

Is improving access to university enough? Socio-economic gaps in the earnings of English graduates*

JACK BRITTON

Institute for Fiscal Studies, London

LORRAINE DEARDEN

Institute for Fiscal Studies, London and Institute of Education, University College London

NEIL SHEPHARD

Department for Economics and Department of Statistics, Harvard University

ANNA VIGNOLES

Department of Education, University of Cambridge

June 29, 2018

Abstract

Much research and policy attention has been on socio-economic gaps in participation at university, but less attention has been paid to socio-economic gaps in graduates' earnings. This paper addresses this shortfall using tax and student loan administrative data to investigate the variation in earnings of English graduates by socio-economic background. We find that graduates from higher income families (with median income of around £77,000) have average earnings which are 20% higher than those from lower income families (with median income of around £26,000). Once we condition on institution and subject choices, this premium roughly halves, to around 10%. The premium grows with age and is larger for men, in particular for men at the most selective universities. We estimate the extent to which different institutions and subjects appear to deliver good earnings for relatively less well off students, highlighting the good performance of medicine, economics, law, business, engineering, technology and computer science, as well as the prominent London-based universities.

Keywords: Administrative data; Graduate earnings; University access; Social mobility.

JEL: I23, I24, I26, J62

*Many civil servants and policy makers have helped us gain access to the data which is the core of this paper. We must thank in particular Daniele Bega, Dave Cartwright, Nick Hillman, Tim Leunig and Lord Willetts who were all crucial in making this project happen. We also thank Jonathan Cribb, Paul Johnson and participants at the RES and AASLE conferences. We solely are responsible for any errors. We also thank the British Academy (grant CH00096.0000) and the ESRC (grant ES00116.0000) and the Nuffield Foundation for financial support. The Nuffield Foundation is an endowed charitable trust that aims to improve social well-being in the widest sense. It funds research and innovation in education and social policy and also works to build capacity in education, science and social science research. The Nuffield Foundation has funded this project, but the views expressed are those of the authors and not necessarily those of the Foundation. HM Revenue & Customs (HMRC) and Student Loans Company (SLC) have agreed that the figures and descriptions of results in the attached document may be published. This does not imply HMRC's or SLC's acceptance of the validity of the methods used to obtain these figures, or of any analysis of the results. Copyright of the statistical results may not be assigned. This work contains statistical data from HMRC which is Crown Copyright and statistical data from SLC which is protected by Copyright, the ownership of which is retained by SLC. The research datasets used may not exactly reproduce HMRC or SLC aggregates. The use of HMRC or SLC statistical data in this work does not imply the endorsement of either HMRC or SLC in relation to the interpretation or analysis of the information.

1 Introduction

Higher education is seen as a potentially crucial tool for social mobility, providing a possible route for students from lower income family backgrounds to achieve labour market success and higher earnings. Consequently there have been numerous government policies around the world focussed on improving access to university degrees for those from poorer households. However, there is relatively little evidence on whether this should be the primary focus of governments trying to improve social mobility.

Consistent with most countries around the world, in England educational achievement and higher education access varies substantially by the level of parental income, with many fewer students from poorer backgrounds attending university, particularly the highest status institutions (Chowdry et al. (2013), Ermisch et al. (2012)). However, little is known about the differences in earnings between graduates from poorer and richer family backgrounds. Further, primarily due to data limitations, the question of whether differences in earnings still exist conditional on university and subject choice, has remained largely unanswered.¹ In this paper we are able to address these shortfalls in the literature by making use of a unique administrative database that tracks the earnings of graduates into their mid thirties.

We use a dataset that consists of anonymised individual level-administrative taxable earnings data supplied by Her Majesty’s Revenue and Customs (HMRC), linked to information on students’ higher education (university or college) from the English Student Loan Company (SLC). The latter is an institution supported by the state to provide loans to students to fund their higher education. The HMRC and SLC data sets are hard linked using a national identification number (National Insurance number²) and we have access to a 10% random sample. We study cohorts of students who entered higher education from 1999-2005, and focus on the same students’ earnings between 2008/09 and 2013/14. This allows us to follow graduates through their most crucial career developing years and well into their thirties. We also use Higher Education Statistics Agency (HESA) data which we can match at the subject-institution (rather than individual) level. This includes the socio-economic background and prior academic achievement of the students studying the same subject in the same institution. This allows us to add further controls that capture differences in the demographics of students in a given university and subject, although we acknowledge that this does not eliminate ability bias in returns or deal with differential selection into courses across individuals from different socio-economic backgrounds.

¹The exceptions include a number of papers that investigate returns to private versus state secondary school education in the UK, conditional on university education (for example Crawford et al. (2016)), and Chetty et al. (2017) which investigates variation in returns to attending university by parental income in the US.

²This is the key individual identifier for all taxes, social security and student loans.

A common problem with administrative data is a limited set of background characteristics for individuals.³ We also face these limitations, and do not directly observe parental income for individuals in our sample. However, we are able to infer a simple binary measure of parental income based a student’s SLC record, which notes the amount each student borrowed in their first year of study. For English students starting university before 2006, the amount individuals were eligible to borrow was linked in a monotonic way to their parental income. We identify people as being from a higher income household if they are borrowing exactly the maximum amount an individual from a higher income household is eligible for in their first year of study.⁴ This consists of approximately 20% of borrowers, which in the paper we refer to as the richer group. The remaining 80% of borrowers are of course relatively poorer, rather than poor in an absolute sense. Indeed, based on a sample of borrowers in the Family Resources Survey, we estimate that the median parental earnings of these groups is around £77,000 for the richer group and around £26,000 for the rest (2018 prices).

Clearly our parental income measure is likely to have issues with measurement; people from poorer households might borrow the rich maximum, people from richer households might not borrow the rich maximum, and we are unable to say anything at all about the roughly 15% of people who attend university but choose not to borrow, which is likely to be weighted towards those from higher income households. Given these measurement issues - all of which are likely to bias down our estimates - we find considerable differences in earnings between graduates from richer and relatively less well off family backgrounds. These differences roughly halve once we condition on subject and institution choices but remain economically important at around 10%, and are statistically significant. These socio-economic differences also exist right through the earnings distribution and are larger at the bottom and top of the earnings distribution, suggesting family wealth is particularly good at both protecting graduates against very poor outcomes and providing them with opportunities for very high earnings. The conditional differences grow with age and are somewhat smaller for Science, Technology, Engineering and Mathematics (STEM) or Law, Economics and Management/Business (LEM) as compared to other subjects. They are also particularly pronounced for men from the most selective universities.

These findings are descriptive, but clearly important for policy. Data limitations mean we are unable to control for: individual-level qualifications;⁵ degree outcomes, such as completion and

³The availability of linked administrative data has improved dramatically in the UK in recent years. The Longitudinal Educational Outcomes data (LEO) allows the linkage of entire education histories of individuals to their earnings records. However, these data are currently available only for government research.

⁴There were subsequent changes to both tuition fees and student support that took effect from 2006 - see Section 3 for more detail. These changes do not affect our results, however, as we focus on the first-year borrowing of people who entered university before 2006.

⁵The period we are looking at was before the big increase in ‘contextualised admissions’ policies whereby universi-

degree classification (i.e. grades); progression onto (and timing of) postgraduate study; and early career occupation choices. These, along with differences in non-cognitive skills and the networks of those from richer and poorer backgrounds should be the subject of future research into understanding the drivers of these earnings differences, and could have important implications for firms, universities and policy.

Finally, we follow Chetty et al. (2017) by estimating ‘social mobility scorecards’, which measure the extent to which different universities appear to help students from relatively poorer backgrounds get into the top fifth of graduate earners (specifically the ‘mobility score’ is the probability of a course admitting a poorer student multiplied by the probability that the student goes on to enter the top fifth of the earnings distribution). Our parental income measure is less rich than that used by Chetty et al. (2017), who focus on the bottom 20% of the observed parental earnings distribution, and consequently our results are not directly comparable. However, unlike Chetty et al. (2017) we are able to estimate mobility scorecards for different subject disciplines. We find that medicine and economics are particularly good at delivering relatively poorer students into the top 20% of the graduate earnings distribution. However, it is not clear that all STEM subjects are broadly effective at delivering this. On the other hand, we find that LEM subjects are effective. More broadly, professional facing subjects (e.g. LEM, computer science, engineering, technology, business) seem to deliver routes to social mobility. At the other end of the scale, biological sciences, mass communication and creative arts subjects do this to a much lesser extent. For institutions, the high-profile London universities - namely the LSE, Imperial College, King’s College and UCL - do very well by this index, while outside of London, Warwick and Manchester are two of the best performing universities from the set we have permission to name.⁶ These results are necessarily descriptive only and come with several caveats. However, they represent the first descriptive evidence on which institutions and subjects are best for encouraging social mobility.

The paper is laid out as follows. In the next section we discuss our contribution to the existing literature. In Section 3 we outline the institutional details of Higher Education in the UK. In Section 4 we describe our data and introduce our measure of parental income. In Section 5 we presents results from our modelling. In Section 6 present social mobility scorecards by subject and institution. Section 7 concludes.

ties make lower offers to students who had attended certain schools, typically those in poorer neighbourhoods. This suggests it would be more important to control for individual qualifications for later cohorts.

⁶Given HMRC rules about anonymity, we can only name institutions which gave explicit permission, which we received from 18 institutions (see Section 6 for the full set of these).

2 Existing literature

This work will contribute to an important literature that has suggested a major impact from higher education on individuals' earnings. We focus on the English graduate labour market (Blundell et al. (2005)), though our findings are also relevant to the large US literature which has looked at the heterogeneity in graduate earnings by subject and institution (see, for example, Dale and Krueger (2014) as well as Webber (2014) and Altonji et al. (2015) for reviews). The evidence of a sizeable graduate wage premium for English graduates is convincing, see for example (Walker and Zhu (2011)). Yet although higher education in England appears to be a good investment for many, as is the case for the US, there is also a sizeable empirical literature that has shown substantial variation in graduate earnings that has increased over time (Chevalier (2011), Hussain et al. (2009), Sloane and O'Leary (2005), Smith and Naylor (2001), Walker and Zhu (2011)). A key question is therefore, given this increased diversity in graduates' earnings, whether students from poorer backgrounds achieve the same earnings gains compared to their similarly qualified counterparts who come from more richer families.

Differences in earnings between graduates from poorer and richer family backgrounds may of course be attributable to differences in the institutions they attend and the subjects they study. Previous work has shown that graduate earnings vary considerably by subject of degree (Sloane and O'Leary (2005), Chevalier (2011), Walker and Zhu (2011), Walker and Zhu (2013), Chowdry et al. (2013)). Walker and Zhu (2013) suggested substantial differences in private returns by degree subject and insignificant differences in returns by institution type (the data were insufficiently granular to analyse at institution level). Britton et al. (2016) also found considerable variation in earnings by both subject and institution, though much of this difference is attributable to different prior achievement levels of the students taking different degree options. Since prior achievement levels are lower, on average, for poorer students, we would expect sorting by subject and institution to depress their earnings.

Even with similar subject field and institution choices, an individual's socio-economic background may have an effect on their labour market outcomes after graduation. This might be because students from more advantaged backgrounds have higher levels of (non-cognitive) skills (see for example Blanden et al. (2007) and Kassenboehmer et al. (2018)) that are not measured by their highest education level, or by their degree subject or institution. Related to this, performance in the degree could be important. Crawford et al. (2016) show that students from poorer socio-economic backgrounds are less likely to complete a degree and are less likely to graduate with a top grade than their more wealthy peers. We do not observe any non-cognitive skills or degree outcomes (grades or whether a student completes the degree) in our dataset.

Alternatively, advantaged graduates may earn more because they have greater levels of social capital and are able to use their networks to secure higher paid employment. The literature in the UK at least does suggest that graduates from more advantaged backgrounds, particularly privately educated students, achieve higher status occupations and there is some evidence that privately educated students earn a higher return to their degree (Bukodi and Goldthorpe (2011b), Bukodi and Goldthorpe (2011a), Macmillan et al. (2013), Crawford and Vignoles (2014)). For example, Crawford and Vignoles (2014) found that graduates who attended private secondary schools earn around 7% more per year, on average, than state school students 3.5 years after graduation, even when comparing otherwise similar graduates and allowing for differences in degree subject, university attended and degree classification. This is consistent with earlier work using data from the 1970s and 1980s by Dolton and Vignoles (2000) that found the earnings return for graduates varied according to whether the individual attended a private school or a state school. This research also found that the private school wage premium for graduates who left university in 1980 was 7% for males but there was no premium for females, conditional on subject of degree and institution. Similar results were found by Naylor (2002) for a cohort of 1993 graduates (3% wage premium) and Green et al. (2012) using the National Child Development Study 1958 cohort and the 1970 British Cohort Study. The latter found that the private school wage premium increased from 4% for the earlier cohort to 10% for the later one. By contrast, work on how graduates' earnings vary by parental income level or parental socio-economic status, rather than by whether they attended private school, is more limited. For example, using the British Cohort Study (BCS) Bratti et al. (2005) found little evidence of variation in the return to a degree by social class.

Beyond the UK, there is an impressive body of work that has drawn on administrative data largely from Scandinavian countries (and some US states) to investigate the relationship between parental income and children's outcomes Figlio et al. (2015). Much of this work estimates causal impacts of parental income or education on children's educational outcomes (e.g. Black et al. (2005)). There is less work on the extent to which parental earnings impact on graduate's earnings, conditional on the nature of the higher education achieved. Perhaps the most relevant paper in this body of literature is Chetty et al. (2017) which looks at this issue for the US using administrative tax data linked to data from the National Student Loan Data System for around 30 million individuals who were university students between 1999 and 2013. Their study has the advantage of granular information on both parent and child income (the former measured when the student was aged 15-19 and the latter when the student was 32-34). From this, they were able to construct intergenerational income correlations for graduates from different institutions. They found stark differences in the likelihood of poor students accessing elite institution. For instance, a student with parents in the

top 1% of the income distribution is 77 times more likely to go to an Ivy League university than those with parents in the bottom fifth of the income distribution. However, they also concluded that students from poorer and richer backgrounds did similarly well if they graduated from the same college. At least for those who are able to gain access, universities appear to be levelling the income playing field in the US. Our study has key differences from Chetty et al. (2017). First, our measure of parental income is binary which is a clear limitation. Second, unlike Chetty et al. (2017) we are able to control for subject of study at the individual level, which is important given the evidence on variation in earnings by subject and the early subject specialisation in the English system which differs markedly from the broader curricula of the average US bachelors degree. Third, in England it is less likely, on average, for wealthy but low achieving students to gain access to elite institutions. This is potentially due to differences in the HE admissions and funding systems, with the English system at this time arguably presenting fewer barriers to access compared to the US system. Admission in England is centralised and regulated, with the probability of entry into elite institutions closely correlated with students' prior achievement in national examinations taken at age 18 (A levels or equivalent). English tuition fees were also comparatively very low during this period and were income contingent, so students from poorer households could be exempt from paying. This point is reflected in evidence for England (Chowdry et al. (2013)) which found that conditional on prior achievement, there was no socio-economic gap on entry into HE and a gap of just a few percentage points on entry into elite universities. These different institutional arrangements may mean that the socio-economic selectivity into HE, and particularly elite institutions, is somewhat different in the two countries which will impact on graduates' earnings, especially given that both Chetty et al. (2017) and our own study are limited by not having individual level measures of skill or IQ. Hence the analyses in both papers is necessarily descriptive.

3 Institutional Background

During the period of study, the minimum school leaving age in England was 16, although comfortably more than half of students stayed in school until age 18. The majority of those that progress on to university do so within the first two years of leaving school. The vast majority of university degrees are in one subject (or sometimes two subjects combined) and take three to four years. Subject specialisation therefore occurs relatively early by international standards (and, in particular, compared to the US). It is very common in England to move out of the family home for university, and the government has been loaning money to students to help with their living costs during study since the 1980s.

The English Student Loan Company (SLC) was introduced in 1990 to administer a reformed version of these ‘mortgage-style’ living cost loans for English students attending a higher education in the UK. There were no tuition fees at the time. The mortgage-style nature of the loans meant that repayments were in equal instalments that were independent of student’s subsequent income.⁷

In 1998, means-tested tuition fees of up to £1,000 per year (1998 prices) were introduced for the first time, with fees payable up front. Alongside this, the living cost (‘maintenance’) loans now became income contingent, so that individuals were automatically deducted 9% of their income above a threshold (initially this threshold was £10,000, though it increased on several occasions since) by the tax authority (HMRC). Any outstanding loans were written off when the individual turned 65.⁸ Interest rates on student loans were set equal to the lowest of the Bank of England base rate plus 1% and the RPI measure of inflation. This is the regime that all of the students in our estimation in this paper were facing, namely the cohorts of students which started university between 1999 and 2005 who borrowed from the English SLC.

Crucially for our design, maintenance loans eligibility was dependent on parental income.⁹ All individuals were eligible for some loan, but people from lower income households could borrow more. People who wanted to borrow more had to prove their income in the previous year by submitting their end of year tax statement (‘P60’) to HMRC. This meant it was difficult for people from higher income households to gain access to the larger loans. People could also borrow more if they lived in London during their studies (due to the higher living costs), while there was a different cap for those living with their parents while studying.¹⁰ See Table 2 in the following section for the non income-assessed maximum loan amounts inside and outside London.

Subsequently, there have been further changes in England. In 2006, fees were increased to £3,000 per year (2006 prices), although students could now borrow this money from the SLC to add to their student loans. Alongside this, there were changes to the rules for maintenance loan eligibility. Prior to 2006, individuals from poorer backgrounds could borrow the most. From 2006, the relationship became non-monotonic, as maintenance loans were increased as grants were tapered, which resulted in students from middle-earning families borrowing the most.

Combined, these changes make it very difficult to identify poorer individuals in the data from

⁷Borrowers were eligible to start making repayments once they started earning more than a certain threshold (85% of average annual earnings for full-time workers). They could also defer payments if they earned less than that amount in a given year.

⁸In 2006 the write-off period was reduced to 25 years from leaving HE. In 2012 it was lengthened again to 30 years from leaving HE.

⁹Other forms of financial support, including cash bursaries and hardship loans were available during this period. These are unobservable to us, but fortunately did not affect loan eligibility, which means they are unimportant for our identification of richer and poorer individuals.

¹⁰See the Statistical First Release series from the Student Loan Company on student support for higher education, for example SLC SFR 03/2005.

2006 onwards. This is primarily because the poorest students could now borrow almost exactly the same amount as the richest students (although they did receive additional living cost support through maintenance grants), but also because not everybody borrowed the tuition fee loans. Individuals who start under a given regime stay in that regime (so, for example, if you start a degree in 2005, you are not eligible to pay £3,000 fees in 2006). In any case, we identify poorer individuals based on the amount they borrow in their first year of study only, which avoids issues with borrowing different amounts in different years, or variable course length. The first year borrowing of people starting university between 1999 and 2005 is therefore completely unaffected by the 2006 reforms.¹¹

There were further considerable changes to the English system in 2012;¹² fees were trebled to £9,000 per year (2012 prices), interest rates were increased to RPI plus up to 3% and there were a number of other changes to the repayment conditions. Subsequently maintenance grants for poorer students were abolished again, which meant a return to the situation of the poorest students borrowing the most. Again, these changes do not affect our results directly, but it is worth keeping in mind that the system is now very different to the one in place during our period of analysis.

We were unable to gain access to equivalent student loan data for the rest of the UK which are administered by separate bodies. We therefore do not observe students from Wales, Scotland or Northern Ireland. Higher Education is a devolved policy area, which means there is now considerable variation in policy across the UK. However, in the period we are interested in (1999-2005 starters), this is less true; at this point the different systems were quite similar. Although non-UK European Union residents were also eligible to borrow from the English SLC, we do not observe them in our dataset either.

4 Data

This is an exciting new dataset for investigating graduate outcomes. Other UK surveys, such as the Labour Force Survey (LFS) and the Destinations of Leavers from Higher Education (DLHE) survey, have information on subject of study and institution. However, information on higher education institution has only recently been collected by the LFS, limiting the sample sizes available to researchers. The LFS also has only very limited data on the parents of graduates. Meanwhile, the DLHE does have information on graduates' earnings by subject and institution but has issues with

¹¹Focussing on first year borrowing also means that we are unaffected by the fact that course length is variable (typically 3-5 years).

¹²Another important change is the large increase in 'contextualised admissions', whereby universities make lower offers to students who had attended certain schools, typically those in poorer neighbourhoods. This was not highly prevalent during the period we are investigating, but it suggests both that earnings gaps might change for later cohorts and also that conditioning on qualifications on entry might be a crucial addition for these cohorts.

sample selection (it is a voluntary online survey) and only captures full time equivalent earnings just three and a half years after graduation. Our data by contrast is able to provide insight into graduates' earnings up to more than a decade after graduation. More extensive detail on the dataset is provided in Britton et al. (2018). We have a 10% sub sample of all borrowers from the English part of the SLC, which means they had to be domiciled in England upon application to university and attend a university in the UK. We have data on those who entered higher education between 1998-2008 but focus on the 1999-2005 entrants (henceforth, 'cohorts') because of the low uptake of loans in 1998 (driven by the slow transition into the income contingent loan system) and the availability of tuition fee loans and maintenance grants after 2005 (see discussion above).

These data provide us with information on gender, first year of study (cohort),¹³ institution attended,¹⁴ field of study,¹⁵ region on application to higher education and a detailed measure of income from employment (Pay As You Earn taxable income) and from self employment (Self Assessment income). We do not observe degree outcomes, which means we do not see degree classification or indeed completion. This could of course be important, although we note that dropout rates are low by international standards at around 10%.¹⁶

We focus on earnings data from the tax years 2008/09 through 2013/14. We use earnings from labour, meaning employment income, profits from partnerships and profits from self-employment are included. We exclude trust income, profits on share transactions, profits from land and property, income from foreign employment, savings, UK dividends, pension income, life policy gains, 'other' income, bank and building society interest. Clearly we are focussing on a period that follows the 2008 recession, which should be kept in mind when considering the results, as it may have implications for the magnitudes of the effects that we see. For example, wealthier students might be more likely to partake in postgraduate study during the start of the recession and that may boost their income subsequently compared to their less well off peers. Unfortunately we are unable to use other years of earnings data.

The sample sizes for our cohorts of interest are given in Table 1, which also shows the gender split. These samples reflect 10% of English borrowers at UK Higher Education Providers. These

¹³For people who switched degrees we observe their second degree course. The total debt figures include previous borrowing, but the 'first year borrowing' that we use is from the first year of the course we observe them studying.

¹⁴Students in officially recognised UK higher education learning institutions are eligible for loans. The government defines these as either 'recognised' or 'listed'. The former can award degrees and the latter can offer courses that lead to a degree from a recognised institution. We observe students at both types of institution meaning some Further Education Colleges will be included. Overall there are several hundred of these, although we observe 170 distinct institutions with the rest classified as 'other' institutions.

¹⁵We observe the first digit of the 'JACS' code, which is a broad subject level classification set by HESA. JACS codes at this level include a heterogeneous range of courses. For example 'biological sciences' ranges from psychology to biology. Whilst we would ideally control for subject of study at a more granular level, this was not possible for disclosure reasons. If there is lots of variation by background within the JACS code measures, this could affect our results (e.g. poorer students might choose courses within the JACS bands that have lower earnings potential).

¹⁶See <https://www.hesa.ac.uk/data-and-analysis/performance-indicators/non-continuation>

sample sizes align with overall numbers from the Higher Education Statistics Agency (HESA) for the same period. There are more women, reflecting the higher participation rates of women in the UK (rather than different borrowing behaviour). Note that we use up to six years of earnings data for each individual throughout the majority of this paper.

Cohort	All	Men	Women
1999	22,621	10,590	12,031
2000	23,506	10,853	12,653
2001	23,924	11,025	12,899
2002	23,891	11,060	12,831
2003	23,972	11,024	12,948
2004	23,577	10,767	12,810
2005	25,103	11,439	13,664

Table 1: Number of graduates (10% sample of loan database), by cohort and gender. Cohort denotes the first year the individual received a loan from the SLC.

The administrative data described above is linked to data from HESA. Whilst we cannot link data at the individual level, we are able to do it at the institution and subject level. This provides a quantitative profile of the characteristics of students in each institution-subject combination. These data enable us to control more effectively for the characteristics of students attending different institutions and taking different subjects. This is important if we are trying to identify the residual correlation between socio-economic background and subsequent earnings after allowing for the fact that poorer students take different degree options. These data also allow us to control for the government region in which the student’s institution is located, which is important since wages vary by region and we do not have data on the graduates’ current location (current region is in any case endogenous since graduates with degrees that are more highly valued in the labour market may be better able to secure high paying jobs in high paying regions). Since a high proportion of graduates remain near their university when they enter the labour market, controlling for region of institution goes some way to account for this issue. We use HESA data from 2002/03. The key characteristics which we can control for (all averaged at the subject-institution level) are: UCAS ‘tariff score’,¹⁷ ethnic composition, gender composition and measures of students’ socio-economic status. The latter include parental occupation, the percentage of students living at home whilst studying, the percentage of students who attended an English state school (i.e., non-private) and the ‘Participation of Local Areas’ (POLAR) classification (neighbourhood level participation in higher education by age 19).

¹⁷The tariff score is a single quantitative summary of the performance of students prior to entering university in national tests taken at Advanced level (A level) or equivalent at age 18.

4.1 Creating our measure of parental income

Our focus is on how graduates' earnings vary by socio-economic background of the student. Unfortunately the data do not include a direct measure of parental income. As discussed above, during this period, the SLC loaned English domiciled students at UK universities money to help with their living costs. Crucially the amount loaned varied by parental income. Our database includes the amount borrowed by each graduate for their student loan overall and in their first year of borrowing. We are able to use this to make an inference about the parental income of each individual because the maximum amount the UK Government was willing to loan a student depended on their parents' income, with individuals from lower income households able to borrow more than their more well-off peers. As discussed, for the 1999-2005 cohorts that we investigate, there was a monotonic relationship between how much individuals could borrow and their parental income, with the students coming from the poorest households able to borrow the most.

There is a lot of noise in the observed amount individuals borrow. However, for each of the 1999-2005 cohorts we observe clear spikes at points in the distribution that we are able to exploit. To explain this, we provide an illustrative density plot in Figure 1. This shows the distribution of the amount individuals borrow in their first year of study, where x is the maximum an individual from a higher income household can borrow, or the 'unassessed maximum'. The plot is normalised so x is set to 0 to allow for the fact that the maximum amount changes each year and differs for individuals studying inside and outside London to allow for the fact that the borrowing limits are higher in London. People borrowing more than x need to provide evidence of their parents' earnings from the previous tax year to the SLC.

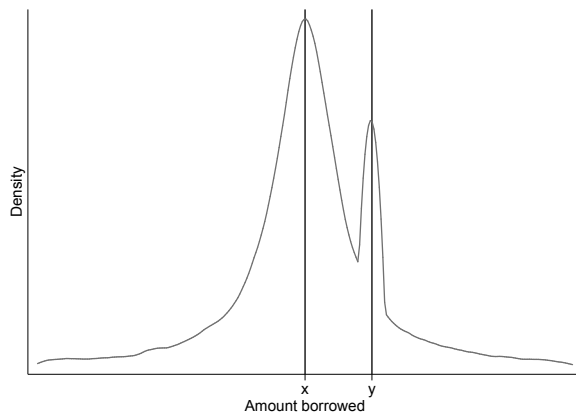


Figure 1: Illustrative density plot of amount borrowed. x represents the higher income maximum, while y represents the lower income maximum. Amounts and densities deliberately excluded for disclosure reasons.

The exact loan amounts for each year, the minimum parental income threshold and the share of individuals at different points in the distribution x are given in Table 2. The biggest spike in the

distribution is at exactly $\pounds x$, with the share at this point increasing from around 15% in the 1999 cohort to around 25% in the 2005 cohort and averages around 20% across all cohorts. Although not shown here, the distributions are very similar when split by gender, but with slightly more men borrowing exactly x . The next biggest spike is at exactly $\pounds y$, where between 10 and 15% are borrowing exactly the overall maximum.

We also see from the table that around one-third of borrowers borrow less than x and around 20% borrowers borrow between $\pounds x$ and $\pounds y$. We also see around 20% of borrowers above the official maximum. These individuals are most likely lower income individuals studying courses with longer than standard term lengths.

Cohort	Parental Income	Loan Amounts**		Distribution of borrowing (b)				
	Threshold* (£)	x (£)	y (£)	$b < x$	$b = x$	$x < b < y$	$b = y$	$y < b$
1999	35,000	2,795 (3,445)	3,635 (4,480)	33.1	14.6	20.9	9.5	21.8
2000	36,000	2,795 (3,445)	3,725 (4,590)	31.8	18.9	17.2	10.8	21.3
2001	38,500	2,860 (3,525)	3,815 (4,700)	30.4	21.4	17.0	10.8	20.4
2002	40,000	2,930 (3,610)	3,905 (4,815)	30.2	21.8	17.1	11.6	19.3
2003	40,000	3,000 (3,695)	4,000 (4,930)	29.5	23.8	14.9	13.2	18.6
2004	40,950	3,070 (3,790)	4,095 (5,050)	27.8	24.8	16.0	14.4	17.0
2005	41,950	3,145 (3,885)	4,195 (5,175)	26.5	25.8	16.1	13.3	18.3

Table 2: Loan availability and borrower distribution by loan amounts. *This the minimum parental income someone can have to qualify to borrow more than x . **Loan amounts for people studying at a university in London are given in the parentheses.

Using this measure of borrowing we infer a blunt measure of parental income that we set equal to one (indicating high parental income) if the individual borrows exactly x in her first year, and zero otherwise (indicating low parental income). Based on data from the Family Resources Survey (FRS), we were able to approximate the average earnings of our two groups. Taking the set of 18-21 year olds living with their parents and borrowing a student loan between 2002 and 2005, we observe that the average parental earnings of those above the threshold for extra loans was around $\pounds 77,400$, while the average parental income of those earnings below the threshold was $\pounds 25,900$ in 2018 prices.¹⁸ Around 30% of borrowers are above the parental income threshold, which is not dramatically more than the 25% we observe at the rich maximum in 2004.

We acknowledge that this measure of parental income does not perfectly identify all student

¹⁸Based on around 1,000 borrowers. These numbers include the parents' income from employment and any other private sources, including private pensions and investments. It does not include state benefits (which are not included in student loan assessment) or state pensions (which are, but only a small fraction of parents in the sample are old enough to be eligible). We thank Jonathan Cribb of the IFS for these calculations.

from higher income households, for a number of reasons. First, those from higher income households may borrow less than the maximum available. Second, individuals from lower income households may choose to only borrow the higher income maximum because they do not want to borrow more or are unable to provide evidence of their parent’s income. Third, we are missing altogether those individuals from the wealthiest households who did not borrow at all. While this figure is around 15% of the overall student population, it is likely to represent considerable fractions of the student populations at some high-status institutions in particular. Fourth, there may be misreporting of parental income to the SLC, though they do require official proof of income to gain access to additional loans. Whilst we cannot completely overcome these weaknesses in our measure, we do provide indicative evidence below that it does indeed identify individuals from more wealthy households. Further, we suggest that most of these issues with the measure are likely to bias our impacts towards zero.

4.1.1 Validation of the parental income indicator

Here we investigate whether our simple indicator is indeed picking up higher income individuals by showing how it relates to university access and voluntary repayments. First, we show the share of higher income students in different types of institutions. We know that poorer students on average access less selective universities where the mean entry tariff score is lower. We divide all the universities in our database up into deciles based on the mean entry scores of their students, taken from HESA data. We split the top 10% of universities into two groups to identify the most elite top 5% of universities since this group is of particular policy interest given their very high earnings (Britton et al. (2016)) and their relatively low shares of poorer students. In Figure 2 we plot the share of higher income students (conditional on being borrowers) in each of these university groups, by gender. It is clear that for both men and women, universities with higher entry criteria have much higher shares of individuals we define as being from a higher income household. In the most selective universities, more than half of students come from the 20% of individuals we define as being higher income.

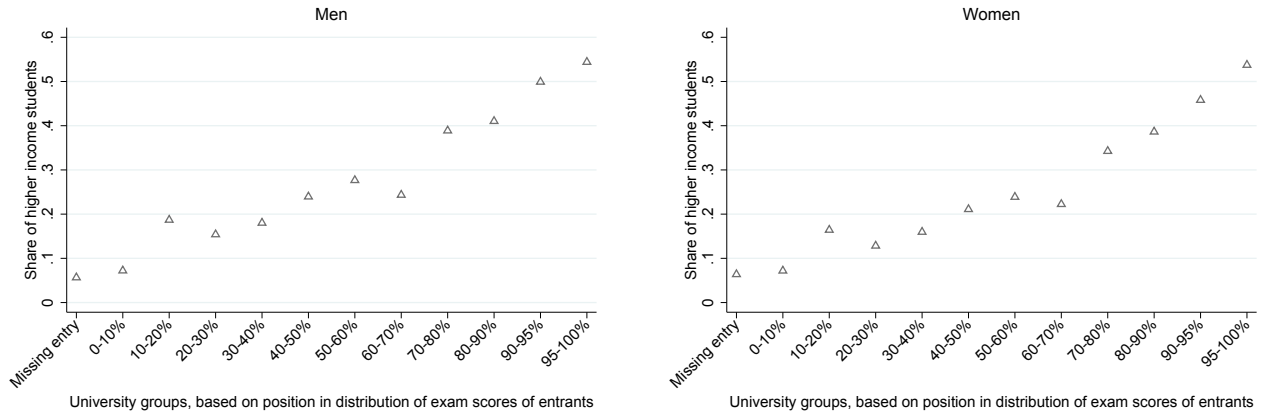


Figure 2: Share of individuals in each university group at the higher income HH borrowing amount. Includes the 1999-2005 cohorts. Shares incorporate borrowers only.

Second, we examine the voluntary repayments of students. These are repayments of student loan amounts that are made direct to the SLC over and above the legally required repayments that are determined by the graduates' income level. Given the loan forgiveness and the low real rate of interest faced by the cohorts we are investigating, voluntary overpayment does not appear to be an optimal strategy for graduates. However, a summary of voluntary repayments is given by gender in Table 3, and clearly a significant amount of repayments occur, possibly due to debt aversion¹⁹ or to avoid overpayment.²⁰ We are interested in these repayments because conditional on the graduates' own level of income, they may be more likely to be made by those from wealthier families who can afford such lump sum payments. From Table 3 we see that around 9% of students make voluntary repayments at some point between starting university and 2011, the final year we have data on voluntary repayments. The mean annual repayment (conditional on making a repayment) amount is around £2,500. A marginally higher share of women make repayments than men, and women on average make more voluntary repayments, with 34% of those making any repayments making more than one, versus 29% for men. However average repayments are typically smaller for women than they are for men.

¹⁹Recently some lenders, including mortgage lenders, do take account of the presence of student debt when making lending decisions, which might make it desirable to pay off student debt more rapidly, but this was not very common in the period we are investigating.

²⁰There have been incidents of this that were widely reported in the British press. They occur due to slow communication about repayments and outstanding debt between HMRC and the SLC. People with variable income are the most vulnerable to this - in practice all overpayments are refunded by the SLC, although the process can be very slow.

	Men			Women		
	Mean	S.D.	N	Mean	S.D.	N
Share ever making repays	0.08	-	76,776	0.09	-	89,818
Share making > 1 repayment	0.29	-	6,460	0.34	-	7,991
Share making > 2 repayments	0.13	-	6,460	0.16	-	7,991
Share making > 3 repayments	0.06	-	6,460	0.08	-	7,991
	£			£		
Average repay overall	2,840	4,071	9,544	2,490	3,785	12,632
Average repay 4 years from starting	5,058	4,696	419	5,041	4,613	608
Average repay 5 years from starting	4,720	5,222	813	4,037	4,913	1,182
Average repay 6 years from starting	3,817	4,795	1,015	3,208	4,399	1,485
Average repay 7 years from starting	3,146	4,422	1,249	2,478	3,802	1,722
Average repay 8 years from starting	2,494	3,668	1,224	2,144	3,491	1,684
Average repay 9 years from starting	2,407	3,715	1,283	2,066	3,383	1,740
Average repay 10 years from starting	2,292	3,413	1,406	1,891	3,071	1,712
Average repay 11 years from starting	1,873	2,956	1,042	1,776	2,906	1,245
Average repay 12 years from starting	1,998	3,269	716	1,503	2,478	831
Average repay 13 years from starting	1,577	2,789	377	1,616	2,791	423

Table 3: Voluntary repayments summary statistics

In Table 4 we estimate the probability of individuals from higher income households making any voluntary repayments. We estimate a probit model with a dummy set equal to one if an individual makes any repayment in a given year. The results show that individuals we classify as being from a higher income household are significantly more likely to make voluntary repayments, even conditional on their current earnings. They are about one percentage point more likely to make voluntary repayments, on a baseline of 3.3%.

	Repayments	Add Gender	Add Earnings	Add HESA
	[1]	[2]	[3]	[4]
Higher Income HH	.189*** (.007)	.191*** (.007)	.182*** (.007)	.142*** (.007)
$P(\text{repays} > 0 \text{Lower income HH})$.033	.033	.033	.033
$P(\text{repays} > 0 \text{Higher income HH})$.048	.048	.047	.044
N	666,376	666,376	666,376	666,376

Table 4: Probit regression predicting ever making repayments. *** indicates significant at the 1% level; ** the 5% level. Controls for cohort, age and year are included in all columns.

Table 5 further investigates voluntary repayments by highlighting differences in the size of individual repayments. The table shows results from regressing the individual voluntary repayments made by students on demographic characteristics and the higher income household indicator. Individual repayments from those from higher income households are considerably larger than for those from lower income households. Again, this holds true when controls for gender and current earnings are added. Among those who make voluntary repayments, those from higher income households make repayments that are around £1,000 larger on average. When HESA controls for subject-institution mix of students doing the same course are included, this estimate reduces to around £600, but remains statistically significant. Finally, in column 5 we show results using a tobit rather

than OLS, with the same specification as in column 4. The broad result - that coming from a higher income household is associated with higher voluntary repayments - is robust, and indeed considerably stronger than in column 4, with individuals from higher income households repaying around £1,200 more, conditional on gender, earnings and university characteristics. This strongly favours the argument that individuals borrowing exactly x are indeed from more advantaged households than those who borrow different amounts.

	Repayments [1]	Add Gender [2]	Add Earnings [3]	Add HESA [4]	Tobit [5]
Higher Income HH	976.9*** (55.819)	958.5*** (55.804)	966.9*** (55.870)	614.5*** (59.938)	1207.0*** (55.170)
Female		-384.6*** (51.319)	-395.0*** (51.436)	-334.2*** (53.884)	462.1*** (48.312)
Earnings			-0.00386** (0.001)	-0.00839*** (0.001)	-0.0119*** (0.001)
Constant	2624.6*** (72.526)	2849.8*** (78.423)	2949.1*** (85.489)	2502.9*** (720.306)	-10595.2*** (646.125)
N	22,176	22,176	22,176	22,176	666,376
Adjusted R^2	0.067	0.069	0.070	0.095	-

Table 5: Size of total voluntary repayments (£), conditional on making them. *** indicates significant at the 1% level; ** the 5% level. Female is a dummy set equal to one for women. Controls for cohort, age and year are included in all columns. Column 5 is run with the same set of controls as column 4.

Column 5 also shows that conditional on making repayments, women make larger repayments by around £460 on average. The sign is flipped compared to the OLS, suggesting differential selection into repayment by gender. Meanwhile, the relation between voluntary repayments and current earnings is economically immaterial, despite being statistically significant (the earnings coefficient in column 5 suggests a £10,000 increase is associated with a reduction in voluntary repayments of just £12). Hence graduates own income levels do not appear to influence whether they make voluntary repayments.²¹

4.1.2 Treatment of those borrowing below the unassessed maximum

We also investigate closely those who borrow less than the unassessed borrowing maximum (i.e. $\leq x$ in Figure 1) to best determine how they should be treated. We repeat the above analysis, splitting out those who borrow below x (Type A, or ‘low borrowers’) and above x (Type B, or ‘high borrowers’) from those who borrow exactly x (Type X - higher income households).

Figure 3 shows the distribution of university attendance for the three groups, split by gender. Note that this differs from Figure 2 by showing the density function for each of the three groups

²¹This result is surprising, although it is important to keep in mind that involuntary repayments are by definition higher for anyone above the income threshold for repayment, meaning higher earnings individuals are paying more on average overall. It is possible that higher earnings are associated with higher financial literacy and better understanding of the system (because the interest rate is so low, it is not obvious why people would want to make involuntary repayments). However, we do not think this is important for our conclusions here.

so that the total for each group sums to one. The most notable feature is the high share of Type A ‘low borrowers’ in the group of universities with missing entry scores. This group of institutions typically consists of smaller, lower-status universities and Further Education colleges. Beyond that, it is clear that Type A ‘low borrowers’ look much more like Type B ‘high borrowers’ than they do typical Type X high income household individuals. A very low share of Type A ‘low borrowers’ and Type B ‘high borrowers’ attend the top 30% of universities, with a tiny fraction going to the top 5%. This contrasts with Type X high income household individuals, of whom a high share go to top institutions.

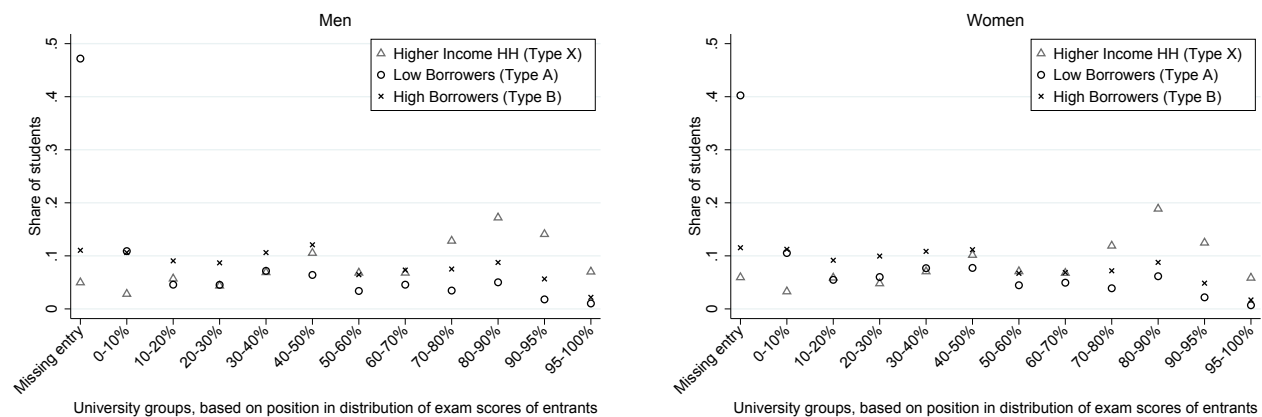


Figure 3: Share of individuals attending different university groups from different parts of the borrowing distribution.

In Table A1 in the Appendix, we investigate voluntary repayments of Type A ‘low borrowers’ and Type B ‘high borrowers’ relative to Type X higher income household individuals. Both make much smaller voluntary repayments than the latter, with Type A ‘low borrowers’ making smaller voluntary repayments than Type B ‘high borrowers’ individuals. Of course, Type A individuals have lower debt, which makes them less likely to make large repayments. However, this is a big difference compared to Type X higher income household individuals, and is suggestive that they again are more like Type B individuals than Type X individuals. Based on the evidence here, we treat Type A and B individuals as our ‘lower income household’ group. We investigate the robustness to this assumption in our subsequent analysis. We now move on to consider the raw earnings differences between individuals from the two groups.

4.2 Descriptive earnings differences

Figure 4 shows the earnings distribution for male and female graduates from higher income households (grey triangles), graduates from lower income households (black circles) and for non graduates (grey line), for the 1999 cohort in 2012/13. The non graduate sample comes from the HMRC

databases (more information is given in Britton et al. (2018)), including a discussion of the relatively high proportion of graduates and non-graduates who have zero or low earnings. In that paper we argue this is a combination of higher earners who are working abroad and hence do not pay tax, lower earners with intermittent attachment to the labour market and part time and self-employed workers who will fall below the tax threshold.). Points to the right of each figure show the mean for each group. The results are striking; graduates from higher income households earn more right across the distribution, from the 20th percentile upwards, for both females and males. Whilst graduates from both lower and higher income households earn more than non graduates, the gap between graduates from lower and higher income backgrounds is also sizeable, particularly at the very top of the distribution. Indeed, whilst around 20% of the graduate population come from higher income households by our definition, of those in the top 1% of the earnings distribution, 45% (men) and 39% (women) come from higher income households.

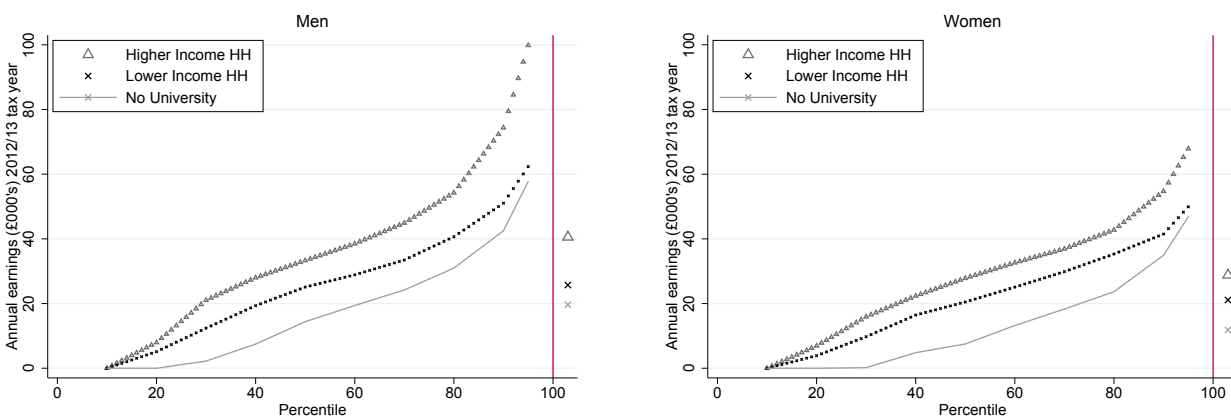


Figure 4: Earnings distribution for individuals in the 1999 cohort in 2012/13 (October 2012 prices) for those from higher income households vs. individuals from lower income households, with the non-university distribution also included. Means given on the RHS of the plot.

As already discussed, students from different socio-economic backgrounds take different degrees, with students from higher income households more likely to attend high status universities. It is possible that this sorting into universities could explain the raw earnings differences between those from high and lower income households. Figure 5 takes the first step to address this by plotting average earnings (conditional on earnings being positive) for graduates by the university groups defined above, by gender. Even within these institution groups, the differences in average earnings between graduates from high and lower income households are clear, suggesting that broadly speaking even when comparing graduates from similar institutions, those from a higher income background go on to do better in the labour market. This appears to be particularly pronounced for men from the most selective universities.

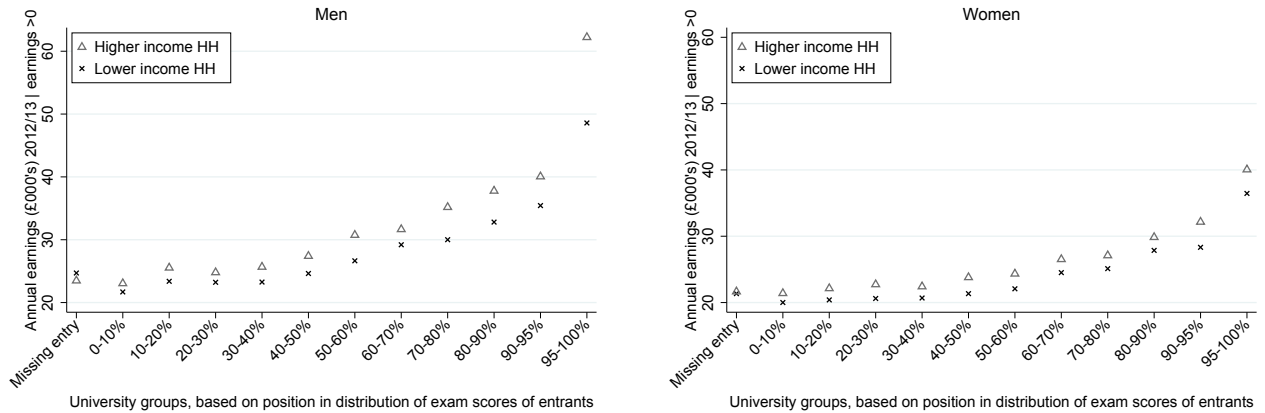


Figure 5: Average earnings (given earnings > 0, October 2012 prices) for graduates from higher and lower income households by university tariff group.

Of course, these figures do not properly control for different degree choices between those from high and lower income backgrounds. Individuals from higher income households might attend the more selective institutions within our coarse university grouping, or might make subject choices that lead to higher earnings. In the next section we try to address this more formally by investigating earnings differences conditional on subject and institution, as well as some other demographic characteristics.

5 Estimation

In Table 6 we estimate the following, conditional on individuals having positive earnings²²

$$\ln(y_{it}) = \alpha + \beta H_i + X'_{it}\gamma + \epsilon_{it} \quad (1)$$

where y_{it} is earnings of individual i at time t , H_i is an indicator for whether an individual is from a higher income household and X_{it} is a vector of controls. We sequentially add additional controls into the vector X .

Columns 1 and 2 indicate the raw differential in earnings between students from higher income households, conditioning only on cohort and year. The differences in earnings are sizeable at around 21% for men and 16% for women. Controlling for subject of degree in columns 3 and 4 reduces these premia by 1-2 percentage points, suggesting that choice of subject explains very little of the

²²Alongside this approach, we also estimated a probit model predicting employment. We find negligible differences in employment between individuals from high and low income households (see Appendix for more details). However, the data does not include an indicator of whether someone is employed or not. We infer employment from whether or not positive earnings are reported. This means we define employment as those with zero earnings. This will unfortunately also include individuals who move abroad. This could be more common for individuals coming from higher income households, again causing some bias and an underestimate of the socio-economic gap in employment.

differences in earnings. Despite the early specialisation of English degrees, it appears that the phenomenon of socio-economic gaps in earnings is not primarily driven by subject choice or sorting of students from higher income households into particular subject areas. By contrast, adding variables which control for the different characteristics of students attending a particular degree course reduces the coefficients considerably (columns 5 and 6 - labelled HESA controls in the table, i.e. including variables describing the course participants from the HESA data). This implies that the nature of the degree course, particularly the entry tariff score, explains more of the variation in earnings between high and lower income students than does their choice of subject. In the final column we include university fixed effects (labelled HEI fixed effects in the table). This does not make an appreciable impact on the coefficients, over and above controlling for the characteristics of the students attending a particular degree course. Overall the results indicate that even allowing for both institution and subject, students from higher income households earn around 10% more than students from lower income households. This suggests that higher education does not fully level the playing field in terms of graduates' earnings.

	Unconditional		Plus subject		Plus HESA		Plus HEI	
	Men [1]	Women [2]	Men [3]	Women [4]	Men [5]	Women [6]	Men [7]	Women [8]
Higher income HH	0.205*** (0.004)	0.157*** (0.003)	0.185*** (0.004)	0.150*** (0.003)	0.108*** (0.004)	0.0970*** (0.004)	0.113*** (0.004)	0.102*** (0.004)
Cohort poly	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subject fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes
HESA controls	No	No	No	No	Yes	Yes	Yes	Yes
HEI fixed effects	No	No	No	No	No	No	Yes	Yes
Adjusted R^2	0.089	0.059	0.110	0.077	0.121	0.083	0.126	0.089
N	399,463	470,272	399,463	470,272	399,463	470,272	399,463	470,272

Table 6: Regression of log earnings on higher income dummy and various controls. Includes the 1999-2005 cohorts and earnings data between 2008/09 and 2013/14. *** indicates significant at the 1 percent level. Standard errors are clustered at the individual level. HESA controls include variables describing the characteristics of students enrolled on the course. HEI fixed effects include fixed effects for each institution.

We assess the robustness of our findings to different definitions of higher income in Table 7.²³ Defining a student as being from a higher income family in a number of different ways, we still obtain the same broad result which is that there remains a wage premium coming from a higher income household of approximately 10%, even conditioning on degree subject and institution. The alternative definitions of higher income student are as follows, where x and y are defined in Figure 1:

- Baseline definition: amount borrowed = x ;

²³We also tested the robustness of our results to including second HESA moments at the course level where possible and found that it made a negligible difference.

- Definition 1: amount borrowed $\leq x$;
- Definition 2: amount borrowed $< y$;
- Definition 3: amount borrowed $= x$, but individuals with amount borrowed $< x$ excluded.

	Baseline		Definition 1		Definition 2		Definition 3	
	Men	Women	Men	Women	Men	Women	Men	Women
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Higher income HH	0.113*** (0.004)	0.102*** (0.004)	0.099*** (0.003)	0.086*** (0.003)	0.089*** (0.004)	0.075*** (0.003)	0.126*** (0.004)	0.115*** (0.004)
Cohort poly	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subject fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HESA controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HEI fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.126	0.089	0.126	0.089	0.126	0.089	0.141	0.101
N	399,463	470,272	399,463	470,272	399,463	470,272	283,904	327,578

Table 7: Robustness of earnings regression to higher income definition. Includes the 1999-2005 cohorts and earnings data between 2008/09 and 2013/14. *** indicates significant at the 1 percent level. Standard errors are clustered at the individual level. HESA controls include variables describing the characteristics of students enrolled on the course. HEI fixed effects include fixed effects for each institution.

Another robustness check is presented in Table 8 which compares OLS regression estimates to those obtained using a nearest-neighbour propensity score matching estimate. To deal with convergence issues, we use the specification from columns 5 and 6 from Table 6 and match on the same set of variables as included in the OLS equation. Again the results are very similar. For men, the coefficient increases marginally by one percentage point, while for women it reduces slightly. Hence even with an alternative, arguably more flexible estimation approach, we find that the wage premium for students from higher income households is around 10%.

	Baseline OLS		Matching	
	Men	Women	Men	Women
	[1]	[2]	[3]	[4]
Higher income HH	0.108*** (0.004)	0.0970*** (0.004)	.118*** (.011)	.094*** (.009)
Cohort poly	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
Subject fixed effects	Yes	Yes	Yes	Yes
HESA controls	Yes	Yes	Yes	Yes
HEI fixed effects	No	No	No	No
N	399,463	470,272	399,463	470,272

Table 8: Matching estimates of differences in log earnings. Includes the 1999-2005 cohorts and earnings data between 2008/09 and 2013/14. *** indicates significant at the 1 percent level. Standard errors are clustered at the individual level. HESA controls include variables describing the characteristics of students enrolled on the course. HEI fixed effects include fixed effects for each institution.

5.1 Heterogeneity

We are interested in potential heterogeneous effects, particularly across different subject areas. It may be that the advantage of higher family income impacts upon some subject-occupation trajectories more than others. Table 9 shows the preferred specification but estimated separately for three different subject areas, namely LEM (Law, Economics and Management courses), STEM (Science, Technology, Engineering and Mathematics courses) and Other (the rest, typically humanities, languages and the arts). The wage premium from coming from a higher income household is similar across all three subject areas except for women who take STEM subjects where interestingly the premium is somewhat lower at around 7%.

	LEM		Other		STEM	
	Men [1]	Women [2]	Men [3]	Women [4]	Men [5]	Women [6]
Higher income HH	0.104*** (0.009)	0.101*** (0.008)	0.117*** (0.006)	0.111*** (0.005)	0.106*** (0.006)	0.0675*** (0.008)
Cohort poly	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Subject fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
HESA controls	Yes	Yes	Yes	Yes	Yes	Yes
HEI fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.161	0.134	0.107	0.077	0.128	0.096
N	70,524	72,915	170,598	271,052	158,341	126,305

Table 9: Regression of log earnings on higher income dummy and controls by subject group. Includes the 1999-2005 cohorts and earnings data between 2008/09 and 2013/14. *** indicates significant at the 1 percent level. Standard errors are clustered at the individual level. HESA controls include variables describing the characteristics of students enrolled on the course. HEI fixed effects include fixed effects for each institution.

Another aspect of heterogeneity we are able to explore is the magnitude of the wage premium from coming from a higher income family for those who attend different institutions. Figure 6²⁴ shows the wage premium for different groups of institutions, split by their average entry score. It is striking that for males only, the wage premium for those from higher income backgrounds is considerably larger if the student attended an institution in the top 5% of the institutional distribution, at around 25%. For women this effect is not evident.

²⁴Full tables for these results are given in the Appendix.

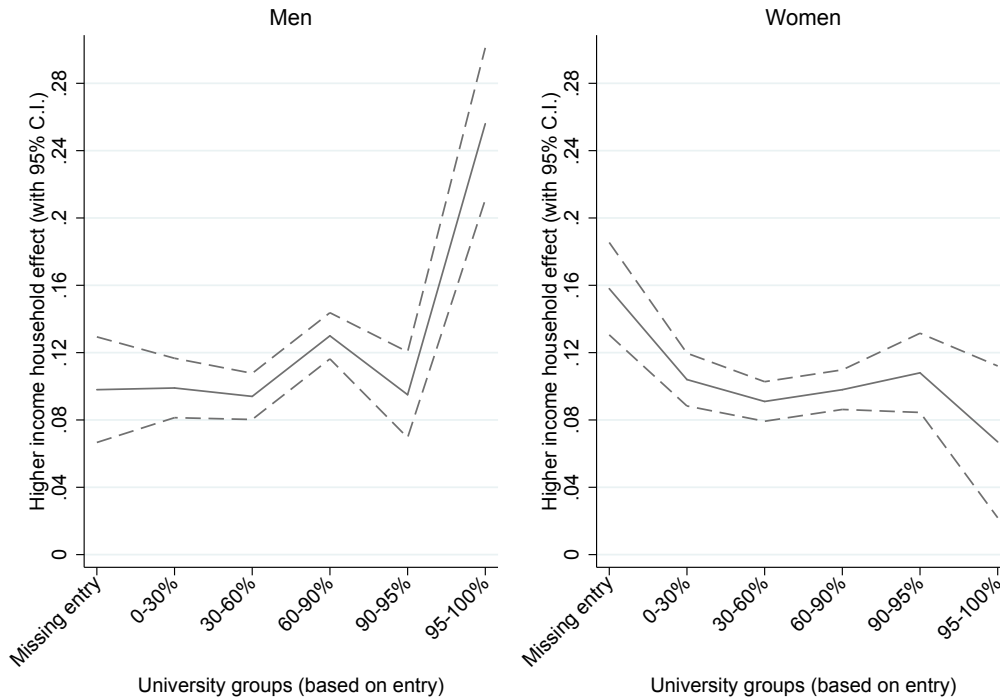


Figure 6: higher income household effect split by university grouping, which is based on average entry scores of their students (as in Figures 2, 3 and 5). Includes the 1999-2005 cohorts and earnings data between 2008/09 and 2013/14. Dashed lines represent 95% confidence intervals. Corresponding tables given in the Appendix.

We also explore heterogeneity by age. Because we cannot disentangle age from cohort and year effects, we show cohort effects holding year fixed in Figure 7, and year effects holding cohort fixed in Figure 8. It is evident that the wage premium from coming from a higher income background increases in both cases, suggesting that the impacts increase with age. Typically they appear to rise to around 14% for men and 12% for women by graduates' early thirties, starting at around half that in each case in graduates' mid twenties. There appear to be gender differences in how the effect changes with age; for women, there appears to be a dip in both figures at points that correspond to their early thirties, potentially due to family formation decisions. For men, this dip is not present, with the effects apparently continuing to rise. It should again be kept in mind that the time period we are investigating here coincides with the recovery from the 2008 recession. However, the fact that we see such similar patterns by both cohort and year suggests that the findings are not entirely driven by the recovery.

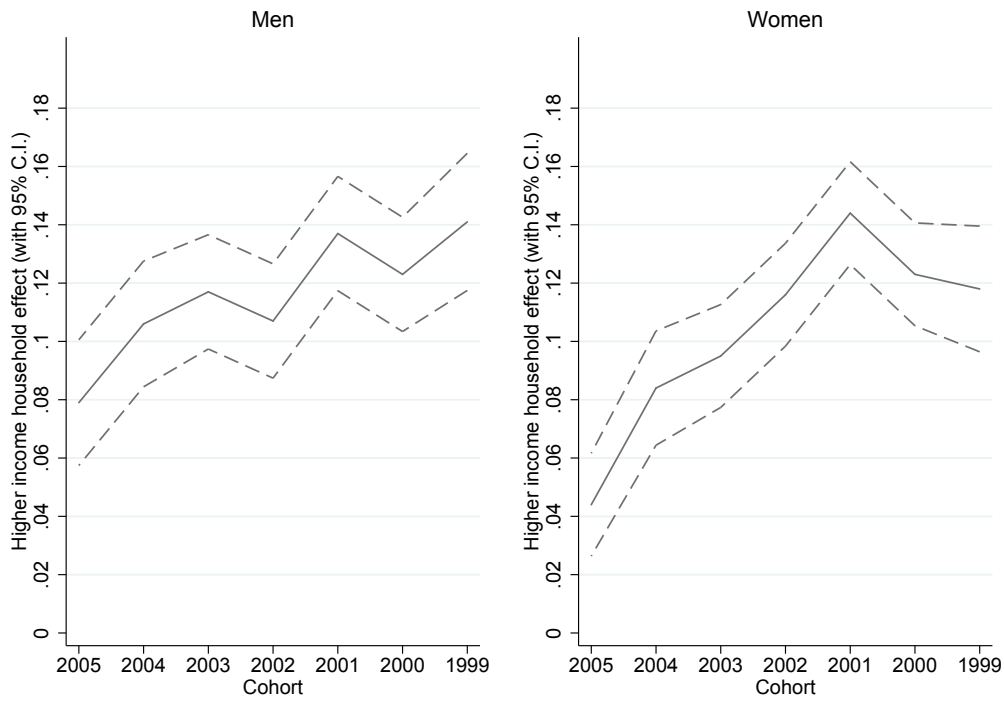


Figure 7: higher income household effect split by cohort (inverted, to show increasing effect with age). Includes earnings data between 2008/09 and 2013/14. Dashed lines represent 95% confidence intervals. Corresponding tables given in the Appendix.

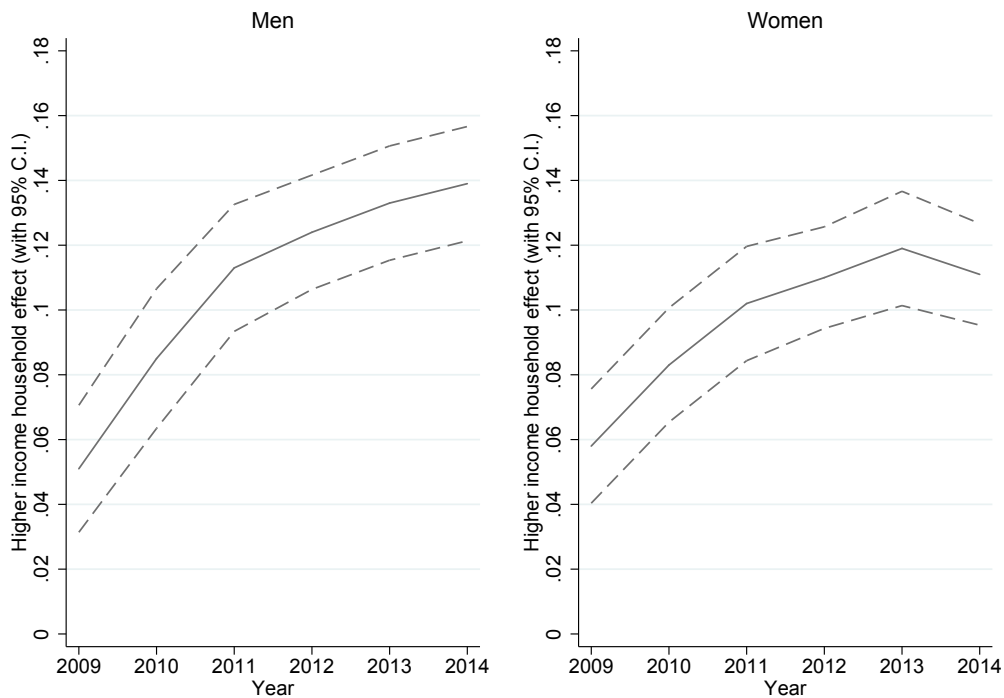


Figure 8: higher income household effect split by year. Includes the 1999-2005 cohorts. Dashed lines represent 95% confidence intervals. Corresponding tables given in the Appendix.

	Raw earnings (£000's)		Raw differences from low family income (£000's)		Conditional differences from low family income (£000's)	
	Men	Women	Men	Women	Men	Women
20th Percentile						
Higher Income HH	20.3	14.0	7.7***	4.4***	3.0***	1.7***
	<i>1.1</i>	<i>.7</i>	<i>.9</i>	<i>.6</i>	<i>.7</i>	<i>.6</i>
Lower Income HH	12.6	9.6				
	<i>.5</i>	<i>.3</i>				
% Wage Premium	61.1	45.8	61.1	45.8	16.1	13.4
50th Percentile						
Higher Income HH	35.0	27.8	8.0***	5.3***	3.3***	2.1***
	<i>1.0</i>	<i>.7</i>	<i>.9</i>	<i>.6</i>	<i>.6</i>	<i>.6</i>
Lower Income HH	27.0	22.5				
	<i>.5</i>	<i>.4</i>				
% Wage Premium	29.6	23.6	29.6	23.6	10.9	8.6
90th Percentile						
Higher Income HH	84.0	54.9	30.8***	13.5***	10.7***	6.2***
	<i>7.0</i>	<i>2.1</i>	<i>6.8</i>	<i>2.0</i>	<i>1.8</i>	<i>1.3</i>
Lower Income HH	53.2	41.4				
	<i>1.5</i>	<i>.8</i>				
% Wage Premium	57.9	32.6	57.9	32.6	19.6	14.3
Years	Yes	Yes	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes	Yes	Yes
Regions	No	No	No	No	Yes	Yes
Age	No	No	No	No	Yes	Yes
Subject fixed effects	No	No	No	No	Yes	Yes
HEI fixed effects	No	No	No	No	No	No
HESA controls	No	No	No	No	Yes	Yes
N	18,038	20,413	18,038	20,413	18,038	20,413

Table 10: Earnings differences for graduates from lower and higher income households at the 20th, 50th and 90th percentiles estimated from quantile regression models. Note that zero earnings are excluded from these regressions. High family income premium indicates the additional earnings for graduates from a higher income household. Low family income earnings indicates earnings of graduates from a lower income background. Percentage wage premium calculates the wage premium for those coming from a higher income household compared to the earnings of those from lower income households, assuming all controls are held constant across the two groups at their means. The first two columns of results show raw estimated earnings for high and low household income earnings. The next two show the difference in earnings from low household income. The final two columns show the conditional difference from low household income - i.e. the difference once controls for region, age, subject and student characteristics are included. All figures are in £000's. Uses 2011/12 and 2012/13 data and the 1999 cohort (estimates are given for 2012/13). Standard errors are clustered at HEP level. * indicates significantly different to the base (lower family income) at 10% level, ** 5% and *** 1%. HESA controls include variables describing the characteristics of students enrolled on the course. HEI fixed effects include fixed effects for each institution.

Finally, in Table 10 we investigate the magnitude of the wage premium at different quantiles of the earnings distribution, motivated in part by the strong policy interest in England in effects through the distribution rather than just at the mean. Due to issues with convergence of the estimator, we use a restricted dataset that includes only the 1999 cohort in 2011/12 and 2012/13. We provide raw earnings and conditional estimates at the 20th, 50th and 90th percentiles of the distribution, separately by gender. What is striking is that although at the median the conditional wage premium for men and women is around 10%, this rises to 16-20% for men at the bottom (20th percentile) and the top (90th percentile) of the distribution. A similar, though less stark

pattern is present for women. This implies that those from higher income households are both better protected against low earnings and more likely to achieve high earnings.

5.2 Summary

In summary, men from higher income households (with median household income of around £77,000, 2018 prices) earn around 21% more than men from lower income households (with median household income of around £26,000, 2018 prices), while the equivalent figure for women is 16%. These estimates roughly halve to around 10% once controls for university, subject and other demographics are included. The differences appear to increase with age, doubling between individuals' mid twenties and their early thirties, before levelling off for women but continuing for to rise for men. This suggests previous work which has focussed on socio-economic differences in early career outcomes (for example, Macmillan et al. (2013); Crawford et al. (2016)) may underestimate earnings gaps. Given our data limitations on the career paths of these individuals, this encourages further research. In particular it would be interesting to determine whether earnings differences by family background in graduates' early thirties are driven by greater participation in postgraduate study or perhaps initial placement into careers with faster earnings trajectories. Whether differences in earnings by parental income exist, even conditional on early career choices, is also an important research question.

The socio-economic gap in graduates' earnings is similar across broad subject groupings, with the exception of women doing STEM courses, for whom the earnings gap is considerably smaller. This latter result may be attributable to the types of occupations pursued by women in STEM, particularly those in medicine and the public sector where salaries are more regulated and hence where coming from a more advantaged family may make less difference to earnings. It could also be a selection effect, if women pursuing STEM are somewhat atypical and if family background makes less of a difference to these atypical women in their career prospects.

We find large earnings variation by university type: men from higher income households attending universities with the most demanding entry requirements earn around 25% more than their relatively less well off peers, even holding institution and subject choice constant. This is a stark finding, suggesting being from a higher earning household is particularly advantageous at the top institutions for men. For women we do not observe this result: a finding that aligns with Crawford and Vignoles (2014) and Dolton and Vignoles (2000) which both find larger effects of private schooling on earnings for men than for women. While we cannot explain this gender difference, we know from other research that earnings from the top institutions are considerably higher than elsewhere (Britton et al. (2016)) and also that there are large differences in the occupational choices of men and women, even comparing those who take the same degree subject (for example, Hakim (2016)).

It is possible that some of the advantages of coming from a wealthy background are particularly pertinent right at the very top of the distribution. We explore this further when we investigate effects through the earnings distribution.

However, our quantile regression results suggest that the advantages of coming from a higher income household are larger at the bottom and top than at the middle of the earnings distribution for *both* men and women. This suggests that there is still an advantage for women going into very high earning occupations.

The overall earnings differences we observe are large, particularly given the bluntness of our parental income measure. Indeed we believe this bluntness is likely to result in an underestimate of the true difference. This is because some lower income individuals will borrow the higher income maximum, while some higher income individuals will borrow less than their full allocation. Further, some individuals will not borrow at all; and we would expect these to be especially from higher income households. All of these issues are likely to bias downward our estimates.

6 Mobility scorecards

The results above suggest that attending university does not appear to be levelling the playing field in terms of earnings. In this section, we follow Chetty et al. (2017) and estimate mobility scorecards to consider which are the best institutions and courses for encouraging social mobility. This has potentially important implications for policymakers trying to reduce the earnings gaps that we have highlighted in this paper.

Specifically we investigate the extent to which different subjects and universities appear to help individuals from lower income backgrounds to become top earners, defined as having earnings in the top quintile of the earnings distribution. We split this analysis out by gender, although we consider the probability of getting to the top 20% of the overall earnings distribution, pooled across genders.²⁵ Chetty et al. (2017) define a mobility score for a given university as follows:

$$P(\text{Child in Q5 and Parent in Q1}) = P(\text{Parent in Q1}) \times P(\text{Child in Q5} \mid \text{Parent in Q1})$$

Where Q5 is the top quintile of the income distribution and Q1 is the lowest quintile. We are limited to a binary indicator of household income and we therefore estimate the probability of a child making it to the top quintile of the earnings distribution given they are from a lower income household. Figures 9 and 10 follow Chetty et al. (2017) by plotting, for men and women respectively, $P(\text{Child in Q5} \mid \text{Lower income household})$ on $P(\text{Lower income household})$ for 21 subject groups we

²⁵Results looking at the gender-specific distributions are available on request from the authors.

observe in our data (see the Appendix for more information on the subject groupings and for the numbers behind the chart). For each subject we give the rank of their overall scorecard,²⁶ and for a subset we show their mobility score. The figures give a sense of how good different subject groups are at delivering individuals that come from lower income households to the top of the graduate earnings distribution. We reiterate that care is needed in interpreting these findings, since our ‘lower income household’ individuals are lower income only in a relative sense and make up 80% of our population of students.

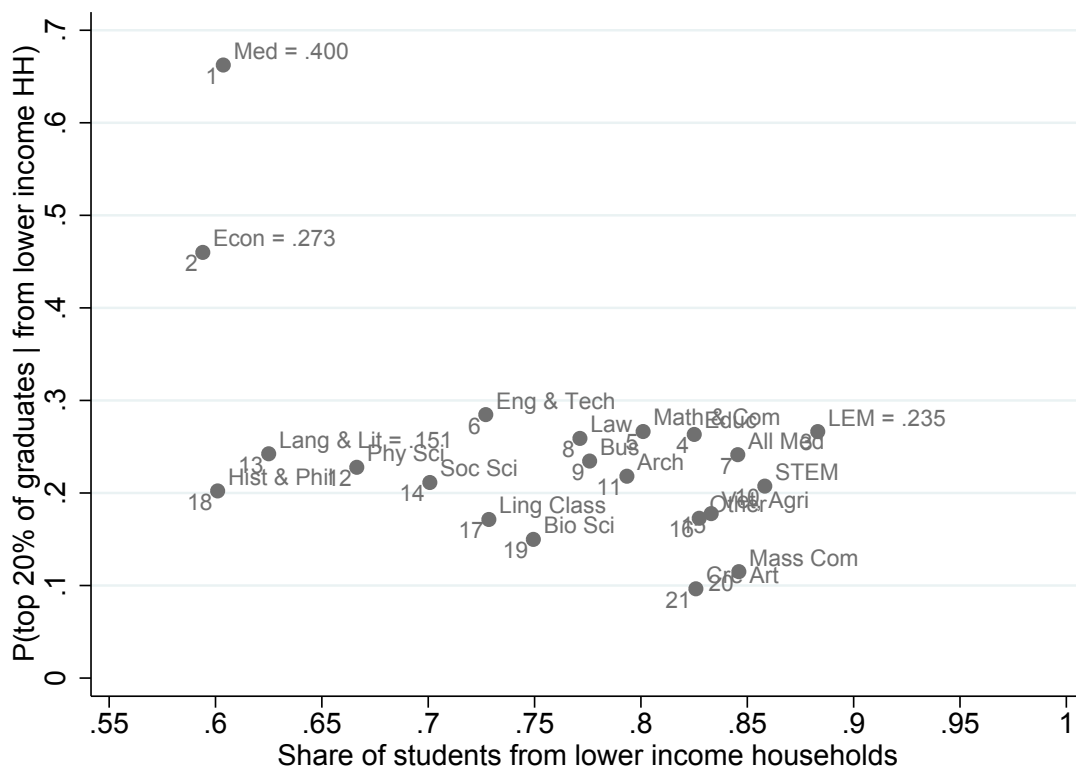


Figure 9: Subject mobility scorecard for men. Earnings rankings use the 2011/12-2013/14 tax years, treating individual observations as independent. The results are not very sensitive to this approach, however. The numbers to the bottom left of each point represent within-gender subject ranking by the mobility index, which is the product of the x and y axes. We also show the mobility score for a subset of subject. See Appendix for full subject definitions and set of mobility scores.

Medicine and economics are the highest performing subject groups by this measure (with scores of 0.4 and 0.273 respectively for men and 0.33 and 0.353 for women). Although these subjects are amongst the worst performing subjects in terms of the proportion of students enrolled from lower income backgrounds (65%), their delivery of students into the top 20% of the earnings distribution is very good. At least 40% of lower income students taking these subjects get into the top 20% of the overall earnings distribution. Other high mobility subject groups are maths and computing

²⁶Note that ‘rank’ is based on multiplying the corresponding points on the x and y axis together. The maximum score is one, which would mean taking 100% poorer students and delivering them all into the top 20%.

and engineering and technology. Miscellaneous law economics and management subjects also do relatively well.²⁷

On the other hand, we see languages and literature, history and philosophy, linguistics and classics and biological and physical sciences all have a relatively low share of lower income students enrolled and also have very poor delivery of those students into the top of the earnings distribution. Though creative arts does far better at enrolling students from lower income backgrounds, it is the worst subject in terms of enabling students from lower income households to reach the top of the earnings distribution. This latter result is because more generally students taking creative arts are less likely to achieve very high earnings, rather than being attributable to some failure within this subject for poorer students to thrive. Nonetheless, from a social mobility perspective, it is clear that some subjects are more likely than others to provide a pathway for poorer students to achieve very high earnings.

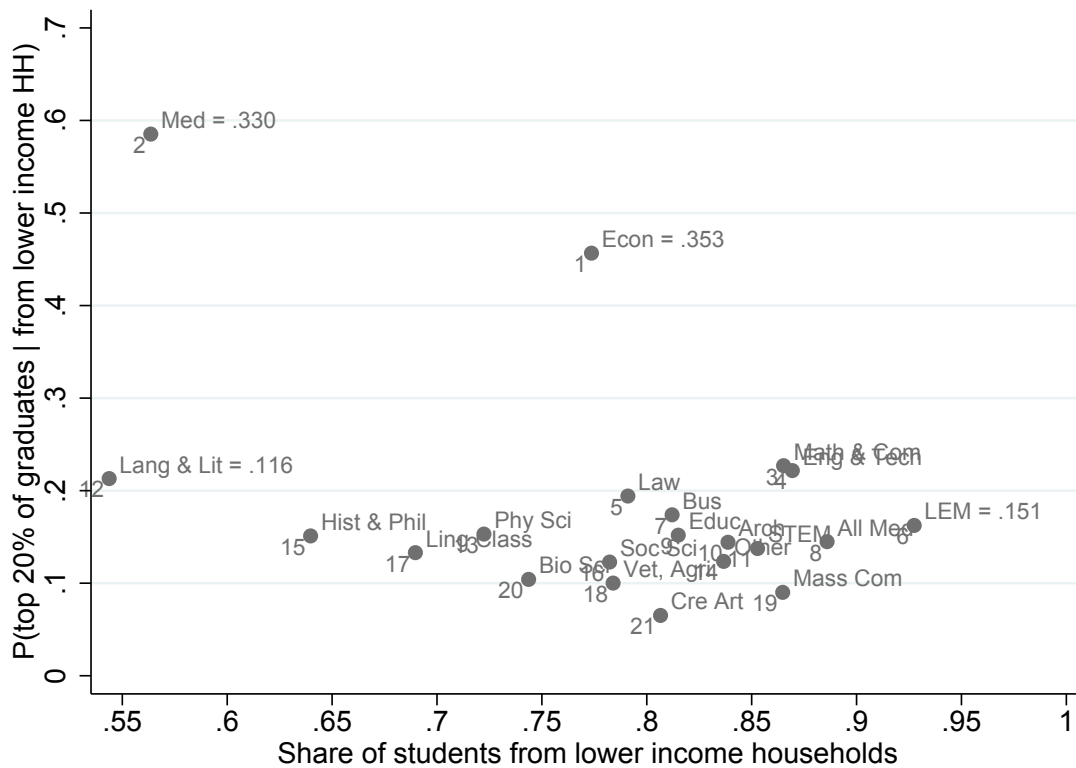


Figure 10: Subject mobility scorecard for women. Earnings rankings use the 2011/12-2013/14 tax years, treating individual observations as independent. The results are not very sensitive to this approach, however. The numbers to the bottom left of each point represent within-gender subject ranking by the mobility index, which is the product of the x and y axes. We also show the mobility score for a subset of subject. See Appendix for full subject definitions and set of mobility scores.

Figures 11 and 12 repeat the same exercise, but for universities. This gives an indication of

²⁷This is a broad subject group and includes students whose subject group at their given institution was too small for us to get their detailed subject grouping.

how each university is delivering individuals that come from lower income households into the top quintile of the graduate earnings distribution. The results show a clear negative relationship between the share of poorer students and the probability of them getting into the top 20% of the earnings distributions. The best performing of the named institutions are clearly those based in London, with the prominent universities of LSE, Kings, Imperial and UCL all performing well for both genders.

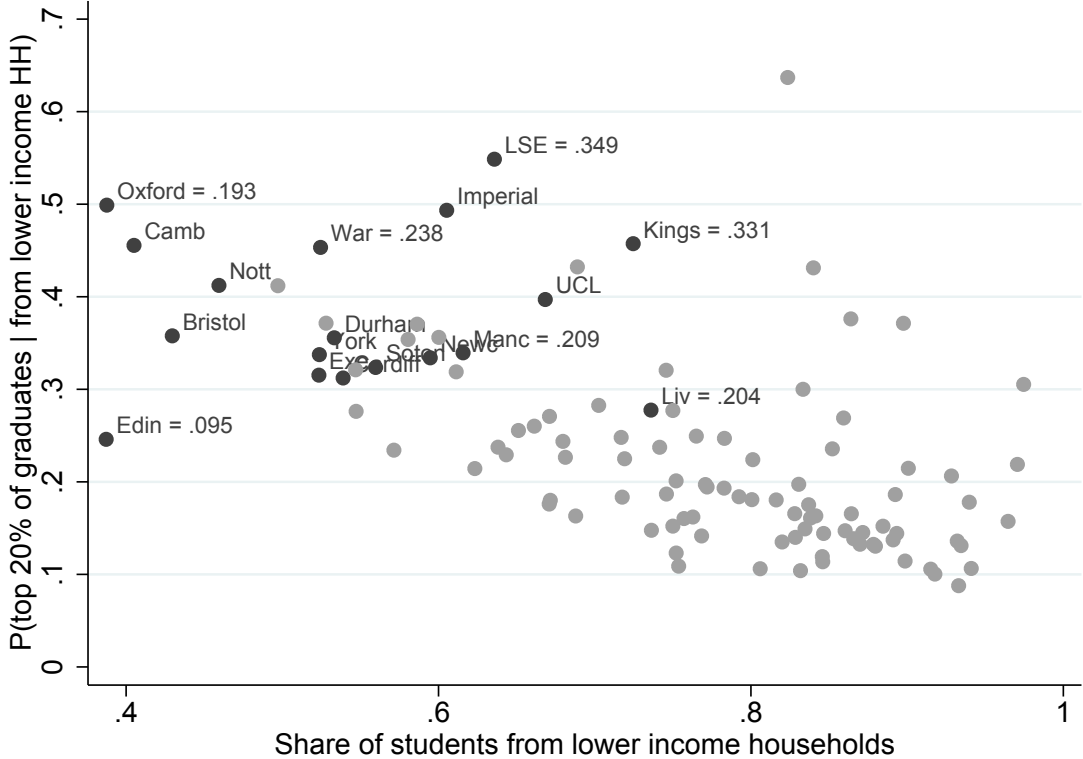


Figure 11: Institution mobility scorecard for men. Earnings rankings use the 2011/12-2013/14 tax years, treating individual observations as independent. The results are not very sensitive to this approach, however. We label a subset of universities we have been granted permission to name. Numbers behind this figure are available in the Appendix.

For men, around 60% of students at the prominent London-based institutions are from lower-income households, compared to less than 50% at Oxford, Cambridge and Bristol. The former all deliver at least 40% of these individuals into the top 20% of the earnings distribution, with the LSE doing the best out of the named institutions by delivering more than 50% into the top. Warwick is the highest performing of the named non-London institution in terms of the mobility score (0.238). It accepts similar shares of poor students to Durham, York, Exeter, Southampton and Cardiff, but is considerably more successful at delivering them to the top of the earning distribution. The worse performing institutions have a delivery rate of under 10%. For women, the LSE and Imperial College appear to have relatively high rates of mobility, accepting similar shares of poorer students

to Newcastle, Manchester, UCL, Southampton and Liverpool, but performing dramatically better in terms of delivery into the top. Oxford, Cambridge and Bristol are again similar, with amongst the lowest shares of students from poorer backgrounds and delivery in to the top of around 30%. Manchester is the best named non-London institution in terms of its overall mobility score (0.160).

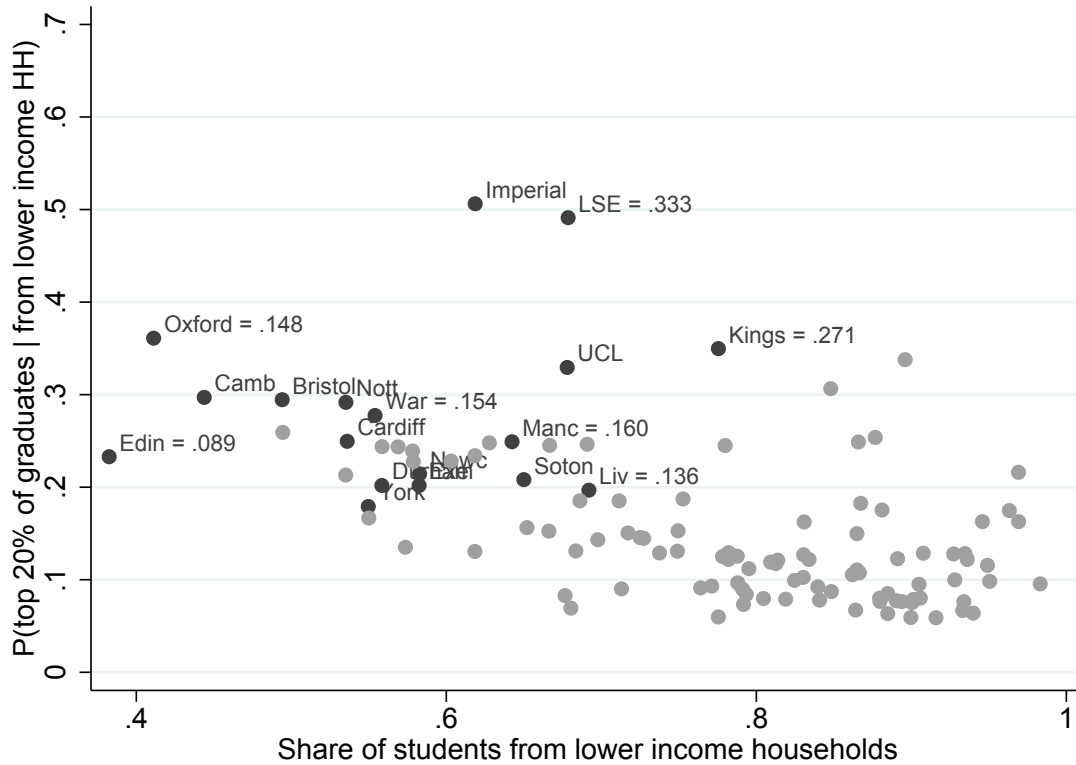


Figure 12: Institution mobility scorecard for women. Earnings rankings use the 2011/12-2013/14 tax years, treating individual observations as independent. The results are not very sensitive to this approach, however. We label a subset of universities we have been granted permission to name. Numbers behind this figure are available in the Appendix.

These mobility scorecards are all clearly descriptive and do not therefore reflect the causal impact of these institutions on students' earnings. For example, the scorecards do not account for subject compositions of courses, or indeed institution compositions of subjects and do not adjust for proximity to the higher wage labour market in London, which is clearly an important factor. The data also refer to courses taken some years ago and hence may not reflect the outcomes from courses currently offered by these institutions. Not least, results are likely to have been affected by policy reform since 2005 and also the increase in contextualised admissions. However, they do illustrate the point that historically at least, some institutions admit a large number of lower income students but such students do not necessarily go on to have high earnings, whereas some institutions admit far fewer but are more successful in delivering such students into the top of the income distribution.

7 Conclusions

Using an innovative administrative data set, consisting of hard linked tax and student loan individual level data, as well as aggregate data on graduates' degree courses, we document how the earnings of graduates from relatively higher and lower income households vary, even after allowing for differences in subject taken and institution attended. The paper is the first of its kind to use such data in the English context to examine the correlation between a measure of higher parental income (i.e. those in the top fifth of the household income distribution of those borrowing from the SLC to attend HE) and graduates' earnings, whilst being able to take account in some detail for the type of higher education experienced.

The main finding from this paper is that graduates' family background - specifically whether they come from a relatively lower or higher income household - continues to influence graduates' earnings long after graduation. The socio-economic gap in graduates' earnings is by no means entirely explained by differences in the subjects studied or institutions attended, though it is approximately halved once we account for these factors. When we take account of different student characteristics, degree subject and institution attended, the gap between graduates from higher and lower income households is still sizeable, at around 10% at the mean and median. Further, we find that the gap is larger at the 20th and 90th percentiles of the graduate earnings distribution, suggesting that coming from a higher income household both protects against low earnings and provides greater opportunity for very high earnings. Men from high income households who attend the most elite universities appear to do particularly well in terms of their earnings.

Clearly, there are caveats to these findings. First, our measure of parental income is blunt, and we miss the roughly 15% of students who are non-borrowers, who are likely to be from the highest income households. We argue that our estimate is likely to be an underestimate of the true earnings differences between the richest and the poorest households. Second, we are analysing earnings in the post-2008 crash period, which may have impacted on the magnitude of our estimates. It is conceivable that the magnitude of the socio-economic gap in graduates' earnings may be affected by the state of the labour market, with students from the wealthiest families being better able to secure the good jobs that become more scarce during a recession. Third, changes to the higher education funding system in England in the intervening period may also mean that the magnitude of the socio-economic gap in earnings could differ going forward, for example due to the large subsequent increases in tuition fees and the increase in contextualised admissions policies.

However, the fact that we observe such robust effects are highly important, and suggest that simply focussing on getting poorer students into university is not enough. Perhaps most importantly, this paper encourages future research into the drivers of the earnings differences that we

observe. Possible explanations include differences in attainment on entry to university, performance at university, progression onto postgraduate study, early career occupation and location decisions and career progression, networks, or non-cognitive skills. Uncovering the most important of these drivers could have significant implications for policy, universities and firms. Researchers seeking solutions for improving social mobility might also find inspiration from the subjects and institutions that are best performing in terms of their social mobility scorecards.

References

- Altonji, J. G., P. Arcidiacono, and A. Maurel (2015). The analysis of field choice in college and graduate school: Determinants and wage effects. Technical report, National Bureau of Economic Research.
- Black, S. E., P. J. Devereux, and K. G. Salvanes (2005). Why the apple doesn't fall far: Understanding intergenerational transmission of human capital. *American economic review* 95(1), 437–449.
- Blanden, J., P. Gregg, and L. Macmillan (2007). Accounting for intergenerational income persistence: noncognitive skills, ability and education. *The Economic Journal* 117(519).
- Blundell, R., L. Dearden, and B. Sianesi (2005). 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* 168, 473–513.
- Bratti, M., R. Naylor, and J. Smith (2005). Variations in the wage returns to a first degree: Evidence from the British Cohort Study 1970.
- Britton, J., L. Dearden, N. Shephard, and A. Vignoles (2016). How English domiciled graduate earnings vary with gender, institution attended, subject and socio-economic background. Technical report, Institute for Fiscal Studies WP1606.
- Britton, J., N. Shephard, and A. Vignoles (2018). Comparing sample survey measures of English earnings of graduates with administrative data. *Journal of the Royal Statistical Society: Series A*. Forthcoming.
- Bukodi, E. and J. Goldthorpe (2011a). Social class returns to higher education: Chances of access to the professional and managerial salaries for men in three British birth cohorts. *Longitudinal and Life Course Studies* 2, 1–71.

- Bukodi, E. and J. H. Goldthorpe (2011b). Class origins, education and occupational attainment in Britain. *European Societies* 13(3), 347–375.
- Chetty, R., J. Friedman, E. Saez, N. Turner, and D. Yagan (2017). Mobility report cards: The role of colleges in intergenerational mobility. Technical report, mimeo, Stanford University.
- Chevalier, A. (2011). Subject choice and earnings of UK graduates. *Economics of Education Review* 30, 1187–1201.
- Chowdry, H., C. Crawford, L. Dearden, A. Goodman, and A. Vignoles (2013). Widening participation in higher education: analysis using linked administrative data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176(2), 431–457.
- Crawford, C., L. Dearden, J. Micklewright, and A. Vignoles (2016). *Family Background and University Success: Differences in Higher Education Access and Outcomes in England*. Oxford University Press.
- Crawford, C. and A. Vignoles (2014). Heterogeneity in graduate earnings by socio-economic background. Unpublished paper: Institute of Fiscal Studies.
- Dale, S. B. and A. B. Krueger (2014). Estimating the effects of college characteristics over the career using administrative earnings data. *Journal of Human Resources* 49(2), 323–358.
- Dolton, P. and A. Vignoles (2000). The incidence and effects of overeducation in the UK graduate labour market. *Economics of Education Review* 19, 179–198.
- Ermisch, J., M. Jantti, and T. M. Smeeding (2012). *From parents to children: The intergenerational transmission of advantage*. Russell Sage Foundation.
- Figlio, D. N., K. Karbownik, and K. G. Salvanes (2015). Education research and administrative data. Technical report, National Bureau of Economic Research.
- Green, F., S. Machin, R. Murphy, and Y. Zhu (2012). The changing economic advantage from private schools. *Economica* 79, 658–679.
- Hakim, C. (2016). *Key issues in women’s work: Female diversity and the polarisation of women’s employment*. Routledge-Cavendish.
- Hussain, I., S. McNally, and S. Telhaj (2009). University quality and graduate wages in the UK. CEE Discussion Paper No. 99.

- Kassenboehmer, S. C., F. Leung, and S. Schurer (2018). University education and non-cognitive skill development. *Oxford Economic Papers* 70(2), 538–562.
- Macmillan, L., C. Tyler, and A. Vignoles (2013). Who gets the top jobs? The role of family background and networks in recent graduates access to high-status professions. *Journal of Social Policy*, 1–29.
- Naylor, R. (2002). Sheer class? The extent and sources of variation in the UK graduate earnings premium. LSE STICERD Research Paper No. CASE054.
- Sloane, P. J. and N. C. O’Leary (2005). The return to a university education in Great Britain. *National Institute Economic Review*, 75–89.
- Smith, J. and R. A. Naylor (2001). Determinants of individual degree performance: Evidence for the 1993 UK university graduate population from the USR. *Oxford Bulletin of Economics and Statistics* 63, 29–60.
- Walker, I. and Y. Zhu (2011). Differences by degree: Evidence of the net financial rates of return to undergraduate study for England and Wales. *Economics of Education Review* 30, 1177–1186.
- Walker, I. and Y. Zhu (2013). The impact of university degrees on the lifecycle of earnings: Some further analysis. *BIS Research Paper No.112*.
- Webber, D. (2014). Is the return to education the same for everybody? *IZA World of Labor*.

Appendix A: Voluntary Repayments

	Repayments [1]	Add Gender [2]	Add Earnings [3]	Add HESA [4]
High Borrowing (Type B)	-837.6*** (61.4)	-817.5*** (61.4)	-826.4*** (61.4)	-505.3*** (63.2)
Low Borrowing (Type A)	-1171.8*** (66.3)	-1155.6*** (66.3)	-1163.2*** (66.3)	-858.2*** (75.0)
Female		-387.0*** (51.3)	-397.3*** (51.4)	-333.4*** (53.9)
Earnings			-0.004** (0.001)	-0.008*** (0.001)
Constant	3598.9*** (83.2)	3807.0*** (87.6)	3913.3*** (95.0)	3167.9*** (720.6)
N	22,176	22,176	22,176	22,176
Adjusted R^2	0.068	0.071	0.071	0.096

Table A1: Size of total voluntary repayments (£), conditional on making them, relative to Type X individuals. *** indicates significant at the 1% level; ** the 5% level. Female is a dummy set equal to one for women. Controls for cohort, age and year are included in all columns.

Appendix B: Heterogeneity estimates

	No Intake Information [1]	Bottom HEPs [2]	Lower Middle HEPs [3]	Middle HEPs [4]	Top HEPs [5]	Very Top HEPs [6]
Higher income HH	0.0982*** (0.016)	0.0993*** (0.009)	0.0945*** (0.007)	0.130*** (0.007)	0.0951*** (0.013)	0.256*** (0.023)
Cohort poly	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Subject fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
HESA controls	Yes	Yes	Yes	Yes	Yes	Yes
HEI fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.095	0.080	0.105	0.155	0.177	0.182
N	79,500	89,436	99,279	94,106	25,875	11,267

Table A2: Estimation of % earnings differences between high and lower income households by HEP group, for men. Includes the 1999-2005 cohorts and earnings data between 2008/09 and 2013/14. *** indicates significant at the 1 percent level. Standard errors are clustered at the individual level. Table provides raw data behind Figure 6 in main text.

	No Intake Information [1]	Bottom HEPs [2]	Lower Middle HEPs [3]	Middle HEPs [4]	Top HEPs [5]	Very Top HEPs [6]
Higher income HH	0.158*** (0.014)	0.104*** (0.008)	0.0913*** (0.006)	0.0986*** (0.006)	0.108*** (0.012)	0.0671** (0.023)
Cohort poly	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Subject fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
HESA controls	Yes	Yes	Yes	Yes	Yes	Yes
HEI fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.079	0.057	0.073	0.109	0.140	0.144
N	88,005	114,069	120,243	111,085	26,750	10,120

Table A3: Estimation of % earnings differences between high and lower income households by HEP group, for women. Includes the 1999-2005 cohorts and earnings data between 2008/09 and 2013/14. *** indicates significant at the 1 percent level. Standard errors are clustered at the individual level. Table provides raw data behind Figure 6 in main text.

	1999 [1]	2000 [2]	2001 [3]	2002 [4]	2003 [5]	2004 [6]	2005 [7]
Higher income HH	0.141*** (0.012)	0.123*** (0.010)	0.137*** (0.010)	0.107*** (0.010)	0.117*** (0.010)	0.106*** (0.011)	0.0797*** (0.011)
Cohort poly	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subject fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HESA controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HEI fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.108	0.098	0.096	0.083	0.081	0.099	0.141
N	54,291	55,965	57,566	57,702	57,855	56,481	59,603

Table A4: Estimation of % earnings differences between high and lower income households by cohort, for men. Includes earnings data between 2008/09 and 2013/14. *** indicates significant at the 1 percent level. Standard errors are clustered at the individual level. Table provides raw data behind Figure 7 in main text.

	1999 [1]	2000 [2]	2001 [3]	2002 [4]	2003 [5]	2004 [6]	2005 [7]
Higher income HH	0.118*** (0.011)	0.123*** (0.009)	0.144*** (0.009)	0.116*** (0.009)	0.0952*** (0.009)	0.0845*** (0.010)	0.0443*** (0.009)
Cohort poly	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subject fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HESA controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HEI fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.086	0.080	0.069	0.063	0.070	0.071	0.130
N	61,806	65,101	67,071	67,607	68,203	68,198	72,286

Table A5: Estimation of % earnings differences between high and lower income households by cohort, for women. Includes earnings data between 2008/09 and 2013/14. *** indicates significant at the 1 percent level. Standard errors are clustered at the individual level. Table provides raw data behind Figure 7 in main text.

	2008/09	2009/10	2010/11	2011/12	2012/13	2013/14
	[1]	[2]	[3]	[4]	[5]	[6]
Higher income HH	0.0515*** (0.010)	0.0857*** (0.011)	0.113*** (0.010)	0.124*** (0.009)	0.133*** (0.009)	0.139*** (0.009)
Cohort poly	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Subject fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
HESA controls	Yes	Yes	Yes	Yes	Yes	Yes
HEI fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.184	0.128	0.105	0.102	0.104	0.107
N	67,244	66,009	66,633	67,106	66,441	66,030

Table A6: Estimation of % earnings differences between high and lower income households by year, for men. Includes the 1999-2005 cohorts. *** indicates significant at the 1 percent level. Standard errors are clustered at the individual level. Table provides raw data behind Figure 8 in main text.

	2008/09	2009/10	2010/11	2011/12	2012/13	2013/14
	[1]	[2]	[3]	[4]	[5]	[6]
Higher income HH	0.0589*** (0.009)	0.0837*** (0.009)	0.102*** (0.009)	0.110*** (0.008)	0.119*** (0.009)	0.111*** (0.008)
Cohort poly	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Subject fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
HESA controls	Yes	Yes	Yes	Yes	Yes	Yes
HEI fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.177	0.106	0.080	0.075	0.072	0.076
N	79,851	78,008	78,744	78,639	77,575	77,455

Table A7: Estimation of % earnings differences between high and lower income households by year, for women. Includes the 1999-2005 cohorts. *** indicates significant at the 1 percent level. Standard errors are clustered at the individual level. Table provides raw data behind Figure 8 in main text.

Appendix D: Employment probabilities

	Men			Women		
	Employment [1]	Add Subject [2]	Add HESA [3]	Employment [4]	Add Subject [5]	Add HESA [6]
Higher Income HH	-.003 (.006)	.010 (.006)	.014* (.006)	.046*** (.006)	.058*** (.006)	.036*** (.006)
$P(earn > 0 \text{Lower income HH})$.867	.867	.867	.873	.873	.873
$P(earn > 0 \text{Higher income HH})$.866	.869	.870	.882	.884	.880
N	460,656	460,656	460,656	538,908	538,908	538,908

Table A8: Probit regression predicting employment. Predicted probability gives the probability of those from lower income households having non-zero earnings. *** indicates significant at the 1% level; ** the 5% level. Controls for cohort, age and year are included in all columns, with sequential addition of subject and HESA controls as indicated.

Appendix E: Mobility scorecard tables

Subject abbreviations used in Figures 9 and 10.

Medicine and Dentistry (Med); Subjects allied to Medicine (All Med); Biological Sciences (Bio Sci); Veterinary Sciences, Agriculture and related subjects (Vet, Agri); Physical Sciences (Phy Sci); Mathematical and Computer Sciences (Math & Com); Engineering and Technologies (Eng & Tech); Architecture, Building and Planning (Arch); Social studies (Soc Sci); Economics (Econ);

Law (Law); Business and Administrative studies (Bus); Mass Communications and Documentation (Mass Comm); Linguistics, Classics and related subjects (Ling Class); European Languages, Literature and related subjects (Lang & Lit); Historical and Philosophical studies (Hist & Phil); Creative Arts and Design (Cre Art); Education (Educ); STEM, LEM and Other represent miscellaneous STEM, LEM and Other courses - where individuals are in classes that are too small for us to be given their fine subject grouping. In these cases, we simply get their broader subject grouping.

Rank	Subject	Lower Income Share (A)	Delivery to top 20% (B)	Mobility Score (A*B)	Mobility Score Standard error	N
1	Medicine & Dentistry	0.604	0.662	0.400	0.007	4284
2	Economics	0.594	0.460	0.273	0.008	3510
3	LEM	0.883	0.266	0.235	0.008	2772
4	Education	0.825	0.263	0.217	0.004	12399
5	Maths & Computing	0.801	0.266	0.213	0.002	28071
6	Engineering & Technology	0.727	0.285	0.207	0.003	17916
7	Allied Medicine	0.846	0.241	0.204	0.005	7011
8	Law	0.771	0.259	0.200	0.004	7950
9	Business	0.776	0.234	0.182	0.002	26262
10	STEM	0.858	0.207	0.178	0.005	6708
11	Architecture	0.793	0.218	0.173	0.006	3993
12	Physical Sciences	0.666	0.228	0.152	0.003	11358
13	Languages & Literature	0.625	0.242	0.151	0.009	1776
14	Social Sciences	0.701	0.211	0.148	0.003	12762
15	Vetinary/Agriculture	0.833	0.178	0.148	0.009	1635
16	Other	0.827	0.173	0.143	0.002	24177
17	Linguistics & Classics	0.729	0.171	0.125	0.005	5094
18	History & Philosophy	0.601	0.202	0.122	0.004	8451
19	Biological Sciences	0.749	0.150	0.112	0.003	13947
20	Mass Communication	0.846	0.115	0.097	0.004	6897
21	Creative Arts	0.826	0.096	0.080	0.002	23355

Table A9: Subject mobility scorecard table for men.

Rank	Subject	Lower Income Share (A)	Delivery to top 20% (B)	Mobility Score (A*B)	Mobility Score Standard error	N
1	Economics	0.774	0.457	0.353	0.014	1206
2	Medicine & Dentistry	0.564	0.585	0.330	0.006	5994
3	Maths & Computing	0.865	0.227	0.196	0.005	7389
4	Engineering & Technology	0.869	0.222	0.193	0.007	2802
5	Law	0.791	0.194	0.154	0.003	14052
6	LEM	0.927	0.162	0.151	0.008	2151
7	Business	0.812	0.174	0.141	0.002	23880
8	Allied Medicine	0.886	0.145	0.128	0.002	19698
9	Education	0.815	0.152	0.124	0.002	35886
10	Architecture	0.839	0.144	0.121	0.010	1116
11	STEM	0.853	0.138	0.117	0.005	4749
12	Languages & Literature	0.544	0.213	0.116	0.005	3945
13	Physical Sciences	0.722	0.153	0.111	0.004	7218
14	Other	0.837	0.124	0.103	0.002	33678
15	History & Philosophy	0.640	0.151	0.097	0.003	8319
16	Social Sciences	0.782	0.123	0.096	0.002	20229
17	Linguistics & Classics	0.690	0.133	0.092	0.003	11862
18	Vetinary, Aggriculture	0.784	0.100	0.078	0.005	3402
19	Mass Communication	0.865	0.090	0.078	0.003	9984
20	Biological Sciences	0.744	0.104	0.077	0.002	21453
21	Creative Arts	0.807	0.065	0.053	0.001	30441

Table A10: Subject mobility scorecard table for women.

Institution	Lower Income Share (A)	Delivery to top 20% (B)	Mobility Score (A*B)	Mobility Score Standard error	N
Bristol	0.430	0.358	0.154	0.008	2193
Cambridge	0.405	0.455	0.184	0.008	2385
Cardiff	0.539	0.312	0.168	0.009	1581
Durham	0.533	0.356	0.190	0.008	2436
Edinburgh	0.387	0.246	0.095	0.010	798
Exeter	0.523	0.315	0.165	0.008	2115
Imperial	0.605	0.493	0.299	0.012	1383
King's College	0.725	0.457	0.331	0.012	1536
LSE	0.636	0.549	0.349	0.022	453
Liverpool	0.736	0.277	0.204	0.008	2409
Manchester	0.616	0.339	0.209	0.006	4278
Newcastle	0.595	0.334	0.199	0.008	2487
Nottingham	0.459	0.412	0.189	0.007	3585
Oxford	0.388	0.499	0.193	0.008	2523
Southampton	0.560	0.324	0.181	0.008	2562
UCL	0.668	0.397	0.265	0.010	1809
Warwick	0.524	0.453	0.238	0.009	2145
York	0.524	0.337	0.177	0.010	1392

Table A11: Institution mobility scorecard table for men. Shows named institutions only, in alphabetical order.

Institution	Lower Income Share (A)	Delivery to top 20% (B)	Mobility Score (A*B)	Mobility Score Standard error	N
Bristol	0.494	0.295	0.146	0.007	2295
Cambridge	0.444	0.297	0.132	0.007	2670
Cardiff	0.536	0.250	0.134	0.007	2205
Durham	0.558	0.202	0.113	0.006	2697
Edinburgh	0.382	0.233	0.089	0.008	1302
Exeter	0.582	0.202	0.118	0.007	2220
Imperial	0.619	0.506	0.313	0.018	645
King's College	0.776	0.350	0.271	0.009	2205
LSE	0.679	0.491	0.333	0.030	252
Liverpool	0.692	0.197	0.136	0.007	2688
Manchester	0.642	0.249	0.160	0.005	4530
Newcastle	0.583	0.214	0.125	0.007	2538
Nottingham	0.535	0.292	0.156	0.006	3408
Oxford	0.411	0.361	0.148	0.008	2196
Southampton	0.650	0.208	0.135	0.006	3060
UCL	0.678	0.329	0.223	0.010	1854
Warwick	0.554	0.278	0.154	0.007	2361
York	0.550	0.179	0.098	0.008	1239

Table A12: Institution mobility scorecard table for women. Shows named institutions only, in alphabetical order.