeboen et al., 2016; Bhuller et al., 2017), but data availability has so far limited such work in the UK.

This paper seeks to build on this work and close the evidence gap for the UK by using rich administrative data to estimate the returns to attending university. This linked administrative data has rich measures of prior attainment and socio economic background for the population of English students, and tracks them into the labour market up to their late 20s. This enables us to control for a rich set of observable characteristics when looking at labour market returns to a degree. The richness of the data and the very large number of individuals we observe enable us to investigate heterogeneity in returns by subject, institution and the interaction between the two, which we refer to as a course (e.g. mathematics at the University of Birmingham).

This work is likely to be of particular interest for policymakers and students seeking to make decisions around higher education. Our results are of significant value to students in informing them on how returns are likely to vary across their potential subjects and institutional choices. Not only do we estimate overall returns, we also investigate returns for specific subgroups based on GCSE attainment and A-level subject choices. Returns to those subgroups will better inform students about the average returns for people with similar educational backgrounds. We further show how these subgroup returns vary by subject and institution type - a significant improvement on the information currently available to students. This is also crucial information for policymakers thinking about the desirability of a further increase in HE participation; positive overall returns driven by groups with already extremely high participation rates, for example, would suggest that returns for those at the margin of attending might be very low. Showing how the returns vary by institution type and - in particular - subject could also inform policymakers on the types of courses that may be particularly beneficial for the students and hence which they may particularly want to encourage students to attend.

There are three main caveats that should be attached to our results, however. First, the results should not be interpreted as definitively causal. Whilst we are able to move beyond the existing literature by making use of rich data and sophisticated estimation techniques, unobservable differences that could affect earnings may remain between individuals taking different education options, such as different preferences over, for example, potential career paths or different levels of passion or enthusiasm for working and studying. Generally speaking, the academic literature that has looked at this issue finds the potential bias to be relatively small when thinking about overall returns (Card, 1999), but to be larger when looking at different subjects and institutions

(Kirkeboen et al., 2016; Andrews et al., 2017). Second, we are only able to look directly at earnings up to age 29, which is clearly very early in the careers of graduates. We provide evidence to show that the earnings differential between graduates and non-graduates is still growing at that point, which suggests our estimated returns are likely to understate the potential *lifetime* differences in earnings between graduates and non-graduates. Third, our estimates of the returns to HE are solely pecuniary. Whilst these are likely to be a major component of the return to HE, we are not estimating non-pecuniary returns, such as improved health, a more pleasant work environment, reduced crime or increased civic participation.

The rest of the report is outlined as follows. Section 2 describes the dataset we use and gives some key descriptives. Section 3 describes our methodology and Section 4 describes the results on the overall returns to HE. Section 5 shows the returns by subject, Section 6 by institutions and Section 7 by course. Section 8 explores how these returns might be likely to change as individuals age. Section 9 concludes.

## 2 Data

We use the Longitudinal Educational Outcomes (LEO) dataset, generated for the first time in collaboration with the Department for Education. This data links the National Pupil Database (NPD) and Higher Education Statistics Agency (HESA) data to Her Majesty's Revenue and Customs (HMRC) earnings and employment data and Work and Pensions Longitudinal Study (WPLS) benefits data. Belfield et al. (2018) provide more information on the LEO dataset. This paper expands on Belfield et al. (2018) by including those who do not attend HE.

### 2.1 Sample selection

In the estimation, we include individuals who took their GCSEs between 2002, the first year such data is included in the NPD, and 2007, as later cohorts will not have been in the labour market for long enough to be included in our analysis. For each of these GCSE cohorts, Table 1 outlines the number of individuals included in our estimation, and the sample selection performed to arrive at this sample. The first column shows all individuals in the NPD who attended school in England, which includes those who were educated in private schools. Our analysis will exclude those with fewer than five A*-C GCSE grades, as this level of attainment is a near-universal prerequisite for entry to university[2] and we only want to include individuals who conceivably had the option of going to HE in our control sample. The effect we are measuring is therefore the impact of HE *for those with at least five A*-C grades at GCSE*. As shown by Columns 1 and 2, around half (52%) of the NPD cohort members have five A*-C GCSE grades.

Of those with five A*-C GCSE grades, 98% have all the background characteristics and exam data that we need for our analysis, as shown in Column 3. In order to perform our analysis, we then need to match the school records of these individuals to their higher education and earnings records. Around 4% of the remaining individuals cannot be matched to their higher education and earnings records, which leaves us with the sample in Column 4. Individuals who cannot be

---

[2]Only around 6% of those with fewer than five A*-Cs at GCSE have started a degree by age 21.

matched may, for example, have missing or conflicting name and address information. The match rate is higher for more recent cohorts (over 97% for the 2007 GCSE cohort, up from 94% for the 2002 cohort).

In Column 5, we exclude those who enter university before 18 or after age 20 and part-time students, as we focus on estimating the impact of entering a full-time HE course within the first three years of leaving school.[3] The impact of HE on earnings may be very different for those who attend HE as mature students or those who attend part-time. In addition in Column 5, we drop a very small number of individuals with missing information on the subject they studied or who study very unusual courses (such as 'combined studies'). We are unable to provide reliable estimates for these courses, and this decision makes no difference to our overall estimates.

Finally, we show the set of people included in the final estimation in Column 6. This includes anyone in sustained employment[4] with positive earnings in the HMRC data at some point between age 24 and age 29, the ages included in our analysis. Around 15% of those in Column 5 are never observed in sustained employment. This proportion is higher for the later cohorts, for whom we have fewer years of observations. The missing individuals here will either be in non-sustained employment, have moved abroad to outside the UK or simply not be filing a PAYE tax return with HMRC in all of the years we possibly observe them in the tax data. Some of these individuals may also be self-employed. We only have access to earnings from self-employment from tax year 2012-13 onwards, hence we do not include self-employment income (or those are self-employment but not in 'sustained PAYE employment') in our main analysis. We will show the sensitivity of our results to including this in the Appendix.

Column 6 shows the final number of individuals included in any of our analysis as a result of all the above restrictions.[5] From a total of 3.8 million English domiciled individuals taking their GCSEs between 2002 and 2007, we include around 1.4 million in the analysis. This is around three-quarters of the 'at least five A*-C GCSEs' sample.[6] Table A2 in the Appendix shows some descriptives of both those in the matched sample for analysis (those in Column 5 of Table 1) and those with at least five A*-C GCSEs who we do not use (i.e. those who are dropped between Columns 3 and 5 of Table 1) to give an idea of the types of individuals who are lost in the process of our sample selection. The groups have broadly comparable characteristics, but the group we include in the analysis has slightly higher prior attainment and comes from a slightly more advantaged background on average. This is mostly driven by individuals who study part-time or start HE after age 20 coming from less well-off backgrounds and having lower prior attainment

---

[3]A significant majority of HE attenders enter university within the first three years after leaving school. While interesting and important, estimating the returns to part-time degrees and to degrees that people start at older ages is beyond the scope of this report, as it presents very different data challenges and consequently requires a different methodological approach.

[4]Sustained employment is a variable generated by the Department for Education and is defined as being in employment in five out of the last six months of the tax year. We include individuals who are in both sustained employment and further study.

[5]Table A1 in the Appendix shows how many of the set of individuals in Column 6 are observed in any given year between ages 24 and 29. We see that for the 2002 GCSE cohort, around 85% of those ever included between 24 and 29 are included at any given age. This share is higher for the later cohorts as there are sequentially fewer years for each cohort to enter the dataset.

[6]In our later analysis we drop individuals with missing or extreme weights and this differs across the analysis performed (see Section 3.1 for more detail.) Consequently, the sample sizes in each estimation will not always sum to the total number of individuals in Column 6.

than the average HE student.

Table 1: LEO sample by GCSE year

| GCSE cohort | Population (1) | ≥5 A*-C GCSEs (2) | No missing NPD (3) | Matched identifiers (4) | Final sample (5) | Final earnings sample (6) |
|---|---|---|---|---|---|---|
| 2002 | 589,663 | 310,897 | 295,898 | 278,229 | 249,699 | 226,102 |
| 2003 | 621,929 | 316,802 | 312,121 | 295,352 | 266,103 | 238,564 |
| 2004 | 644,873 | 331,168 | 325,936 | 311,780 | 282,349 | 248,367 |
| 2005 | 644,345 | 336,754 | 333,069 | 320,042 | 292,506 | 250,741 |
| 2006 | 653,971 | 337,950 | 334,921 | 325,139 | 300,209 | 247,207 |
| 2007 | 662,225 | 345,119 | 342,019 | 332,053 | 310,069 | 233,444 |
| Total | 3,817,006 | 1,978,690 | 1,943,964 | 1,862,595 | 1,700,935 | 1,444,425 |
| % of Column (2) | - | 100% | 98% | 94% | 86% | 73% |

Note: Column 1 is the full sample of English domiciled pupils in the NPD. Column 2 excludes the individuals with fewer than five A*-C grades at GCSE. Column 3 shows the number of all individuals with have all the exam data and background characteristics we use in the analysis. Column 4 shows the individuals whom we have also been able to match to the identifiers needed to link school records to HE and earnings records. Column 5 excludes anyone studying part-time or entering university before 18 or after age 20, and in addition excludes anyone who went to HE but for whom we do not know the subject they studied or who studied a very unusual HE course (e.g. 'combined studies'). Column 6 shows the number of individuals who are observed in sustained employment and with positive earnings at least once between ages 24 and 29; this includes individuals in both sustained employment and further study.

Table 2 shows how the final sample given in Column 6 of Table 1 breaks down. There are just over 80,000 non-HE people per cohort in our final sample, which is around one-third of the total. About 8% of people who start a degree course do not graduate and instead 'drop out' (these individuals are included in our analysis as discussed in Section 3). This rate is slightly lower than in Belfield et al. (2018) due to using slightly different sample restrictions and due to us here conditioning on having positive earnings and being in sustained employment (i.e. being present in Column 6 of Table 1). Individuals who attended university full-time and started prior to age 21, as well as individuals in sustained employment, are less likely to have dropped out than the overall population of HE attendees.

Table 2: Analysis sample

| GCSE cohort | Analysis sample | No HE | Dropout | Graduate |
|---|---|---|---|---|
| 2002 | 226,102 | 82,296 | 11,934 | 131,872 |
| 2003 | 238,564 | 84,146 | 12,490 | 141,928 |
| 2004 | 248,367 | 88,652 | 12,430 | 147,285 |
| 2005 | 250,741 | 89,923 | 12,400 | 148,418 |
| 2006 | 247,207 | 84,398 | 11,799 | 151,010 |
| 2007 | 233,444 | 80,608 | 11,092 | 141,744 |
| Total | 1,444,425 | 510,023 | 72,145 | 862,257 |

Note: This sample aligns with our final earnings sample in Column 6 of Table 1.

Table 3 shows some basic descriptive statistics for our treatment group (those who attended HE and have at least five A*-C GCSEs) and for our control group (those who have at least five A*-C GCSEs and did not go to HE). The descriptive variables come from the NPD, to which we have merged in some neighbourhood measures of deprivation which we have combined with free school meal eligibility in order to generate a continuous measure of SES for those who went to state secondary schools (following Belfield et al. (2018)). Here we group that continuous SES measure into quintiles. For reference, we also show the descriptives for those with fewer than five A*-C GCSEs, although these individuals are not included in our analysis.[7]

| | Women | | | Men | | |
|---|---|---|---|---|---|---|
| | <5 A*-C | ≥5 A*-C | | <5 A*-C | ≥5 A*-C | |
| | | no HE | HE | | no HE | HE |
| *Background (share)* | | | | | | |
| FSM | 0.21 | 0.07 | 0.05 | 0.18 | 0.06 | 0.05 |
| EAL | 0.09 | 0.05 | 0.10 | 0.10 | 0.04 | 0.10 |
| SEN | 0.20 | 0.03 | 0.02 | 0.27 | 0.05 | 0.03 |
| State school | 0.98 | 0.93 | 0.86 | 0.98 | 0.93 | 0.84 |
| *of which:* | | | | | | |
| SES Q1 - least deprived | 0.10 | 0.23 | 0.32 | 0.12 | 0.25 | 0.34 |
| SES Q2 | 0.15 | 0.24 | 0.25 | 0.16 | 0.26 | 0.25 |
| SES Q3 | 0.19 | 0.22 | 0.20 | 0.20 | 0.22 | 0.19 |
| SES Q4 | 0.26 | 0.18 | 0.14 | 0.24 | 0.17 | 0.13 |
| SES Q5 - most deprived | 0.31 | 0.13 | 0.10 | 0.27 | 0.11 | 0.08 |
| *Ethnicity (share)* | | | | | | |
| White | 0.83 | 0.90 | 0.81 | 0.82 | 0.90 | 0.81 |
| Black | 0.04 | 0.02 | 0.04 | 0.04 | 0.01 | 0.03 |
| Asian | 0.05 | 0.03 | 0.08 | 0.06 | 0.03 | 0.09 |
| Other | 0.07 | 0.06 | 0.07 | 0.07 | 0.06 | 0.07 |
| *Prior attainment (share)* | | | | | | |
| KS2 maths level 5+ | 0.03 | 0.23 | 0.38 | 0.06 | 0.33 | 0.49 |
| KS2 English level 5+ | 0.06 | 0.33 | 0.50 | 0.04 | 0.24 | 0.40 |
| KS4 maths A*/A | <0.01 | 0.09 | 0.31 | <0.01 | 0.10 | 0.38 |
| KS4 English A*/A | <0.01 | 0.16 | 0.46 | <0.01 | 0.09 | 0.35 |
| N | 698,214 | 370,033 | 678,059 | 835,569 | 338,068 | 557,804 |

Table 3: Background characteristics by attainment group

Note: The sample includes all individuals with no missing NPD information who we can link to higher education and/or earnings files, and hence the total sample with at least five A*-C GCSEs aligns with Column 3 in Table 1 and averages over the individuals in the six GCSE cohorts. EAL = English as an additional language, FSM = free school meals, SEN = non-statemented special educational needs. All numbers are proportions over the respective sample, except socio-economic status quintiles, ethnicity and Key Stage 4 scores, which are only defined for state school pupils.

The descriptive variables that we show are a subset of the control variables included in our estimation regressions. Amongst those with at least five A*-C GCSEs, on average HE partici-

---

[7]These descriptives include all English domiciled individuals with no missing NPD information who we can link to higher education and/or earnings files and hence relate to the total sample in Column 3 of Table 1.

pants have higher prior attainment and are more likely to come from higher socio-economic backgrounds than those who do not go to HE. They are also more likely to be non white and less likely to have attended a state secondary school. This highlights the importance of adjusting for composition of the HE intake when estimating the returns to HE, as there are several variables that are likely to be correlated with both HE attendance and subsequent earnings.

The table shows that while the differences between the treatment and control groups are large, they are much smaller than the differences between individuals with and without at least five A*-C GCSEs. For both genders, the individuals who did not obtain at least five A*-C GCSEs come from considerably more deprived backgrounds than those with higher prior attainment. In particular, the difference in the SES gradient between those in HE and those without five A*-C GCSEs is particularly stark: around 33% of those in HE come from the richest 20% and 9% come from the poorest 20%, while for those without five A*-C GCSEs just 11% are from the richest 20% and nearly 30% are from the poorest 20%. In general, this highlights just how different those with fewer than five A*-C GCSEs are from those who conceivably had the opportunity to go to HE, which is our justification of the decision not to include those individuals in the analysis.

## 2.2 Data descriptives

We now show some simple descriptives from the dataset, based on the sample in the final column of Table 1. Whenever those with fewer than five A*-C GCSEs are included, this includes all individuals with non-missing NPD information who can be matched to their HMRC records and are in sustained employment with non-zero earnings. We are using 'raw' earnings in the tax year (conditional on being in sustained employment) rather than annualised earnings. Age is proxied for by assuming individuals are 16 when they sit their GCSEs: for example, we assume the 2007 GCSE cohort will be 21 in the 2012-13 tax year. In practice, individuals from this cohort are likely to be between 20 and 22 during that tax year.

Figure 1 shows male and female average earnings by age for those in sustained employment for the different education groups - again we include those without five A*-C GCSEs for reference. For both genders, we see a clear ordering of average earnings by the time education group by individuals are in their late 20s. Among those who do not attend HE, those with at least five A*-C GCSEs earn over £5,000 more than those without. In their early 20s, graduates have lower earnings than those who did not attend HE, but they have much stronger earnings growth and soon overtake non-graduates. This differential earnings growth is particularly noticeable for women where, unlike for men, the lowest-educated group sees very little earnings growth throughout their 20s. These different growth rates by education level highlight the importance of when the returns to HE are estimated. For this reason we focus on the returns at age 29 - the oldest ages available in our data.

## Figure 1: Real earnings by education level
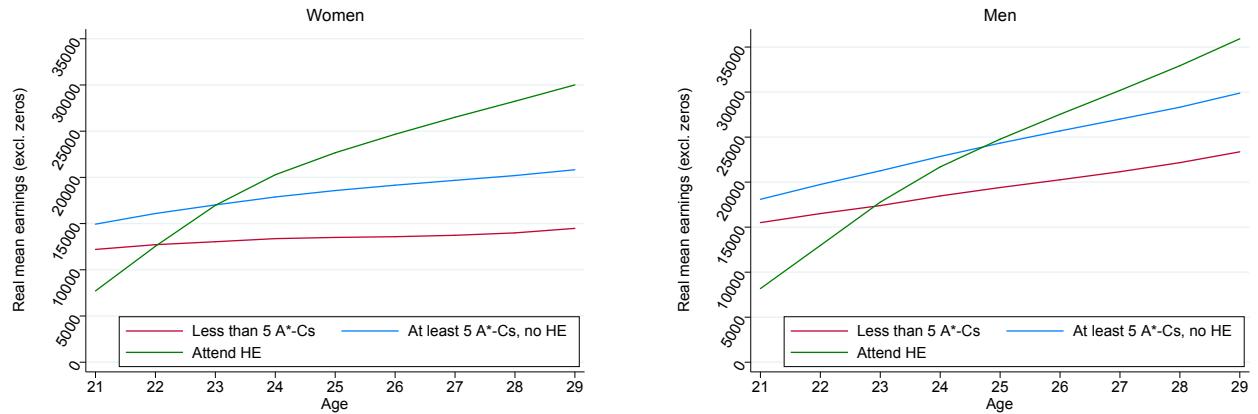


Note: The graph shows averages over GCSE cohorts 2002-2007 (when data is available) for those in sustained employment with positive earnings. We proxy for age by assuming age = midpoint of tax year - GCSE year + 16 (the age at the end of the school year individuals take their GCSEs). For example, the midpoint of the 2015/16 tax year is 2015, so individuals in the 2002 GCSE cohort are labelled as being 2015 - 2002 + 16 = 29. Earnings are in 2018 prices. Earnings are capped at the 1st and 99th percentiles to avoid sensitivity to outliers. Self employment income is excluded.

The HE groups in Figure 1 include everyone who started a degree course, which fails to acknowledge that not all students complete their degree course. Figure 2 separates out these individuals from those who do graduate, showing that for both genders the outcomes are considerably worse for dropouts than for those who graduate. However, dropouts do see stronger earnings growth than individuals who did not enter HE at all: for men, dropouts have lower average earnings throughout their 20s than men with at least five A*-C GCSEs who did not enter HE, though this gap is closing over time, while for women, the average earnings of dropouts have overtaken those of the five A*-C non-HE group by their late 20s. Although we cannot say conclusively from these figures, these earnings outcomes seem to imply that the majority of those dropping out are doing so not due to having better outside labour market opportunities. Later in our analysis, we will show our estimation results both including and excluding this dropout group and show the impact of this decision on our results.

Figure 2: Real earnings by education level
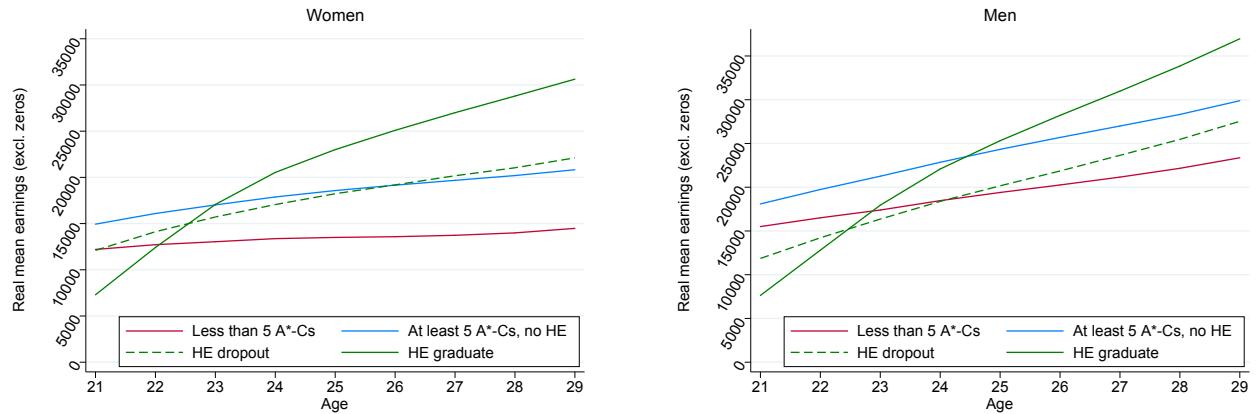


Note: Uses 2002-2007 GCSE cohorts. Figures condition on being in sustained employment with positive earnings, cap earnings at the 1st and 99th percentiles and exclude self-employment income. Dropouts are those who start a degree and are never observed completing a degree up until 2015-16. Earnings in 2018 prices.

The graphs above have shown how raw earnings vary across education levels for those in sustained employment. However, education can also have an impact on the probability of being in employment. Figure 3 shows the probability of being in sustained employment for the different education groups and shows that graduates are indeed the group most likely to be in sustained employment from around age 24. Among non-graduates, those with at least five A*-C GCSEs are significantly more likely to be in work than those without. Outcomes for dropouts are very similar to those for the five A*-C non-HE group by their late 20s.

Figure 3: Probability of being in sustained employment by education level
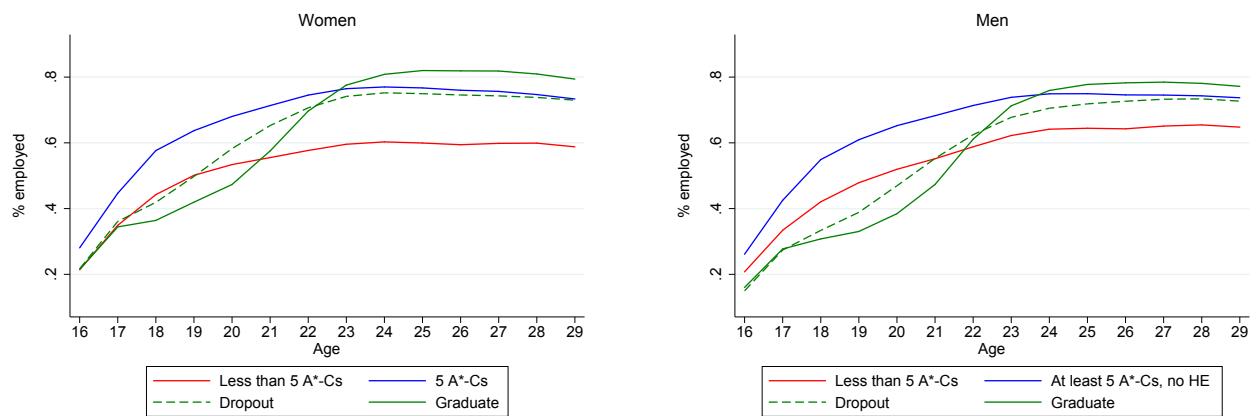


Note: Uses 2002-2007 GCSE cohorts. Employment rates are based on a comparison of Columns 5 and 6 of Table 1 (or equivalent).

Not only do very large differences in earnings exist between those who did and did not attend HE, but we also see large earnings differences among those from different socio-economic backgrounds. In Figure 4, we look at these earnings differences by socio-economic background based on the SES measure introduced in the previous subsection. Students who attended independent

16

(private) schools are shown separately. The differences in earnings are extremely large, with privately educated women earning more than twice as much as women from the 20% most deprived socio-economic backgrounds by age 29, highlighting the potential importance of controlling for background when investigating returns to HE.

The HE attendance across each of these groups varies considerably, with individuals from independent schools and from the higher SES groups being much more likely to attend HE than those from lower SES backgrounds. This is likely to explain a significant fraction of those very large earnings differences, and also aligns with independent-schooled individuals actually having some of the lowest earnings in their early 20s, but much steeper earnings growth, just as we saw when comparing graduates and non-graduates in the previous figures.

Figure 4: Real earnings by SES
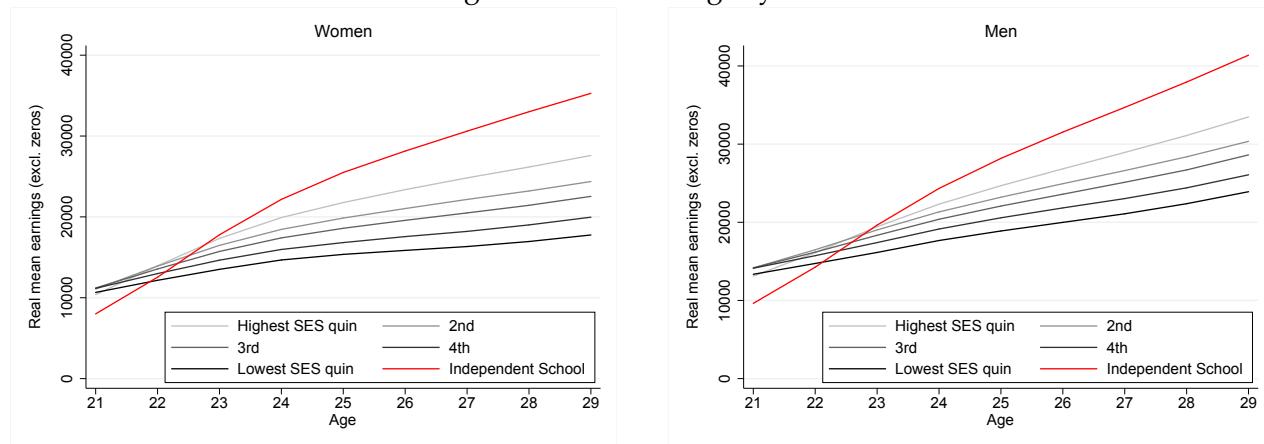


Note: Uses 2002-2007 GCSE cohorts. Figures include everyone (including non-HE with and without at least five A*-C GCSEs) conditional on being in sustained employment, with earnings capped at the 1st and 99th percentiles and excluding self-employment income. Earnings are in 2018 prices. SES quintiles are based on the whole NPD sample excluding those in independent schools. See Chowdry et al. (2013, p. 439-440) for more details on how this index has been constructed.

When looking at students with different maths and English GCSE grades, we also see large heterogeneity in earnings, as shown in Figures 5 and 6 respectively . As with SES, HE attendance varies significantly across these groups and we see some evidence of a 'crossing over effect' in earnings in the early 20s. Students with an A* in maths see particularly rapid earnings growth compared with those with lower grades, a difference which is less pronounced for those with an A* in English. For both maths and English, the outcomes of those with less than a C at GCSE are particularly poor, with little earnings growth at all throughout their 20s, especially for women. While these differences in earnings between individuals with different prior attainment and with different SES do not show a causal effect of GCSE grades or family background on earnings, they do highlight the importance of controlling for prior attainment when estimating returns to HE.

Figure 5: Real earnings by maths GCSE grade
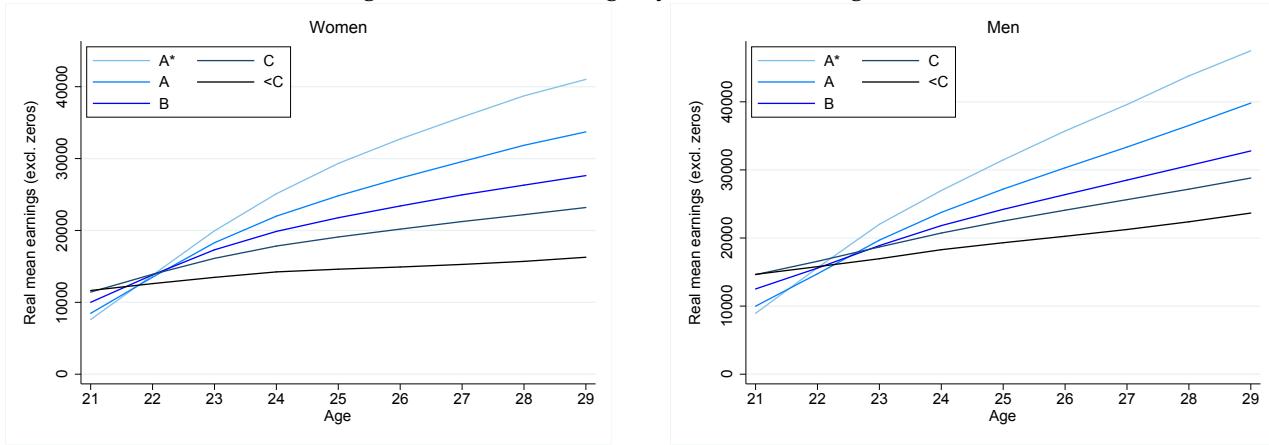


Note: Uses 2002-2007 GCSE cohorts. Figures include everyone (including non-HE with and without at least five A*-C GCSEs) conditional on being in sustained employment, with earnings capped at the 1st and 99th percentiles and excluding self-employment income. Earnings are in 2018 prices.

Figure 6: Real earnings by English GCSE grade
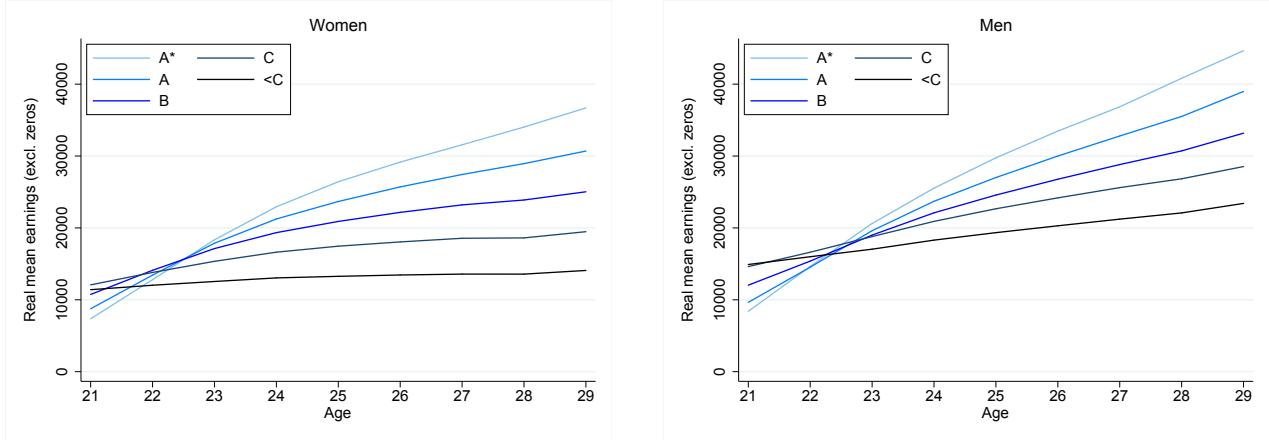


Note: Uses 2002-2007 GCSE cohorts. Figures include everyone (including non-HE with and without at least five A*-C GCSEs) conditional on being in sustained employment, with earnings capped at the 1st and 99th percentiles and excluding self-employment income. Earnings are in 2018 prices.

We now turn to looking at how raw earnings vary amongst the HE group. Figure 7 shows average earnings (conditional on sustained employment) for 29-year-old men by subject, while Figure 8 shows the equivalent for women. These figures are very similar to those in Belfield et al. (2018) but with the addition of two horizontal lines which show the average earnings of the two non-HE groups, splitting them out between those with and without at least five A*-C GCSE grades.[8]

---

[8]They also differ slightly because these are the earnings of the 2002 GCSE cohort in 2015-16 (which we loosely term 29-year-olds), where Belfield et al. (2018) show earnings five years after graduation. This is also true for the earnings by HEI shown in Figures 9 and 10. Following Belfield et al. (2018), we use the CAH2 definition of subjects, with chemistry combined with biosciences. Celtic studies, humanities & liberal arts, and 'combined' subjects have been excluded for sample size reasons.

Figure 7: Real earnings by degree subject at age 29, men
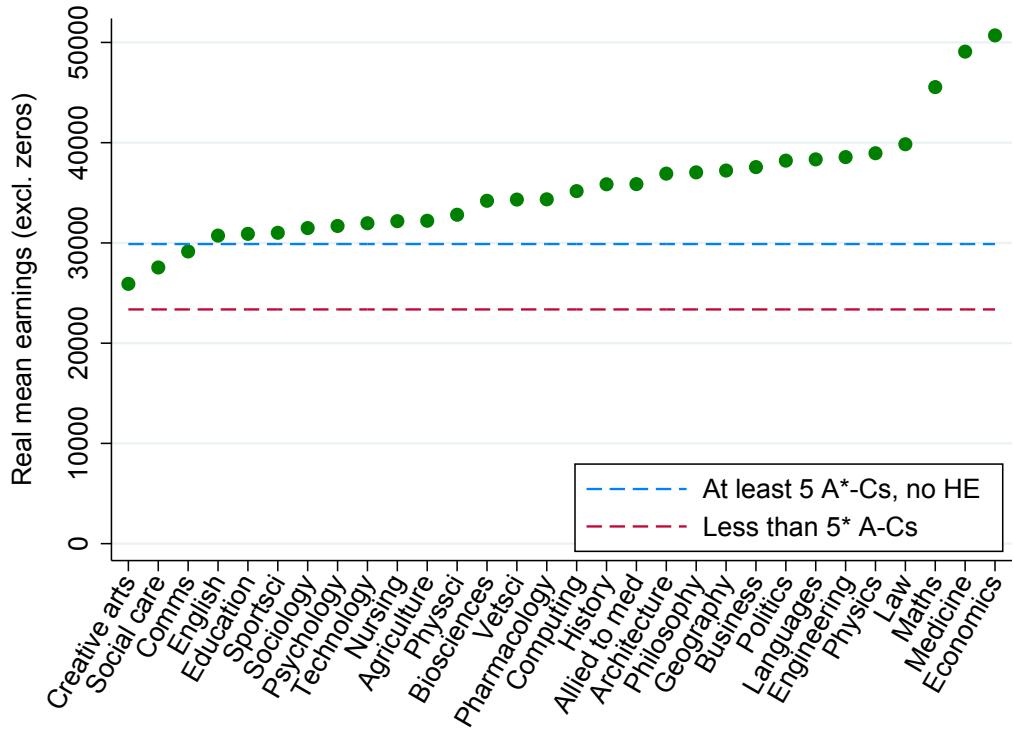
Note: Figure shows average earnings (2018 prices) by subject studied (including dropouts) for the 2002 GCSE cohort in 2015-16, conditional on being in sustained employment, with earnings capped at the 1st and 99th percentiles and excluding self-employment income.

We see a very similar ranking of subjects for both men and women. For both genders, medicine, economics, maths and physics graduates have some of the highest earnings, and social care and creative arts graduates have the lowest earnings. Gender differences in earnings are much smaller among graduates than among non-graduates. For men, but not for women, there are a number of degree subjects with average age 29 earnings below those of the group that has at least five A*-C GCSE grades but did not attend HE.

Figure 8: Real earnings by degree subject at age 29, women
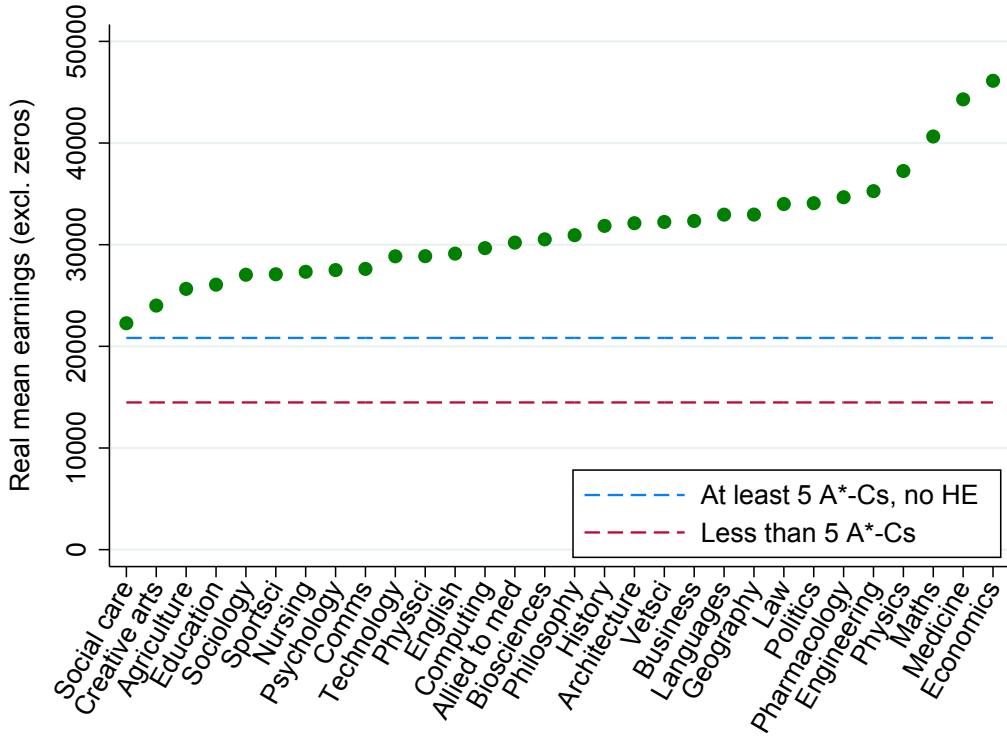


Note: Figure shows average earnings (2018 prices) by subject studied (including dropouts) for the 2002 GCSE cohort in 2015-16, conditional on being in sustained employment, with earnings capped at the 1st and 99th percentiles and excluding self-employment income.

A similar pattern emerges when we investigate earnings by higher education institution (HEI). Figures 9 and 10 show average earnings for those in sustained employment at age 29 by HEI attended for men and women respectively. For men, more than one-third of institutions have average earnings below the five A*-C non-HE group, and a handful of institutions even have average earnings below the fewer than five A*-C group. For women, only a few institutions have average earnings below the five A*-C non-HE group, and the graduates of all institutions have average earnings above the fewer than five A*-C group.

It is important to emphasise that these figures show differences in age 29 earnings, which may not always be reflective of lifetime earnings differences. In particular, the stronger earnings growth of HE graduates (and to a lesser extent, HE dropouts) compared with those who did not attend HE could imply that while there are institutions with average earnings below those of the non-HE group at age 29, this may not necessarily still be the case at later ages.

Figure 9: Real earnings by HEI at age 29, men
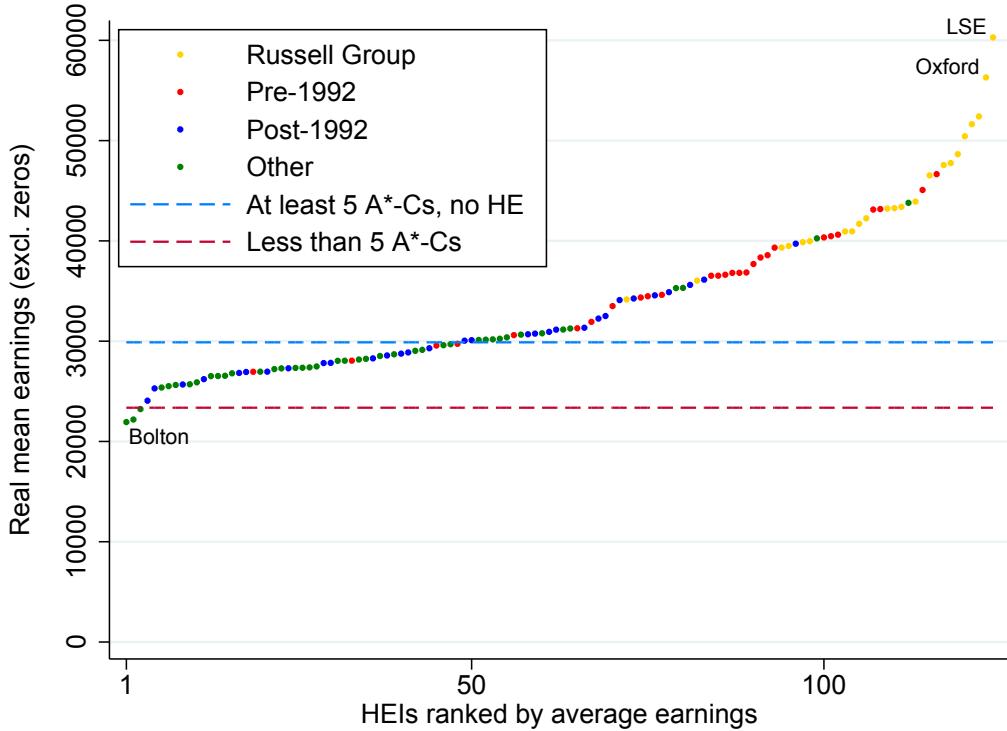
Note: Figure shows average earnings (2018 prices) by higher education institution attended (including dropouts) for the 2002 GCSE cohort in 2015-16, conditional on being in sustained employment, with earnings capped at the 1st and 99th percentiles and exclude self-employment income.

As in Belfield et al. (2018), we label the top and bottom institutions and also split out some broad institution groupings.[9] For both men and women it is clear that while there is considerable variation around this trend, the Russell Group dominates in terms of average earnings, with Pre-1992 universities generally also ranked quite highly, and Post-1992 and Other universities typically towards the bottom of the distribution.

Of course, as when we showed differences in earnings by socio-economic background and GCSE grades, these earnings differentials by subject and institution are purely descriptive and may be driven by differences in background and prior attainment of their student intake. Later in the report, we will move away from these simple descriptive differences and will investigate differences in earnings allowing for differences in student composition. We will discuss our methodology for doing this in Section 3.

---

[9]We divide institutions into four groups: 'Russell Group', which is a self-selected association of elite research institutions; 'Pre-1992' institutions, which are traditional universities that obtained university status prior to the 1992 Further and Higher Education Act; 'Post-1992' institutions with polytechnic or central institution roots; and 'Other' institutions, which are Post-1992 institutions without polytechnic or central institution roots. See the Online Appendix for a list of the institutions in each group.

Figure 10: Real earnings by HEI at age 29, women
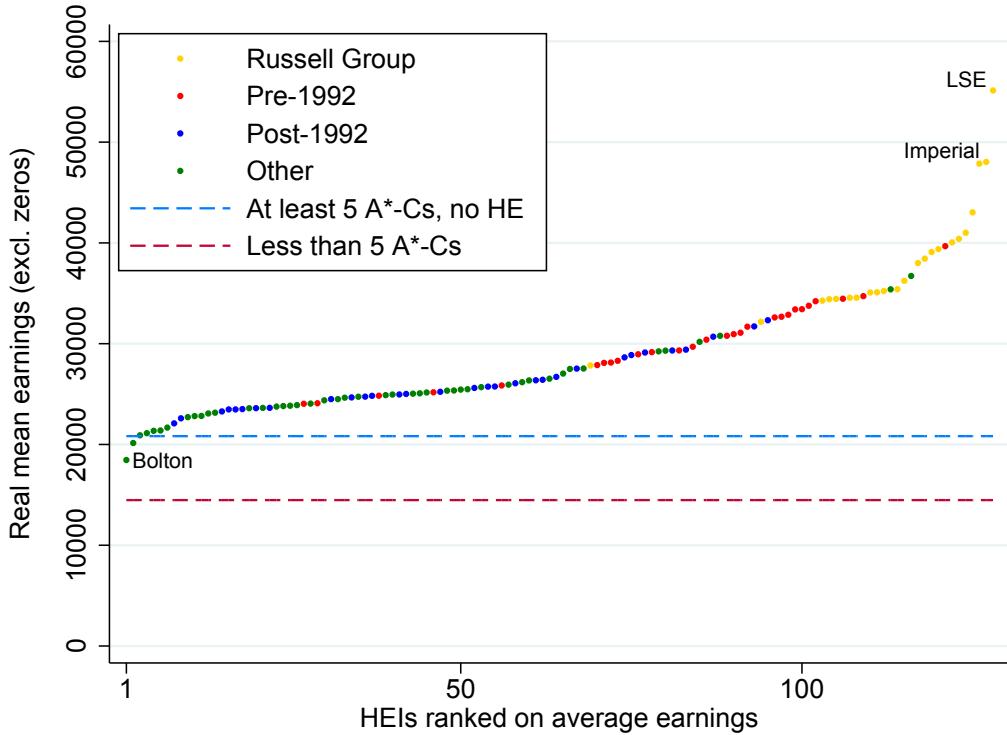
Note: Figure shows average earnings (2018 prices) by higher education institution attended (including dropouts) for the 2002 GCSE cohort in 2015-16, conditional on being in sustained employment, with earnings capped at the 1st and 99th percentiles and exclude self-employment income.

## 2.3 Subgroup analysis

In this report, we are interested not only in the return for the average person but also in how these returns vary based on the background characteristics of individuals that we observe *prior to entry to HE*. Specifically, we will perform subgroup analysis by dividing our sample into lower, middle and higher prior attainment groups based on their GCSE performance,[10] as we might expect the returns to HE to vary considerably across this metric. We also split these three groups again by whether individuals have at least one maths or science A- or AS-level (henceforth, 'STEM A-level'),[11] which is also likely to interact importantly with returns: people who have not chosen STEM A-level subjects are likely to be able to access a different set of courses from those who have, and those courses may differ significantly in terms of their earnings potential.

---

[10]These groups correspond respectively to the bottom, middle and top tercile of KS4 point scores in our sample. As such, the classification into 'lower', 'middle' and 'higher' prior attainment groups is relative, and individuals in all groups, including the bottom group, will have obtained at least five A*-C GCSE grades. The KS4 score cut-off between the lower and middle groups is 350 points, while the cut off between the middle and higher groups is 395 points. For our cohorts, GCSE grades translate into KS4 points as follows: A* = 58 points, A = 52 points, B = 46 points, C = 40 points, D = 34 points, E = 28 points, F = 22 points, G = 16 points and U = 0 points. The total KS4 points score is then obtained by adding up the points for the different subjects. For someone taking 8 GCSEs, they need to achieve the equivalent of 5 As and 3 Bs or higher to be in the top group, or equivalent of 5 Bs and 3 Cs or higher to be in middle group.

[11]A science A- or AS-level is defined as an A- or AS level in biology, human biology, chemistry, physics, science, electronics, geology or computer studies.

We do this subgroup analysis separately for men and women, meaning there are six subgroups for each. Table 4 shows some basic descriptives. Around three-quarters of women and two-thirds of men do not have a STEM A-level. Very few individuals in the lower prior attainment group take a STEM A-level, but despite the small population shares, we still observe over 10,000 individuals in our smallest subgroup, highlighting the advantages of the administrative data.

Table 4: Sample sizes and HE participation rates by prior attainment

|  | Prior attainment | Women | | | Men | | |
|  |  | N | Share | Share in HE | N | Share | Share in HE |
|---|---|---|---|---|---|---|---|
| No STEM | Lower | 251,262 | 0.32 | 0.37 | 251,166 | 0.38 | 0.33 |
| A-level | Middle | 208,771 | 0.27 | 0.68 | 134,182 | 0.20 | 0.64 |
|  | Higher | 124,101 | 0.16 | 0.89 | 55,123 | 0.08 | 0.87 |
| STEM | Lower | 10,746 | 0.01 | 0.71 | 27,236 | 0.04 | 0.68 |
| A-level | Middle | 49,451 | 0.06 | 0.85 | 75,878 | 0.11 | 0.84 |
|  | Higher | 133,302 | 0.17 | 0.95 | 123,207 | 0.18 | 0.95 |
| Total |  | 777,633 | 1.00 | 0.67 | 666,792 | 1.00 | 0.62 |

Note: This table includes all individuals in the analysis sample, pooled over the 2002-2007 GCSE cohorts. The samples add up to the final column in Table 1. Prior attainment is based on tercile of GCSE attainment conditional on having at least five A*-C GCSEs and being in Column 5 of Table 1. STEM A-level means the individual has at least one A- or AS-level in maths, biology, human biology, chemistry, physics, science, electronics, geology or computer studies.

The share of HE participants in these six subgroups varies considerably, with the HE rate increasing with prior attainment and the presence of a STEM A-level.[12] When considering the results it should be kept in mind that the control group for the higher prior attainment STEM A-level group is small, due to the extremely high HE participation rate (95%) amongst this group.

Individuals in the different subgroups not only differ in their HE participation rate, they are also likely to choose different subjects and attend different institutions conditional on being in HE. A full list of the subject choices for each subgroup is provided in Appendix Tables A3 and A4, but Tables 5 and 6 highlight some of the most striking differences between the groups.

---

[12]The prior attainment groups correspond respectively to the bottom, middle and top tercile of KS4 points in our sample. As such the classification into 'low', 'middle' and 'high' prior attainment groups is relative, and individuals in all groups, including the bottom group, will have obtained at least 5 A*-C GCSE grades. On average, the STEM group has higher KS4 scores within each tercile. The average KS4 score for individuals without a STEM A-level is 327, 372 and 422 for the lower, middle and higher prior attainment groups respectively, while the equivalent for the STEM A-level group is 335, 377 and 432 points.

Table 5: Degree subject by prior attainment

| | No STEM A-level | | | STEM A-level | | | |
| | Lower | Middle | Higher | Lower | Middle | Higher | All |
|---|---|---|---|---|---|---|---|
| *Men* | | | | | | | |
| Business | 0.19 | 0.17 | 0.10 | 0.15 | 0.13 | 0.07 | 0.13 |
| Creative arts | 0.16 | 0.15 | 0.12 | 0.05 | 0.05 | 0.03 | 0.09 |
| Computing | 0.10 | 0.06 | 0.02 | 0.13 | 0.09 | 0.05 | 0.07 |
| History | 0.04 | 0.06 | 0.13 | 0.02 | 0.02 | 0.04 | 0.05 |
| Engineering | 0.05 | 0.03 | 0.02 | 0.13 | 0.14 | 0.14 | 0.08 |
| Sportsci | 0.08 | 0.06 | 0.02 | 0.05 | 0.05 | 0.02 | 0.05 |
| Biosciences | 0.01 | <0.01 | <0.01 | 0.10 | 0.10 | 0.10 | 0.05 |
| Other | 0.37 | 0.45 | 0.58 | 0.38 | 0.42 | 0.55 | 0.47 |
| **All subjects** | **1.00** | **1.00** | **1.00** | **1.00** | **1.00** | **1.00** | **1.00** |
| *Women* | | | | | | | |
| Business | 0.14 | 0.12 | 0.07 | 0.12 | 0.10 | 0.06 | 0.10 |
| Creative arts | 0.20 | 0.20 | 0.17 | 0.04 | 0.05 | 0.04 | 0.14 |
| Computing | 0.02 | 0.01 | <0.01 | 0.03 | 0.02 | <0.01 | 0.01 |
| History | 0.02 | 0.04 | 0.09 | 0.01 | 0.02 | 0.04 | 0.04 |
| Engineering | <0.01 | <0.01 | <0.01 | 0.02 | 0.02 | 0.02 | 0.01 |
| Sportsci | 0.03 | 0.02 | 0.01 | 0.04 | 0.04 | 0.02 | 0.02 |
| Biosciences | 0.01 | <0.01 | <0.01 | 0.12 | 0.12 | 0.13 | 0.05 |
| Other | 0.57 | 0.60 | 0.65 | 0.61 | 0.63 | 0.68 | 0.63 |
| **All subjects** | **1.00** | **1.00** | **1.00** | **1.00** | **1.00** | **1.00** | **1.00** |

Note: Table shows the proportion of students taking each of a set of subjects by groups of prior attainment (full set of subjects is shown in Appendix Tables A3 and A4). Figures are calculated based on the 2002-2007 GCSE cohorts, for those who achieved at least five A*-Cs and attended HE in the UK.

We see from Table 5 that the choice of business degrees is very clearly negatively correlated with prior attainment, for both the STEM and non-STEM A-level groups and for both men and women. Meanwhile, creative arts is a dramatically more popular choice amongst those without a STEM A-level and biological science is a much more popular choice for those with a STEM A-level. For men, engineering is also a common choice for those with a STEM A-level, but the same is not really true for women.

Table 6: University type by prior attainment

| | No STEM A-level | | | STEM A-level | | | |
|---|---|---|---|---|---|---|---|
| | Low | Middle | High | Low | Middle | High | Total |
| *Men* | | | | | | | |
| Russell | 0.01 | 0.09 | 0.46 | 0.07 | 0.18 | 0.60 | 0.19 |
| Pre-1992 | 0.04 | 0.13 | 0.18 | 0.13 | 0.22 | 0.19 | 0.12 |
| Post-1992 | 0.18 | 0.27 | 0.14 | 0.37 | 0.33 | 0.11 | 0.20 |
| Other | 0.77 | 0.51 | 0.22 | 0.44 | 0.26 | 0.09 | 0.48 |
| **Total** | **1.00** | **1.00** | **1.00** | **1.00** | **1.00** | **1.00** | **1.00** |
| *Women* | | | | | | | |
| Russell | 0.01 | 0.08 | 0.40 | 0.05 | 0.14 | 0.55 | 0.19 |
| Pre-1992 | 0.04 | 0.11 | 0.18 | 0.11 | 0.18 | 0.18 | 0.11 |
| Post-1992 | 0.18 | 0.29 | 0.18 | 0.39 | 0.37 | 0.14 | 0.22 |
| Other | 0.77 | 0.52 | 0.25 | 0.46 | 0.31 | 0.12 | 0.48 |
| **Total** | **1.00** | **1.00** | **1.00** | **1.00** | **1.00** | **1.00** | **1.00** |

Note: Table shows the proportion of students attending each type of university by groups of prior attainment. Figures are calculated based on the 2002-2007 GCSE cohorts, for those who achieved at least five A*-Cs and attended HE in the UK.

We see from Table 6 that there are very large differences in HEI type attended across the six groups. There is a very clear correlation between prior attainment and attendance at Russell Group or Pre-1992 universities versus Post-1992 or Other HEIs. Between 55 and 60% of the higher prior attainment STEM A-level group attend Russell Group universities, compared with just 1% of the lower prior attainment no STEM A-level group. Those without a STEM A-level are much more likely to attend Other HEIs, which is unsurprising as these HEIs are much more likely to offer subjects such as creative arts that we know this group are much more likely to study.

Finally, we turn to the earnings differences across the six groups in Figures 11 and 12. The figures show average earnings by age between 24 and 29 for those within each of the six groups by HE attendance. This enables us to better understand how earnings profiles differ descriptively between the treatment and control groups in each case.

Figure 11: Real earnings by prior attainment group, men
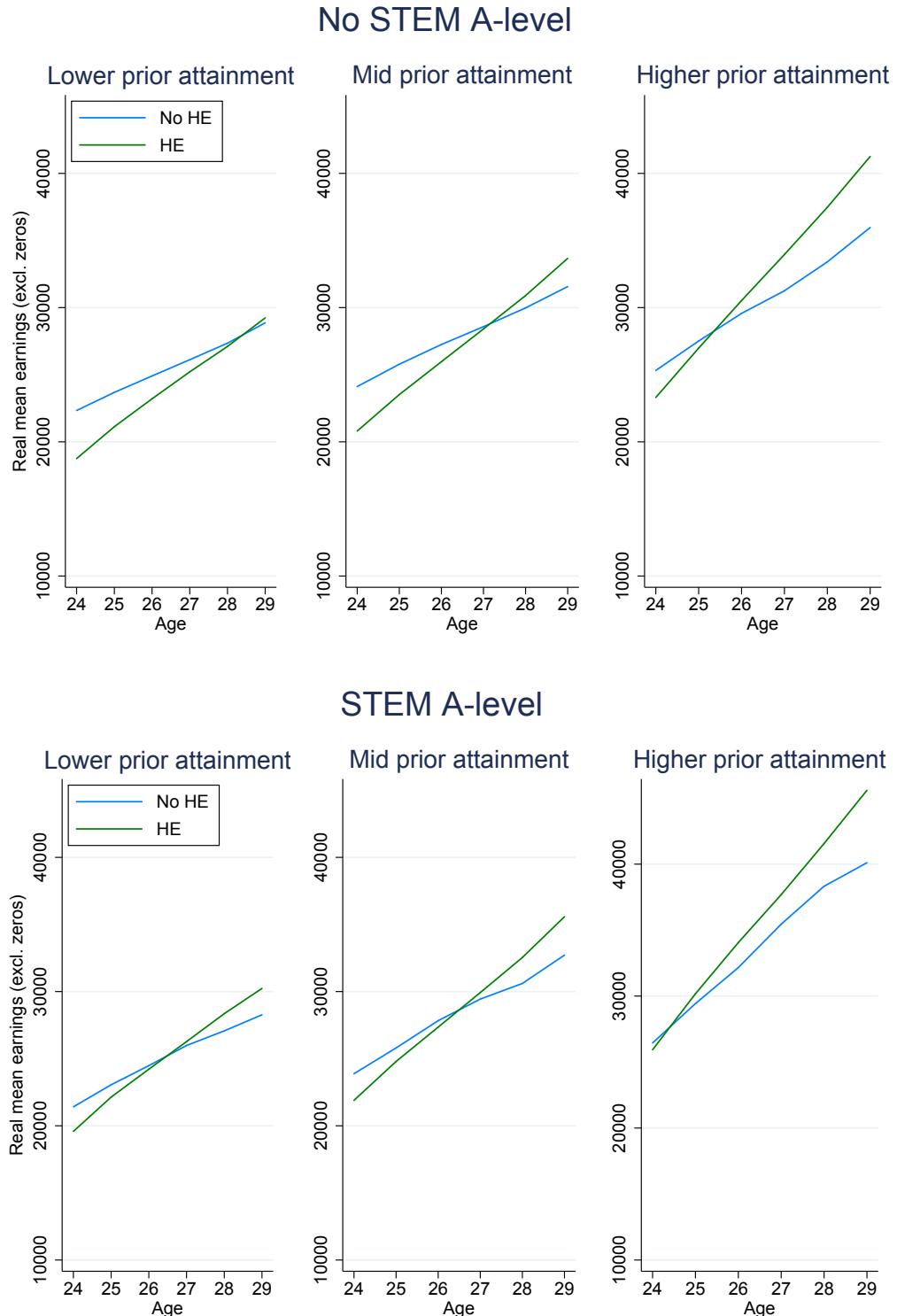
## No STEM A-level



## STEM A-level



Note: Uses 2002-2007 GCSE cohorts. Average earnings are in 2018 prices and are based on those in sustained employment. Earnings are capped at the 1st and 99th percentile. Prior attainment is based on GCSE results.

Figure 12: Real earnings by prior attainment group, women
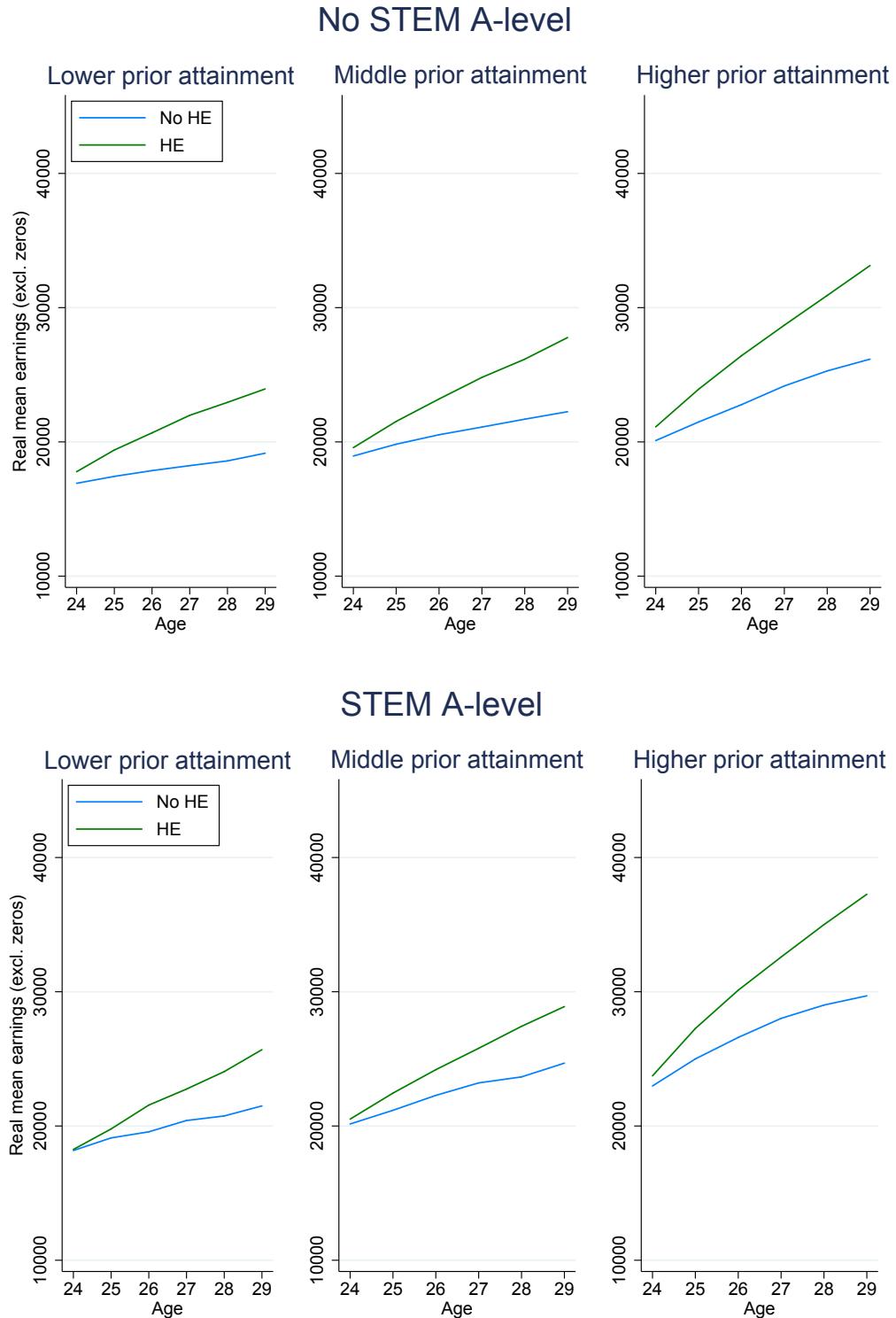
Note: Uses 2002-2007 GCSE cohorts. Average earnings are in 2018 prices and are based on those in sustained employment. Earnings are capped at the 1st and 99th percentile. Prior attainment is based on GCSE results.

We see that there is a clear pattern of average earnings being higher at any given age for those with higher prior attainment, for both the HE and non-HE groups. It is also true that the slope of earnings is steeper for those with higher prior attainment, which means that earnings gaps between those with high and low prior attainment grow with age. However, within group, it is important to note that the earnings profile is always steeper for the treated (HE) group than for the control (non-HE) group. This suggests first that returns will vary with age within the sample that we have, and second that the returns may be large at later ages. Indeed, there is very little evidence of a 'levelling-off' effect in earnings growth amongst the HE groups. Finally, it is also interesting to note that for all six groups there is a 'cross-over effect' for men - whereby average earnings for the HE group overtake the non HE groups in the mid to late-20s - that does not exist for women. In all cases for women, the HE group starts slightly ahead at age 24 and accelerates away. Of course, as with the descriptive differences in Figure 1, these are descriptive effects only; in our later analysis, we will attempt to adjust for various differences between treatment and control groups that enable us to move towards estimating causal effects of attending HE for each of the six subgroups.

## 3   Methodology

The decisions on whether to attend HE, and on which subject to take and where to study are not random. Family background and prior attainment influence these choices, meaning these characteristics are likely to differ between graduates and non-graduates and across different degrees. Many of these characteristics are also likely to impact individuals' earnings, which may confound the impact of a degree on graduates' earnings.

The methodological challenge in estimating the earnings return or employment impact of a university course is therefore to account for all the differences between individuals that might affect earnings or employment prospects. We want to isolate the true impact of studying a given course and remove the effect of background characteristics which may mean that individuals on that course may have done better or worse in the labour market regardless of attending the course. In this section, we set out our method for controlling for these differences, along with the assumptions that this implies.

We start by defining our sample as the set of individuals who could possibly enter HE. Individuals with very low prior attainment are very unlikely to ever enter HE and these individuals are likely to be very different from those who do attend HE. Some of these large differences are likely to be unobserved, which means we would be unable to fully control for all differences that are likely to impact earnings. Including those individuals in our sample would then have the potential to bias our estimates. We therefore include in our estimation sample only individuals who achieve at least five A*-C GCSEs. As having at least this level of prior attainment is virtually a prerequisite for attending an HE course, this effectively excludes individuals for whom going to HE was never a feasible option.

Previous work has used two alternative control groups when estimating the returns to HE. First, Blundell et al. (2005) use individuals with at least one A-level as a control group. This is a

tighter control group than our methodology as these individuals could be considered more likely to attend HE. However, around 16% of individuals who do attend HE do not have any A-levels and so this approach would lead to a selective sample of HE attendants.[13]

Second, Bibby et al. (2014) estimate the returns to further education by comparing those who completed their course with those who started but did not finish; the equivalent here would be to compare HE graduates with HE dropouts. This control group is arguably more similar to the group of HE graduates as they also self-selected into HE. However, dropping out of HE is inherently not random. If the reasons these individuals did not complete their degrees also affect their earnings later in life (and this is not controlled for in the observable characteristics), this will bias the estimates. We believe this to be very likely, as dropout might be correlated with things such as a lack of motivation or long-term health problems, which are also likely to be correlated with subsequent earnings. We therefore do not want to use dropouts as our control in our main estimates. We will, however, apply these two alternative approaches to the LEO dataset as a robustness check of our results.

Figure 13: Distribution of KS4 points score by university
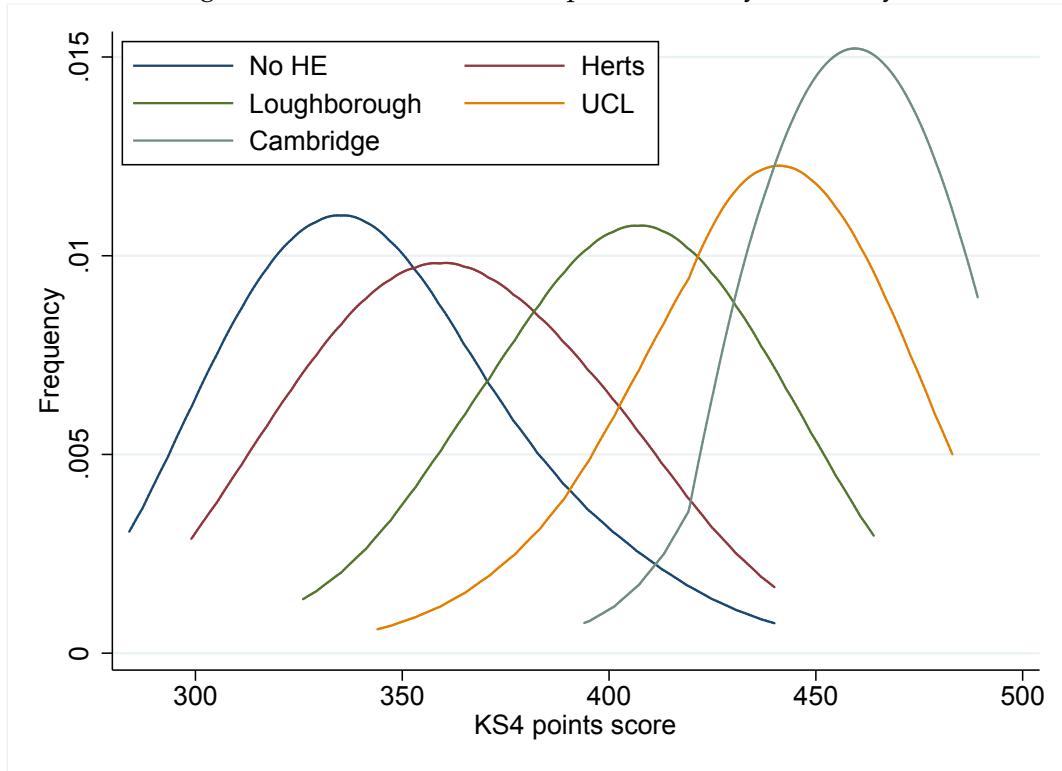


Notes: Figures calculated based on those students who took GCSEs in England between 2003 and 2007 and achieved at least five A*-Cs (GCSE points scores in 2002 are not consistent with those later - this is accounted for in the rest of the analysis). GCSE grades translate into KS4 points as follows: A* = 58 points, A = 52 points, B = 46 points, C = 40 points, D = 34 points, E = 28 points, F = 22 points G = 16 points and U = 0 points. The total KS4 points score is then obtained by adding up the points for the different subjects.

Even after selecting a control group of individuals who had the possibility of attending HE,

---

[13]The decision to take A-levels could also be consider endogenous to the HE participation decision and so focusing solely on this group may lead to biased estimates of HE returns.

there are still considerable differences between those who do and do not attend HE, and among students who study different subjects or study at different institutions. Figure 13 shows the extent of sorting of students into institutions based on ability. Clearly, students with high prior attainment (measured here by KS4 points score) typically attend higher-status universities such as the University of Cambridge or University College London.

Similarly, there are considerable differences in the educational background of students who study different subjects. As shown in Figure 14, almost all men studying for maths and physics degrees have a maths A- or AS-level compared with around 5% of students studying communication and sociology. Figure 15 shows the same pattern for women.

Figure 14: Share of students studying each subject with a maths A- or AS-level, men
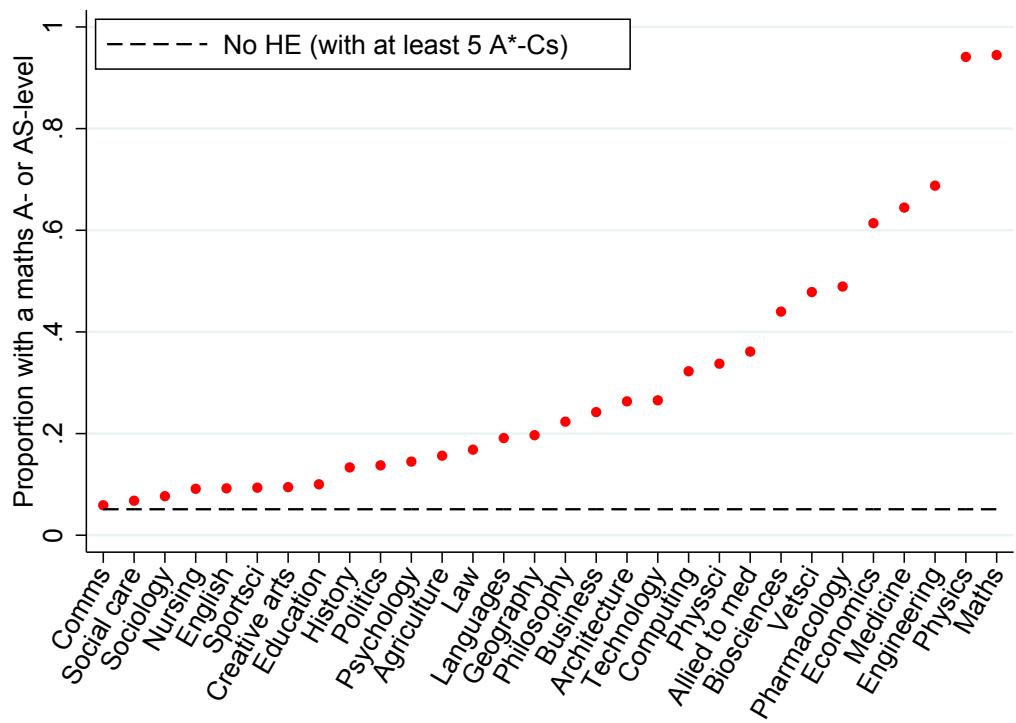


Note: Figures calculated based on those students who took GCSEs in England between 2002 and 2007 and achieved at least five A*-Cs.

Importantly, these differences between students are observable in the data, and there is considerable overlap between different subjects, institutions and those not attending HE (i.e. it is not the case that all HE entrants have a higher KS4 points score than those who do not attend). It is therefore possible to control for the effect of these characteristics on earnings when estimating the impact of university degrees. This highlights the value of the rich background characteristics observed in the NPD and HESA data we are using in this work.[14]

---

[14]Although note that the very low number of individuals with a maths A-level who do not attend HE highlights the value of the subgroup analysis we discuss later on.

Figure 15: Share of students studying each subject with a maths A- or AS-level, women
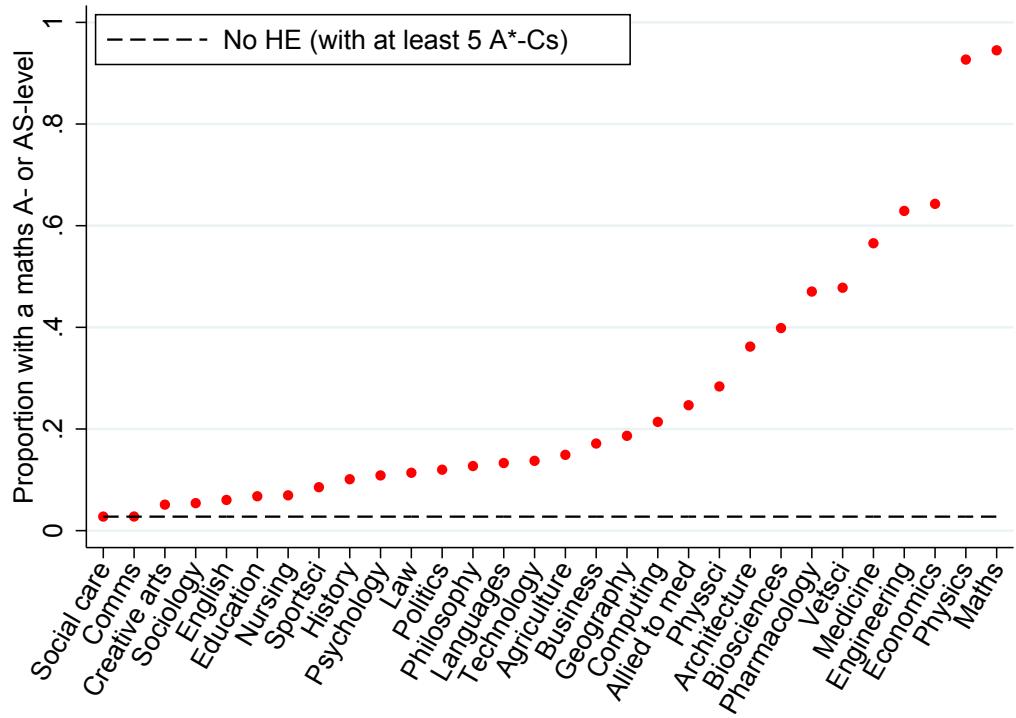


Note: Figures calculated based on those students who took GCSEs in England between 2002 and 2007 and achieved at least five A*-Cs.

The simplest approach for controlling for observed differences between students taking different courses is to control for observable characteristics using Ordinary Least Squares (OLS) estimation using a linear functional form, as in the following:

$$Y_i = X_i'\gamma + Course_i'\beta + \epsilon_{it} \tag{1}$$

where $Y_i$ is the outcome measure of earnings, $Course_i'$ is a vector of indicators for the course (subject-HEI combination)[15] studied and $X_i'$ is a vector of observable characteristics. In our estimation, $X_i'$ includes:

- prior attainment - measured by GCSE and A-level points score;[16]

- whether attended sixth form;

- A-level subject mix;[17]

- school type (independent or state school);[18]

---

[15]We provide more detail on the subject and HEI specifications below.

[16]GCSE and A-level measures are raw point scores, including all GCSEs and A-levels taken.

[17]We control for a set of indicators of whether students take an A-level in a given subject (maths, sciences, social science, arts, humanities, languages, other) and an indicator for whether they have taken a vocational qualification at 18. We do not control for the subject of the vocational qualification.

[18]We include a separate dummy for independent schools, but we cannot separately control for ethnicity and SES for individuals in independent schools.

- SES background;

- ethnicity;

- region of applicant;

- cohort of graduation;

- age started university.

$\gamma$ gives the impact of these background characteristics on earnings and the vector of coefficients $\beta$ shows the impact of degree courses on earnings. The omitted category here is 'not attending HE', so all estimates are relative to someone with similar characteristics who does not enter HE. Note that we intentionally do not include degree classification in the $X$ variables, as this is on the causal path. That is, an individual might be more likely to get a first-class degree if they choose one university or subject over another, and that should be reflected in our outcomes rather than adjusted for.

This methodology makes a number of important assumptions. First, there is a strong functional form assumption. The model implies that the characteristics in $X_i'$ have a linear impact on earnings. In practice, this may not be the case. Characteristics may have non-linear impacts or interactions between some characteristics may be important. This is always an issue when comparing groups that are observationally disparate.

A possible solution to this issue is to use a matching methodology, which only compares individuals in each group who look similar on observable characteristics. However, this is clearly impractical when considering a large number of treatment groups as in this analysis. Instead, we improve on the standard OLS estimation by using an Inverse Probability Weighted Regression Adjustment (IPWRA) following Wooldridge (2007). This process weights individuals in each treatment group (university-subject combination) who look more similar to individuals not in that treatment group more heavily. Hence, the final sample is weighted so that the treatment and control groups look similar. This is done by explicitly modelling the probability that each individual studies the course they take using observable characteristics and weighting individuals with the inverse of this probability. In practice, therefore, we use a slightly more elaborate methodology than the simple OLS regression shown in Equation (1). We elaborate further on the IPWRA methodology in the following subsection.

The second potential issue is that some of the differences between students that can affect earnings are not observed in the data, such as non-cognitive skills and preferences for different types of work. If these characteristics differ systematically between institutions or subjects then they might bias our estimates. For example, students who go to the London School of Economics might have a preference for working in finance jobs, which would have increased their earnings regardless of the university they attended. To the extent that this occurs, our estimates reflect the impact of the specific course combined with the effect of systematic differences in unobservable characteristics.

Third, this proposed methodology has implicitly assumed that the returns to specific courses are homogeneous - that is, that all types of students, on average, could be expected to receive the

same return from studying a given course (even if the levels of earnings differ due to differences in background characteristics). This is a strong assumption. For example, we might consider that a student with lower prior attainment or without a science A-level might get less benefit from a medicine or physics degree because they might understand less of the material. If the treatment effects are heterogeneous, our estimates will be some weighted average of the different treatment effects. This should be kept in mind when interpreting our overall results. We address this issue by additionally providing results for the six subgroups defined in Section 2.3. For this analysis, we re-estimate the our models completely separately, conditioning on being in the specific subgroups. The characteristics of these subgroups are described in detail in the previous section.

## 3.1 Inverse Probability Weighting First Stage

In our IPWRA methodology, we follow Belfield et al. (2018) and weight individuals by the inverse of the probability of them studying a specific subject at a given institution. Typically, in the existing applications of the method (e.g. Walker and Zhu, 2018) the IPWRA first stage is modelled using a single multinomial logit model. However, with more than 1,600 possible degree courses, this is computationally impractical. To circumvent this issue, we use a four step nested multinomial logit model. We first model the probability an individual enrols in HE as a function of observable characteristics $Z'$ (Step 0). For those that do enter HE, we then estimate the probability an individual studies each subject (Step 1).[19] Then, conditional on the subject they study we estimate the probability they attend one of four types of institution - Russell Group, Pre-1992, Post-1992 or Other (Step 2). Finally, conditional on the subject studied and type of university, we estimate the probability of attending each specific university (Step 3). This estimation is given by the following equations:

$$Pr(HE = Enrol_i) = \frac{1}{1 + e^{-Z_0'\beta_i^S}}; \tag{2}$$

$$Pr(Sub = s_j | HE = Enrol_i) = \frac{exp(Z_1'\beta_j^S)}{\sum_k exp(Z_1'\beta_k^S)} \quad \forall \quad k \neq j \in Sub; \tag{3}$$

$$Pr(UniType = UT_h | HE = Enrol_i, Sub = s_j) = \frac{exp(Z_1'\beta_h^T)}{\sum_i exp(Z_1'\beta_i^T)} \quad \forall \quad i \neq h \in UniType; \tag{4}$$

$$Pr(Uni = U_q | HE = Enrol_i, Sub = s_j, UniType = UT_h) = \frac{exp(Z_2'\beta_q^U)}{\sum_p exp(Z_2'\beta_p^U)} \quad \forall \quad p \neq q \in Uni. \tag{5}$$

Multiplying the predicted probabilities from these four equations gives the probability each individual studies the course that they do. The weights are then calculated as the inverse of this probability. As a result, an individual who looks more similar to those who do not take the same course is weighted more heavily. We drop any individuals with very extreme weights[20] (i.e. those

---

[19]The set of observable characteristics varies slightly between the steps of the nested logit. For Step 0, $Z'$ includes: KS4 attainment, SES, gender, ethnicity and school type. For Steps 1 and 2, $Z'$ adds KS5 attainment and subject choices. For Step 3, $Z'$ excludes subject choices as these are unlikely to be an important determinant of university conditional on subject and university type.

[20]We drop those with weights above the 95th percentile of the distribution. We also cap the bottom of the distribution

who had an extremely small probability of doing the course they study based on their observable characteristics) to avoid placing too heavy a weight on a small number of (potentially very unusual) students. In practice, we are interested in estimating the returns to HE at a number of different levels (overall returns and returns by subject, institution and course). We use the relevant IPWRA weights at each level. For example, when measuring the subject-level returns, we use the inverse of the probability an individual studies the subject they do; similarly, when measuring the returns to a given institution, we use the inverse of the probability an individual attends that institution (for any subject). These are simply different combinations of the probabilities derived from the above regressions.

## 3.2 Model of earnings

Our main results focus on the impact of university subjects, institutions and specific courses (institution-subject combinations) on earnings at age 29. We also investigate the sensitivity to estimating returns at other ages. This estimation focuses on those who are in *sustained employment*, defined as working five out of the last six months of the tax year. We include those who are in both sustained employment and further study. We exclude earnings data before age 24 (the first two years of earnings after graduation for HE entrants) as these are likely to be very noisy and not representative of the long-run returns to subjects or institutions. This means the final sample for our earnings model includes all individuals for whom we have data at age 24 or later, which means we include the cohorts that took their GCSEs between 2002 and 2007. Our estimates include those who drop out from university in the HE group, so we are measuring the expected impact of *enrolling* in HE on earnings.[21] We also report the estimates excluding dropouts to show the estimated return to *graduating* from university, rather than the return to *attending* university but not necessarily graduating.

We include observed earnings in a pooled cross-sectional model. This extends the model given by Equation 1 to include multiple earnings observations per individual. In doing so, we increase the efficiency of our estimates by reducing the randomness resulting from the transitory component of earnings. Our outcome is real log earnings between ages 24 and 29.[22] We only have Self-Assessment data between 2013-14 and 2015-16; hence including Self-Assessment earnings would mean that we would be unable to use earnings information from earlier years. This would severely reduce our sample size and hence reduce the reliability and precision of our estimates. As a result, our main estimates will exclude self-employment income, although we investigate the robustness of the results to this restriction.

We model log real earnings as a function of observable characteristics, $X_i'$, the treatment of interest ($Subject_i'$, $HEI_i'$ or $Course_i'$) and a treatment-specific quadratic time trend ($Subject_i' * f(t)$, $HEI_i' * f(t)$, or $Course_i' * f(t)$).[23] The time trend is included to allow different courses to have

---

at the 5th percentile.

[21]Dropouts are defined as individuals who never achieve an undergraduate qualification in the data. Those who switch course are counted as having taken the course in which they graduate (or the last course they study).

[22]Earnings are capped at the 1st and 99th percentiles to reduce the sensitivity to outliers.

[23]We add a quadratic time trend because we are modelling earnings right at the end of our sample period and we do not want our estimates to be sensitive to the functional form assumptions. Note this differs slightly from the method in Belfield et al. (2018) because in that work we were estimating the returns at a slightly earlier age.

different age-earnings profiles. We also interact KS5 attainment and subject choice variables with quadratic time trends to allow, for example, maths A-level to have an impact on earnings that grows over time. All results are estimated separately by gender. We estimate the following equations:

$$Y_{it} = X_i' \gamma + \omega_1 t + \omega_2 t^2 + HE_i' \alpha_1 + (HE_i * t)' \alpha_2 + (HE_i * t^2)' \alpha_3 + \epsilon_{it}; \tag{6}$$

$$Y_{it} = X_i' \gamma + \omega_1 t + \omega_2 t^2 + Subject_i' \delta_1 + (Subject_i * t)' \delta_2 + (Subject_i * t^2)' \delta_3 + \epsilon_{it}; \tag{7}$$

$$Y_{it} = X_i' \gamma + \omega_1 t + \omega_2 t^2 + HEI_i' \rho_1 + (HEI_i * t)' \rho_2 + (HEI_i * t^2)' \rho_3 + \epsilon_{it}; \tag{8}$$

$$Y_{it} = X_i' \gamma + \omega_1 t + \omega_2 t^2 + Course_i' \beta_1 + (Course_i * t)' \beta_2 + (Course_i * t^2)' \beta_3 + \epsilon_{it}. \tag{9}$$

In order to avoid results being masked by the early-career volatility in earnings, we use returns at age 29 as our main outcome of interest. From the above equations:[24]

- **overall returns** are given by $\alpha_1 + 5\alpha_2 + 25\alpha_3$;

- **subject returns** are given by $\delta_1 + 5\delta_2 + 25\delta_3$;

- **HEI returns** are given by $\rho_1 + 5\rho_2 + 25\rho_3$;

- **course returns** are given by $\beta_1 + 5\beta_2 + 25\beta_3$ .

It is important to note that, unlike in Belfield et al. (2018), we do not control for the effects of subject studied when estimating the institution returns and vice versa. This is because the control group - not attending HE - is perfectly collinear for each set. This slightly changes the interpretation of our estimates; the institution coefficients are the average return to studying at that institution given the set of subjects they offer - rather than holding subject studied fixed.

## 4 Overall results

We now turn to our results. We start by estimating the average impact of attending HE on earnings at age 29 compared with not attending (for individuals with at least five A*-C GCSE grades), and then show how this impact varies for different types of students. Sections 5, 6 and 7 will show how this average return is broken down into the return to different subjects, institutions and courses (subject-institution combination) respectively, before Section 8 explores how these returns are likely to evolve over the lifecycle.

### 4.1 Average overall returns

Table 7 shows our main estimates of the impact of attending HE on earnings at age 29, in log points.[25] Column 1 displays the raw differences in average earnings between those who do and do not attend HE. As shown in Figure 1, individuals who attend HE earn considerably more at

---

[24]Note that the time trend is set to zero at age 24.

[25]Estimates of the coefficients on the background characteristics we include ($X_i'$) are shown in the Online Appendix Tables 9 and 10.

age 29 on average: men who attend HE earn 0.19 log points (21%) more than men who do not, while the equivalent figure for women is 0.44 log points (55%). In Column 2, we control for the age students start university and report the earnings differential between those who start university at age 18 and those who do not go. This increases the estimated earnings difference because those who start university later typically earn less at age 29 due to having less post-HE work experience. All the estimates in this report will control for the age students start their course and so should interpreted as the *impact of attending HE at age 18 on earnings at 29*.[26]

Some of this gap in earnings is a result of HE and non-HE students differing in background characteristics that affect their earnings potential. We know, for example, that HE students typically have higher prior attainment and are more likely to come from better-off families. Columns 3 and 4 add controls for background characteristics and prior attainment respectively. Controlling for different background characteristics such as socio-economic status, ethnicity and region does reduce the estimated return to HE, but only slightly. Controlling for prior attainment, however, has a dramatic effect on the estimated returns for both men and women. Finally, Column 5 applies the IPWRA weights discussed in Section 3, which has only a limited impact on the estimates. The final estimates, in Column 5, imply that attending HE increases earnings at age 29 by 6% for men and around 26% for women.[27]

Table 7: Overall returns to HE at age 29

|  | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Men | 0.19*** | 0.25*** | 0.22*** | 0.04*** | 0.06*** |
|  | (0.00) | (0.00) | (0.00) | (0.01) | (0.00) |
|  |  |  |  |  |  |
| *No. of observations* | *2,183,120* | *2,183,120* | *2,183,120* | *2,183,120* | *2,183,120* |
| *No. of individuals* | *629,138* | *629,138* | *629,138* | *629,138* | *629,138* |
|  |  |  |  |  |  |
| Women | 0.44*** | 0.50*** | 0.46*** | 0.23*** | 0.23*** |
|  | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
|  |  |  |  |  |  |
| *No. of observations* | *2,619,982* | *2,619,982* | *2,619,982* | *2,619,982* | *2,619,982* |
| *No. of individuals* | *731,200* | *731,200* | *731,200* | *731,200* | *731,200* |
|  |  |  |  |  |  |
| Cohort/Age start controls | No | Yes | Yes | Yes | Yes |
| Background charcteristics | No | No | Yes | Yes | Yes |
| Prior attainment | No | No | No | Yes | Yes |
| IPWRA weight | No | No | No | No | Yes |

Note: Table reports derived estimates of the overall impact of HE on annual earnings at age 29 based on the 2002-2007 GCSE cohorts, conditional on at least five A*-C GCSEs and on being in sustained employment. Table sequentially adds age, background and prior attainment controls, and finally IPWRA weights. Estimates are in log points, which can be converted into percentage points using the transformation $100 * (e^x - 1)$, where $x$ is the log points estimate.

The finding that returns to HE are much higher for women than for men echoes the existing literature. However, there are three important caveats to bear in mind when interpreting these re-

---

[26]Controlling for individuals starting after age 18 may be particularly important when looking at subject and institution returns as some subjects or institutions may be more likely than others to have students who did not start straight out of school. We remove this effect from the impact on earnings.

[27]As discussed in Sectio3.2 we exclude earnings from self-employment from our analysis because we only observe this variable for three years. In Table A5 in the Appendix, we show that when we restrict the sample to this subset of years and compare the results with and without earnings from self-employment, the estimates are almost identical.

sults. First, these estimates show the impact on annual earnings at age 29 for all individuals who are in 'sustained employment', including both those who work full-time and those who work part-time.[28] Evidence from the Labour Force Survey suggests that around 40% of non-graduate working women work part-time in their late 20s, compared with only 20% of the graduate women in work. Some of the large impact of HE on earnings we find is hence likely to result from graduate women working more hours, rather than purely being an effect on hourly wages. Increased hours may be a causal impact of attending HE if, for example, HE increases women's full-time job opportunities. However, there may also be unobserved factors that affect both the likelihood of attending HE and the likelihood of working, or the likelihood of working more hours: for example, women with a strong desire to work full-time may be more likely to attend HE. This has the potential to bias our estimates of the earnings return to HE upwards.

Second, even assuming that this impact on hours is causal, it is important to recognise that increased working hours may represent a trade-off. For example, while longer working hours may result in higher earnings, it may also be related to delaying child-bearing - it is well established that highly educated women are more likely to have children later in life. While in this work we are unable to look at other outcomes such as hours and child-bearings, it is important to recognise earnings are only one of many outcomes that HE can affect.[29]

Our estimates for men are somewhat lower than many of the results in the existing literature on the returns to HE. Blundell et al. (2000), for example, estimate the return to HE for men to be around 14%. One of the primary reasons for this difference is that we are observing individuals earlier in the lifecycle - Blundell et al. (2000) estimate the effect on earnings at age 33.[30] At the earlier age of 29, the impact of lost work experience while at university will be relatively more important. Furthermore the earnings growth of HE participants is stronger than that of those who did not attend HE, as indicated by Figure 1. We address these issues further in Section 8.

Another important difference between our estimates and much of the existing literature is that we estimate the return to *entering* HE, rather than *graduating* from HE. That is, individuals who enter HE but drop out before graduation are included in our treatment group. Estimating the impact of entering HE provides students who are deciding whether to go to university with vital information on the potential impact of this decision. Similarly, when evaluating universities' performance, it is important that they are accountable for the outcomes of all their students, not merely those who graduate. Previous work has mostly used survey datasets, which often record qualifications obtained and hence cannot identify dropouts. In our data, we are able to identify both graduates

---

[28]Our data do not contain hours worked so we are not able to distinguish full and part-time workers or estimate the impact on hourly wages.

[29]This labour supply channel may have differential impacts at different points in the lifecycle. In fact, recent work by Belfield and van der Erve (2018) shows that HE has a large impact on women's labour supply in their early 30s but this effect dissipates by their early 40s.

[30]There are likely to be a number of other reasons. First, our graduates have entered the labour market during the recession and earnings growth has been negligible for much of the period that we observe them. Whilst the recession has affected non-graduates too, we need to be mindful that this might be a particularly difficult period for graduates entering the labour market. Another key issue is that Blundell et al. (2000) estimate returns for individuals born in 1958, who went through a very different educational system from more recent cohorts, and had a much lower HE participation rate. Further, as discussed, we condition on being in 'sustained employment'. Finally, our data allows us to control, with a higher degree of accuracy and granularity than hitherto possible, for differences in individuals' pre-HE academic achievement. This may provide a partial explanation as to why our estimates of the return to a degree are lower for men than previous literature suggests.

and dropouts, which provides a unique opportunity to estimate the return to entering university.

As we saw in Figure 2, those who drop out from HE typically earn significantly less than those who graduate. As a result, we would expect estimates of the impact of graduating from university to be larger than the impact of merely attending university. Table 8 shows how our estimates change under different assumptions about the treatment and control groups. Column 2 shows the impact of graduating from HE (relative to not attending) on earnings at age 29 (Column 1 is a repeat of the main specification results for comparison). Focusing on graduates increases the returns for men from 0.06 to 0.08 log points (6 to 8%), a considerable increase when we consider that only a small proportion of HE entrants drop out. Similarly, the return for women increases from 0.23 to 0.25 log points (26 to 28%).

Table 8: Overall returns with alternative sample definitions

| Control group:<br>Treatment: | 5+ A*-C<br>(all) | 5+ A*-C<br>(grads) | 1+ A-level<br>(all) | 1+ A-level<br>(grads) | Dropouts<br>(grads) |
|---|---|---|---|---|---|
| Men | 0.06*** | 0.08*** | 0.04*** | 0.08*** | 0.16*** |
| | (0.00) | (0.00) | (0.01) | (0.00) | (0.01) |
| | | | | | |
| *No. of observations* | *2,183,120* | *2,051,067* | *1,594,834* | *1,476,860* | *1,468,352* |
| *No. of individuals* | *629,138* | *593,974* | *433,559* | *403,055* | *397,276* |
| | | | | | |
| Women | 0.23*** | 0.25*** | 0.19*** | 0.21*** | 0.22*** |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| | | | | | |
| *No. of observations* | *2,619,982* | *2,502,474* | *1,992,425* | *1,889,447* | *1,909,839* |
| *No. of individuals* | *731,200* | *700,553* | *525,063* | *499,156* | *502,308* |

Note: Table reports derived estimates of the overall impact of HE on annual earnings at age 29 based on the 2002-2007 GCSE cohorts, conditional on being in sustained employment. Column 1 repeats the results in the last column of Table 7. Column 2 excludes anybody who drops out from HE before graduating (estimating the impact for graduating versus not attending). Columns 3 and 4 repeat the analysis on the sample of individuals who have at least one A-level. Column 5 uses the same methodology but uses dropouts as a control group for graduates. Estimates are in log points.

A further consideration when estimating the returns to HE is which control group to use. The choice of control group is important as it determines the set of individuals for whom the estimates apply and the range of counterfactual non-HE outcomes that are considered. As already discussed, in our main estimates we limit our sample to only include individuals with at least five GCSEs at grades A*-C. The vast majority (around 93%) of individuals who enter HE have at least five A*-C GCSEs, so level of prior attainment can be seen as a prerequisite for entering. Attending HE is essentially not a relevant option for those who do not have at least five A*-C GCSEs and hence they are excluded from the analysis.

In Table 8, we consider two alternative controls that have been used in the literature: individuals who have at least one A-level and individuals who drop out from HE. Columns 3 and 4 repeat the estimation in Columns 1 and 2 using those with at least one A-level as the control group, estimating on a treatment group of HE entrants and graduates respectively. This slightly reduces the estimated return of HE, likely due to the higher educational outcomes of the control group, but it does not significantly alter our conclusions. Using individuals who drop out of HE as the control group, on the other hand, does significantly impact our results (for men at least). Column 5 shows that when using drop outs as a control group the return to graduating from HE is estimated to

be 0.16 log points (17%) for men. However, for the reasons set out above we do not believe this provides a robust estimate of the impact of HE.

We do not use these definitions for a number of reasons. First, the number of individuals who enter HE without any traditional A-level qualification has increased in recent years, and so excluding this group misses an important population for which HE is a relevant option. Second, dropping out of HE is not a random event. As a result, there will likely be unobserved differences between dropouts and graduates that will also affect earnings and will hence bias the estimates of the return to HE. Third, dropping out of HE may present a signal to employers of the 'quality' of potential employees and so directly impact their earnings. Fourth, individuals who enter HE and drop out receive some of the treatment; that is, they may have gained some of the human capital that HE provides before dropping out or they may have lost labour market experience while in HE. As a result, using dropouts as a control group will not give the true return to graduating from HE. We therefore prefer to use the results shown in Column 5 of Table 7 as our main specification.

## 4.2   Heterogeneity in overall returns by background characteristics

The analysis up to this point has focused on the average return to attending HE for all individuals. Tables 9 and 10 investigate how this return varies across different types of student, based on their prior attainment and A-level choices (see Section 2.3).[31]

Table 9 shows our estimated returns for the six groups for men. We see that for those without a STEM A-level, the returns dramatically increase with the individual's prior attainment level. For the lowest attainment group (which still conditions on having at least five A*-C GCSE grades), attending HE appears to only increase earnings at age 29 by 0.04 log points (4%), while for the higher prior attainment group the impact is 0.18 log points (20%).

This finding of low returns for students with lower prior attainment without a STEM A-level is particularly important when considering expansion of the HE system. Across all other prior attainment groups, nearly two-thirds of students already attend HE, while in this group only one-third of students currently attend university. If HE participation were to increase further, it is likely that a significant proportion of the additional students would come from this prior attainment group.

For students with a STEM A-level there is no equivalent pattern of higher returns for higher-prior attainment students. The lower and middle prior attainment groups have returns of 0.10 and 0.09 log points (11 and 9% respectively), while in fact the higher prior attainment group has a return of 0.05 log points (5%). This is perhaps surprising, but it should be noted that the control group in this case is quite small, and probably quite unusual. However, it is probably true that those with high prior attainment and a STEM A-level have very good labour market opportunities even if they do not go to HE.

---

[31]These heterogeneity estimates come from the estimation of six separate specifications. As a result, the weighted average of the estimates does not necessarily align with the overall average.

Table 9: Overall returns at age 29 by student type, men

| | | Prior attainment | | |
| --- | --- | --- | --- | --- |
| | | Lower | Middle | Higher |
| No STEM A-level | Returns to HE | 0.04*** | 0.08*** | 0.18*** |
| | | (0.01) | (0.01) | (0.02) |
| | | | | |
| | *No. of observations* | *763,319* | *432,219* | *180,351* |
| | *No. of individuals* | *234,614* | *124,441* | *49,498* |
| | | | | |
| With STEM A-level | Returns to HE | 0.10*** | 0.09*** | 0.05*** |
| | | (0.02) | (0.01) | (0.01) |
| | | | | |
| | *No. of observations* | *69,165* | *247,562* | *406,871* |
| | *No. of individuals* | *19,276* | *67,659* | *112,231* |

Notes: Table reports derived estimates of the overall impact of HE on annual earnings at age 29 based on the 2002-2007 GCSE cohorts, conditional on being in sustained employment. Estimates are done separately for the different subgroups. 'Lower', 'middle' and 'higher' prior attainment correspond respectively to the bottom, middle and top tercile of KS4 points in the sample of those with at least five A*-C GCSE grades.


Table 10: Overall returns at age 29 by student type, women

| | | Prior attainment | | |
| --- | --- | --- | --- | --- |
| | | Lower | Middle | Higher |
| No STEM A-level | Returns to HE | 0.21*** | 0.22*** | 0.27*** |
| | | (0.01) | (0.01) | (0.01) |
| | | | | |
| | *No. of observations* | *783,092* | *706,047* | *420,384* |
| | *No. of individuals* | *234,694* | *194,351* | *112,073* |
| | | | | |
| With STEM A-level | Returns to HE | 0.20*** | 0.15*** | 0.21*** |
| | | (0.03) | (0.01) | (0.01) |
| | | | | |
| | *No. of observations* | *24,549* | *161,501* | *430,293* |
| | *No. of individuals* | *6,651* | *42,766* | *119,244* |

Notes: Table reports derived estimates of the overall impact of HE on annual earnings at age 29 based on the 2002-2007 GCSE cohorts, conditional on being in sustained employment. Estimates are done separately for the different subgroups. 'Lower', 'middle' and 'higher' prior attainment correspond respectively to the bottom, middle and top tercile of KS4 points in the sample of those with at least five A*-C GCSE grades.

For women, Table 10 shows high returns for all groups, although similar patterns are present: amongst those without a STEM A-level, the returns increase with the level of prior attainment, while there is no clear pattern amongst those who do have a STEM A-level.

These differences raise the question of what is driving these returns. It could be that different types of students benefit differently from HE, or it could be that the non-HE alternative (the outside option) differs for these different groups. Alternatively, these students may study different subjects or attend different institutions when in HE. The estimation in Tables 9 and 10 does not

control for subject studied or institution attended, showing instead the average effect of HE on these groups given the set of courses they take. We explore the impact of subject and institution choice in more detail in the following sections.

Grouping students by prior attainment is only one of many ways we could break down the average returns. Another way is to split students by their family background. Table 11 divides students into those who went to private school and, amongst those who went to state school, quintiles of the measure of socio-economic status introduced in Section 2. For women, all SES groups have broadly similar returns. However, for men, it is clear that the returns are highest for those students who went to private school. These individuals have an average return of 0.16 log points (17%), compared with between 0.04 and 0.09 log points (4% and 9%) for state school students. Again these different returns could be driven by differential impacts of HE or by different subjects taken or institutions attended.

Table 11: Overall returns at age 29 by SES background

|  | Ind school | High SES | 2 | 3 | 4 | Low SES |
|---|---|---|---|---|---|---|
| Men | 0.16*** | 0.07*** | 0.04*** | 0.06*** | 0.04*** | 0.09*** |
|  | (0.02) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
|  |  |  |  |  |  |  |
| *No. of observations* | *276,848* | *611,503* | *484,806* | *382,782* | *261,550* | *165,631* |
| *No. of individuals* | *76,210* | *173,603* | *140,211* | *111,571* | *77,511* | *50,032* |
|  |  |  |  |  |  |  |
| Women | 0.23*** | 0.21*** | 0.24*** | 0.21*** | 0.23*** | 0.21*** |
|  | (0.02) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
|  |  |  |  |  |  |  |
| *No. of observations* | *294724* | *702503* | *582209* | *474315* | *338683* | *227548* |
| *No. of individuals* | *79728* | *192071* | *162013* | *133796* | *97072* | *66520* |

Note: Table reports derived estimates of the overall impact of HE on annual earnings at age 29 based on the 2002-2007 GCSE cohorts, conditional on being in sustained employment. Estimates are done separately by group. SES quintile is based on the whole NPD sample, by cohort (hence the variation in sample sizes).

# 5 Subject results

## 5.1 Overall returns by subject

The previous section showed the average impact of attending HE. In this section, we break this down to look at the earnings returns to studying a degree in a specific subject area, relative to not attending HE. Figures 16 and 17 show the raw percentage earnings difference between HE and the five A*-C non HE group at age 29 (shown in black). The subjects are ranked by these raw differences.[32]

The figures also show in red our estimates of the estimated impact these subjects have on earnings at age 29 once we account for differences in background characteristics (such as prior attainment and socio-economic status) and implement the IPWRA. In all cases, these estimated

---

[32]Differences in the rank ordering from Figures 7 and 8 arise because these raw earnings differences are estimated using a model including several GCSE cohorts, controlling for the age individuals started their course.

returns are lower than the raw differences. This is because, across all subjects, the background characteristics of HE students - who are more likely to have higher prior attainment and to come from more privileged backgrounds - mean that, on average, we would expect them to earn more than non-HE students regardless of whether they went to HE.

Figure 16: Raw earnings differences and estimated returns at age 29 by subject, men
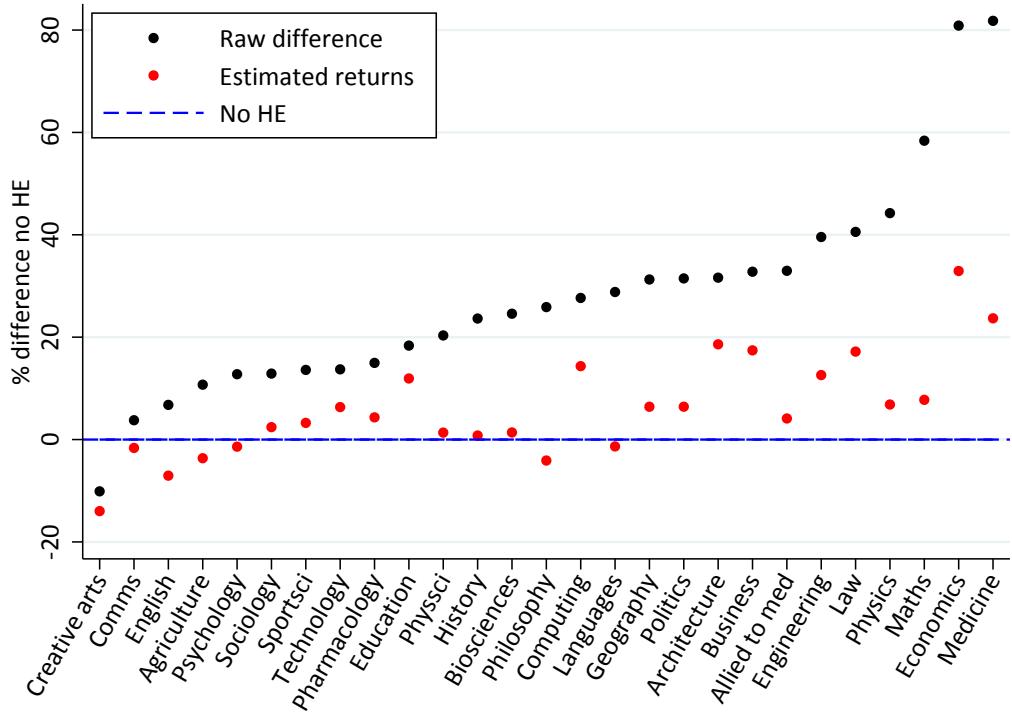


Note: Figure reports derived estimates of the impact of studying different subjects on annual earnings at age 29 based on the 2002-2007 GCSE cohorts, conditional on being in sustained employment. Raw differences only account for age at which course, while estimated returns control for age, background and prior attainment and use IPWRA weights. Subjects are defined using the CAH2 definition. Results have been converted to percentage differences using a log-point conversion. Subjects are ranked based on raw earnings differences.

The difference between the raw and estimated returns highlights how important it is to control for observable differences between HE and non-HE students when estimating returns. It is also clear that controlling for these characteristics does not affect all subjects equally. The returns to very selective subjects such as medicine and economics are reduced by over 40 percentage points, much more than the returns to creative arts and communications degrees, which are only reduced by around 5 percentage points. This is because the differences in background characteristics between those taking medicine degrees and those not attending HE are far greater than those between creative arts students and those not attending HE.

Once we control for observable differences between students, there remains significant variation in the impact of different degree subjects on earnings. For men, studying economics increases earnings at age 29 by 33% on average, while studying creative arts actually reduces earnings by an average of 14% relative to not attending university. For women, studying medicine increases

earnings by 75%, compared with a 9% increase for those studying creative arts. These results illustrate that the choice of which subject to study at university is at least as important as the decision of whether to attend at all.[33]

Figure 17: Raw earnings differences and estimated returns at age 29 by subject, women
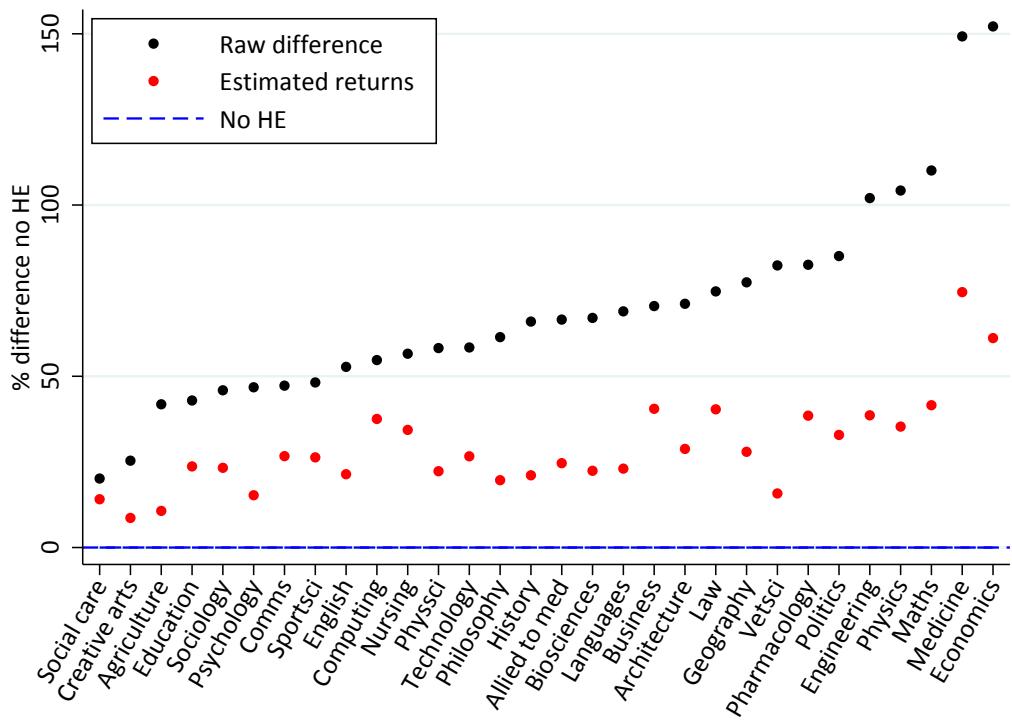


Note: Figure reports derived estimates of the impact of studying different subjects on annual earnings at age 29 based on the 2002-2007 GCSE cohorts, conditional on being in sustained employment. Raw differences only account for age at which course, while estimated returns control for age, background and prior attainment and use IPWRA weights. Subjects are defined using the CAH2 definition. Results have been converted to percentage differences using a log-point conversion. Subjects are ranked based on raw earnings differences.

For men, a number of subjects exhibit negative and very low returns at age 29. We show that creative arts, English and philosophy have a statistically significant negative impact on age 29 earnings, while agriculture, communications, psychology, languages, history, biosciences (which includes chemistry) and physical sciences have an effect that is not statistically significantly different from zero.[34] For several of these subjects, the finding of a zero return is particularly interesting because their raw earnings are more than 20% higher than for the non-HE group. This again highlights how different the conclusions that might be drawn from the data are once the composition

---

[33]These estimates are compared with the 'relative returns' estimates of Belfield et al. (2018) in Appendix Figures A7 and A8, showing that, shows despite a number of methodological differences, the estimates are very highly correlated. In addition, Figures A3 and A4 compare the estimates with a model that incorporates Self-Assessment tax data (but based on fewer tax years) and show this generally does not significantly impact the majority of our estimates. The exception is that the inclusion of Self-Assessment data increases the estimated return to medicine degrees by around 10 percentage points, likely due to doctors having very good self-employment outcomes (e.g. GPs) or receiving self-employed income in addition to PAYE earnings.

[34]The estimated returns by subject are provided with confidence intervals (and re-ordered based on the returns) Figures A3 and A4 in the Appendix.

of students taking a certain course is adjusted for.

As discussed above, one should note that these negative and low returns are likely to be at least in part the result of lost work experience while at university; we discuss this in more detail in Section 8. Furthermore, some students studying these subjects may deliberately select into certain occupations which are low-paying but yield other benefits such as job satisfaction.

## 5.2 Heterogeneity in subject returns

In Section 4.2 we showed that different types of students experience very different returns to HE. In particular, higher prior attainment men without a STEM A-level have significantly higher returns on average than lower prior attainment men without a STEM A-level.

Given the large variation in returns across subjects that we have just shown, the large differences across the different subgroups could possibly be driven by different subject choices. Indeed, we know from Section 2.3 that the six groups do make very different choices. However, referring back to Table 5 with more knowledge of how returns vary by subject, it does not appear to be the case that the lower returns are entirely driven by different subject choices. In fact, amongst men without a STEM A-level, those with lower prior attainment are more likely than those with higher prior attainment to study business or computing degrees, both of which are subjects with quite high returns. We actually even see that higher prior attainment students without a STEM A-level are more likely to study subjects with low returns such as history and languages. Hence, the differences in returns between students with different prior attainment levels is likely to be driven by HE having different impacts on the earnings of different types of students.

In Figures 18 to 21, we investigate subject-specific returns for each of our six subgroups, separately by gender. In several cases - for example, maths and physics degrees for those without a STEM A-level - an insufficient number of individuals take the subject to enable estimation, and consequently we do not provide an estimate in these cases.

For men without a STEM A-level, Figure 18 clearly shows that the returns to nearly every degree subject are the largest for higher prior attainment individuals. Notably, there is no subject, other than creative arts, with a negative impact on earnings at age 29 for the higher prior attainment group.

On the other hand, there are several subjects that yield zero or negative returns by age 29 for the lower prior attainment group (without a STEM A-level). The contrast with the higher prior attainment group is particularly stark for philosophy, history and biosciences: overall, studying these courses has virtually no impact on men's earnings at age 29, but when we break this down, we observe large positive returns for those with higher prior attainment and statistically significant *negative* returns for those with lower prior attainment. What is driving these differences remains unclear. On the one hand, it may be that higher prior attainment men studying history actually learn more or use this knowledge more effectively, or that higher prior attainment men are more likely to be studying history at more selective universities that have higher returns. This second point is addressed directly in the following section.

Figure 18: Subject returns at age 29 by prior attainment (no STEM A-level), men
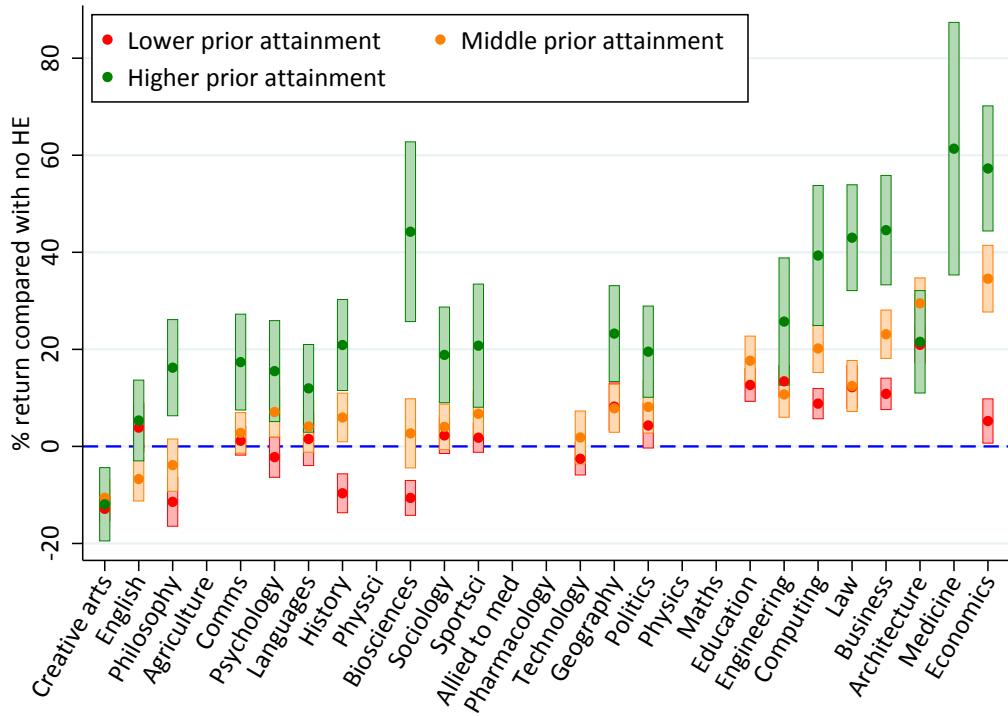


Note: Figure reports derived estimates of the impact of studying different subjects on annual earnings at age 29 based on the 2002-2007 GCSE cohorts, conditional on being in sustained employment. The bars show 95% confidence intervals. Prior attainment is based on GCSE results. Estimates are done separately by subgroup and are only shown for groups where the sample size is sufficient for estimation. Estimates are converted to percentage differences using a log-point conversion. Subjects are ranked on estimated overall subject returns for men.

Figure 19 shows the equivalent figures for higher, medium and lower prior attainment men *with* a STEM A-level. Here there is a less clear pattern of returns across the prior attainment groups, reflecting the results in Table 9 that, on average, students with a STEM A-level have more similar returns to HE across prior attainment groups. However, it is striking that, particularly for higher prior attainment men with a STEM A-level there are several subjects with negative impacts on earnings at age 29. These are typically humanities subjects such as English, communications and languages. It therefore appears that for those with high prior attainment and a STEM A-level that some arts, humanities and social science degrees are not necessarily a good option in terms of earnings (although economics, business and law degrees are notable exceptions), as their outside option of not going to HE is so strong.

It is worth noting that to some extent higher prior attainment men with a STEM A-level seem to select into the subjects that give them high returns, such as business, engineering and economics. However, there does not seem to be perfect self-selection based on high returns: around 10% of all men with a STEM A-level take biosciences degrees, which have a return close to zero. This acts to lower the overall returns for these groups.

Figure 19: Subject returns at age 29 by prior attainment (with STEM A-level), men
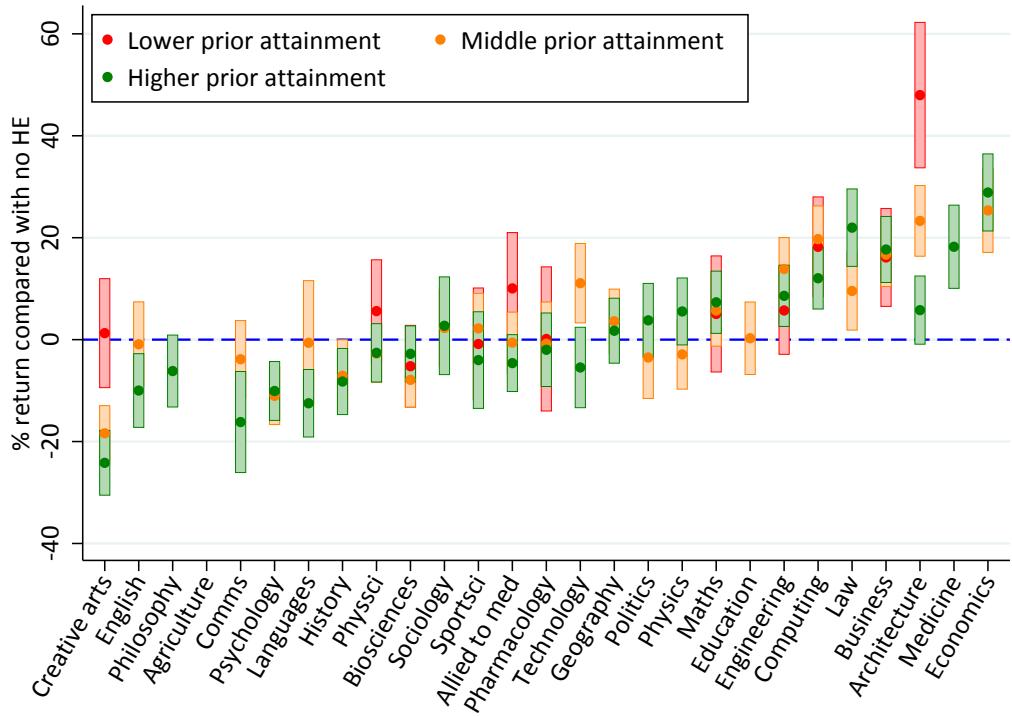


Note: Figure reports derived estimates of the impact of studying different subjects on annual earnings at age 29 based on the 2002-2007 GCSE cohorts, conditional on being in sustained employment. The bars show 95% confidence intervals. Prior attainment is based on GCSE results. Estimates are done separately by subgroup and are only shown for groups where the sample size is sufficient for estimation. Estimates are converted to percentage differences using a log point conversion. Subjects are ranked on estimated overall subject returns for men.

The equivalent results for women are in Figures 20 and 21. They are quite different. First, across all student group types, there are very few subjects that have negative returns for any prior attainment group. Second, Figure 20 shows there is a less clear pattern of certain subjects having substantially different returns for different types of students, with the exception of the very highest return subjects (medicine, economics, business and law) which have substantially higher returns for higher prior attainment students.

Unlike men, women with lower prior attainment without a STEM A-level do appear slightly more likely to select into subjects with lower returns. Lower prior attainment women are more likely to study social care, sociology and education than their higher prior attainment counterparts, although they are still much more likely to study business degrees, which have high average returns (see Appendix Table A4).

When we look at women with a STEM A-level, as for men the returns to humanities subjects such as English, history, sociology and languages are low, although in the vast majority of cases these returns are still positive. Taking these subjects still gives higher earnings at age 29 than not attending HE.

Figure 20: Subject returns at age 29 by prior attainment (no STEM A-level), women
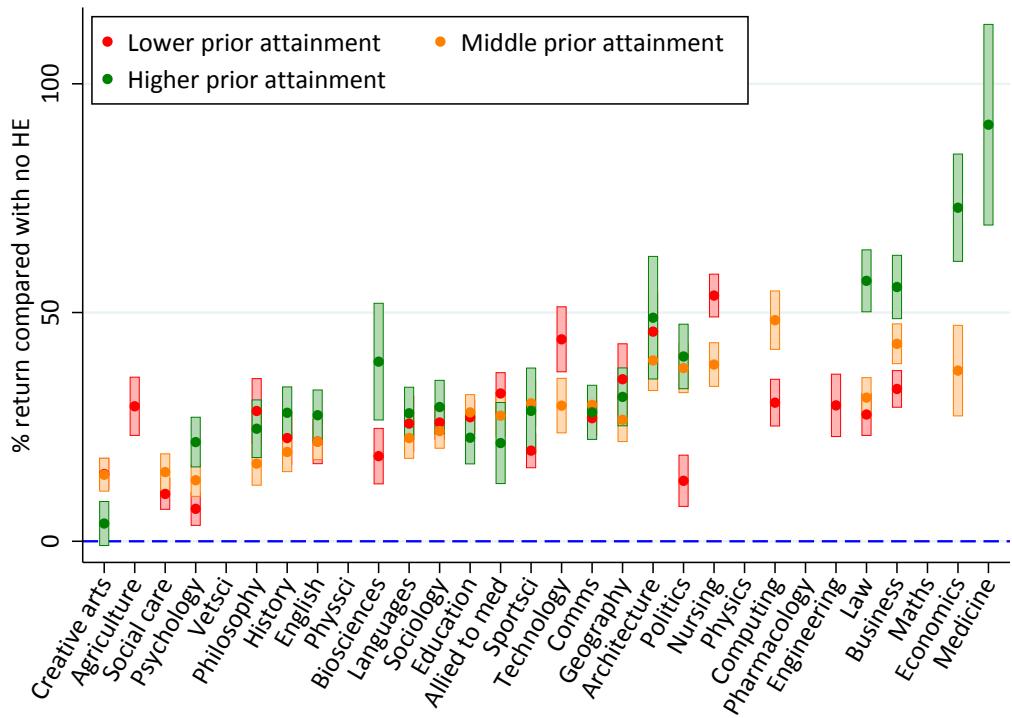


Note: Figure reports derived estimates of the impact of studying different subjects on annual earnings at age 29 based on the 2002-2007 GCSE cohorts, conditional on being in sustained employment. The bars show 95% confidence intervals. Prior attainment is based on GCSE results. Estimates are done separately by subgroup and are only shown for groups where the sample size is sufficient for estimation. Estimates are converted to percentage differences using a log point conversion. Subjects are ranked on estimated overall subject returns for women.

These estimates of the returns to specific subjects for different types of individual provide valuable information to students when making their subject choice decision. Students know their prior attainment level and whether they have a STEM A-level, so these figures provide estimates on the impact of specific subjects for individuals like them.

It is also clear that not all subjects are available options for all students. There are a large number of science-based subjects that individuals without a STEM A-level will be excluded from, while only the very highest prior attainment students will be able to take medical degrees. This is important, because although medicine is the highest- or second-highest-returning subject overall, it is not an option to most students.

Figure 21: Subject returns at age 29 by prior attainment (with STEM A-level), women
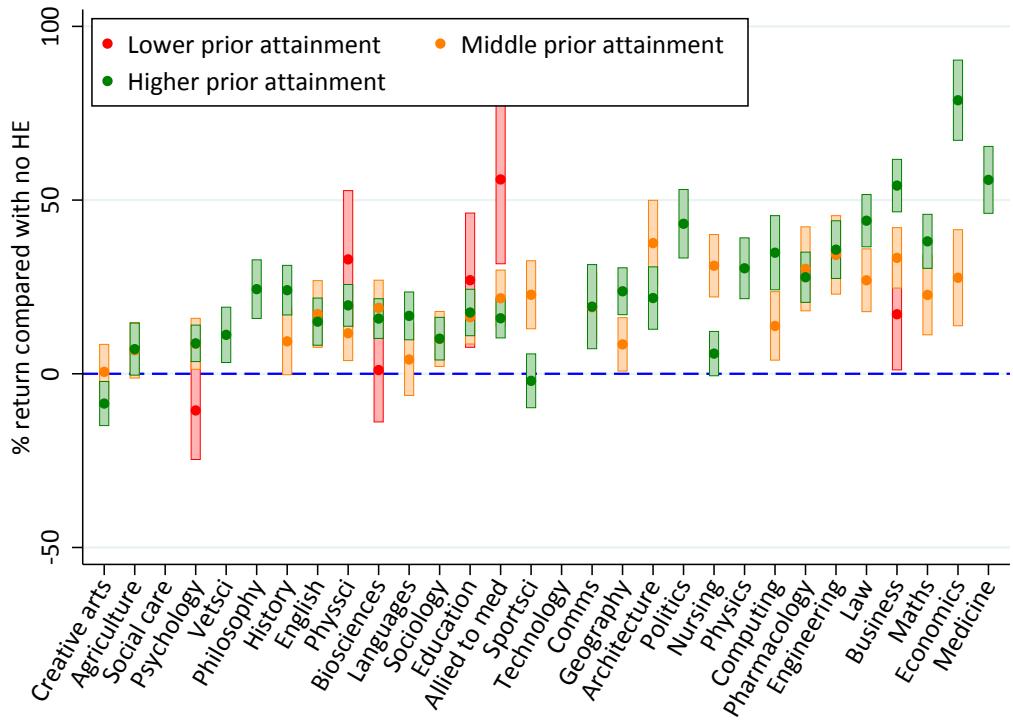
Note: Figure reports derived estimates of the impact of studying different subjects on annual earnings at age 29 based on the 2002-2007 GCSE cohorts, conditional on being in sustained employment. The bars show 95% confidence intervals. Prior attainment is based on GCSE results. Estimates are done separately by subgroup and are only shown for groups where the sample size is sufficient for estimation. Estimates are converted to percentage differences using a log point conversion. Subjects are ranked on estimated overall subject returns for women.

Instead our results show that for any subject that has negative returns on average, there is another subject which similar students take that has positive returns. For example, men with lower prior attainment without a STEM A-level experience a negative return on average if they study philosophy or history, but these students are also likely to have had the option to study business or politics (as many similar students do) and these subjects exhibit positive returns. Similarly, men with higher prior attainment and a STEM A-level experience negative returns if they take many humanities subjects, but many similar students instead take engineering, medicine or economics, which have very high returns. Finally, it is important to reiterate that here we solely focus on the impact of a subject on earnings, and students may have many other reasons for choosing to study a specific subject.

# 6 HEI results

## 6.1 Overall returns by HEI

In this section, we present estimates of the return to studying a degree at different Higher Education Institutions (HEIs), comparing the earnings of graduates from a particular HEI with similar

individuals who did not attend HE at all. As discussed in Section 3, this estimation does not control for the subject students take at the given university. This means the estimates report the average impact of taking any degree at the university (rather than the returns if a student took the overall average degree at the university).[35] Figure 22 compares the raw differences in earnings between those attending different institutions and non-HE students (in black), and the estimated returns once background characteristics are controlled for (in red) for men and Figure 23 plots the equivalent estimates for women.

Figure 22: Raw earnings differences and estimated returns at age 29 by HEI, men
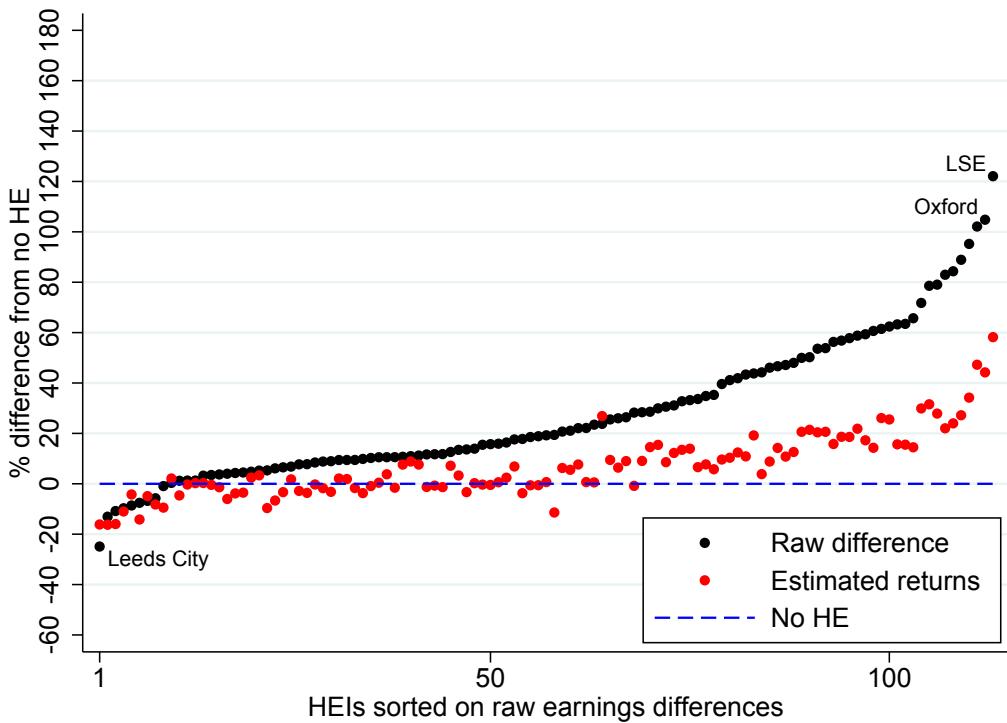


Note: Figure reports derived estimates of the impact of studying at different institutions on annual earnings at age 29 based on the 2002-2007 GCSE cohorts, conditional on being in sustained employment. Raw differences only account for age at which course started, while estimated returns control for age, background and prior attainment and use IPWRA weights. Results have been converted to percentage differences using a log-point conversion. HEIs are ranked based on raw earnings differences.

As with the subject returns estimated in the previous section, once background characteristics are controlled for the estimated return to nearly every institution falls. This means that, at the vast majority of HEIs, the differences in background characteristics between their students and those who did not go to HE are such that we would expect their students to earn more even if they had not gone to university.

For many universities, background characteristics can account for over half of the earnings gap between those attending specific universities and those not attending HE. However, there are

---

[35]One implication of this is that if a university only offered high-returning subjects, this might increase their estimated return. The benefit is that our estimates accurately represent the average impact of a university on its students even if this is due to the courses it chooses to offer.

a small number of institutions whose returns increase for men when accounting for background characteristics (although not significantly), implying students at these institutions are expected to earn less than who do not go to HE at all based on their characteristics prior to entry.

Figure 23: Raw earnings differences and estimated returns at age 29 by HEI, women
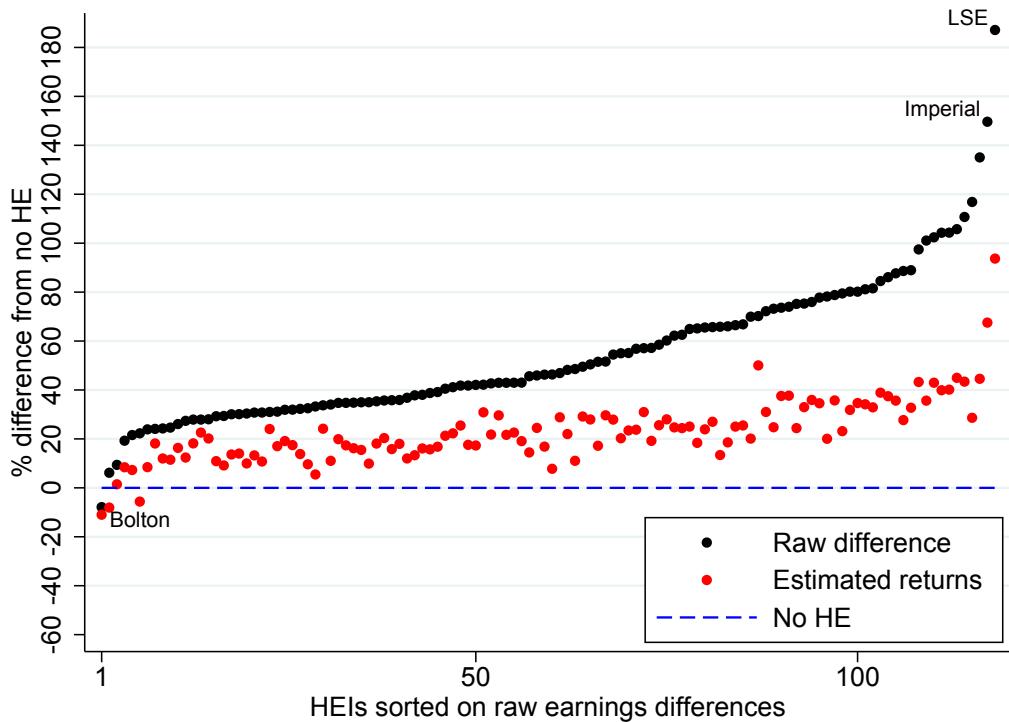


Note: Figure reports derived estimates of the impact of studying at different institutions on annual earnings at age 29 based on the 2002-2007 GCSE cohorts, conditional on being in sustained employment. Raw differences only account for age at which course started, while estimated returns control for age, background and prior attainment and use IPWRA weights. Results have been converted to percentage differences using a log-point conversion. HEIs are ranked based on raw earnings differences.

Figures 24 and 25 show the HEI estimates after controlling for background characteristics (re-ordering the estimates given in red in Figures 22 and 23), colouring institutions by type (Russell Group, Pre-1992, Post-1992 and Other) and with confidence intervals on the estimates. As we saw in the previous figures, there is a wide variation in returns by HEI, reflecting the importance of institution choice for earnings. Russell Group and Pre-1992 universities on average still tend to have very high returns after controlling for background characteristics, but there is a reasonable amount of variation around this trend, with some Pre-1992 universities having low (or even negative) returns and a few Other and Post-1992 universities doing very well. Across all HEIs male earnings returns at age 29 vary from -16% to 58% with the median individual attending an institution with an average return of 10%.[36] Female earnings returns vary from -11% to 94% with the median woman attending a university with a return of 24%. The figures show that 39 of the institutions have negative returns for men (three for women), but only 12 (two for women) of these

---

[36]This implies that 50% of students attend a HEI with an average return above 10%, and 50% attend a HEI with an average return below 10%.

are statistically significantly negative. This compares to 74 of the institutions that have positive returns for men (115 for women), of which 54 (110 for women) of these are statistically significantly positive. [37]

Figure 24: Estimated returns at age 29 by HEI, men
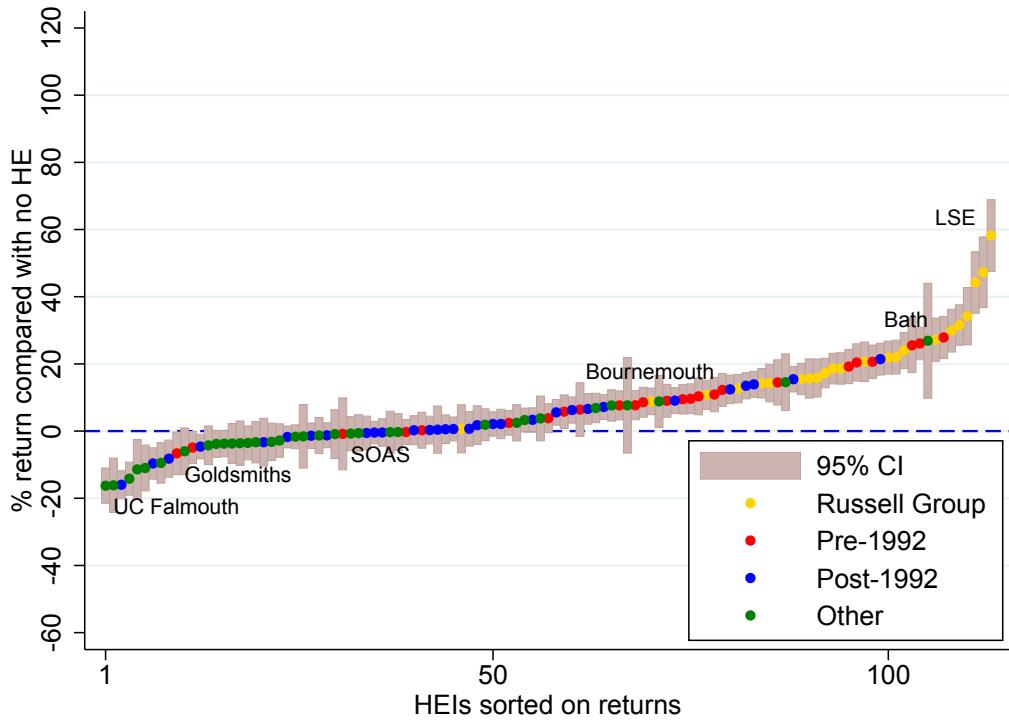


Note: Figure reports derived estimates of the impact of studying at different institutions on annual earnings at age 29 based on the 2002-2007 GCSE cohorts, conditional on being in sustained employment, controlling for age, background and prior attainment, and using IPWRA weights. Results have been converted to percentage differences using a log point conversion. Dots are coloured by institution type classification: Russell Group, established Pre-1992, established Post-1992, and Other institutions (largely specialist universities). HEIs are ranked on estimated returns.

[37]These estimates are compared with the 'relative returns' estimates of Belfield et al. (2018) in Appendix Tables A9 and A10, showing that, despite a number of methodological differences, the estimates are very highly correlated.

Figure 25: Estimated returns at age 29 by HEI, women
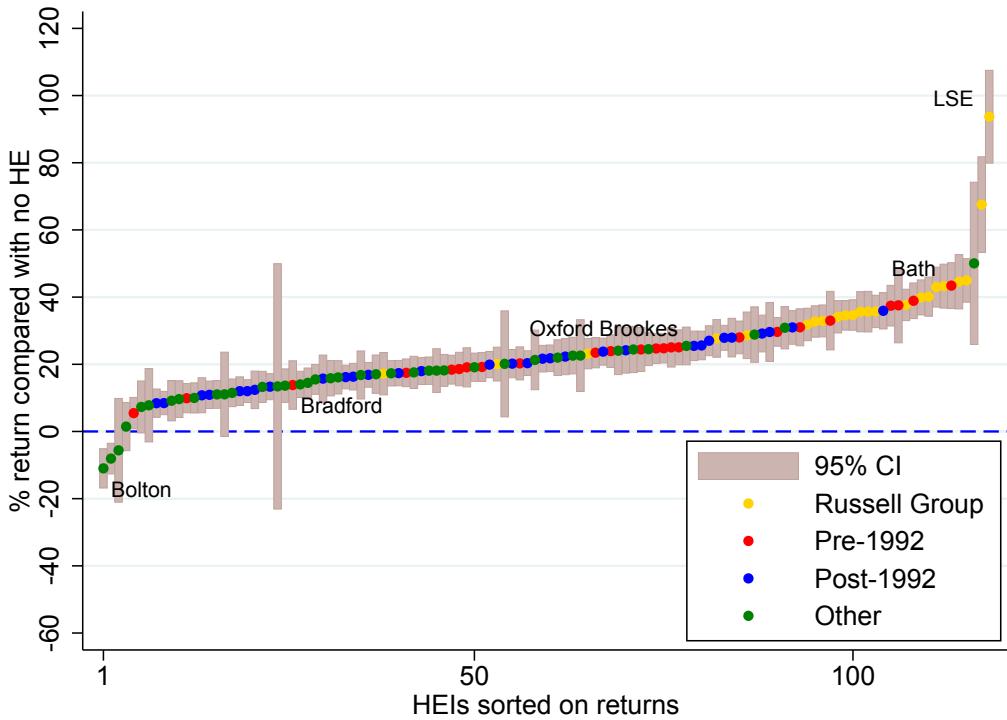


Notes: Figure reports derived estimates of the impact of studying at different institutions on annual earnings at age 29 based on the 2002-2007 GCSE cohorts, conditional on being in sustained employment, controlling for age, background and prior attainment, and using IPWRA weights. Results have been converted to percentage differences using a log point conversion. Dots are coloured by institution type classification: Russell Group, established Pre-1992, established Post-1992, and Other institutions (largely specialist universities). HEIs are ranked on estimated returns.

Figure 26 shows the cumulative density plot of returns assuming every student gets the average return from their HEI. This shows that the low-returning institutions are typically smaller; while 39 (35%) of all HEIs have negative returns for men, these only represent 19% of the male student population. Similarly, the three (2.5% of) institutions with negative returns for women only represent 0.5% of the female student body. In fact, the institutions with *statistically significant* negative returns only represent 4% of all male students and 0.4% of the female students. Similarly, 67% of men and 99% of women (85% overall) attend universities with statistically significant *positive* returns by age 29.

The plot also highlights that returns to HE at age 29 are not just higher for women on average, but also higher across the distribution of HEIs. Table 7 showed that returns for women are 19 percentage points higher on average. Figure 26 shows that even when we focus on the 20% of men and women who attend the highest-returning institutions, female returns are still higher. The 20% of women attending the highest returning institutions have average returns of at least 34%, compared with 20% for the top 20% of men.[38]

---

[38]This does not mean that every institution has higher returns for women than men, just that this is true for each *rank* of HEI.

Figure 26: Cumulative density of student numbers by average HEI returns
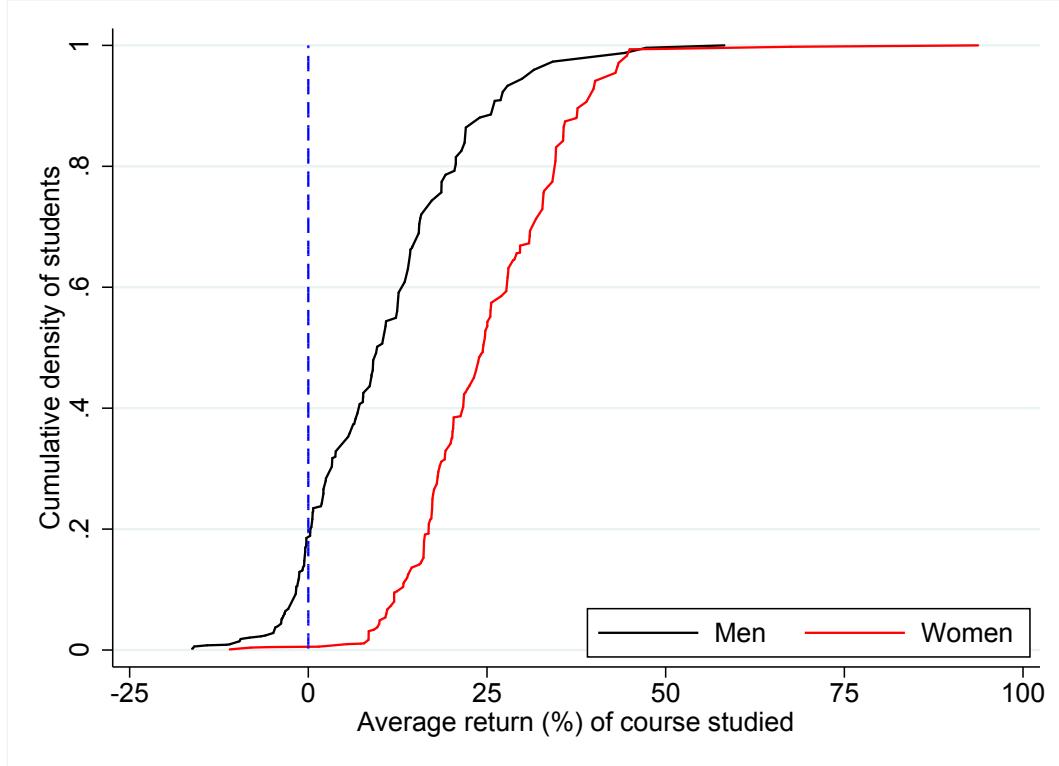


Note: Figure depicts the cumulative proportion of students attending an institution with an average age 29 return at least as high as each given point, based on estimates from Figures 24 and 25.

## 6.2 Heterogeneity in HEI returns

In Section 5.2, we explored whether the variation in returns to different types of students (split by prior attainment and A-level choices) was the result of individuals taking different subjects or experiencing different returns when they take the same subject. Here we examine the extent to which these differential returns across groups may be driven by students attending different types of institutions or whether it is due to them experiencing different returns at these institutions.[39]

In Figures 27 and 28, we plot our estimated returns to attending each of the different university types for our three prior attainment groupings, for men without and with a STEM A-level respectively. The size of the markers reflects the sample size in each group, with the marker sizes being comparable across plots.

For men without a STEM A-level (Figure 27), quite a large share of those with lower prior attainment attend Other universities, which have lower returns on average: indeed, they are close to zero for all prior attainment groups. Higher prior attainment men, on the other hand, are much more likely to attend the highest-return Russell Group universities. This indicates that the type of institution attended will explain some of the higher returns to HE that men with higher prior attainment have compared with men with lower prior attainment.

---

[39]The sample size is not sufficient to explore specific institutions by attainment group, so we instead follow Walker and Zhu (2018) and focus on institution type.

Figure 27: HEI type returns at age 29 by prior attainment (no STEM A-level), men
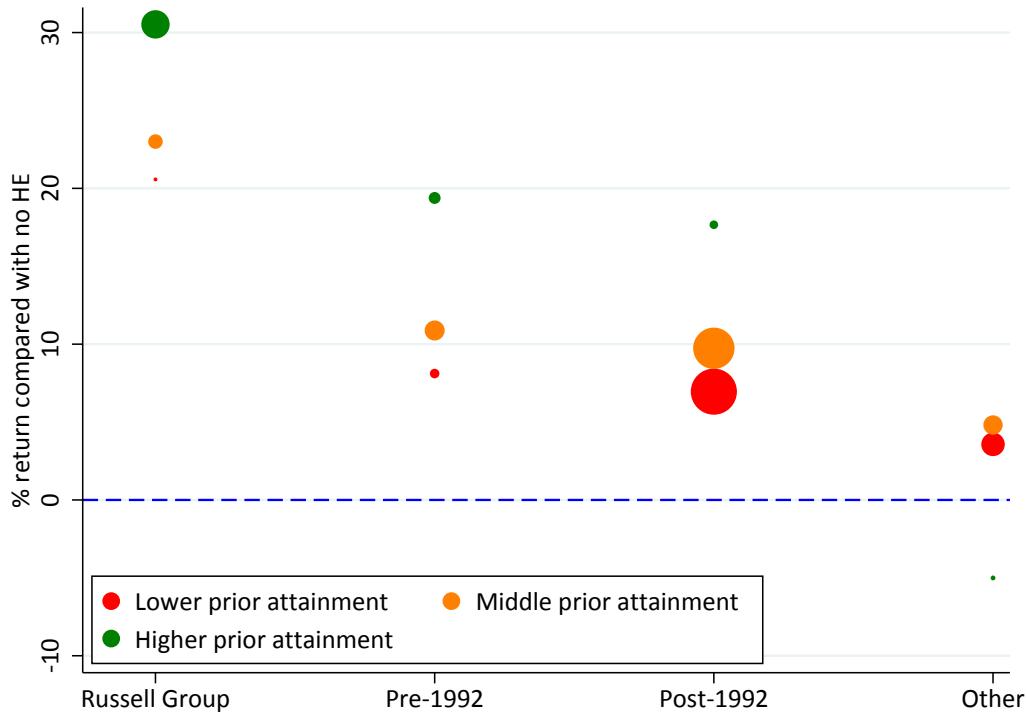


Note: Figure reports derived estimates of the impact of studying at different HEI types on annual earnings at age 29 based on the 2002-2007 GCSE cohorts, conditional on being in sustained employment. Prior attainment is based on GCSE results. Estimates are done separately by subgroup, with the size of the markers reflecting the number of students. Estimates are converted to percentage differences using a log-point conversion.

However, men with higher prior attainment do also have notably higher returns when they attend any university type except Other universities - with an average return at each university type around 10 percentage points higher than the lower attainment group. This indicates that some of the low returns for these lower prior attainment groups may also be due to having a different learning experience from the higher prior attainment groups, even conditional on the university they attend. As we are only able to show heterogeneity in returns for broad university groups, this difference could be in part driven by attending different universities within these broad groups.

The same patterns do not hold true when we look at men with a STEM A-level (Figure 28). In fact, the higher prior attainment group experiences the lowest average returns (near zero or negative) at all university types other than the Russell Group. This is the direct corollary of the finding that higher prior attainment men with a STEM A-level have very low, and often negative, returns to studying humanities or arts subjects. This is a group whose prospects outside of HE are very high; that is, higher prior attainment men with a STEM A-level who do not go to HE have considerable earning potential - indeed, on average they earn around £40,000 at age 29, which is over 30% more than the average non-HE man with at least five A*-C GCSEs. They can go to HE and increase their earnings further if they choose the right subject and institution; however, many of the choices available to them can end up reducing their earnings at age 29.

Figure 28: HEI type returns at age 29 by prior attainment (with STEM A-level), men
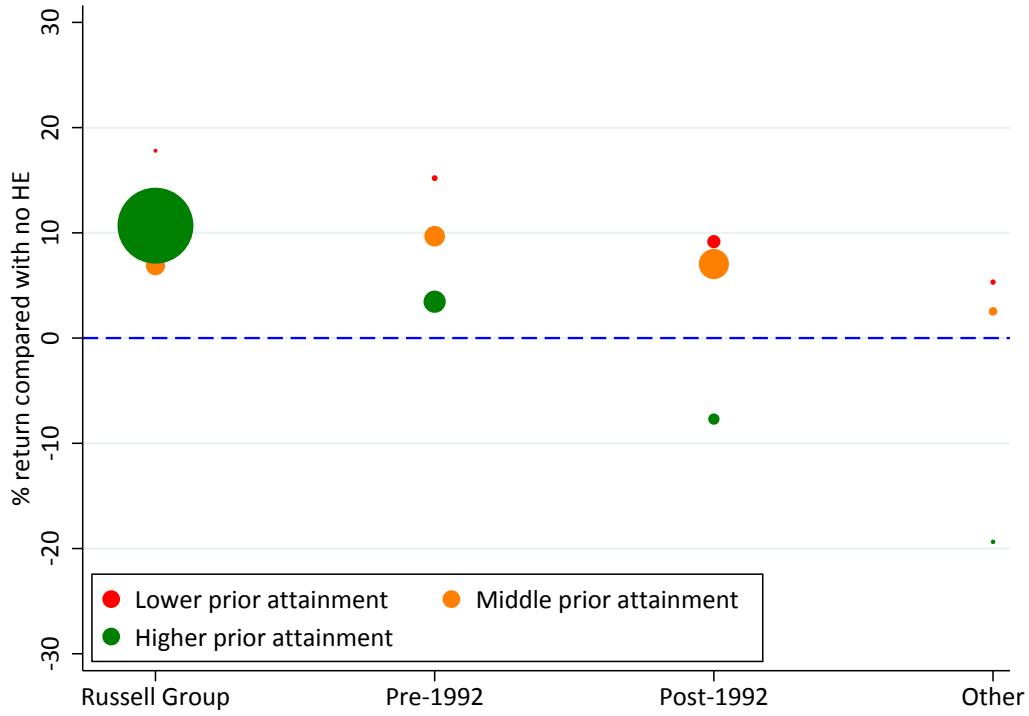


Note: Figure reports derived estimates of the impact of studying at different HEI types on annual earnings at age 29 based on the 2002-2007 GCSE cohorts, conditional on being in sustained employment. Prior attainment is based on GCSE results. Estimates are done separately by subgroup, with the size of the markers reflecting the number of students. Estimates are converted to percentage differences using a log-point conversion.

Figures 29 and 30 show the equivalent estimates for women. For women without a STEM A-level (Figure 29), there appears to be some self-sorting into university type. Lower prior attainment women have the highest returns at Post-1992 and Other universities which they are most likely to attend, while high prior attainment women have the highest return to Russell Group universities which account for 49% of their HE student body. This optimal selection into universities with the highest return for their prior attainment group acts to reduce the differences in returns between the groups. The higher prior attainment group has the highest return overall as Russell Group universities have the biggest impact on earnings for all groups.

Finally, amongst women with a STEM A-level (Figure 30), the lower prior attainment group have the highest return at all university types. However, it should be noted that this is a particularly small portion of the population, making up just 1% of all the women included in the analysis. For middle and higher prior attainment women, similar patterns exist: the higher prior attainment group have the highest returns in Russell Group universities while they have similar or lower returns than the middle prior attainment group when attending Pre-1992, Post-1992 and Other universities. Notably, unlike for men, all university groups have positive returns for all types of women at age 29, reflecting the larger overall benefit to attending HE we find for women.

Figure 29: HEI type returns at age 29 by prior attainment (no STEM A-level), women
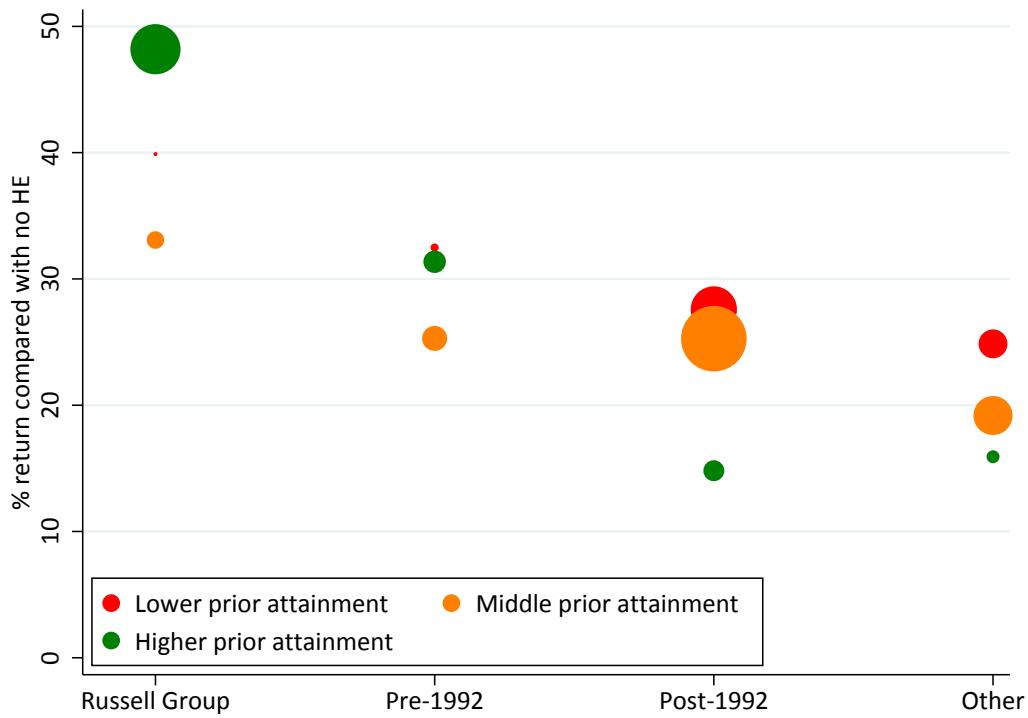


Note: Figure reports derived estimates of the impact of studying at different HEI types on annual earnings at age 29 based on the 2002-2007 GCSE cohorts, conditional on being in sustained employment. Prior attainment is based on GCSE results. Estimates are done separately by subgroup, with the size of the markers reflecting the number of students. Estimates are converted to percentage differences using a log-point conversion.

Figure 30: HEI type returns at age 29 by prior attainment (with STEM A-level), women
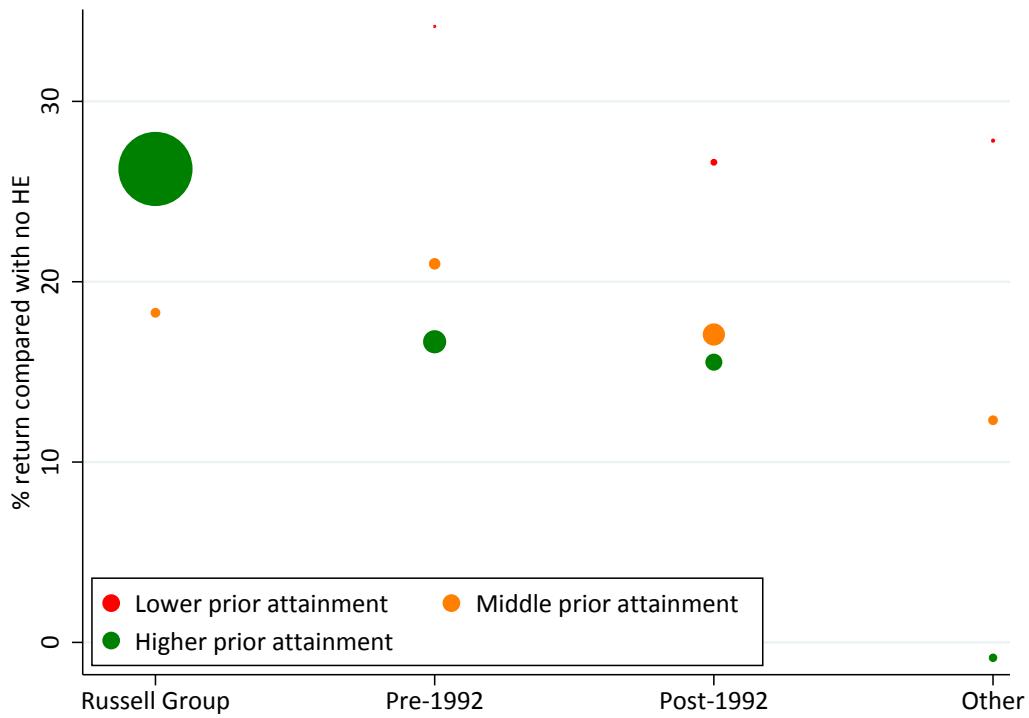


Note: Figure reports derived estimates of the impact of studying at different HEI types on annual earnings at age 29 based on the 2002-2007 GCSE cohorts, conditional on being in sustained employment. Prior attainment is based on GCSE results. Estimates are done separately by subgroup, with the size of the markers reflecting the number of students. Estimates are converted to percentage differences using a log-point conversion.

# 7 Course results

The previous two sections explored how returns to HE vary by degree subject and institution. However, this may miss some important components of the returns to HE. Some institutions specialise in providing specific courses which may have particularly high returns; this is unlikely to be captured in either the institution or subject estimates. Further, we have shown that institution choice is an important determinant of future earnings, but there remains a question of how important institution choice is once focusing on a specific subject (or, equivalently, how important subject choice is once considering a specific university).

Fortunately, our administrative data is sufficiently rich to enable us to explore these issues further. Here we investigate the impact of individual courses - i.e. studying a specific subject at a specific university - on earnings at age 29. For sample size reasons, not all courses offered are included in this analysis. We require at least 30 earnings observations at age 27, 20 at age 28 and 10 at age 29 for inclusion.[40] This leaves 1,235 courses for men and 1,332 courses for women for which we can estimate the returns.

---

[40]This is to ensure data disclosure requirements are met and that we are not predicting earnings returns 'out-of-sample', which would significantly increase the uncertainty and importance of the underlying assumptions in our estimates.

Figure 31: Estimated returns at age 29 by course, men
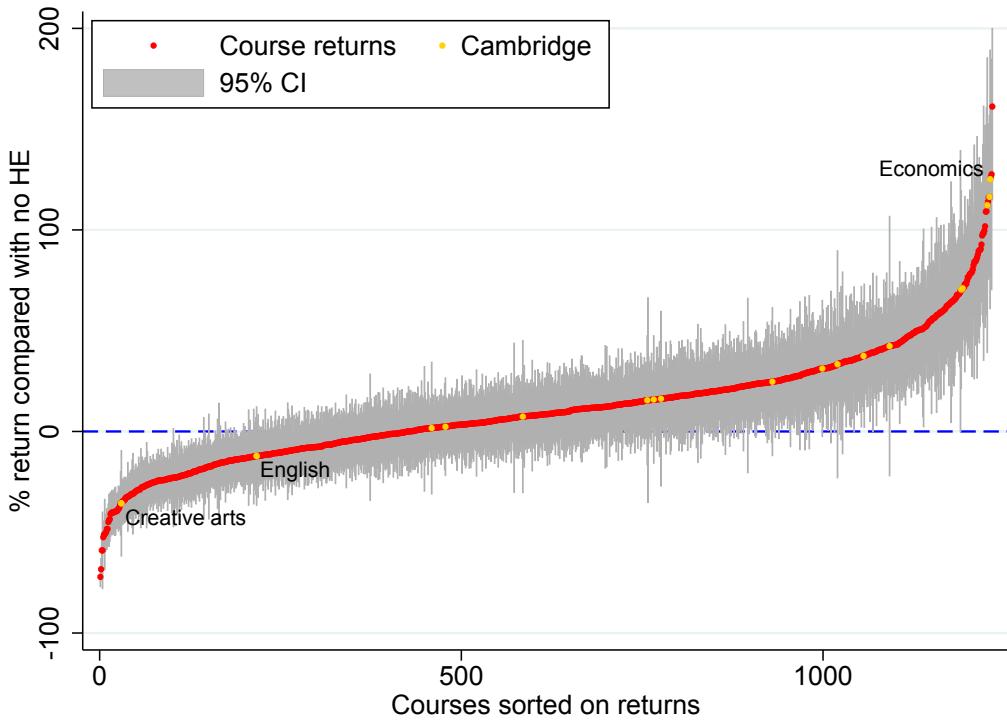
Note: Figure reports derived estimates of the impact of studying different courses (subject-institution combinations) on annual earnings at age 29 based on the 2002-2007 GCSE cohorts, conditional on being in sustained employment, controlling for age, background and prior attainment, and using IPWRA weights. Results have been converted to percentage differences using a log-point conversion. Courses are only included if they have at least 30 earnings observations at age 27, 20 at age 28 and 10 at age 29. Courses are ranked based on estimated returns. All estimates are shown in Table 7b of the Online Appendix

Figure 31 shows the estimated impacts of each of these courses on earnings at age 29 for men and Figure 32 shows the equivalent for women. The variation in the return across courses is striking, with the top courses more than doubling, and the bottom courses more than halving earnings for both men and women. The Figures both highlight the individual course estimates at the University of Cambridge, which is a high-status university and a high performer in our individual institution estimates in Section 6.1 (with returns of around 30% for both men and women). This shows that it is not simply a case of the best institutions clustering at the top in terms of course performance; indeed, even at this very-high performing university, there are courses that offer negative returns for both men and women. While studying law or economics at Cambridge gives returns of over 100% for both genders, a number of subjects have very low returns, including creative arts which significantly reduces age 29 earnings for both men and women.

Figure 32: Estimated returns at age 29 by course, women
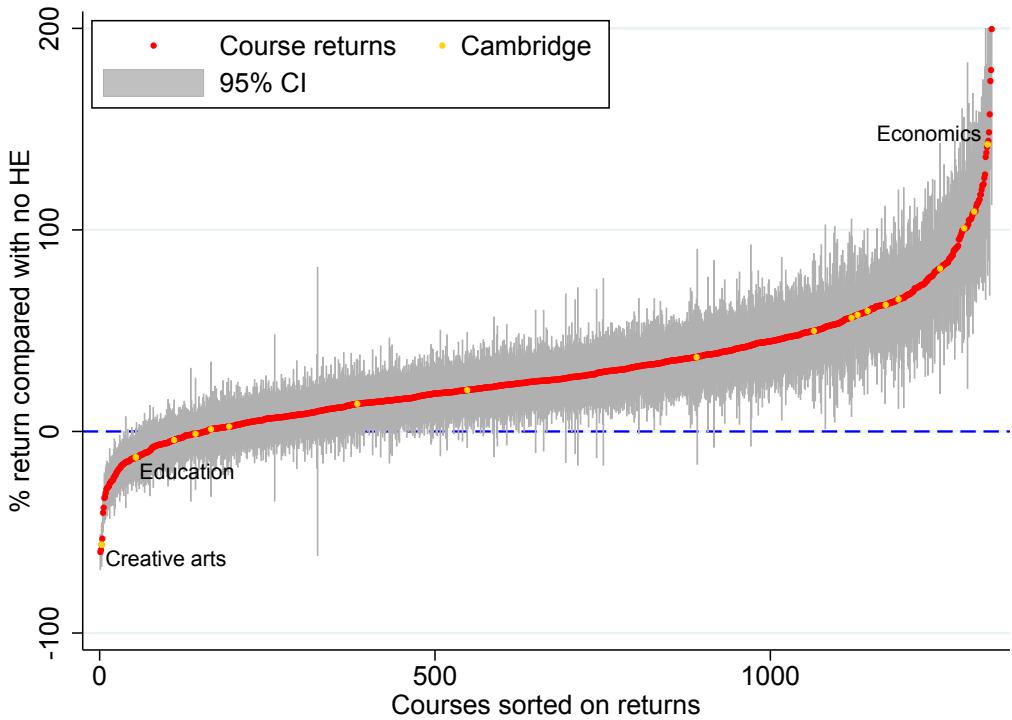


Note: Figure reports derived estimates of the impact of studying different courses (subject-institution combinations) on annual earnings at age 29 based on the 2002-2007 GCSE cohorts, conditional on being in sustained employment, controlling for age, background and prior attainment, and using IPWRA weights. Results have been converted to percentage differences using a log-point conversion. Courses are only included if they have at least 30 earnings observations at age 27, 20 at age 28 and 10 at age 29. Courses are ranked based on estimated returns The y axis is cut at 200%; two courses have female returns higher than this, as shown in Table 7a of the Online Appendix.

Figure 33 shows the proportion of students who study courses with average returns above given thresholds. This clearly shows that while there is wide variation in the returns - 10% of courses have returns below -21% and 10% have returns above 46% for men - the vast majority of courses offer a positive wage return compared with not attending any HE. However, some 34% of courses - accounting for 29% of male students - attract a negative absolute return for men, suggesting that students who took these courses would have potentially earned more if they had not gone to university (or, alternatively, if they had chosen a different course). In fact, once the uncertainty in our estimates is accounted for, only 15% of courses accounting for 13% of students - have statistically significantly negative returns. For women, the cumulative density plot clearly shows that returns are higher on average, with only 12% of courses - or 10% of female students - attracting a negative absolute return. Once uncertainty is accounted for, only 4% of courses representing 2% of female students study courses with negative returns. On the other hand, 58% of all students (71% of women and 42% of men) study a course with statistically significantly positive returns.

Again, it is important to acknowledge that this ignores other benefits from HE, focusing rather narrowly on just one measure of labour market success. This also only looks at the return at age

29, which is likely to be an underestimate of lifetime earnings returns to HE, as discussed in the following section.

Figure 33: Cumulative density of student numbers by average course returns
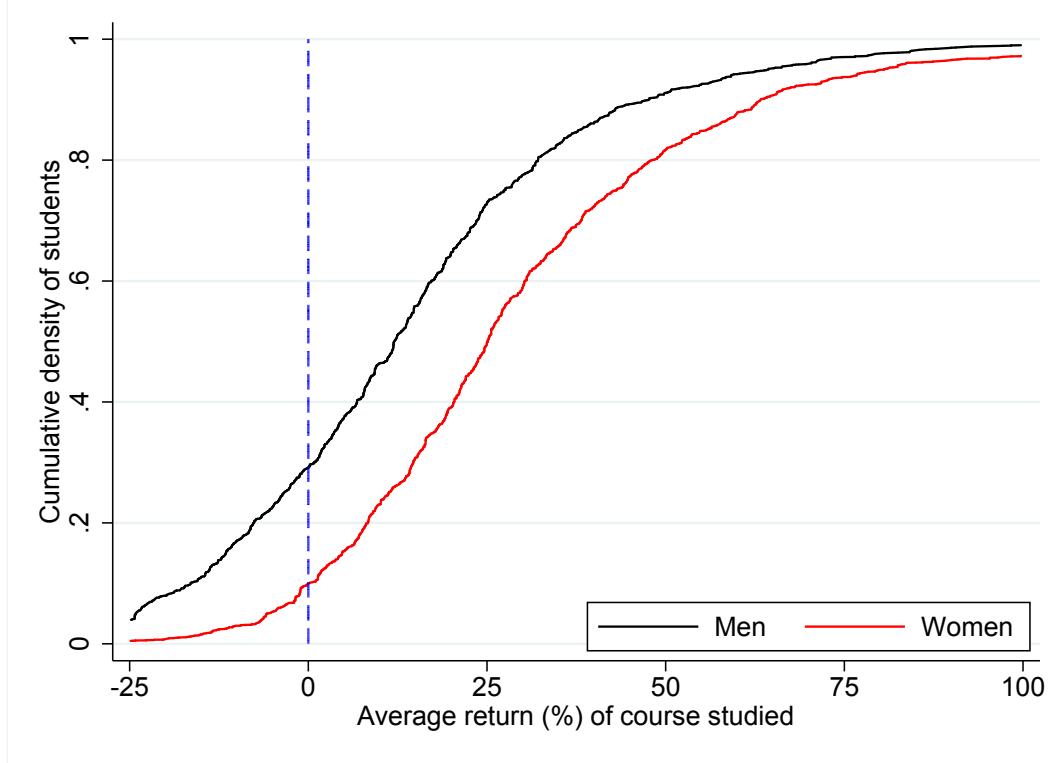


Note: Figure depicts the cumulative proportion of students attending an institution with an average age 29 return at least as high as each given point, based on estimates from Figures 31 and 32.

However, when considering this wide range of returns, it is important to consider that not all students have the option to study each of these courses. Many subjects require specific ('facilitating') A-levels as a prerequisite and many institutions are selective based on prior attainment. Table 13 in the Online Appendix shows the A-level subjects and GCSE scores that at least 95% of the student body on each course possess, as a way of providing an indication of the set of courses students with different prior attainments might be able to take.

With this in mind, these estimates enable us to consider the importance of institution choice even after focusing on a specific subject. Figures 34 and 35 show the range of returns within a given subject across institutions for all the subjects we consider. The whiskers show the highest and lowest returns for each subject, the box shows the range of the middle 50% of institutions (25th to 75th percentile), and the horizontal line inside each box shows the median return for the subject. For example, for men studying economics at the highest-returning institution (Bristol) increases almost doubles male attendants' earnings at age 29, but the lowest returning university (Kingston) reduces earnings by 23%. In fact, every subject for men and most subjects for women have at least one institution that offers a negative return and one with a positive return.

There is also evidence of universities specialising in specific subjects. For example, studying biosciences (including chemistry) at the University of Bradford increases men's earnings at age 29

by around 70% despite the average male return to degrees at Bradford being negligible and the average return to biosciences degrees also being close to zero. There are many other examples of courses outperforming the returns implied by the average subject and institution return, including physical sciences at Oxford, computer science at Royal Holloway and economics at Bristol for men and maths at Oxford, veterinary sciences at the Royal Veterinary College and agriculture at Reading for women.

Figure 34: Boxplot of course returns, men
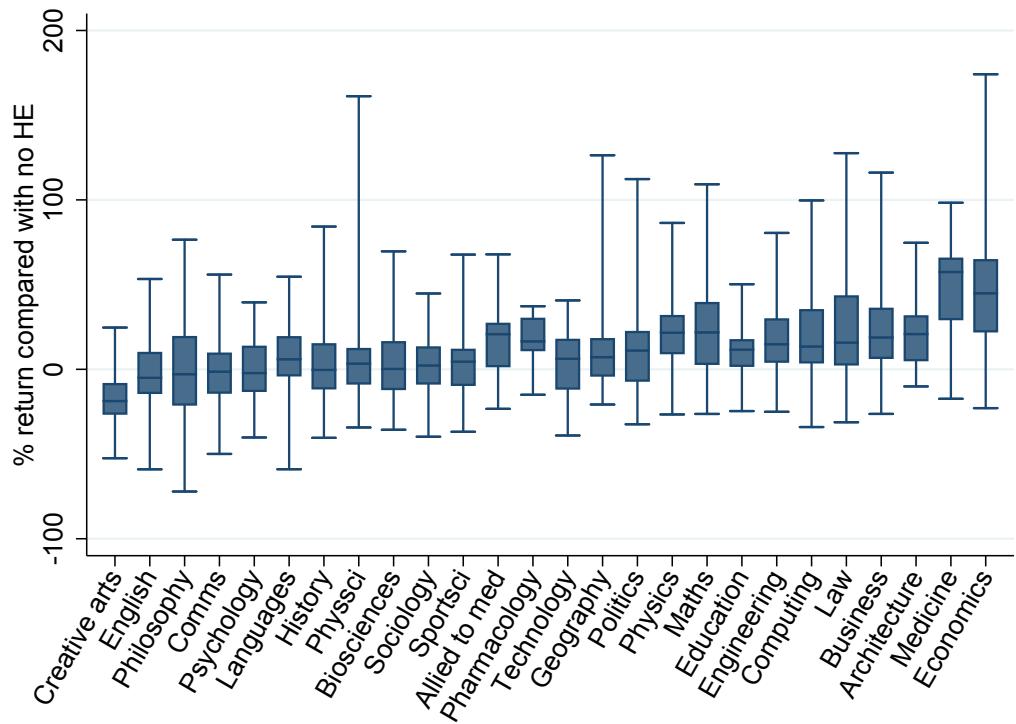


Note: Figure shows a boxplot of course estimates by subject, taken from the results in Figure 31. Whiskers show the highest- and lowest- returning institution in each subject, the box shows the interquartile range of institutions and the dash shows the median. Only subjects with at least eight course estimates are included.

Figure 35: Boxplot of course returns, women
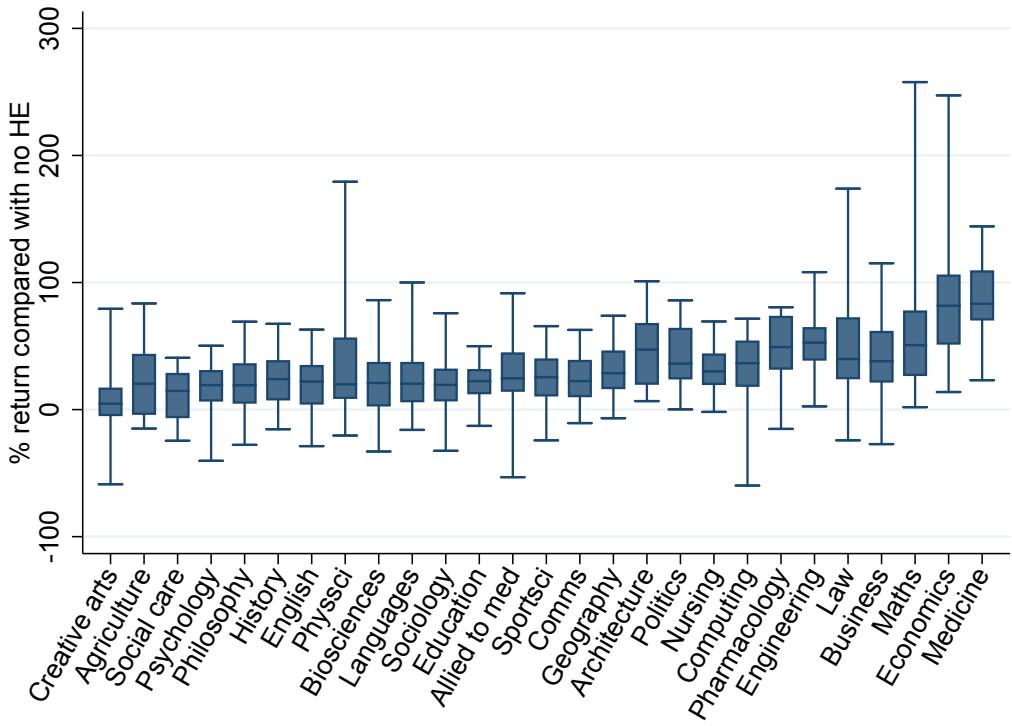


Note: Figure shows a boxplot of course estimates by subject, taken from the results in Figure 32. Whiskers show the highest- and lowest-returning institution in each subject, the box shows the interquartile range of institutions and the dash shows the median. Only subjects with at least eight course estimates are included.

We have shown the large heterogeneity in returns across institutions for the same subject, but an alternative approach is to compare what the return to different subjects is at the same institution. In Figures 31 and 32, we showed how much variation there is in course returns at the University of Cambridge. There is also a lot of variation for other universities, suggesting that the average estimate for a given university masks a wide variation in the possible returns from different subject choices.

# 8 Returns at different ages

Our estimates of absolute returns focus on earnings at age 29. This is primarily due to data limitations: the earliest cohort for which we observe full prior attainment records in the NPD took their GCSEs in 2002 (aged 16), and the latest we observe their outcomes is in the tax year 2015-16 when these individuals were approximately aged 29. This focus on early-career earnings might be one of the reasons our estimated returns are lower than findings in the existing literature, particularly for men. By comparison, Blundell et al. (2000) and Blundell et al. (2005) look at earnings at age 33, while Walker and Zhu (2011) look at earnings across the whole lifecycle. In this section, we consider the implications of estimating returns at such an early age and explore whether the returns are likely to be larger at older ages.

To investigate the impact of age on the returns to university, Table 12 repeats the overall returns to HE estimates in Section 4 at earlier ages also observed in the LEO data.[41] This shows clear evidence of returns to HE increasing with age. In fact, at younger ages, attending HE actually reduces the earnings of men. For both men and women, the returns increase dramatically (by around 20 percentage points) between ages 24 and 29.

A possible explanation for this pattern is lost labour market experience. At age 24, an individual who left school at 18 has six years of potential labour market experience, whereas an individual who studied a three-year degree only has three years of potential labour market experience. Assuming there are returns to work experience, this would boost the earnings of those who do not attend HE. At early ages, this lost experience makes up a significant proportion of total labour market experience but this proportion declines as people spend longer in the labour market. We showed in Figure 3 that over 60% of the non-HE five A*-C GCSEs control group were in 'sustained employment' in the years between ages 18 and 21, which confirms that this group is indeed gaining more labour market experience.

Table 12: Overall returns to HE at various ages

|  | Age 24 | Age 26 | Age 28 | Age 29 |
|---|---|---|---|---|
| Men | -0.13*** | -0.03*** | 0.04*** | 0.06*** |
|  | (0.00) | (0.00) | (0.00) | (0.00) |
| *No. of observations* | *2,183,120* | *2,183,120* | *2,183,120* | *2,183,120* |
| *No. of individuals* | *629,138* | *629,138* | *629,138* | *629,138* |
|  |  |  |  |  |
| Women | 0.04*** | 0.13*** | 0.20*** | 0.23*** |
|  | (0.00) | (0.00) | (0.00) | (0.00) |
| *No. of observations* | *2,619,982* | *2,619,982* | *2,619,982* | *2,619,982* |
| *No. of individuals* | *731,200* | *731,200* | *731,200* | *731,200* |

Note: Table reports estimates based on the 2002-2007 GCSE cohorts, conditional on at least five A*-C GCSEs and on being in sustained employment. Estimates control for background and prior attainment, and use IPWRA weights. Estimates are in log-points.

One potential way to deal with this effect is to control for the years of potential labour market experience, instead of controlling for the age of the individuals.[42] Estimates of the overall returns to HE after six years of potential labour market experience are reported in Table 13. This shows that for men, controlling for experience significantly increases the estimated returns to HE. This is not surprising, as we saw in Figure 1 that earnings grow rapidly during individuals' careers, and controlling for potential experience essentially compares the non-HE sample with HE individuals three years older than them (after they have had three years to make up for the lost experience). For women we see that controlling for experience does not significantly increase the returns, because the non-HE control group does not exhibit rapid growth in their earnings during their mid

---

[41]These are derived from the same regression model, using the quadratic age trends to calculate marginal effects at different points in time.

[42]It is not possible to control for both as, in general, age, continuous education and potential labour market experience are linear combinations of one another.

20s.[43]

While we believe these results are informative about the importance of lost labour market experience, we strongly argue that they do not fairly reflect the 'returns' to HE. We argue that it is more intuitive and transparent to present the impact on earnings at a specific age rather than at a specific number of years after leaving education. This is a more important metric for young people choosing whether to attend HE, for whom lost labour market experience is a very real cost of attending that should be accounted for in the estimated returns.

Table 13: Overall returns to HE controlling for potential labour market experience

|  | Age 29 | 6 years of experience |
| --- | --- | --- |
| Men | 0.06*** | 0.19*** |
|  | (0.00) | (0.00) |
|  |  |  |
| Women | 0.23*** | 0.24*** |
|  | (0.00) | (0.00) |
|  |  |  |
| Cohort/Age start controls | Yes | Yes |
| Background characteristics | Yes | Yes |
| Prior attainment | Yes | Yes |
| IPWRA weight | Yes | Yes |

Note: Table reports derived estimates of the impact of HE on annual earnings at age 29 and after six years of potential labour market experience. Estimates are based on the 2002-2007 GCSE cohorts, conditional on at least five A*-C GCSEs and on being in sustained employment. Experience estimates control for quadratic trends in years of potential experience rather than age. Estimates are in log-points.

A remaining question is how we might expect our estimated returns to differ if we looked at later ages than 29. For descriptive evidence on this we turn to the Labour Force Survey (LFS) in which we can observe HE and non-HE individuals at older ages. We are not able to control for detailed background characteristics to estimate causal returns, but Figure 36 presents the median age-earnings profiles for graduates and non graduates, separately for men and women. This clearly shows that for men even after age 29, the earnings differential between graduates and non-graduates continues to grow. This aligns with the findings of previous literature estimating the return to HE over the lifecycle (Walker and Zhu, 2011; Bhuller et al., 2017) and suggests that our estimates of the returns to HE at age 29 are likely to be underestimates of the lifetime returns to HE. The same pattern is not clear for women, which suggests that looking at this point in the lifecycle is less of an issue for women than it is for men.

---

[43]Figures A5 and A6 in the Appendix show the same results for subjects, highlighting the increase in returns across nearly all subject areas, particularly the longer courses (e.g. medicine), which result in more lost experience. The impact of this specification change may differ across subjects as graduates of different subjects are likely to end up working in different occupations and hence may have very different earnings growth. For example, we might expect graduates working in finance to have much steeper earnings growth than those working in teaching.

Figure 36: Median earnings across the lifecycle in the Labour Force Survey
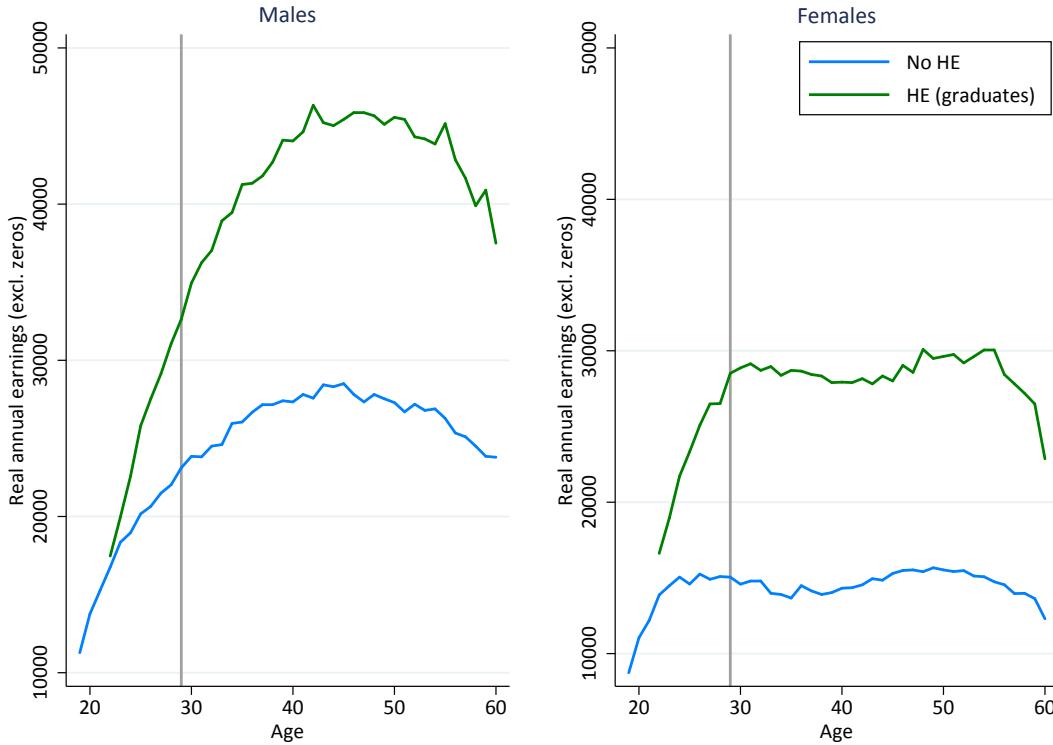


Note: Male and female earnings using 2004-2017 data from the LFS, with earnings in 2017 prices. No HE includes all individuals with highest qualifications of NVQ level 2 or above (equivalent to at least five A*-C grades at GCSE), but no degree-level qualifications. HE includes the individuals with a degree-level qualification.

## 9  Conclusion

It is widely documented that individuals who attend HE earn more, on average, than those who do not. We show that at age 29, men who attend HE earn around 25% more than those who do not, and women who attend HE earn 50% more. However, those who attend HE also typically have higher prior attainment than those who do not and, on average, come from higher socio-economic backgrounds. Once we account for these differences in characteristics we estimate the average impact of attending HE on earnings at age 29 to be 6% for men and 26% for women.

These estimates are lower than those in much of the existing literature, particularly for men, so it is important to put these findings in context. We are measuring the returns to *entering* HE rather than graduating from HE. If we remove individuals who drop out from HE, we find a return to *graduating* from HE of 8% for men and 28% for women. It is also important to note that we are estimating the impact of HE on earnings at a relatively young age, a point at which those who do not attend HE will have had the opportunity to acquire considerably more work experience than those who do go to HE. We find significant evidence that the returns to HE increase with age and so it is entirely plausible - and potentially quite likely - that the overall impact of HE on *lifetime* earnings is considerably higher than the returns we find at age 29.

It is important to highlight one of the drawbacks of our data: we are not able to observe hours worked, and so instead we investigate annual earnings. This is likely to be particularly important in our estimates of the returns to HE for women, which are likely to at least partly reflect the fact that women who attend HE are much less likely to work part-time and so have higher earnings directly as a result of working more hours.

Our estimates of the overall return to HE are interesting and important, but this focus masks important variation in the returns to particular degree courses. In line with Belfield et al. (2018), we show that subject and institution choices are very important: for men, studying creative arts, English or philosophy actually reduces average earnings at age 29, and 4% of men attend institution that have statistically significant negative average returns. However, studying medicine or economics increases earnings at age 29 by more than 20% on average, and over 50% of men attend institutions with an average return of more than 9%. We also show that institution choice is still vital even after students know what they want to study; for men, in every subject there is at least one institution with negative and one with positive average returns. Simply choosing the right subject or the right university is not enough to guarantee a high average return to HE.

Subject and institution choice are also important, for women. While there are no subjects with negative average returns and fewer than 1% of women attend institutions with negative returns, studying economics or medicine increase women's earnings by over 60% on average and more than 20% of women attend universities with average returns in excess of 30%.

HE also has very different impacts on different types of students. While the estimated returns to HE are low for men on average, this is not true for all individuals. For higher prior attainment men without a STEM A-level, the average return at age 29 is 20%, while for lower prior attainment men without a STEM A-level it is only 4%. This low return is driven by a greater tendency of lower prior attainment students to take low returning subjects such as creative arts, communications and sport science, being more likely to attend lower-returning Post-1992 and Other institutions, and experiencing lower returns even when they study the same subject or at the same type of university as their higher prior attainment peers. Understanding the impact of HE on those with lower prior attainment and no STEM A-level - and the drivers of this impact - is particularly important when considering expansion to the HE system. Only one in three men in this group attend HE compared with more than two in three across all other student types. In fact, individuals with lower prior attainment and no STEM A-level make up almost 70% of all those students with at least five A*-C GCSEs who do not attend university. Any further expansion to the HE system is very likely to involve more students in this group attending.

High prior attainment men with a STEM A-level are another interesting group of students; 95% of these students attend HE, making up nearly 30% of all HE attendants. This is a group for whom the average returns to HE at age 29 are quite low. On closer inspection it appears that these low average returns are the result of a long-tail of choices that men with a STEM A-level can make that actually reduce their earnings. For this group, studying arts or humanities subjects such as English, history or languages reduces their earnings at age 29, as does attend Post-1992 or Other universities, while studying law, medicine or economics increases earnings by around 20% and the average returns to a degree from a Russell Group university for them is around 10%.

Before starting HE, this group of students have qualifications that are well-rewarded in the labour market, and choosing the right subject or right institution can increase these earnings further, but there are a whole set of degrees that have the potential to reduce their early-career earnings. However, it should be noted that these results are based on comparing higher prior attainment individuals with a STEM A-level who attended HE with their counterparts who did not. Deciding not to attend HE is rare for this group - only made by 5% of individuals - and this may represent a select sample. Caution should be executed when interpreting this result to widely.

Some of the same patterns emerge when looking at different types of women: higher prior attainment women without a STEM A-level have higher returns than their lower prior attainment peers, but it is not the case that high prior attainment women with a STEM A-level have especially low returns. Unlike for men, there is little evidence of lower prior attainment women without a STEM A-level experiencing lower returns when studying the same subject or at the same type of institution as their higher-attaining peers. Instead, the lower returns appear to be driven by a higher tendency to take low returns subjects such as social care or psychology and to attend lower returning Post-1992 and Other universities.

These results provide an insight into the potential returns different types of students may experience when choosing whether to go to university and, if so, what and where to study. They also provide valuable information to policymakers on the effectiveness of different parts of the HE system and the potential issues when considering further expansion. However, it should be kept in mind that these results focus on the impact of HE on earnings at age 29; the returns are likely to increase as individuals age, and earnings are just one of many outcomes HE has the potential to affect. These findings should be taken in the context of evidence on the activities HEIs provide and the other impacts universities have on their students.

# References

**Andrews, Rodney J, Scott A Imberman, and Michael F Lovenheim**, "Risky Business? The Effect of Majoring in Business on Earnings and Educational Attainment," Technical Report, National Bureau of Economic Research 2017.

**Belfield, Chris and Laura van der Erve**, "The impact of higher education on the living standards of female graduates," *IFS Working Paper WP18/25*, 2018.

— **, Jack Britton, Franz Buscha, Lorraine Dearden, Matt Dickson, Laura van der Erve, Luke Sibieta, Anna Vignoles, Ian Walker, and Yu Zhu**, "The returns of different subjects and university choices in England," *Department for Education Report*, 2018.

**Bhuller, Manudeep, Magne Mogstad, and Kjell G Salvanes**, "Life-cycle earnings, education premiums, and internal rates of return," *Journal of Labor Economics*, 2017, *35* (4), 993–1030.

**Bibby, David, Franz Buscha, Augusto Cerqua, Dave Thomson, and Peter Urwin**, "Estimation of the labour market returns to qualifications gained in English Further Education," *BIS Research Paper*, 2014, *195*.

**Blundell, Richard, David Green, and Wenchao Jin**, "The UK wage premium puzzle: how did a large increase in university graduates leave the education premium unchanged," *Institute for Fiscal Studies Working Paper*, 2016, *16* (1).

＿ , **Lorraine Dearden, Alissa Goodman, and Howard Reed**, "The returns to higher education in Britain: evidence from a British cohort," *The Economic Journal*, 2000, *110* (461), 82–99.

＿ , ＿ , **and Barbara Sianesi**, "Evaluating the effect of education on earnings: models, methods and results from the National Child Development Survey," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2005, *168* (3), 473–512.

**Card, David**, "The causal effect of education on earnings," *Handbook of labor economics*, 1999, *Vol. 3*, 1801–1863.

**Chowdry, Haroon, Claire Crawford, Lorraine Dearden, Alissa Goodman, and Anna Vignoles**, "Widening participation in higher education: analysis using linked administrative data," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2013, *176* (2), 431–457.

**Goldin, Claudia Dale and Lawrence F Katz**, *The race between education and technology*, Harvard University Press, 2009.

**Goos, Maarten, Alan Manning, and Anna Salomons**, "Job polarization in Europe," *American economic review*, 2009, *99* (2), 58–63.

**Harmon, Colm, Hessel Oosterbeek, and Ian Walker**, "The returns to education: Microeconomics," *Journal of economic surveys*, 2003, *17* (2), 115–156.

**Heckman, James J, Lance J Lochner, and Petra E Todd**, "Earnings functions, rates of return and treatment effects: The Mincer equation and beyond," *Handbook of the Economics of Education*, 2006, *1*, 307–458.

**Kirkeboen, Lars J, Edwin Leuven, and Magne Mogstad**, "Field of study, earnings, and self-selection," *The Quarterly Journal of Economics*, 2016, *131* (3), 1057–1111.

**Michaels, Guy, Ashwini Natraj, and John Van Reenen**, "Has ICT polarized skill demand? Evidence from eleven countries over twenty-five years," *Review of Economics and Statistics*, 2014, *96* (1), 60–77.

**Psacharopoulos, George and Harry Anthony Patrinos**, "Returns to investment in education: a further update," *Education economics*, 2004, *12* (2), 111–134.

**Walker, Ian and Yu Zhu**, "Differences by degree: Evidence of the net financial rates of return to undergraduate study for England and Wales," *Economics of Education Review*, 2011, *30* (6), 1177–1186.

＿ **and** ＿ , "The impact of university degrees on the lifecycle of earnings: some further analysis," 2013.

__ **and** __ , "University selectivity and the relative returns to higher education: Evidence from the UK," *Labour Economics*, 2018.

**Wooldridge, Jeffrey M**, "Inverse probability weighted estimation for general missing data problems," *Journal of Econometrics*, 2007, *141* (2), 1281–1301.

# Appendix

Table A1: Analysis sample by age

| GCSE cohort | Analysis sample | Age 24 | Age 25 | Age 26 | Age 27 | Age 28 | Age 29 |
|---|---|---|---|---|---|---|---|
| 2002 | 226,102 | 186,711 | 194,024 | 191,470 | 193,935 | 192,750 | 187,622 |
| 2003 | 238,564 | 205,023 | 203,805 | 207,136 | 206,514 | 201,055 | |
| 2004 | 248,367 | 214,487 | 219,270 | 219,550 | 214,100 | | |
| 2005 | 250,741 | 225,684 | 227,646 | 221,971 | | | |
| 2006 | 247,207 | 234,009 | 228,726 | | | | |
| 2007 | 233,444 | 233,444 | | | | | |
| Total | 1,444,425 | 1,299,358 | 1,073,471 | 840,127 | 614,549 | 393,805 | 187,622 |

Note: The full sample aligns with our final analysis sample in Column 6 of Table 1.

Table A2: Background characteristics comparing matched and unmatched individuals

| | Unmatched | Matched |
|---|---|---|
| *Background sShare)* | | |
| FSM | 0.07 | 0.05 |
| EAL | 0.08 | 0.08 |
| SEN | 0.04 | 0.03 |
| State school | 0.87 | 0.88 |
| *of which:* | | |
| SES Q1 - least deprived | 0.25 | 0.30 |
| SES Q2 | 0.24 | 0.25 |
| SES Q3 | 0.21 | 0.20 |
| SES Q4 | 0.17 | 0.15 |
| SES Q5 - most deprived | 0.12 | 0.10 |
| *Ethnicity (share)* | | |
| White | 0.84 | 0.84 |
| Black | 0.03 | 0.02 |
| Asian | 0.05 | 0.06 |
| Other | 0.07 | 0.07 |
| *Prior attainment (share)* | | |
| KS2 maths level 5+ | 0.31 | 0.38 |
| KS2 English level 5+ | 0.34 | 0.40 |
| KS4 maths A*/A | 0.18 | 0.26 |
| KS4 English A*/A | 0.24 | 0.32 |
| N | 243,029 | 1,700,935 |

Notes: The samples add up to Column 3 in Table 1, and average over the individuals in the six GCSE cohorts. The matched sample includes those who we could potentially include in our analysis, or Column 5 in Table 1, while the unmatched sample are those with non-missing information who are dropped in Columns 4 or 5. EAL = English as an additional language, FSM = free school meals, SEN = non-statemented special educational needs. All numbers are proportions over the respective sample, except socio-economic status quintiles, ethnicity and Key Stage 4 scores, which are only defined for state school pupils.

## Prior attainment subgroups

Table A3: Subjects studied by prior attainment subgroup, men

| | No STEM A-level | | | STEM A-level | | | |
| | Lower | Middle | Higher | Lower | Middle | Higher | All |
|---|---|---|---|---|---|---|---|
| Agriculture | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 |
| Architecture | 0.04 | 0.04 | 0.02 | 0.03 | 0.04 | 0.03 | 0.03 |
| Biosciences | 0.01 | <0.01 | <0.01 | 0.10 | 0.10 | 0.10 | 0.05 |
| Business | 0.19 | 0.17 | 0.10 | 0.15 | 0.13 | 0.07 | 0.13 |
| Comms | 0.07 | 0.06 | 0.03 | 0.02 | 0.01 | <0.01 | 0.04 |
| Computing | 0.10 | 0.06 | 0.02 | 0.13 | 0.09 | 0.05 | 0.07 |
| Creative arts | 0.16 | 0.15 | 0.12 | 0.05 | 0.05 | 0.03 | 0.09 |
| Economics | 0.01 | 0.02 | 0.03 | 0.03 | 0.04 | 0.07 | 0.04 |
| Education | 0.02 | 0.02 | <0.01 | 0.01 | 0.01 | <0.01 | 0.01 |
| Engineering | 0.05 | 0.03 | 0.02 | 0.13 | 0.14 | 0.14 | 0.08 |
| Geography | 0.03 | 0.05 | 0.08 | <0.01 | 0.01 | 0.01 | 0.03 |
| English | 0.02 | 0.03 | 0.04 | 0.03 | 0.04 | 0.03 | 0.03 |
| Social care | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 |
| History | 0.04 | 0.06 | 0.13 | 0.02 | 0.02 | 0.04 | 0.05 |
| Languages | <0.01 | 0.03 | 0.09 | <0.01 | <0.01 | 0.02 | 0.03 |
| Law | 0.04 | 0.05 | 0.07 | 0.02 | 0.02 | 0.03 | 0.04 |
| Maths | <0.01 | <0.01 | <0.01 | 0.04 | 0.05 | 0.08 | 0.03 |
| Medicine | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | 0.07 | 0.02 |
| Nursing | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 |
| Pharmacology | <0.01 | <0.01 | <0.01 | 0.02 | 0.02 | 0.01 | <0.01 |
| Philosophy | <0.01 | 0.02 | 0.04 | <0.01 | <0.01 | 0.02 | 0.02 |
| Physsci | <0.01 | <0.01 | <0.01 | 0.04 | 0.04 | 0.02 | 0.02 |
| Physics | <0.01 | <0.01 | <0.01 | 0.02 | 0.03 | 0.05 | 0.02 |
| Politics | 0.02 | 0.04 | 0.07 | <0.01 | 0.01 | 0.02 | 0.03 |
| Psychology | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.01 | 0.02 |
| Sociology | 0.03 | 0.03 | 0.02 | 0.01 | 0.01 | <0.01 | 0.02 |
| Sportsci | 0.08 | 0.06 | 0.02 | 0.05 | 0.05 | 0.02 | 0.05 |
| Allied to med | <0.01 | <0.01 | <0.01 | 0.03 | 0.03 | 0.03 | 0.02 |
| Technology | 0.02 | 0.01 | <0.01 | 0.01 | 0.01 | <0.01 | 0.01 |
| Vetsci | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 |

Note: Table shows the proportion of students taking each of a set of subjects by groups of prior attainment. Figures are calculated based on the 2002-2007 GCSE cohorts, for those who achieved at least five A*-Cs and attended higher education in the UK.

Table A4: Subjects studied by prior attainment subgroup, women

| | No STEM A-level | | | STEM A-level | | | |
| | Lower | Middle | Higher | Lower | Middle | Higher | All |
|---|---|---|---|---|---|---|---|
| Agriculture | <0.01 | <0.01 | <0.01 | 0.02 | 0.02 | 0.01 | <0.01 |
| Architecture | <0.01 | <0.01 | <0.01 | 0.01 | 0.01 | 0.02 | 0.01 |
| Biosciences | 0.01 | <0.01 | <0.01 | 0.12 | 0.12 | 0.13 | 0.05 |
| Business | 0.14 | 0.12 | 0.07 | 0.12 | 0.10 | 0.06 | 0.10 |
| Comms | 0.07 | 0.06 | 0.04 | 0.02 | 0.01 | <0.01 | 0.04 |
| Computing | 0.02 | 0.01 | <0.01 | 0.03 | 0.02 | <0.01 | 0.01 |
| Creative arts | 0.20 | 0.20 | 0.17 | 0.04 | 0.05 | 0.04 | 0.14 |
| Economics | <0.01 | <0.01 | <0.01 | 0.01 | 0.01 | 0.03 | 0.01 |
| Education | 0.10 | 0.09 | 0.04 | 0.06 | 0.06 | 0.02 | 0.06 |
| Engineering | <0.01 | <0.01 | <0.01 | 0.02 | 0.02 | 0.02 | 0.01 |
| Geography | 0.04 | 0.07 | 0.13 | 0.02 | 0.03 | 0.03 | 0.06 |
| English | <0.01 | 0.02 | 0.03 | 0.02 | 0.03 | 0.04 | 0.02 |
| Social care | 0.04 | 0.02 | <0.01 | 0.01 | <0.01 | <0.01 | 0.01 |
| History | 0.02 | 0.04 | 0.09 | 0.01 | 0.02 | 0.04 | 0.04 |
| Languages | 0.01 | 0.03 | 0.11 | <0.01 | 0.01 | 0.04 | 0.05 |
| Law | 0.06 | 0.06 | 0.08 | 0.04 | 0.03 | 0.04 | 0.06 |
| Maths | <0.01 | <0.01 | <0.01 | 0.03 | 0.03 | 0.05 | 0.02 |
| Medicine | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | 0.10 | 0.03 |
| Nursing | 0.04 | 0.03 | <0.01 | 0.05 | 0.05 | 0.02 | 0.03 |
| Pharmacology | <0.01 | <0.01 | <0.01 | 0.03 | 0.03 | 0.03 | <0.01 |
| Philosophy | <0.01 | 0.01 | 0.03 | <0.01 | <0.01 | 0.01 | 0.02 |
| Physsci | <0.01 | <0.01 | <0.01 | 0.05 | 0.04 | 0.02 | 0.01 |
| Physics | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | 0.01 | <0.01 |
| Politics | <0.01 | 0.01 | 0.03 | <0.01 | <0.01 | 0.01 | 0.02 |
| Psychology | 0.06 | 0.08 | 0.07 | 0.09 | 0.09 | 0.07 | 0.07 |
| Sociology | 0.07 | 0.07 | 0.04 | 0.04 | 0.03 | 0.02 | 0.05 |
| Sportsci | 0.03 | 0.02 | 0.01 | 0.04 | 0.04 | 0.02 | 0.02 |
| Allied to med | 0.02 | 0.02 | 0.01 | 0.09 | 0.10 | 0.09 | 0.04 |
| Technology | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 |
| Vetsci | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | 0.02 | <0.01 |

Note: Table shows the proportion of students taking each of a set of subjects by groups of prior attainment. Figures are calculated based on the 2002-2007 GCSE cohorts, for those who achieved at least five A*-Cs and attended higher education in the UK.

## Subject returns estimates

Figure A1: Estimated returns at age 29 by subject, men
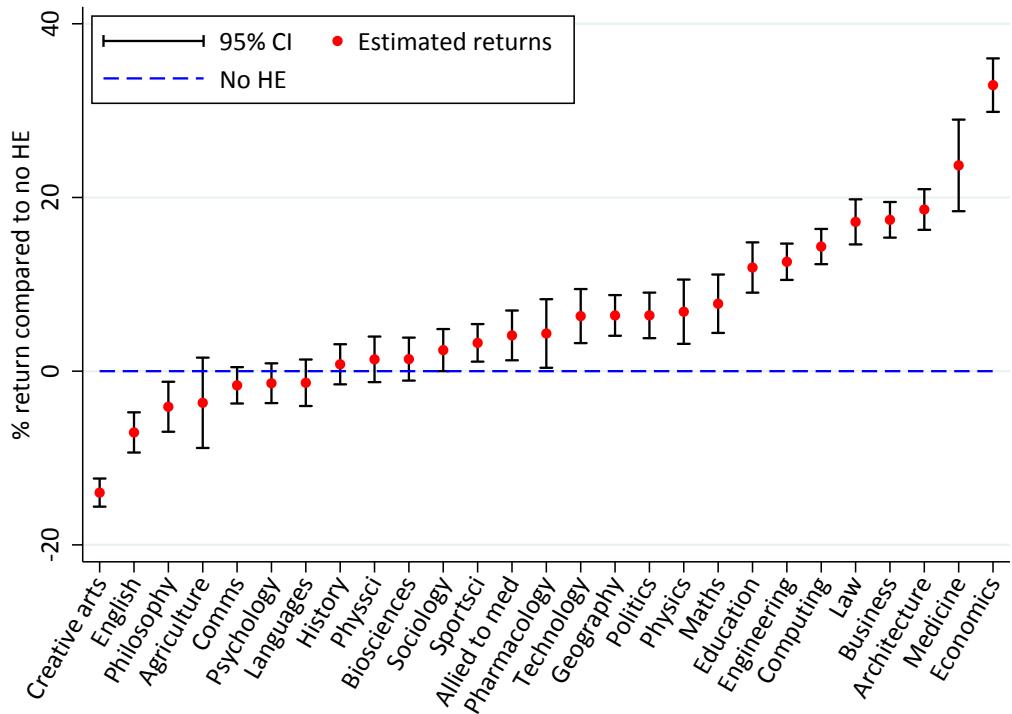


Note: Figure reports derived estimates of the impact of studying different subjects on annual earnings at age 29 based on the 2002-2007 GCSE cohort conditional on being in sustained employment, controlling for age, background and prior attainment, and using IPWRA weights. Subjects are defined using the CAH2 definition with chemistry combined with biosciences. Celtic studies, humanities & liberal arts and combined & general studies have been excluded for sample size reasons. Results have been converted to percentage differences using a log point conversion. Subjects are ranked based on estimated returns.

Figure A2: Estimated returns at age 29 by subject, women
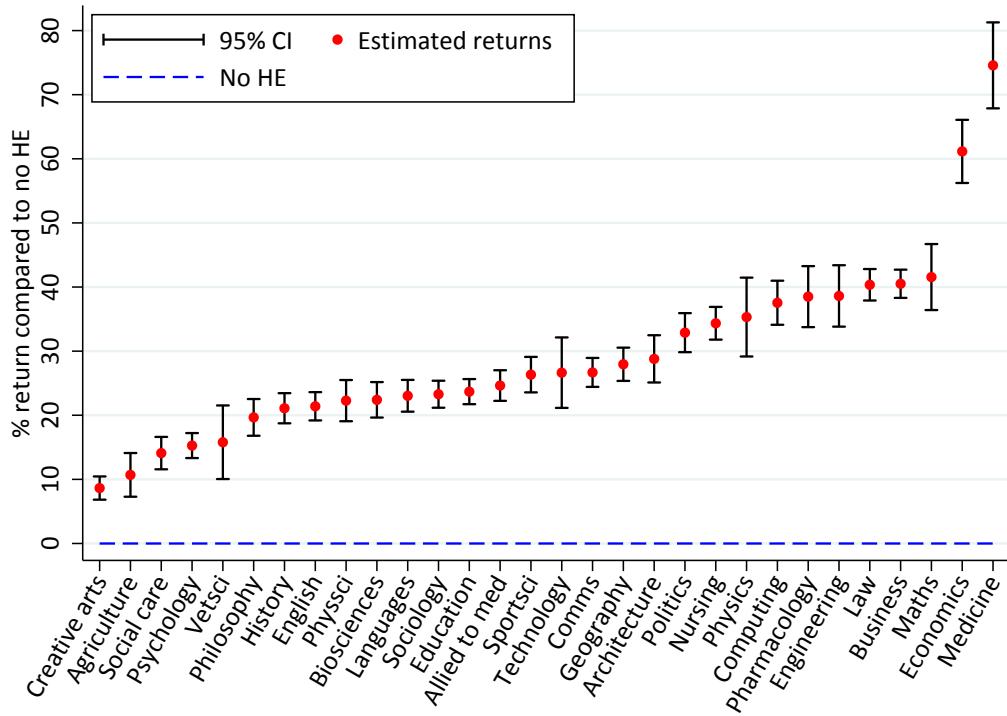


Note: Figure reports derived estimates of the impact of studying different subjects on annual earnings at age 29 based on the 2002-2007 GCSE cohort conditional on being in sustained employment, controlling for age, background and prior attainment, and using IPWRA weights. Subjects are defined using the CAH2 definition with chemistry combined with biosciences. Celtic studies, humanities & liberal arts and combined & general studies have been excluded for sample size reasons. Results have been converted to percentage differences using a log point conversion. Subjects are ranked based on estimated returns.

## Additional results

Table A5: Overall returns with and without Self-Assessment data

|  | PAYE only | PAYE and SA |
|---|---|---|
| Men | 0.05*** | 0.05*** |
|  | (0.01) | (0.01) |
| *No. of individuals* | *83,591* | *90,030* |
| Women | 0.20*** | 0.20*** |
|  | (0.01) | (0.01) |
| *No. of individuals* | *93,855* | *98,593* |

Note: Table reports derived estimates of the impact of HE on annual earnings with and without Self-Assessment data for individuals in sustained employment. This controls for HE start age, background and prior attainment controls, and uses PWRA weights. Estimates are in log points. The estimation only uses tax years 2013-14 to 2015-16 and GCSE cohorts 2002 and 2003. The estimates are different from the headline estimates because of these different sample restrictions.

Figure A3: Returns with and without Self-Assessment data, by subject, men
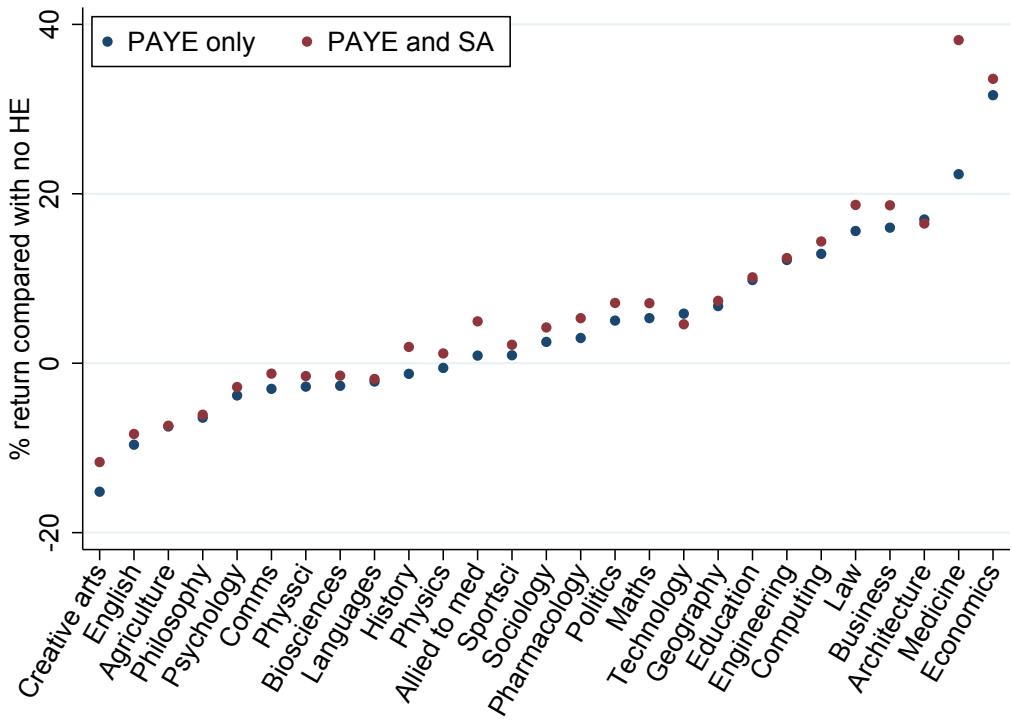


Figure A4: Returns with and without Self-Assessment data by subject, women
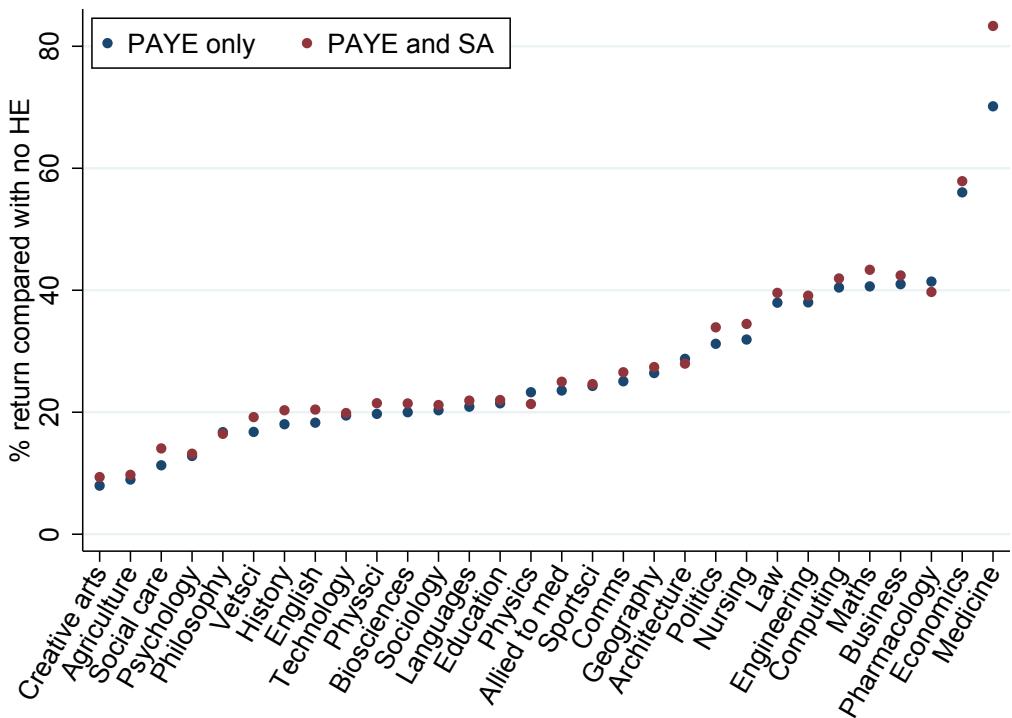
Figure A5: Age and experience returns by subject, men
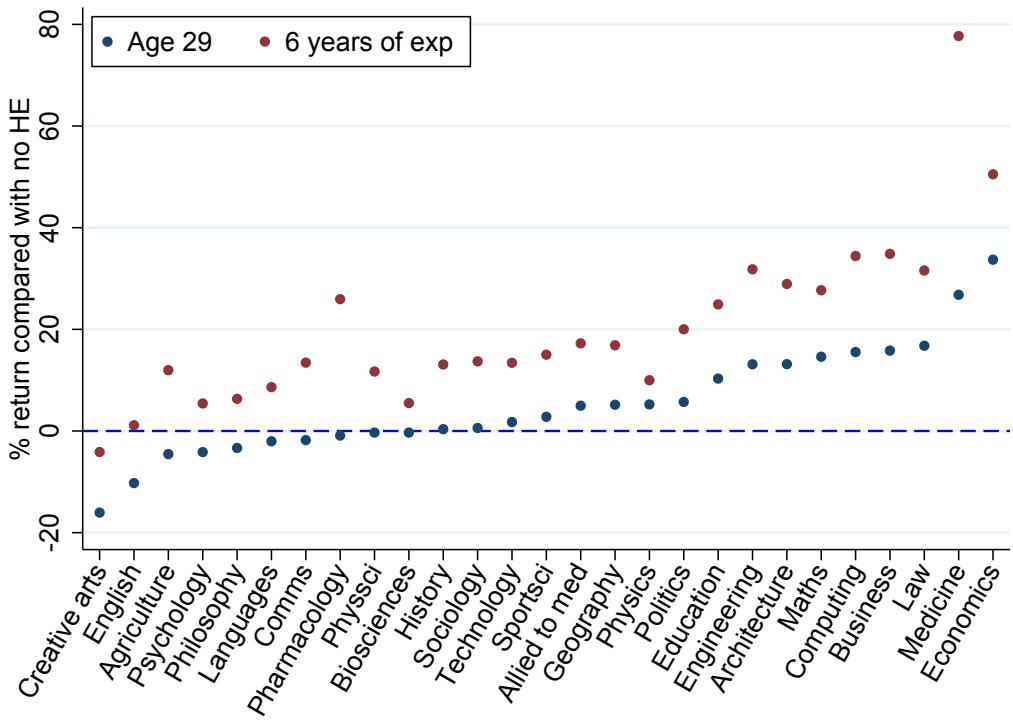


Figure A6: Age and experience returns by subject, women
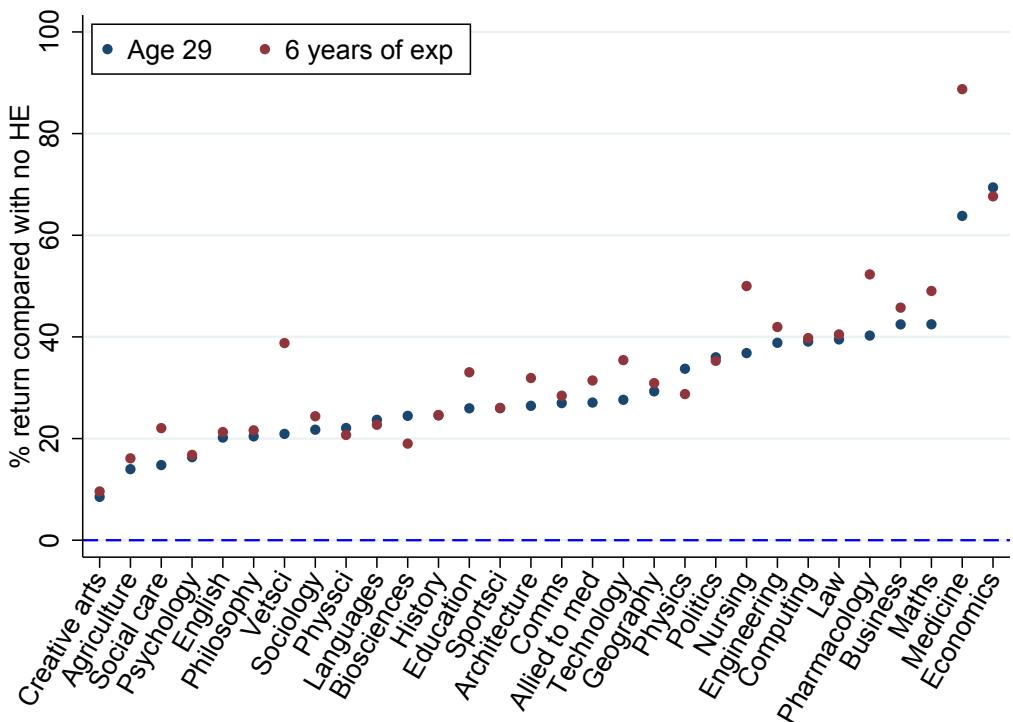
Figure A7: Comparison of our subject estimates and Belfield et al. (2018), men
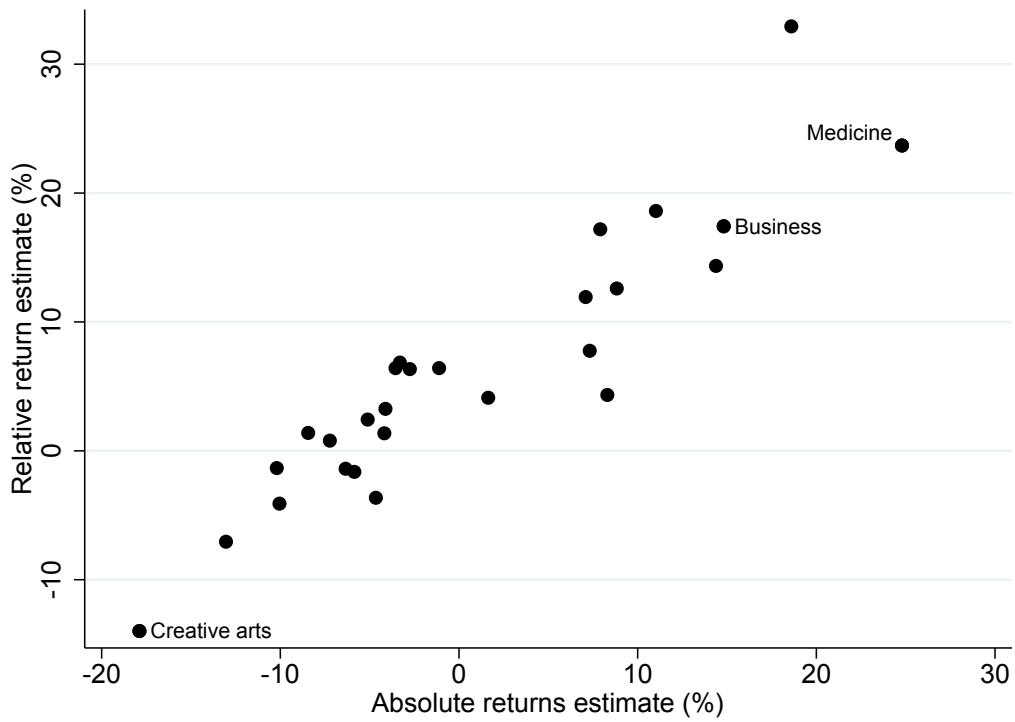
Note: Figure plots the 'absolute returns' estimates reported in Figure 16 and the 'relative returns' estimates reported in Belfield et al (2018)

Figure A8: Comparison of our subject estimates and Belfield et al. (2018), women
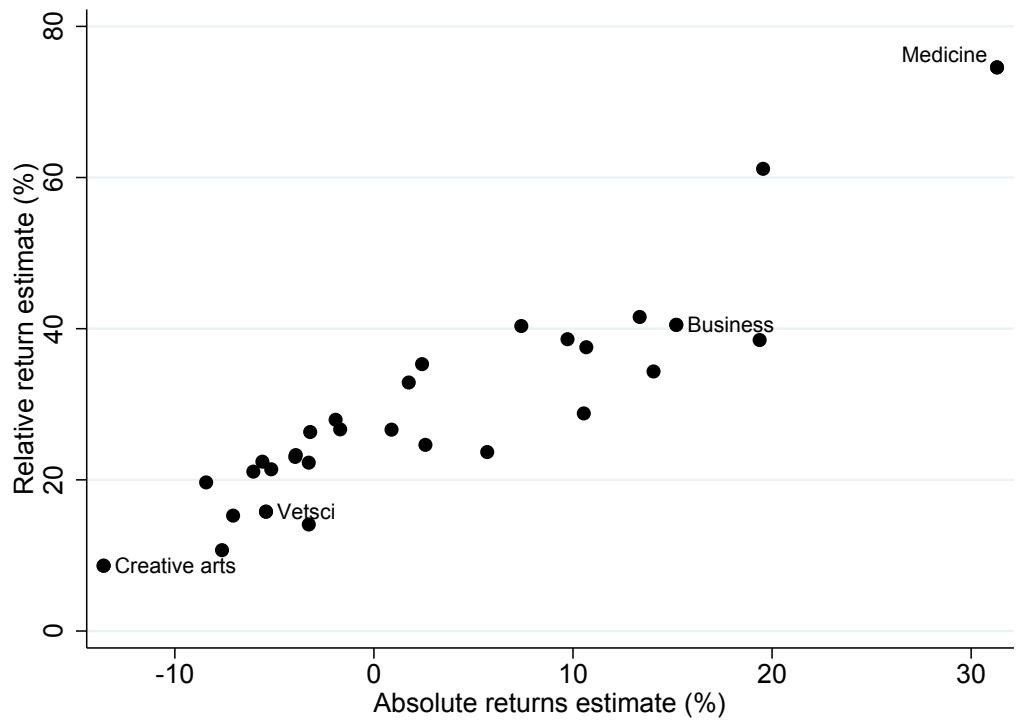


Note: Figure plots the 'absolute returns' estimates reported in Figure 17 and the 'relative returns' estimates reported in Belfield et al (2018)

Figure A9: Comparison of our HEI estimates and Belfield et al. (2018), men
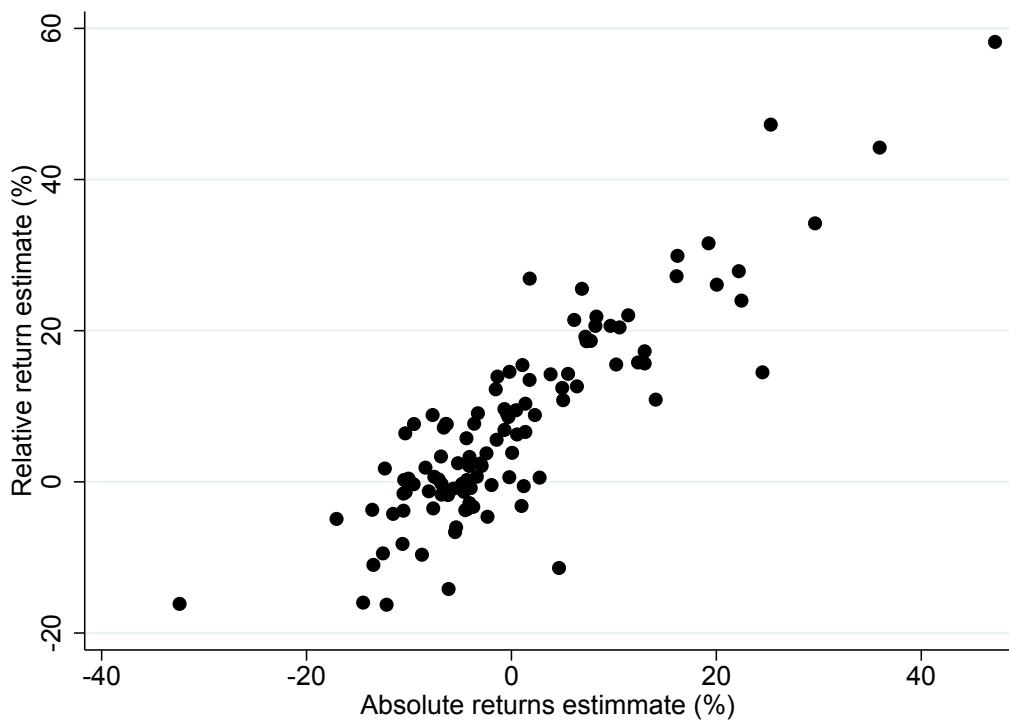
Note: Figure plots the 'absolute returns' estimates reported in Figure 24 and the 'relative returns' estimates reported in Belfield et al (2018)

Figure A10: Comparison of our subject estimates and Belfield et al. (2018), women
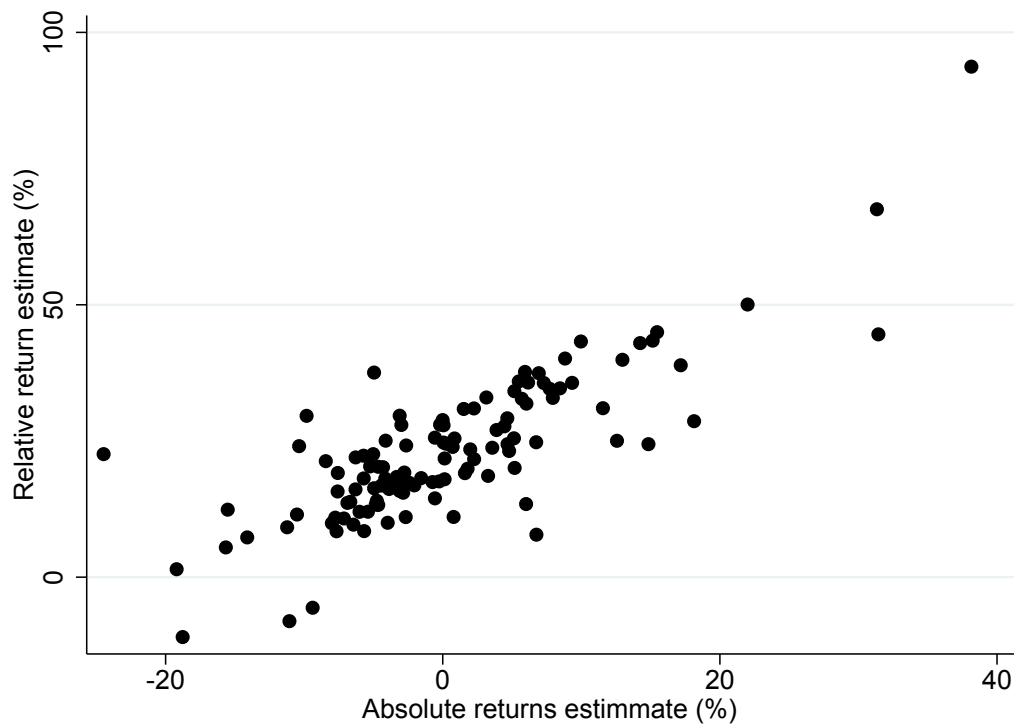


Note: Figure plots the 'absolute returns' estimates reported in Figure 25 and the 'relative returns' estimates reported in Belfield et al (2018)

Any enquiries regarding this publication should be sent to us at: jack_b@ifs.org.uk, laura_v@ifs.org.uk  or www.education.gov.uk/contactus

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