# Essays in Labour Economics

*Laura Margretha van der Erve*

Thesis submitted to the Department of Economics

in partial fulfilment of the requirements for the degree of

**Doctor of Philosophy**

**University College London**

London

February 14, 2022

# Declaration

I, Laura van der Erve, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. Chapters 1 and 5 are sole-authored. Chapter 2 was undertaken as joint work with Pedro Carneiro, Sarah Cattan, Lorraine Dearden, Sonya Krutikova and Lindsey Macmillan. Chapter 3 was undertaken as joint work with Jack Britton, Chris Belfield, Anna Vignoles, Matt Dickson, Yu Zhu, Ian Walker, Lorraine Dearden, Luke Sibieta and Franz Buscha. Chapter 4 was undertaken as joint work with Jack Britton and Elaine Drayton.

Signature:

Date: February 14, 2022

# Statement of conjoint work

Chapter 1 is sole-authored.

Chapter 2 was undertaken as joint work with Pedro Carneiro, Sarah Cattan, Lorraine Dearden, Sonya Krutikova and Lindsey Macmillan. All authors contributed equally to the development of the research idea and analysis plan. I performed all analysis except for Appendix Figure A.1. All the writing was done by me.

Chapter 3 was undertaken as joint work with Jack Britton, Chris Belfield, Anna Vignoles, Matt Dickson, Yu Zhu, Ian Walker, Lorraine Dearden, Luke Sibieta and Franz Buscha. Jack Britton and I contributed equally to the work in this chapter. The remaining authors contributed to Belfield et al. (2018a), a policy report which laid the foundation for the academic work contained in this chapter.

Chapter 4 was undertaken as joint work with Jack Britton and Elaine Drayton. All authors contributed equally.

Chapter 5 is sole-authored.

# Abstract

This thesis studies the interactions between parental background, education and later life outcomes. The first chapter analyses differences across England in the early career earnings of children from low-income families, and the role educational differences play in explaining this variation. Children from low-income families who grew up in the lowest mobility areas are expected to end up around fifteen percentiles lower in the earnings distribution at age 28 than similar children from the highest mobility areas. Differences in educational achievement across areas can explain 25% of this variation for men, and more than 45% for women. This indicates that education policy can potentially play an important role in equalising opportunities for children from low-income families. A second chapter estimates the impact of different higher education degrees on earnings, controlling for the impact of parental background and prior attainment. It finds substantial variation in earnings returns within subjects and across universities with very similar selectivity levels, suggesting degree choices matter a lot for later-life earnings. These returns are poorly correlated with observable degree characteristics, implying students have to make potentially life changing degree choices based on limited information. The third chapter estimates "mobility rates" for all English universities, subjects and degrees, by combining access rates and labour market success of students from low-income families. It finds that less selective institutions outperform the most prestigious universities on this measure. Mobility rates are mostly uncorrelated with average earnings returns, which implies that any policies which restrict funding or access to courses with lower earnings returns can have negative implications for mobility. The final chapter looks at the intergenerational impact of parental unemployment. It finds a strong and persistent negative impact of paternal unemployment on the educational achievement and home ownership rates of women, though not of men.

# Impact statement

The research contained in this thesis has the potential to have an impact both in academia and for public policy. The thesis helps build our understanding of the links between parental background, education and adult outcomes. It helps inform the design of policies trying to equalise opportunities across individuals, and those trying to improve the higher education system.

Chapter 2 estimates the differences in income mobility across areas in the UK, and shows the importance of geographical differences in educational attainment to account for this variation. This adds to a small but growing literature exploring how mobility varies within countries, and improves our understanding of the potential drivers of this variation. The results from this work can be used as the basis of further research investigating potential drivers of income mobility in more depth. These results also provide policymakers with important information on where to target investments to increase mobility, and how effective educational policy can be at equalising opportunities across the country.

Chapter 3 and 4 estimate, for individual institutions, subjects and degrees (institution-subject combinations), average earnings returns and the contribution to social mobility. These estimates are valuable for policy-makers making decisions on the 'value' of certain degrees, and on any changes in funding arrangements or regulation of the higher education sector. Indeed, an earlier version of the work included in Chapter 3 was cited in the Post-18 Review of Education and Funding ('Augar review'). The results from this work can also be used as the foundation for further research investigating which university policies are particularly successful at increasing student earnings and access for students from lower-income backgrounds.

Chapter 5 improves our understanding of the long-run effects of paternal job loss in childhood. It further adds to the evidence on the long-run impacts of the decline of the mining industry, once one of the major industries in the UK. This provides policy-makers with information on the expected long-term costs of industrial decline, and informs them on the potential impact of unemployment benefits to mitigate these costs on children's outcomes.

I have had regular contact with the Department for Education and the Social Mobility Commission throughout my PhD to discuss the findings of my research, and have presented this work widely to academic audiences, policy-makers and the general public. I will seek to continue to disseminate the findings of my research through presenting this work, engaging directly with policy-makers, and publishing this work in academic journals.

# Acknowledgements

I would like to thank my supervisors, Imran Rasul and Pedro Carneiro, for their input and guidance during my PhD.

I am also grateful to the Institute for Fiscal Studies, and my many wonderful colleagues there, for enabling me to pursue my PhD and providing such a supportive, stimulating and enjoyable environment.

While neither of them is still here today, I will forever be grateful for the unwavering support of my parents. I would not be where I am today were it not for the help and encouragement they provided for many years.

# Data provision

The permission of the Office for National Statistics to use the Longitudinal Study, used in Chapter 5 of this thesis, is gratefully acknowledged, as is the help provided by staff of the Centre for Longitudinal Study Information & User Support (CeLSIUS). CeLSIUS is funded by the ESRC under project ES/V003488/1. The author alone is responsible for the interpretation of the data. This work contains statistical data from ONS which is Crown Copyright. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

It is widely perceived as unfair for children's opportunities to be constrained by their parents' circumstances, and increasing social mobility has long been a key policy objective in many countries. Yet, research across the world has documented large gaps in education, health, labour market and other outcomes between children born in low-income families and those from more affluent backgrounds. To make progress on reducing these gaps, a thorough understanding of when they emerge, how they vary across place and what drives them is crucial. This thesis consists of four self-contained chapters, which aim to contribute to this understanding by studying the interactions between parental background, education and later life outcomes.

In Chapter 2 I start by estimating how intergenerational income mobility varies across England. Historically, much of the work on intergenerational mobility focused on cross-country comparisons and national trends. More recently, the availability of large linked administrative datasets has enabled the study of intergenerational mobility within countries. This work, pioneered by Chetty et al. (2014), has shown important variation within countries in the outcomes of children from lower income backgrounds. Documenting this variation both allows us to establish where policy efforts to improve social mobility should be targeted, and provides important variation to help better understand the main mechanisms through which parental background affects a child's later outcomes. Due to the highly demanding data requirements, the work of Chetty et al. (2014) in the US has so far only been replicated in a handful of countries, including Italy (Acciari et al., 2019), Canada (Corak, 2019) and Sweden (Heidrich, 2017). This paper adds to this literature by providing the first estimates of intergenerational income mobility at a detailed geographical level in England. Previous work in the UK has had to rely on survey data (e.g. Blanden et al., 2004; Gregg et al., 2017), which does not provide large enough samples to estimate mobility at a detailed geographical level, or on a 1% sample of the census (Bell et al., 2018), which does not contain measures of income or earnings. I use a newly linked administrative dataset, the Longitudinal Education Outcomes (LEO) data, to estimate income mobility for children born 1985-1988 for more than 150 Local Authorities in England. I focus on the outcomes of children from low-income backgrounds, as measured by their eligibility for Free School Meals (FSM) at age 16, and estimate their average income rank, and the probability of them reaching the top quintile of the income distribution, at age 28. I find substantial differences across the country, with a strong North-South gradient. Children from low-income families who grew up in the lowest mobility areas - overwhelmingly in the North - are expected to end up around fifteen percentiles lower in the income distribution as adults compared to those from the highest mobility areas - overwhelmingly in the South-East.

The rest of the chapter then builds on this analysis by investigating the role educational attainment at different ages plays in explaining these geographical differences in income mobility, and exploring what other area characteristics relate to better outcomes for children from low-income backgrounds. It finds that differences in average educational achievement across areas can explain 25% of the variation in income mobility within the country for men, and more than 45% of the variation for women. For men, nearly all these differences can be explained by geographical differences in age 16 education. For women differences in later educational attainment, especially in higher education, play a much bigger role. These findings indicate that education policy has an important role to play in equalising opportunities of children from low-income families across the country, though will not be sufficient to fully do so on its own. The final part of the chapter explores what other area characteristics potentially play a role in an area having higher income mobility. It finds that high mobility is related to stronger labour markets, more stable families and higher median income.

Having established the importance of educational attainment for explaining variation in labour market outcomes of children from low-income families in Chapter 2, the next two chapters of this thesis then focus in on the education system. Specifically, they use the same linked administrative LEO dataset to explore the impact of different higher education institutions and "degrees" (subject-institution combinations) on earnings, and their contribution to social mobility.

Chapter 3 explores the variation in earnings returns across institutions and degrees, and investigates the predictability of those returns based on observable characteristics. In the UK, as in many other countries, students choose specific subject-university combinations for their degree when applying to university. Despite this, data limitations have meant that most existing papers have focused on the average earnings returns to subjects, institutions or even groups of institutions, and have not been able to explore the heterogeneity across subjects within institutions. To the best of our knowledge, this is the first paper to estimate returns for individual degree courses for an entire higher education market.

The chapter uses the rich LEO dataset which links together administrative school, university and tax records for the population of individuals born since September 1985 and educated in England. This data allows us to track all students in all higher education institutions in the country, and allows us to control for an extremely detailed set of prior attainment measures, as well as parental background and school. Conditioning on this rich set of background characteristics, we find substantial variation in earnings returns. The standard deviation of the degree-level fixed effects is 22 percentage points (ppts) and the 90:10 range is 52 ppts. Substantial variation in earnings returns remains when we look within selectivity bands and subject. Among the least selective degrees, we find a 15 ppts standard deviation in returns, around half of which is within subject. The standard deviation of returns increases to 29 ppts amongst the most selective degrees, around 70% of which is within subject. This suggests that degree choice is crucial for later life earnings outcomes, even for individuals choosing within a specific subject and a relatively narrow selectivity band.

In light of the importance of degree choice, the final part of chapter 3 explores how well students can identify high return degrees based on the information currently available to them. We find that existing measures of degree quality such as publicly available subject-specific university rankings, degree performance and completion rates are only weakly related to returns, except for business, economics and law. Even then, this correlation almost completely disappears once we control for selectivity, suggesting these measures of degree performance contain little information over and above a simple measure of selectivity. Student satisfaction ratings and early career earnings are not well correlated with returns, even unconditionally. This has important implications for social mobility, as students from lower-socioeconomic backgrounds are more likely to have to rely on public information when choosing their degree.

Chapter 4 focuses on the contribution of different institutions and degrees to social mobility. We first explore how access to these different higher education institutions and degrees varies by socio-economic background and find that gaps in access are hugely variable depending on university selectivity. While students from low-income families are as likely to attend the least selective institutions as their wealthier peers, they are far less likely to attend the top universities: in the mid-2000s, private school students were around 100 times more likely to attend Oxford or Cambridge than pupils on free school meals. We then combine the share of low-income students who attend with the labour market outcomes after university of those students to create a "mobility rate". The most prestigious universities perform badly on this measure. While their graduates from lower socio-economic backgrounds perform extremely well in the labour market, these institutions take in so few of these students that they are making very little contribution to social mobility. Instead, the highest mobility institutions are often less selective institutions based in big cities, such as London and Birmingham. These institutions take in very high shares of disadvantaged students, and have reasonable labour market outcomes. We also find important variation in mobility rates within institutions - many universities are in the top 10% of the mobility rankings for some subjects and in the bottom 10% for others. Comparing our mobility rates to estimates of average earnings returns we find no correlation at the university level, and only a small positive correlation at the course level. Many courses that do poorly in terms of boosting average earnings do a lot to promote mobility. This highlights the importance of not solely focusing on average earnings returns when determining the "value" of degrees. Restricting funding to universities based on their average earnings returns alone could be costly in terms of social mobility.

In chapter 5, I move away from the education system to explore another channel affecting children's later life outcomes. Using UK census data, I estimate the intergenerational impact of paternal job loss. Job loss has been shown to have large and persistent negative effects on those affected, not only on earnings but also in terms of health, mortality (Sullivan and von Wachter, 2009; Kuhn et al., 2009), and even an increased probability of divorce (Charles and Stephens, 2004). Despite this, there is only limited evidence on how parental job loss might affect the later life outcomes of children. This chapter adds to this literature by estimating the intergenerational impact of paternal job loss on later life outcomes for a cohort of children born between 1973 and 1981.

I make use of the large scale mine closures that took place in England and Wales in the late twentieth century to identify an exogenous shock to parental employment. While virtually all active coal mines in the UK closed during the course of the twentieth century, the exact timing of those was driven largely by geological factors such as the condition and location of the coal seam (Glyn, 1988; Glyn and Machin, 1997). Mine workers were further a relatively low-skilled group, usually drawn from the surrounding rural areas, making strong selection into mines closing at different times unlikely. I show descriptives which support this.

Using a 1% linked sample of the population census in England and Wales, I identify a cohort of children born between 1973 and 1981, all of whom have fathers who worked in mining in 1981. I then compare the outcomes of the children whose father lost his job in mining between 1981-1991 to the outcomes of children whose father lost his job after this period. I first document the large and persistent losses in family resources after paternal layoffs, and then estimate the impact of this paternal job loss during childhood on children's later life outcomes. My estimates show significant impacts on later life outcomes for daughters, but not for sons. Women are 13 percentage points (around 30%) less likely to have a degree after paternal job displacement, and are around 12 percentage points less likely to own their home. Around half of the effects of paternal job loss on educational attainment, and around 20% of that on home ownership, can be explained by the loss in father's earnings following displacement. This suggests an important role for unemployment insurance in mitigating the negative consequences of parental job loss on children.

# Chapter 2

# Intergenerational income mobility in England and the importance of education

## 2.1 Introduction

In recent years, the increasing availability of large scale linked administrative datasets has enabled a new literature to emerge which compares intergenerational mobility across small areas within countries. This work, pioneered by Chetty et al. (2014) has shown substantial heterogeneity within countries both in the outcomes of children from poor backgrounds and in the gap in outcomes between children from poor and rich backgrounds, not only in the US, but also in smaller and more centralized countries such as Italy (Acciari et al., 2019) and Canada (Corak, 2019). Subsequent work (Chetty and Hendren, 2018) has shown that much of these differences in mobility between areas are due to causal effects of place rather than selection. While still little is known on the causal mechanisms driving mobility differences across areas, attempts at answering this question have been made by looking at which area characteristics best predict mobility.

This paper adds to this recent and growing literature by providing the first estimates of intergenerational income mobility at a detailed geographical level in England. Previous work in the UK (e.g. Blanden et al., 2004; Gregg et al., 2017) has had to rely on longitudinal surveys, which rely on self-reported earnings, and have samples which are too small to estimate how mobility varies across the country. A recent paper (Bell et al., 2018) uses a 1% sample of the linked census to estimate mobility across broad areas in England, but does not have access to any measures of income and is therefore only able to estimate mobility in terms of occupation, education and homeownership. We use new linked administrative data on the whole population of individuals born since September 1985 and educated in England to estimate income mobility for the 1985-1988 birth cohorts for the approximately 150 local authorities in England. We focus on absolute mobility for individuals from low-income backgrounds using two main measures of mobility. The first looks at the average income rank of children on Free School Meals (which indicates that their parents are on means-tested benefits and roughly identifies the children from the 12.5% of families with the lowest incomes), and the second focuses on bottom to top mobility by looking at the proportion of children on FSM who make it to the top 20% of the income distribution. We combine the 1985 to 1988 birth cohorts and measure income as total earned income at age 28, ranking children within their birth cohort.

We first estimate these two measures of absolute mobility at the national level and find that children on Free School Meals are expected to end up at the 37th percentile of the income distribution in their cohort. This implies that these children move up considerably in the income distribution compared to their parents (who are around the 6th percentile on average), yet do remain substantially below the median. The level of mobility estimated for this group of children from very low-income backgrounds is roughly between that estimated in the US and in Italy.

We then explore how mobility varies according to where a child grows up, focusing on the 152 local authorities in England. We assign children to the local authority where they lived at age 16 and estimate their income rank in the national distribution. Differences across the country are substantial. FSM children in the lowest mobility areas are expected to end up around fifteen percentiles lower on average at age 28 than similar children from the highest mobility areas. This gap in earnings rank is remarkably similar for both men and women. There is a strong North-South gradient in terms of mobility, with the North having the lowest mobility, and virtually all of the highest mobility areas being located in the South-East. There is no clear urban-rural distinction, with areas in and around London performing very well, but Northern cities being among the lowest mobility areas in the country. Geographical patterns of mobility for men and women exhibit some interesting differences, with cities, and particularly inner London, performing relatively better in terms of mobility for women than for men.

Having estimated how mobility varies across the country, we then ask how much of this variation can be explained by differences in human capital across areas. This question is of particular policy importance in the UK, as an important part of the current government's social mobility agenda focuses on improving educational outcomes in the local authorities identified as being most in need of additional support.[1] Our analysis will help inform whether these policies have the potential to equalise earnings mobility across the country.

This analysis is related to that in a recent paper by Rothstein (2019) in the US, but expands on it in a few important ways. Firstly, we focus on absolute mobility, while Rothstein focuses on explaining differences in relative mobility across areas. Second, unlike Rothstein, we use the same samples to estimate mobility and human capital across areas, and can estimate human capital at the

[1]These areas, so called 'Opportunity Areas', were identified based on a range of measures, but educational attainment of children from low-income families played an important part in this selection.

area level. Rothstein combines mobility measures at the Commuting Zone level from Chetty et al. (2014) with a sample of around 15,000 indidividuals from the Education Longitidunal Study, which does not enable him to estimate human capital transmission at the area level. Finally, we have much richer measures of educational achievement. Rothstein uses maths scores at age 18 as his main measure of human capital. We instead make use of rich school and university records which give us test scores in individual subjects at ages 11, 16 and 18, as well as subject of study, institution attended, GPA and degree completion for those who attend university. As we show, using less detailed measures of educational attainment understates the importance of differences in human capital for explaining variation in mobility across areas. Our analysis shows that differences across areas in the educational achievement of children from low-income background can explain 25% of the across-area variation in absolute income mobility for men, and more than 45% for women. This suggests a critical role for improving educational outcomes of children from low-income backgrounds in low mobility areas to equalise opportunities across the country. However, despite the important role of education, a substantial part of mobility differences across areas are unexplained by gaps in educational attainment, suggesting policy should also focus on other potential drivers of mobility differences across areas.

As a first step in investigating what these other drivers of mobility are likely to be, we correlate our mobility measures with area characteristics. We do this both with overall mobility $\overline{R}_a^{FSM}$ and with differences in income rank across areas when we hold educational achievement constant. Areas with stronger labour markets, more stable families, higher median income and better schools have higher mobility.

The remainder of the paper is organised as follows. We begin in Section 2 by describing the data. Section 3 lays out our national and regional mobility estimates. Section 4 discusses how differences in human capital might explain differences in mobility across areas, and shows the results of our decomposition of the variance in absolute mobility. Section 5 reports correlations of absolute mobility with area characteristics, both before and after accounting for educational differences. Finally, Section 6 concludes.

## 2.2 Data

We make use of a new UK linked administrative dataset, the Longitudinal Education Outcomes (LEO) dataset. This dataset consists of three component datasets: school records from the National Pupil Database (NPD), university records from the Higher Education Statistics Agency (HESA) and earnings records from Her Majesty's Revenue and Customs (HMRC). These datasets were linked by the UK's Department for Work and Pensions before we got access to the data. Where National Insurance Numbers[2] were available, these were used to hard link education and tax records. Where no National Insurance Numbers were available fuzzy matching using first name, surname, date of birth, postcode and gender was used. In what follows we briefly summarise the main variables and describe our sample.

### 2.2.1 Parental background

Due to the lack of common identifiers, and the absence of up to date address information in UK tax records, it is not possible to reliably link children to their parents in the UK. Fortunately for our purposes however, the NPD school records document whether a child is eligible for Free School Meals (FSM). Children are eligible for FSM when their parents are in receipt of means tested benefits,[3] and have annual gross income below a given threshold, currently £16,190.

[2]National Insurance Numbers are unique person identifiers assigned to individuals at age 16, or upon starting a first job, which are broadly the UK equivalent of US Social Insurance numbers.

[3]These benefits are: Income Support, income-based Jobseeker's Allowance, income-related Employment and Support Allowance, support under Part VI of the Immigration and Asylum

These eligibility criteria, including the income cutoff, do not vary across the country. In our cohorts of analysis, 12.5% of students are eligible for Free School Meals. FSM eligibility therefore broadly identifies the 12.5% of children coming from the lowest income families.[4] The main focus of this paper will be on the outcomes of children who were eligible for Free School Meals at age 16.

### 2.2.2 Child income

Our measure of child income will be earned income at age 28, combining income from employment and self-employment. Data on earnings come from two complementary records, combined in the HMRC tax data: Pay As You Earn (PAYE) records, which record income from employment, and Self Assessment (SA) records, which incorporate income from self-employment. We combine income from both sources and use total earnings in our analysis. We have this data up to the 2016/17 tax year,[5] which means we observe our first cohort (those born between 1st September 1985 and 31st August 1986) up until around age 30. In order to look at earnings as late as possible in the lifecyle, yet have sufficiently large samples to estimate absolute mobility at the the small local area level, we combine three cohorts in our analysis and make use of their earnings at age 28, the last age at which we observe earnings from all three cohorts.

### 2.2.3 Educational attainment

The English education system is characterised by frequent standardized and externally marked tests. This makes the UK administrative education records extremely comprehensive in detailing the educational trajectory of children. At age 11, the end of primary school, children sit Year 6 Standard Assessment Tests

Act 1999, the guaranteed element of Pension Credit and Child Tax Credit (provided the parents are not also entitled to Working Tax Credit).

[4]There may be families with gross income just above the eligibility cut-off but net income (post taxes and benefits) below than that of some of the families who are eligible for FSM.

[5]Tax years in the UK run from 6th April of one year to the 5th of April of the next year.

(SATs) in English, maths and science. At age 16, students sit General Certicate for Secondary Education (GCSE) exams. Students in our cohorts of study sat these exams in English, maths, science, and in typically between five and seven additional subjects. We observe each subject taken, and the grade obtained in each exam. End of high school exams, so called 'A levels' are taken at age 18, in typically between three and four subjects. Students may also sit vocational exams in addition to, or instead of, academic A-levels, in courses such as hospitality or retail. We again have indicators for the subjects taken, as well as grades obtained. Finally, we also have detailed data on higher education attendance from HESA records. We observe subject of study at a granular level, institution attended, whether an individual completes their course and degree classification.

For the purposes of our analysis we will use this rich educational data to construct an overall index of human capital, by combining the information on subjects taken and grades achieved into a continuous measure as follows. Writing these multiple measures as vector $S_i$, we first regress individual earnings on $S_i$:

$$Y_i = \alpha + S_i'\beta + \epsilon_i \tag{2.1}$$

Using the coefficients obtained in this regression, we then generate our measure of human capital $H_i$ as predicted earnings $\widehat{Y}_i$:

$$H_i = \widehat{Y}_i = \widehat{\alpha} + S_i'\widehat{\beta} \tag{2.2}$$

The measures we include are: at age 11, quadratics in science, maths and English scores; at age 16, a quadratic in total scores, and separate indicators of grades in maths, science and English, total number of subjects taken and a dummy for whether achieved 5 A*-C grades; at age 18, an indicator for taking any A-levels, a quadratic in total score and dummies for taking maths, science and social science A-levels at age 18; for post-18 education, indicators for university attended and subject(s) taken, and a dummy for whether they dropped out of their course, and a dummy for whether they achieved a first class or upper second class degree. In addition to our main human capital index, which includes all the above measures, we also construct human capital indices at different ages, to investigate the importance of having this rich set of educational attainment measures. We construct measures of human capital at ages 11, 16 and 18, where for each age we include all educational attainment measures up to that point.

### 2.2.4 Sample

Our full sample includes all individuals who 1) are born between 1st September 1985 and 31st August 1988, 2) went to school in England, and 3) were linked to HMRC or DWP records at any point between 2004/05 and 2016/17.[6]

The focus of this paper is on absolute upward mobility, as defined by the outcomes of children who were on Free Schools Meals at age 16. Table 1 below gives some basic descriptives of children on FSM and compares them to the average characteristics of the non-FSM pupils in the same cohorts, for men and women separately. Children from this group of low-income families are around three times more likely to have English as an additional language, and are more likely to be Black or of Asian ethnicity. They obtain far lower test scores than the average student, and this gap increases between age 11 and age 16. Only around one in four FSM pupils stays in school past age 16, compared to around half of non-FSM pupils. Consequently FSM students are also much less likely to attend university and obtain a degree. This lower educational attainment translates into adult outcomes considerably worse than those of their non-FSM peers. Average age 28 earnings of this group are around £8,000 lower than those

[6]We require individuals to have been in touch with the tax and benefit system at some point between 2004/05 and 2016/17 for them to be recorded in the data. This does not mean individuals have to have had positive earnings at any point in this time frame.

of their peers from more affluent backgrounds, and they are around twice as likely to have no earned income.

Table 2.1: Sample descriptives

| | Women | | Men | |
|---|---|---|---|---|
| | non-FSM | FSM | non-FSM | FSM |
| *Cohort* | | | | |
| 2001/02 GCSE cohort | 0.31 | 0.31 | 0.32 | 0.31 |
| 2002/03 GCSE cohort | 0.33 | 0.34 | 0.34 | 0.34 |
| 2003/04 GCSE cohort | 0.35 | 0.35 | 0.35 | 0.35 |
| *Background characteristics* | | | | |
| English as additional language | 0.07 | 0.20 | 0.06 | 0.20 |
| White | 0.79 | 0.70 | 0.80 | 0.71 |
| Black | 0.02 | 0.07 | 0.02 | 0.06 |
| Asian | 0.05 | 0.14 | 0.05 | 0.14 |
| Missing/Other ethnicity | 0.14 | 0.10 | 0.14 | 0.10 |
| *Educational attainment* | | | | |
| Age 11 test score pctile | 52.89 | 37.57 | 51.00 | 37.86 |
| Age 16 test score pctile | 56.85 | 33.50 | 48.62 | 26.44 |
| Stay in school past 16 | 0.56 | 0.28 | 0.46 | 0.20 |
| Start UG | 0.48 | 0.25 | 0.38 | 0.19 |
| Graduate from UG | 0.41 | 0.18 | 0.31 | 0.12 |
| *Age 28 outcomes* | | | | |
| Mean earnings (£s) | 17,700 | 9,900 | 22,300 | 14,700 |
| Median earnings (£s) | 16,000 | 6,600 | 21,200 | 13,500 |
| Has self-employment income | 0.06 | 0.03 | 0.11 | 0.09 |
| No earned income | 0.15 | 0.31 | 0.13 | 0.22 |
| N | 710,036 | 100,709 | 764,556 | 108,181 |

Notes: The first column shows descriptives of the full set of individuals who were born between 1st September 1985 and 31st August 1988, went to school in England, and were linked to HMRC or DWP records at any point between 2004/05 and 2016/17. The second column shows the same descriptives for the subset of those individuals who were eligible for Free School Meals at age 16. "Attend UG" shows the proportion who start a undergraduate degree, while "UG degree" shows the proportion who actually obtain an degree. Mean and median earnings are defined including zero earnings. "Has self-employment income" shows the proportion of people who have any self-employment income at age 28, regardless of whether this is their only source of income, or whether they also have employment income. Not in employment is defined as individuals who report no employment or self-employment income at age 28.

## 2.3 Absolute mobility in England

We begin by documenting absolute upward mobility at the national level. This will provide some context for the regional estimates of mobility, and will allow us to compare the adult outcomes of children from low-income families in England with those from children from similar backgrounds in other countries. We will then show the variation in mobility across local areas in England.

### 2.3.1 National estimates

We estimate two measures of mobility, one looking at the average outcomes of children from low-income backgrounds, and one looking at how many children from low-income backgrounds make it to the top of the income distribution. The first measure, which we will call $\overline{R}_a^{FSM}$, is the average income rank at age 28 of children from low-income families, as defined by being on Free School Meals at age 16. This is similar to the $r^{25}$ measure of absolute mobility used in Chetty et al. (2014), and the AUM measure used in Acciari et al. (2019), but using a narrower definition of "disadvantage". Where those papers focused on the average outcomes of children from families with below median income, we focus on the average outcomes for those with parental income roughly in the bottom 12.5% nationally. The second measure of mobility we use focuses on the share of children from low-income backgrounds who attain real "success" in the labour market, as defined by reaching the top quintile of the income distribution in their cohort. We define this measure of mobility, which we will call $P(Q5|FSM)$, as the proportion of children on FSM who make it to the top 20% of the income distribution in their cohort at age 28.

Table 2.2 brings together our estimates of absolute mobility for the whole sample and for men and women separately. On average, the parents of children on free school meals were around the sixth percentile in the income distribution.[7] While they still end up considerably below the median, their children do still move up considerably in the income distribution, reaching the 37.5 percentile of the income distribution at age 28 on average. While there are some differences in our income measures, we can get an idea of how this compares internationally by comparing this to the average income rank of children with parents in the bottom 12% of income in Italy from Acciari et al. (2019) and in the US from Chetty et al. (2014). A child with parents in the bottom 12% of income

[7] Families on FSM broadly correspond to the families with income in the bottom 12.5% nationally.

is expected to end up at the 39th percentile of the child income distribution in Italy, and at the 35th percentile in the US. Absolute upwards mobility for this group in England is nearly exactly in between that of Italy and the US.

Turning now to $P(Q5|FSM)$, our measure of bottom-to-top mobility, we see that 8.4% of children on FSM reach the top quintile of the child income distribution at age 28. This shows a considerably degree of immobility - this is less than half of the 20% we would see if parental income had no impact on child outcomes. The equivalent figure is the US from Chetty et al. (2014) is even lower at 6.8%.

In the two right hand side columns of Table 2.2 we split the sample by gender and estimate mobility for men and women separately. We find that men from low-income families end up 10 percentiles higher in the income distribution on average than women. The gender difference is even more pronounced in terms of the probability of moving to the top quintile of the income distribution. This probability is 5.5% for women, compared to 11% for men. As these gender differences are likely driven to a large extent by differences in average incomes and differential labour market participation of men and women at this age,[8] we also rank individuals in their gender specific income distribution. Even in terms of their position in their gender-specific income distribution, men from low-income backgrounds have higher mobility than women from similar backgrounds, both in terms of their average income rank and in terms of their probability of reaching the top of the income distribution. Exploring the labour market participation of men and women, we find much larger socio-economic gaps in labour participation for women than men. At age 28, 13% of non-FSM men has no earned income at age 28, compared to 22% for FSM men. For women, these shares are 15% and 31% respectively. Some of these differences are likely to be driven by differential patterns of fertility for women from different socio-

[8] We cannot observe hours in our data, but do observe individuals with zero earnings. Among our group of children from low-income families, 31% of women have no earned income, compared to 22% of men.

economic backgrounds at this age. If these patterns were to differ significantly across areas, we might be concerned that the ranking of areas in terms of mobility at age 28 might be very different to that at later ages. We investigate this in Appendix Figure A.1, where we use the British Household Panel Survey (BHPS) to estimate the relationship between age 28 and later earnings for women by region. We don't find evidence that this relationship differs significantly across regions, reassuring us that our mobility ranking of different areas in the next section is unlikely to differ substantially to that of later ages.

Table 2.2: National estimates of absolute mobility

| | Overall | Women | Men |
|---|---|---|---|
| $\overline{R}^{FSM}$ | 37.5 | 32.3 | 42.3 |
| $\overline{R}^{FSM}$ - gender specific | | 36.0 | 38.3 |
| $P(Q5\|FSM)$ | 0.084 | 0.055 | 0.111 |
| $P(Q5\|FSM)$ - gender specific | | 0.075 | 0.087 |
| N | 208,915 | 100,716 | 108,199 |

Notes: $\overline{R}^{FSM}$ is the average income rank at age 28 of children from low-income families, as defined by being on Free School Meals at age 16. $P(Q5|FSM)$ shows the proportion of children on FSM who make it to the top 20% of the income distribution in their cohort at age 28. Sample consists of English educated individuals born between 1st September 1985 and 31st August 1988. Income ranks defined within each school cohort (1st September to 31st August of each year). We assign children with zero income the average income rank of that group.

### 2.3.2 Geographical variation

Having looked at mobility at the national level, we now explore how outcomes of low-income children vary depending on where in the country they grow up. We estimate this for around all local authorities in England.[9] The average population in a local authority is just over 320,000 individuals, though these vary in size from less than 50,000 individuals for the smallest local authorities, to close to 1.3M for the largest local authority.[10]

[9] We use upper-tier local authorities as defined in the most recent census, which took place 2011. We only show results for local authorities with at least 100 children on FSM.

[10] We measure population size from the 2001 population census, which is the closest census to when we measure location of residence of our sample (2002-2004).

For each area, we estimate the same two measures of absolute mobility as used at the national level. We first estimate, for each area $a$, the average income rank at age 28 of children on Free School Meals. We call this measure $\overline{R}_a^{FSM}$. In the Appendix we also show the results for $P_a(Q5|FSM)$, the proportion of children on FSM who grew up in area $a$ and make it to the top 20% of the national income distribution at age 28. We always define children's income ranks at the national level and within their cohort, and assign children to the local authority where they lived at age 16.

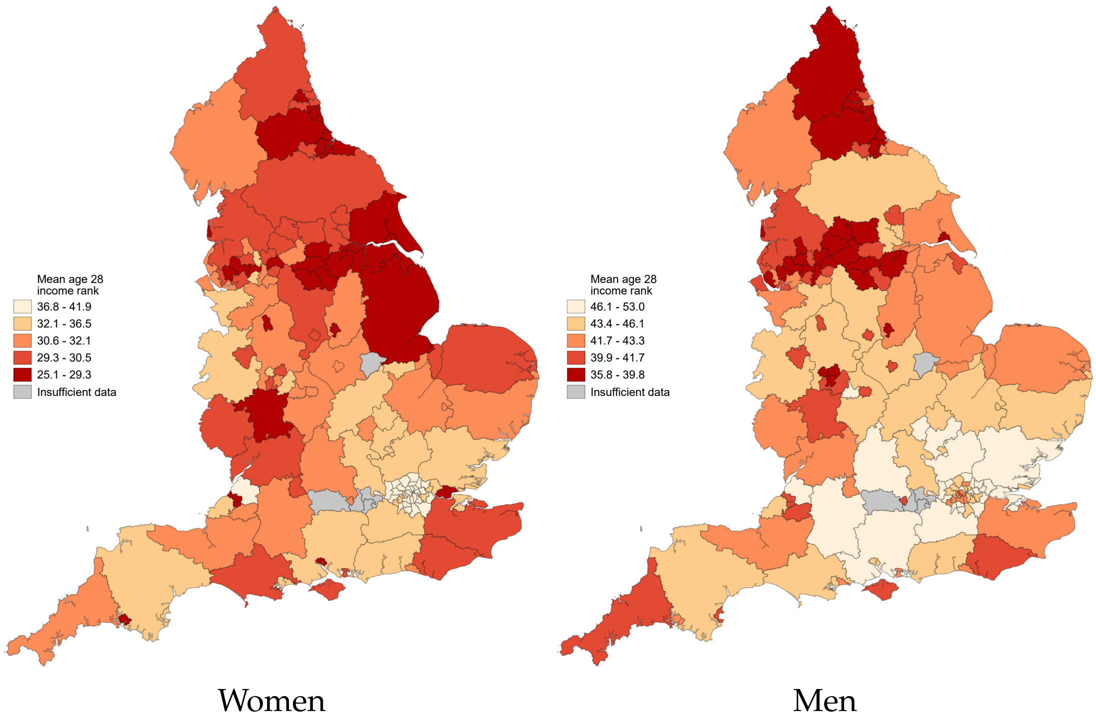
We show the estimates of $\overline{R}_a^{FSM}$ for all local authorities in Figure 2.1. The first thing to note is the large differences in average earnings ranks. Both for men and women, children on FSM who grow up in the highest mobility areas are expected to end up around fifteen percentiles higher in the income distribution than those from the lowest mobility areas. Some broad geographical patterns emerge. There is a clear North to South East gradient for both genders, though there are also some interesting differences across gender. Women from low-income backgrounds who grew up in London have the highest average income ranks at age 28, and those growing up in the Northern cities the lowest. For men, outcomes in the North East are relatively much better, and many areas in inner London do not perform as well in terms of mobility. There seems to be something about these inner city areas that produces bad outcomes for men, but not for women. The overall correlation between absolute mobility for men and women is only 0.65, which suggests that there might be important differences in the characteristics of areas which lead to good outcomes for men, and those that lead to good outcomes for women.

We explore whether these differences between men and women are driven by differences in participation, and show how results differ when excluding individuals with zero earnings in Figure A.2 in the Appendix. Urban areas tend to have higher non-participation rates than more rural areas, especially for men.[11]

[11] See Figure A.4 in the Appendix.

As a result, when we exclude individuals with zero earnings from our mobility measure, a few rural areas do slightly worse, and some urban areas slightly better. Overall, however, this does not seem to substantially change the geographical patterns we observed and the two measures are highly correlated (corr = 0.96 for women and 0.94 for men).

Figure 2.1: Average earnings rank age 28 - by gender




Notes: Figure shows average earnings rank at age 28 of individuals who were on FSM at age 16. Individuals are assigned to the area they lived at age 16. Only areas with at least 100 children on FSM in our analysis sample are shown. Local authorities are split into quintiles based on the mean earnings rank of FSM children. The darkest colour shows the areas with the lowest average earnings rank. We assign children with zero income the average income rank of that group.

In Figure A.3 in the Appendix we additionally look at mobility to the top, as measured by $P_a(Q5|FSM)$. For women, the areas where FSM children do well on average are nearly indistinguishable from those where they are most likely to reach the top of the income distribution. For men, rural areas in the North and South West do relatively better in terms of average earnings than in terms of the share of children reaching the top of the income distribution. Inner London on the other hand does better on this measure, perhaps unsurprisingly by the high incomes at the top in London. Both measures are still highly related (corr=0.89 for men compared to 0.94 for women). This indicates that broadly areas which have good outcomes for FSM children on average are also those where these children have the best chance of making it to the top of the income distribution.

## 2.4 The importance of education

In the previous section, we established that substantial variation exists in the earnings outcomes of children from low-income backgrounds, depending on where in the country they grew up. In this section, we investigate whether differences in educational attainment can explain this variation. If differences in educational attainment are an important driver of differences in absolute income mobility across areas, we would expect children from areas with high absolute income mobility to have high levels of human capital. Conversely, if we find that the average income ranks and human capital levels of individuals across areas are not highly correlated, or if these human capital differences do not matter much in the labour market, we would expect other mechanisms, such as labour market policies, to be more important in driving the differences in income mobility across areas. We investigate this more formally by decomposing the variance in absolute income mobility across areas in a component which can be explained by differences in human capital accumulation and the return to this human capital in the labour market, and a remainder which cannot be explained by human capital differences across areas. This decomposition helps inform us as to the potential for educational interventions to equalise opportunities for children from low-income backgrounds across the country.

We discuss this decomposition in Section 2.4.1, before showing the results in Section 2.4.2.

### 2.4.1 Methodology

Write the adult income rank of FSM eligible child $i$ who grew up in area $a$ as $R_{i,a}^{FSM}$. We can then write their earnings as a function of their human capital $H_{i,a}^{FSM}$, the returns to this in the labour market, $\beta$, and the earnings impact of the location they grew up in controlling for their level of human capital, $\eta_l$:

$$R_{i,a}^{FSM} = \beta H_{i,a}^{FSM} + \eta_a + w_{i,a} \tag{2.3}$$

We can also write their level of human capital as a function of the average human capital in the area they grew up in, and an individual level error term:

$$H_{i,a}^{FSM} = \overline{H}_a^{FSM} + v_{i,a} \tag{2.4}$$

Plugging in equation 2.4 into equation 2.3 and rearranging, we get:

$$R_{i,a}^{FSM} = (\beta \overline{H}_a^{FSM} + \eta_a) + (\beta v_{i,a} + w_{i,a}) \tag{2.5}$$

In the previous section, we estimated $\overline{R}_a^{FSM}$, the average adult income rank of FSM children in each area. We can write individual earnings rank of individual $i$ who is on Free School Meals and grows up in area $a$ as:

$$R_{i,a}^{FSM} = \overline{R}_a^{FSM} + u_{i,a} \tag{2.6}$$

Comparing equations 2.6 and 2.5 we can then see that $\overline{R}_a^{FSM} = (\beta \overline{H}_a^{FSM} + \eta_a)$. This implies we can decompose the average earnings rank in an area, $\overline{R}_a^{FSM}$, into a part which is explained by the level of human capital in an area ($\overline{H}_a^{FSM}$) and the return to this in the labour market ($\beta$), and a remainder which is not explained by human capital levels in the area.

We use this to decompose the variance in absolute mobility (as defined by $\overline{R}_a^{FSM}$) across areas into a part which is explained by differential human capital accumulation across areas, and a part which cannot be explained by differences in human capital accumulation. Using $\overline{R}_a^{FSM} = (\beta\overline{H}_a^{FSM} + \eta_a)$ we get:

$$Var(\overline{R}_a^{FSM}) = \beta^2 Var(\overline{H}_a^{FSM}) + Var(\eta_a) + 2\beta Cov(\overline{H}_a^{FSM}, \eta_a) \tag{2.7}$$

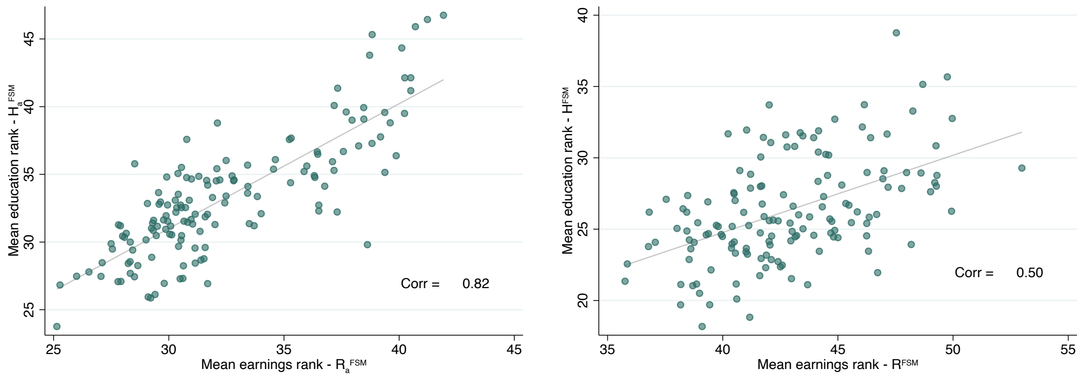
The ratio $(\beta^2 Var(\overline{H}_a^{FSM}))/Var(\overline{R}_a^{FSM})$ tells us what proportion of the variation in absolute income mobility can be explained by differences in human capital accumulation across areas. This will inform us as to the importance of education policy in improving earnings outcomes of children from disadvantaged background across the country.

### 2.4.2 Results

We first look at the variation in average human capital across areas, and how this relates to the variation in average income rank across areas. We plot the average value of the human capital index in each area[12] against the average income rank in each area for FSM eligible children in Figure 2.2. We can see a strong positive correlation between the average human capital index and the average income rank in an area, but there is still considerably variation in income ranks among areas with similar average human capital.

[12]As described in Section 2.2, the human capital index is predicted earnings rank based on test scores and educational attainment at ages 11, 16, 18 and in university.

Figure 2.2: Correlation of $\overline{R}_a^{FSM}$ mobility measure and human capital index $\overline{H}_a^{FSM}$




Women Men

Notes: $\overline{H}_a^{FSM}$ shows the average level of the human capital index in an area, which is created by predicting income rank of FSM eligible children based on their educational achievement at ages 11, 16, 18 and in university. $\overline{R}_a^{FSM}$ is the average income rank of FSM children who grew up in area $a$. The dotted line shows the results from an anweighted regressions of $\overline{H}_a^{FSM}$ on $\overline{R}_a^{FSM}$.

Table 2.3 shows the results of the decomposition of the variance in $\overline{R}_a^{FSM}$. The top rows of Table 2.3 show that if we include all measures of educational achievement from age 11 to university, for women 46% and for men 25% of the variation in absolute income mobility across areas can be explained by variation in the educational achievement of children from low-income backgrounds across areas. This points towards a meaningful role for improving educational achievement of low-income students in low mobility areas in order to equalise opportunities across the country. Yet other channels, for example differences in labour market practices across areas, are clearly at least as important in explaining variation in mobility across the country. In the rest of the table, we look at the impact of including fewer measures of educational achievement. We show the impact of only including student achievement up to ages 11, 16 and 18 in the human capital index, which highlights the importance of including rich measures of educational achievement. Solely including measures at younger ages considerably understates the proportion of the variance in absolute mobility which can be explained by differences in educational achievement. Sections A.6.1 and A.6.2 in the Appendix show that these findings are robust to relaxing our implicit assumption of constant returns to human capital across areas, and to changing how we construct measures of human capital.

Table 2.3: Decomposition of $Var(\overline{R}_a^{FSM})$

| | | Share of $Var(\overline{R}_a^{FSM})$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | **Men** | | | **Women** | | |
| | | Main | Excl. London | Excl. £0s | Main | Excl. London | Excl. £0s |
| *HC index* | $\beta^2 Var(\overline{H}_a^{FSM})$ | 0.25 | 0.12 | 0.18 | 0.46 | 0.28 | 0.30 |
| | $Var(\eta_a)$ | 0.75 | 0.80 | 0.73 | 0.27 | 0.53 | 0.33 |
| | $2\beta Cov(\overline{H}_a^{FSM}, \eta_a)$ | 0.00 | 0.08 | 0.09 | 0.30 | 0.19 | 0.37 |
| *HC index - up to age 11 only* | $\beta^2 Var(\overline{H}_a^{FSM})$ | 0.04 | 0.04 | 0.03 | 0.04 | 0.09 | 0.02 |
| | $Var(\eta_a)$ | 0.92 | 0.95 | 0.96 | 0.89 | 0.95 | 0.96 |
| | $2\beta Cov(\overline{H}_a^{FSM}, \eta_a)$ | 0.04 | 0.01 | 0.01 | 0.07 | -0.04 | 0.02 |
| *HC index - up to age 16 only* | $\beta^2 Var(\overline{H}_a^{FSM})$ | 0.20 | 0.12 | 0.12 | 0.23 | 0.24 | 0.12 |
| | $Var(\eta_a)$ | 0.79 | 0.88 | 0.82 | 0.45 | 0.65 | 0.58 |
| | $2\beta Cov(\overline{H}_a^{FSM}, \eta_a)$ | 0.01 | 0.00 | 0.06 | 0.33 | 0.11 | 0.30 |
| *HC index - up to age 18 only* | $\beta^2 Var(\overline{H}_a^{FSM})$ | 0.22 | 0.12 | 0.13 | 0.30 | 0.24 | 0.16 |
| | $Var(\eta_a)$ | 0.80 | 0.89 | 0.82 | 0.38 | 0.63 | 0.51 |
| | $2\beta Cov(\overline{H}_a^{FSM}, \eta_a)$ | -0.02 | 0.00 | 0.05 | 0.32 | 0.12 | 0.34 |
| | Total $Var(\overline{R}_a^{FSM})$ | 12.6 | 10.3 | 11.9 | 17.9 | 7.1 | 22.7 |
| | Number of areas | 143 | 112 | 139 | 145 | 113 | 140 |

Notes: Local Authorities with fewer than 250 individuals included in the analysis are dropped from the analysis of that gender. Results are shown for multiple measures of human capital, using educational attainment up to age 11, and up to age 16, up to age 18 and up to age 21. Results are shown for men and women separately and human capital measures and area fixed effects are constructed completely separately by gender.

## 2.5 What do mobile areas look like?

We have shown that some areas in England exhibit much higher absolute mobility than others, and that part, but not all of this, can be explained by low-income children in high mobility areas having higher educational achievement. As a further step in trying to explain why certain areas have higher mobility than others, we correlate our measures of mobility with other area characteristics.

While these correlates cannot be interpreted as causal determinants of mobility, they can help guide future research towards potential policies which may help improve mobility.

We look at different sets of area characteristics which have been suggested as important for mobility in previous work in economics or sociology. The first column in Tables 2.4 and 2.5 show, for women and men respectively, the unweighted correlations of the $\overline{R}_a^{FSM}$ measure of mobility and area characteristics at the Local Authority level for all areas with at least 250 FSM eligible children in our sample. Columns (2) to (4) add in an indicator for being a London LA, regional fixed effects, and controls for median income in the LA respectively. Most of our correlations follow the expected direction. Labour market characteristics seem to play an important role, with areas with a stronger labour market having higher mobility, as well as areas with a higher share of skilled jobs. Areas with lower family stability, as measured by the share of single parent families and the divorce rate, have lower mobility. We do not have those measures at the individual level, so cannot disentangle whether this relationship is driven by children of single parents families having lower mobility, or all children in areas with lower family stability having lower mobility. Median income and school quality are also strongly positively associated with mobility. Interestingly, and contrary to previous work in other countries, we find a positive relationship between inequality and mobility. This positive correlation seems to be driven by the good performance of areas in London, where inequality is relatively high. Once we add an indicator for being in London this correlation disappears. When we control for median income, the relationship actually reverses, and we find a strong negative relationship between inequality and mobility, consistent with previous work.

Comparing the factors which predict mobility for men and women, we find significant differences in terms of immigration and ethnic make up of areas. There is a much stronger positive relationship between a higher share of immigration and non-white population for women than for men. Controlling for region fixed effects does not change this for women, but reverses the relationship for men. We also find a stronger association between mobility and the share of skilled jobs, median income and population density, and a weaker association with measures of labour market strength for women than for men. These differences suggests that the policies and institutions which are most effective at promoting mobility for men may not always be those most effective for women. Heterogeneity across groups in mobility patterns, and in their drivers, is an important area for further study.

We also look at correlations between area characteristics and the income rank of FSM children, holding educational achievement constant ($\eta_a$ in equation 2.5) in Columns (5) to (8) in Tables 2.4 and 2.5. For many area characteristics this does not meaningfully alter their correlation with mobility. The largest changes can be found in terms of immigration and ethnicity. Controlling for education significantly reduces the strength of association between higher immigration and a higher non-white share of the population and mobility. Once we control for region fixed effects or median income of the area, there is no relationship for women, while for men we now see strong negative correlations with the share of non-white children and the share of immigration in the area. This suggest that the strong performance of areas with high shares of immigration and a high share of non-white individuals is largely due to the strong educational performance of children in those areas. This aligns with the findings from previous work which find that in England, unlike in many other countries, ethnic minority children outperform their white counterparts in school (though not always in the labour market).

Many of the areas characteristics we have looked at so far will be highly correlated with each other. In Tables A.7 and A.8 in the Appendix we therefore run multivariate regressions where we include all area characteristics[13] to assess

[13]Closely related area characteristics are combined into indices as described in Appendix Sec-

Table 2.4: Correlations of area characteristics and mobility (women)

| | Raw area effects | | | | Controlling for education | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| *Labour market* | | | | | | | | |
| Economically | 0.363*** | 0.246*** | 0.129** | 0.010 | 0.446*** | 0.370*** | 0.255*** | 0.247** |
| active | (0.082) | (0.046) | (0.049) | (0.062) | (0.079) | (0.066) | (0.075) | (0.080) |
| Unemployment | 0.026 | -0.198*** | -0.066 | 0.084 | -0.157 | -0.324*** | -0.169* | -0.118 |
| | (0.088) | (0.049) | (0.053) | (0.055) | (0.087) | (0.070) | (0.081) | (0.074) |
| Professional | 0.726*** | 0.321*** | 0.238*** | 0.132 | 0.496*** | 0.214* | 0.115 | 0.009 |
| jobs | (0.061) | (0.060) | (0.064) | (0.125) | (0.077) | (0.096) | (0.105) | (0.167) |
| Manufacturing | -0.674*** | -0.178* | -0.044 | -0.292*** | -0.519*** | -0.234* | -0.074 | -0.280** |
| share | (0.065) | (0.069) | (0.079) | (0.068) | (0.076) | (0.102) | (0.124) | (0.094) |
| *Immigration and ethnicity* | | | | | | | | |
| Share white | -0.781*** | -0.365*** | -0.325*** | -0.479*** | -0.406*** | 0.056 | 0.126 | -0.099 |
| | (0.055) | (0.070) | (0.062) | (0.057) | (0.081) | (0.112) | (0.107) | (0.095) |
| Share Asian | 0.567*** | 0.249*** | 0.231*** | 0.371*** | 0.279** | 0.033 | -0.018 | 0.130 |
| | (0.073) | (0.051) | (0.043) | (0.048) | (0.085) | (0.082) | (0.075) | (0.077) |
| Share Black | 0.672*** | 0.087 | 0.107 | 0.297*** | 0.368*** | -0.173 | -0.118 | 0.033 |
| | (0.065) | (0.078) | (0.078) | (0.066) | (0.082) | (0.114) | (0.122) | (0.096) |
| Share foreign | 0.825*** | 0.450*** | 0.373*** | 0.546*** | 0.440*** | -0.100 | -0.241 | 0.068 |
| born | (0.050) | (0.081) | (0.074) | (0.067) | (0.079) | (0.131) | (0.125) | (0.111) |
| *Family stability* | | | | | | | | |
| % single parent | 0.055 | -0.225*** | -0.185*** | 0.021 | -0.038 | -0.239** | -0.128 | -0.063 |
| families | (0.088) | (0.049) | (0.048) | (0.055) | (0.088) | (0.074) | (0.078) | (0.074) |
| % married | -0.223** | 0.165** | 0.104* | -0.104 | -0.020 | 0.276*** | 0.165* | 0.067 |
| families | (0.086) | (0.054) | (0.052) | (0.055) | (0.088) | (0.077) | (0.082) | (0.075) |
| *Income distribution* | | | | | | | | |
| Median earnings | 0.781*** | 0.357*** | 0.246** | 0.781*** | 0.551*** | 0.280* | 0.147 | 0.551*** |
| | (0.055) | (0.072) | (0.076) | (0.055) | (0.074) | (0.112) | (0.122) | (0.074) |
| 90:10 ratio | 0.329*** | 0.127* | 0.076 | -0.070 | 0.156 | -0.001 | -0.061 | -0.186 |
| | (0.098) | (0.063) | (0.062) | (0.074) | (0.108) | (0.093) | (0.095) | (0.100) |
| 90:50 ratio | 0.330*** | 0.050 | 0.046 | -0.312*** | 0.056 | -0.155 | -0.152 | -0.477*** |
| | (0.097) | (0.060) | (0.059) | (0.077) | (0.099) | (0.084) | (0.089) | (0.098) |
| *Urbanity* | | | | | | | | |
| Urban | 0.232** | -0.065 | -0.015 | 0.092 | 0.134 | -0.074 | -0.012 | 0.034 |
| | (0.086) | (0.053) | (0.048) | (0.056) | (0.088) | (0.078) | (0.074) | (0.075) |
| Share rural | -0.303*** | 0.034 | 0.000 | -0.095 | -0.218* | 0.012 | -0.032 | -0.072 |
| | (0.084) | (0.055) | (0.049) | (0.057) | (0.086) | (0.080) | (0.076) | (0.077) |
| *Segregation* | | | | | | | | |
| Segregation by | 0.390*** | 0.178*** | 0.103* | 0.089 | 0.327*** | 0.186* | 0.093 | 0.124 |
| KS4 score | (0.081) | (0.050) | (0.045) | (0.060) | (0.084) | (0.074) | (0.070) | (0.080) |
| Index of dissim | -0.041 | 0.132** | 0.090* | -0.035 | 0.044 | 0.165* | 0.100 | 0.048 |
| - FSM | (0.088) | (0.050) | (0.044) | (0.055) | (0.088) | (0.073) | (0.069) | (0.074) |
| Index of dissim | -0.375*** | -0.058 | 0.081 | -0.048 | -0.340*** | -0.138 | -0.020 | -0.127 |
| - ethnicity | (0.082) | (0.054) | (0.051) | (0.061) | (0.083) | (0.079) | (0.080) | (0.081) |
| *School quality* | | | | | | | | |
| % schools rated | 0.368*** | 0.168*** | 0.146*** | 0.076 | 0.169 | 0.026 | 0.003 | -0.053 |
| outstanding | (0.082) | (0.050) | (0.042) | (0.060) | (0.087) | (0.076) | (0.069) | (0.080) |
| Avg school | 0.084 | 0.175*** | 0.070 | 0.044 | 0.147 | 0.210** | 0.058 | 0.118 |
| value-added | (0.088) | (0.048) | (0.048) | (0.055) | (0.087) | (0.071) | (0.076) | (0.073) |
| Population weights | No | No | No | No | No | No | No | No |
| London dummy | No | Yes | No | No | No | Yes | No | No |
| Region FE | No | No | Yes | No | No | No | Yes | No |
| Control for median inc | No | No | No | Yes | No | No | No | Yes |

Notes: Each column shows the coefficients univariate regressions of Local Authority level mobility measures separately on each area characteristics listed in the rows. Both mobility and area characteristics are standardized and coefficients can therefore be interpreted as correlations. Area characteristics are described in more detail in the Appendix.

which characteristics are the strongest predictors of absolute mobility. Columns (1) to (3) show this for absolute upwards mobility as measured by $\overline{R}_a^{FSM}$. The

tion A.8.

Table 2.5: Correlations of area characteristics and mobility (men)

| | Raw area effects | | | | Controlling for education | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| *Labour market* | | | | | | | | |
| Economically | 0.553*** | 0.495*** | 0.282*** | 0.365*** | 0.516*** | 0.512*** | 0.312*** | 0.522*** |
| active | (0.073) | (0.065) | (0.067) | (0.079) | (0.075) | (0.076) | (0.081) | (0.087) |
| Unemployment | -0.347*** | -0.485*** | -0.257*** | -0.282*** | -0.460*** | -0.513*** | -0.293*** | -0.435*** |
| | (0.082) | (0.068) | (0.072) | (0.069) | (0.078) | (0.078) | (0.086) | (0.077) |
| Professional | 0.414*** | 0.197* | 0.073 | -0.324* | 0.113 | 0.082 | -0.018 | -0.488** |
| jobs | (0.080) | (0.100) | (0.093) | (0.151) | (0.087) | (0.114) | (0.111) | (0.173) |
| Manufacturing | -0.435*** | -0.217* | 0.020 | -0.129 | -0.163 | -0.182 | -0.011 | 0.004 |
| share | (0.079) | (0.107) | (0.112) | (0.095) | (0.086) | (0.121) | (0.133) | (0.112) |
| *Immigration and ethnicity* | | | | | | | | |
| Share white | -0.254** | 0.195 | 0.283** | 0.134 | 0.130 | 0.450*** | 0.560*** | 0.454*** |
| | (0.085) | (0.114) | (0.092) | (0.091) | (0.087) | (0.123) | (0.102) | (0.099) |
| Share Asian | 0.191* | -0.020 | -0.073 | 0.022 | -0.109 | -0.191* | -0.268*** | -0.207* |
| | (0.086) | (0.087) | (0.068) | (0.077) | (0.087) | (0.096) | (0.078) | (0.087) |
| Share Black | 0.138 | -0.488*** | -0.499*** | -0.325*** | -0.204* | -0.646*** | -0.652*** | -0.578*** |
| | (0.087) | (0.110) | (0.099) | (0.088) | (0.086) | (0.120) | (0.116) | (0.095) |
| Share foreign | 0.322*** | -0.167 | -0.344** | -0.138 | -0.081 | -0.487*** | -0.682*** | -0.520*** |
| born | (0.083) | (0.134) | (0.107) | (0.102) | (0.087) | (0.145) | (0.117) | (0.111) |
| *Family stability* | | | | | | | | |
| % single parent | -0.336*** | -0.531*** | -0.345*** | -0.359*** | -0.442*** | -0.524*** | -0.292*** | -0.453*** |
| families | (0.082) | (0.067) | (0.066) | (0.066) | (0.078) | (0.079) | (0.083) | (0.075) |
| % married | 0.294*** | 0.584*** | 0.424*** | 0.370*** | 0.478*** | 0.625*** | 0.441*** | 0.518*** |
| families | (0.083) | (0.068) | (0.065) | (0.066) | (0.077) | (0.078) | (0.080) | (0.072) |
| *Income distribution* | | | | | | | | |
| Median earnings | 0.554*** | 0.482*** | 0.348** | 0.554*** | 0.254** | 0.428** | 0.325* | 0.254** |
| | (0.073) | (0.113) | (0.108) | (0.073) | (0.084) | (0.130) | (0.130) | (0.084) |
| 90:10 ratio | 0.292** | 0.187 | 0.127 | 0.064 | 0.124 | 0.115 | 0.091 | 0.070 |
| | (0.102) | (0.098) | (0.081) | (0.106) | (0.106) | (0.110) | (0.103) | (0.121) |
| 90:50 ratio | 0.093 | -0.073 | 0.021 | -0.263** | -0.104 | -0.155 | -0.025 | -0.300** |
| | (0.099) | (0.092) | (0.085) | (0.094) | (0.098) | (0.103) | (0.104) | (0.108) |
| *Urbanity* | | | | | | | | |
| Urban | -0.088 | -0.278*** | -0.159* | -0.201** | -0.219* | -0.287** | -0.166* | -0.278*** |
| | (0.087) | (0.079) | (0.065) | (0.072) | (0.085) | (0.089) | (0.078) | (0.083) |
| Share rural | 0.053 | 0.279*** | 0.179** | 0.227** | 0.200* | 0.284** | 0.186* | 0.296*** |
| | (0.087) | (0.081) | (0.067) | (0.073) | (0.086) | (0.092) | (0.080) | (0.085) |
| *Segregation* | | | | | | | | |
| Segregation by | 0.269** | 0.153 | 0.005 | 0.052 | 0.150 | 0.132 | -0.016 | 0.055 |
| KS4 score | (0.084) | (0.080) | (0.065) | (0.080) | (0.086) | (0.090) | (0.077) | (0.093) |
| Index of dissim | 0.050 | 0.111 | 0.069 | -0.013 | 0.044 | 0.058 | 0.006 | 0.015 |
| - FSM | (0.087) | (0.078) | (0.062) | (0.073) | (0.087) | (0.088) | (0.074) | (0.085) |
| Index of dissim | -0.339*** | -0.204* | -0.015 | -0.145 | -0.283*** | -0.282** | -0.149 | -0.217* |
| - ethnicity | (0.082) | (0.081) | (0.072) | (0.078) | (0.084) | (0.090) | (0.085) | (0.090) |
| *School quality* | | | | | | | | |
| % schools rated | 0.228** | 0.126 | 0.062 | 0.058 | 0.021 | -0.002 | -0.071 | -0.066 |
| outstanding | (0.085) | (0.079) | (0.062) | (0.077) | (0.087) | (0.090) | (0.074) | (0.089) |
| Avg school | 0.464*** | 0.502*** | 0.307*** | 0.413*** | 0.449*** | 0.459*** | 0.247** | 0.428*** |
| value-added | (0.077) | (0.064) | (0.063) | (0.064) | (0.078) | (0.078) | (0.079) | (0.077) |
| Population weights | No | No | No | No | No | No | No | No |
| London dummy | No | Yes | No | No | No | Yes | No | No |
| Region FE | No | No | Yes | No | No | No | Yes | No |
| Control for median inc | No | No | No | Yes | No | No | No | Yes |

Notes: Each column shows the coefficients univariate regressions of Local Authority level mobility measures separately on each area characteristics listed in the rows. oth mobility and area characteristics are standardized and coefficients can therefore be interpreted as correlations. Area characteristics are described in more detail in the Appendix.

next three columns do the same thing with the absolute mobility measure controlling for differences in educational achievement ($\eta_a$ in equation 2.5).

Higher immigration, more stable families, lower inequality and higher share of professional jobs all are strong predictors of absolute mobility, but most other factors lose their significance. Controlling for education differences gives very similar results, but the share of foreign born individuals loses its significance for women, and becomes negative for men, in line with the univariate correlations shown above.

## 2.6 Conclusion

This is the first paper which estimates intergenerational income mobility at the detailed geographical level in England. It finds considerable differences in absolute upwards mobility across the country. Children from low-income families who grew up in the highest mobility areas end up on average fifteen percentiles higher up the income distribution than those from similar backgrounds who grew up in the lowest mobility areas.

More than 45% of this variation in absolute mobility across areas can be explained by differences in educational attainment of children from low-income backgrounds across areas for women, while the equivalent for men is 25%. This suggests that current government policies focused at improving mobility through improving educational outcomes, are promising interventions to equalise opportunities across areas. For a government truly committed to 'levelling-up' across areas however, solely focusing on education will not be enough. While we cannot interpret these as causal drivers, local labour market conditions, average incomes in an area, and stable families are strong predictors of mobility, and those would be promising areas to look for potential policies to increase mobility.

Our work also tentatively suggests that the factors promoting mobility for men from low-income backgrounds may not be the same as those which are most effective at promoting mobility for women from the same backgrounds. Investigating heterogeneity across different groups in mobility patterns, and in the mechanisms driving those, is an important direction for future research.

# Chapter 3

# How much does degree choice matter?

## 3.1 Introduction

As in many countries around the world, prospective higher education students in the United Kingdom (UK) choose between a vast number of different degree options when entering university. This paper exploits a pioneering new administrative dataset to look at labour market outcomes at the degree level - that is, the interaction of subject field and institution. We explore the variation in earnings returns and investigate the predictability of those returns based on other observable characteristics of the degree. To the best of our knowledge, this paper is the first to estimate returns for individual degrees across an entire higher education market.

We find substantial variation in returns, even within relatively tight selectivity bands and within subject. This implies degree choice matters much more than some of the previous evidence has suggested. We find only a weak relationship between selectivity and returns through much of the distribution, but a strong overall positive relationship at the top end of the selectivity distribution, suggesting that there is a large payoff to high ability students attending elite universities. However, this is not true for all subject areas - for some, such as creative arts, there is only a very weak relationship between selectivity and returns throughout the distribution. Finally, aside from selectivity, we find that existing measures of degree quality are not well related to returns. This matters because these measures influence both student choices and university behaviour.

We exploit a new administrative data linkage that was developed in partnership with the UK Department for Education. The dataset links together administrative school, university and tax records for the more than three million individuals who completed secondary school in England between 2002 and 2007. The tax records include annual earnings from 2005/06 to 2016/17, meaning we observe the oldest cohort in the data up until age 30. The school records allow us to condition on an extremely detailed set of prior attainment controls that include exam grades in specific subjects at ages 11, 16 and 18, as well as rich information on student background and secondary (high) school fixed effects. Unlike some of the recent papers in this literature, the dataset tracks all students through all of the available higher education institutions in the country, and captures anyone who is filing for taxes anywhere in the country.

Our data contains more detailed background information on students than many previous papers have been able to use. We exploit this to test the likely role of unobservable factors in driving our results. We show that our headline findings are robust to the exclusion of subsets of our control variables, suggesting that unobservable factors are not likely to affect our main conclusions. We also show that the main findings are not sensitive to reasonable changes in the sample or regression specification.

We start by estimating overall returns to higher education, before looking at how returns vary by institution and subject, within the set of people who go. We find fairly low overall returns to higher education for men, but much higher returns for women. However, when we look within the set of individuals who go to higher education, gender differences are less important. Across institutions, we find a weak association between selectivity and returns through much of the selectivity distribution, but a much stronger relationship at the top end of the distribution, suggesting large payoffs to attending the most elite universities in the UK, in particular the Universities of Oxford and Cambridge, the London School of Economics, and Imperial College London. We estimate big differences by subject. Medicine, economics and law do particularly well and social care and creative arts courses perform poorly in terms of earnings returns. In general, these findings are consistent with the previous literature, which is reassuring for our degree-level estimates.

We then turn to the most novel contribution of the paper by estimating returns at the 'degree' level, which is the interaction of institution and subject. We are able to estimate returns for almost 2000 subject-university combinations (for example, mathematics at the University of Warwick). This is a natural level of granularity to focus on for the UK, and many other countries, where people choose specific subject-university combinations for their degrees prior to starting, and is only viable because of the unique dataset at our disposal. There is substantial variation in raw earnings outcomes across different degrees: the standard deviation of the degree-level fixed effects, without any controls, is 32 percentage points (ppts) and the 90:10 range is 75 ppts. These figures drop to 22 ppts and 52 ppts respectively once we include the full set of controls for prior attainment, student characteristics, and secondary school fixed effects.

There is still substantial variation in returns, even when looking within relatively tight selectivity bands. Amongst the least selective degrees, the standard deviation in returns is still more than 15 ppts, increasing to 29 ppts amongst the most selective set of degrees. It is also the case that a large share of the variation in returns is *within* subject, even within our selectivity bands. Roughly 50% of the variation in degree returns for the least selective band of degrees is within subject, rising to more than 70% of the variation for the most selective degrees. Combined, these results strongly suggest that degree choice is crucial for subsequent earnings outcomes, right across the selectivity distribution, even holding subject choice fixed. For example, it is not at all uncommon to see differences in returns of 40 ppts between degrees in the same subject at similarly selective universities.

Given the importance of degree choice in determining earnings outcomes, in the final part of the paper we consider the predictability of returns across different institutions, within subject. We find that existing measures of degree quality are not well correlated with returns. As with the institution estimates, on average there is only a weak relationship between degree selectivity and returns through much of the distribution but a much stronger relationship at the top end. However, this varies a lot by subject area: for economics, law and business, returns increase rapidly with university selectivity, while for others, such as sociology and the creative arts, they do not.[1] We then see that other measures of degree quality including publicly available subject-specific university rankings, completion rates and degree performance are all correlated with returns, but this almost completely disappears once we control for selectivity. This suggests that observable measures of degree performance contain little information over and above a simple measure of selectivity. Student satisfaction ratings and early career earnings are not well correlated with returns, even unconditionally.

These observable degree characteristics matter. For example, Gibbons et al. (2015) shows that public league table rankings are a key driver of student choices, while many of the other measures we look at (such as very early career earnings and student survey scores) are used as inputs for centralised evaluation of teaching quality in the UK, through the 'Teaching Excellence Framework'. The result that public information on degrees is not well correlated with the earnings outcomes of students has several important implications. First, it will matter for productivity if students select degrees that are not highly valued

[1]We also see no relationship between returns and selectivity for medicine and education, which is not surprising, as so many graduates from these subjects go into careers with centrally regulated wages.

in the labour market. Second, it will affect inequality, as students from more disadvantaged backgrounds are more likely to have to rely on public information when making their higher education choices. Indeed, Campbell et al. (2019) highlight that poorer students are more likely to choose degrees associated with lower earnings outcomes, conditional on prior attainment. Third, it is likely to incentivise universities to focus on metrics that may not be beneficial to the long-term outcomes of students, as doing well on those metrics helps them to achieve good scores in teaching evaluations and attract students.

This rest of this paper is set out as follows. Section 3.2 reviews the related literature and discusses how our paper fits into it. Section 3.3 then describes the dataset we use and gives more detail on the institutional background in the UK. Section 3.4 outlines our methodology. Our results are then presented in Sections 3.5 and 3.6. Section 3.5 provides estimates of the overall earnings returns of attending university versus not attending, and looks at heterogeneity in returns across institutions and subjects. Section 3.6 then focuses on the degree level returns estimates and shows the relationship between degree level returns and selectivity, as well as with other observable characteristics of the degree. Finally, Section 3.7 concludes.

## 3.2 Literature

Our work draws upon and contributes to a substantial academic literature which investigates returns to higher education. This literature can be divided into three main branches. The first investigates the overall returns to higher education using selection-on-observables, finding that returns are high on average (Webber, 2016; Walker and Zhu, 2011; Blundell et al., 2000). Our findings also suggest good overall returns to university and the magnitudes of the estimates are consistent with the previous UK literature, based on estimates before the rapid expansion of higher education that occurred in the UK between during the 1980s and 1990s. This suggests the returns held up well through this expansion, consistent with descriptive evidence from Blundell et al. (2016).

The second strand of this literature investigates heterogeneity in returns by university attended. Many of these papers look at heterogeneity across broad groups of institutions (Chevalier and Conlon, 2003; Andrews et al., 2017; Walker and Zhu, 2018) or at the relationship between returns and a continuous measure of university quality or selectivity (Hussain et al., 2009; Broecke, 2012; Black and Smith, 2006; Dale and Krueger, 2002, 2014; Dillon and Smith, 2020). However, more recent papers have started to investigate heterogeneity in returns across individual institutions (Cunha and Miller, 2014; Mountjoy and Hickman, 2020; Chetty et al., 2020). Many of these papers, like ours, identify returns based on OLS estimation with rich background characteristics. Some of them attempt to address selection issues by controlling for the set of colleges students applied to or were accepted at. While we do not observe application sets, our data contains much more detailed background information on students than previous work has been able to use. In particular, we observe full and detailed academic histories of each student, including specific grades in specific subjects based on national tests taken at ages 11, 16 and 18, alongside rich background characteristics allowing us to control for the local area in which people grow up and the school they attended. Hastings et al. (2013) and Hastings et al. (2018) instead exploit discontinuities in university entry cutoffs to identify returns to different institutions. They find their results to be consistent with those obtained using OLS conditioning on rich observables, without controlling for application sets. Drawing on evidence from their own experimental work as well as that of Wiswall and Zafar (2014) they argue that students do not know much about earnings outcomes and select their university largely based on factors that are unlikely to be correlated with later outcomes. Dillon and Smith (2020) make a similar argument in a recent paper that focuses on match effects in higher education. These papers further strengthen our confidence in our identification strategy.

The evidence from this literature on the relationship between returns and selectivity is mixed. The UK evidence consistently finds a strong relationship between university selectivity and returns (for example, Walker and Zhu, 2018), as do many of the previous papers from the United States. Dale and Krueger (2002, 2014) and Mountjoy and Hickman (2020) however, which control for the application sets of students, all suggest a very weak relationship. However, the work by Cunha and Miller (2014), which exploits very similar data and uses a similar approach to Mountjoy and Hickman (2020) find a strong relationship for universities in Texas, while several papers which have exploited discontinuities in university entry cutoffs to identify returns to specific universities have also found large effects (Anelli, 2018; Hoekstra, 2009; Hastings et al., 2013; Saavedra, 2008; Zimmerman, 2019).[2] We find a weak association between selectivity and returns throughout much of the selectivity distribution, but this becomes much stronger at the top end of the distribution. This suggests very large payoffs to attending the most elite universities in the UK.[3]

The third strand of related literature investigates heterogeneity in returns by subject studied. Altonji et al. (2012) reviews the evidence to that date, highlighting that the majority of papers estimating returns assume selection on observables (Walker and Zhu, 2011, 2018; Blundell et al., 2000; Chevalier, 2011). However, again there are some papers which have exploited discontinuities in

[2]Our paper also relates more tenuously to papers that have estimated effects for students at the margin of going and not going to university. For example, Zimmerman (2014) finds very large earnings returns for academically marginal students. While our estimates are in relative terms and are therefore not compared to the outside option of not going, we still see that the average returns for the least selective institutions are no lower than returns for middling universities. This suggests that our results are consistent with the idea that returns are reasonably good for universities that accept students with low prior attainment who are likely to be close to the margin of going and not going.

[3]The discrepancy between this finding and that of Mountjoy and Hickman (2020) could be explained by the fact that they do not observe anything like the range in university quality that we do. They look at 27 four-year colleges in Texas, where the top institution is University of Texas, Austin. This is a considerably less selective, and less elite institution than the top UK universities. We also note that they suffer from out-of-state selection problems (both for university and for work), while this is dramatically less important in the UK, where very few students study abroad or work abroad after graduation.

entry cutoffs to identify returns to different subject choices (Kirkeboen et al., 2016; Hastings et al., 2013). Kirkeboen et al. (2016) presents a compelling case that there is a large amount of selection into different subjects based on comparative advantage, suggesting returns based on OLS regressions would overstate the causal effects. This finding leads us to be cautious about our cross-subject returns estimates, although we note that our data on subject-specific prior attainment is a considerable improvement on much of the literature. With this caution in mind, we find that economics and medicine are the highest returning subjects, with conditional returns of more than 30% relative to history (our base case). Computing, business and architecture also do well, with returns of 15-20% above the base. At the other end of the scale, social care, creative arts, agriculture and veterinary sciences are the lowest-returning subjects with returns of 10-15% below history (social care is the lowest). Psychology, English, languages and biological sciences also perform poorly (notably, many of these subjects are much more likely to be chosen by women). Our results are more mixed for STEM degrees, as we find a lot of variation in returns within this broad subject area, which is an important result given the context of large pro-STEM agendas in several countries.

As described above, we believe our main contribution to the literature is to investigate returns at the degree level. The only previous paper that has had sufficiently high quality data do this is Hastings et al. (2013), which is able to exploit discontinuities in entry cutoffs to around 1,100 different degree programmes in Chile in order to identify returns. However, the focus of that paper is not on individual degree returns, but rather the relationship between returns and university selectivity, the returns by subject, and the returns by subject interacted with a binary indicator for high selectivity. Like in our paper, selectivity is found to be strongly related to subject returns for some subjects but not for others. To our knowledge, our paper is the first to estimate individual returns for individual university degrees across the whole of a higher education system. This exercise is extremely revealing about the extent to which degree choices can potentially impact later-life outcomes.[4]

Looking at individual degree level returns also enables us to look more carefully at the relationship between observable degree-level characteristics and returns than any previous paper.[5] This enables us to consider the relevance of the *ex-ante* information on degree quality available for students and regulators. Our finding that the measures of degree quality that we consider are unrelated to returns is highly pertinent as this information influences student's degree choices, regulator's ratings of teaching quality, and also the priorities of universities.

## 3.3 Data

We use the Longitudinal Educational Outcomes (LEO) dataset, which was developed in collaboration with the UK Department for Education for the purposes of this paper. In this section, we define our analysis sample, give more information about each of the composite datasets of LEO, and show summary statistics of our analysis sample.

### 3.3.1 Sample

Our base sample of students consists of all individuals who: (1) attended school in England; (2) passed their age 16 exams between 2002 and 2007,[6]; (3) are linked to UK tax records for any of the tax years 2013-14 to 2016-17; and (4) started an

[4]Our institutional setting is also quite different to that of Hastings et al. (2013). The UK has a much larger higher education sector than Chile (OECD, 2014), with a much broader range of institutions, including many that cater to students with relatively low prior attainment as well as several internationally renowned institutions that regularly feature in the top ten of world university rankings. Our findings are therefore likely to be more relevant to higher education systems of countries with more advanced economies such as the US, Australia and several European countries.

[5]Mountjoy and Hickman (2020) and Chetty et al. (2017) look at this, but their comparisons are quite limited. They are only able to look at overall university characteristics of the university, rather than characteristics at the subject-institution level.

[6]We define "passing" age 16 exams as obtaining at least 5 A*-C grades in GCSE exams - see below for more detail on these exams.

undergraduate degree in the UK between the ages of 17 and 21 as a full-time student. This gives us between 161,000 and 204,000 individuals in each cohort (as defined by the year they took their age 16 exams), giving a total of over one million individuals.

When estimating the overall returns to a degree, we will compare these individuals to a control group of individuals who satisfy criteria (1) to (3) above, but did not attend university at any point in our dataset (we drop part time and mature students from the analysis completely). We identify individuals in each group by taking all individuals who appear in administrative records of age 16 exams, and linking them to administrative tax and university records. More information on match rates and sample selection is provided in Appendix B.1.

### 3.3.2 Demographics and school attainment

We obtain information on background characteristics and school attainment of individuals from the National Pupil Database (NPD), which contains exams files as well as a census of English schools.

In England, students take national, externally marked examinations at age 11, 16 and 18, and we have all three records in our data. The age 11 tests, taken at the end of primary school, are the Year 6 Standard Assessment Tests (SATs). They are taken in three subjects - English, mathematics and science - and we have detailed scores from each. The age 16 tests are based on 'General Certificate of Secondary Education' (GCSE) exams, the majority of which are taken in the summer of the school year people turn 16 (Year 11).[7] GCSEs during this period were taken in English (literature and language), mathematics and science plus typically five to seven additional subjects and were graded from A*-G. A grade C was generally considered to be a pass - indeed, a key metric for progression onto further education or training was often whether an individual had at

[7]The school year in England runs from September 1 to August 31.

least five GCSEs graded between A* and C. We observe all of the subjects taken and the grades achieved in each. For presentational purposes, the GCSE exam grades are converted into a single points index (where an A* is worth 58 points, an A is worth 52 points and so on down to the lowest scored grade of G, which is worth 16 points). The age 18 assessment data are based primarily on scores in A-level exams, which are usually taken two years after GCSEs (Year 13). For A-levels, students take exams in the typically three or four subjects they chose to study after GCSEs. A-levels were graded from A-F during this period, with a D grade often considered to be the minimum pass. Again, we observe the subjects taken and the grades achieved. Students during this period can take equivalent vocational qualifications instead of (or as well as) A-levels, such as courses in retail or hospitality, which we also observe.

The school census contains school identifiers and student level demographics, including gender, age, ethnicity, special educational needs and an indicator for English not being the student's first language. We further observe whether a student is eligible for Free School Meals (FSM) and have access to detailed measures of deprivation in the small local area (approximately 130 households) where the child lives at age 16.[8] Following several previous papers (e.g. Chowdry et al., 2013), we combine these multiple measures into one continuous index of socio economic status (SES) at age 16 using principal components analysis. The approximately 7% of pupils who attend private secondary schools are missing the school census data (but we do observe their exam records).[9] We keep this group in the analysis and include missing dummies for any missing

[8]In order for a pupil to be eligible for free school meals their family has to be on means-tested benefits. FSM eligible pupils therefore approximately represent pupils from the poorest 12-13% of families. Local area level deprivation measures include the proportion of individuals in the pupil's local area of residence with a degree, with no qualifications, in managerial and professional jobs, in routine occupations, long-term unemployed, homeowners, in social housing as well as the proportion of children living in income deprived households (IDACI). All these measures are included at the Output Area level (containing 130 households on average), except IDACI and the proportion of individuals living in social housing, which are both measured at the Lower Super Output Area (around 670 households on average).

[9]Some students who also attended a private primary school have no age 11 exam records, but these students do all have age 16 and age 18 exam records.

school census information.

The earliest cohort for whom we have individual level school records are the students who took their age 16 exams in 2002. The vast majority of these individuals were born between 1 September 1985 and 31st August 1986.

### 3.3.3 University attendance

We obtain information on higher education attendance from the Higher Education Statistics Authority (HESA) data. For each year an individual attends a university in the UK this administrative dataset records the type of degree, subject studied, university attended, course intensity (part-time vs full-time) and degree performance. We link individuals over time to determine whether they graduate from their degree.

Students who apply to university typically do so in the the academic year they take their A-level (or equivalent) exams. About half of students who go to university do so within a few months after their A-level (or equivalent) exams, while another 30-40% go within the next two years. We focus on university entrants within this three-year window, meaning that the majority of the HESA records we use are from the 2004/05 - 2009/10 academic years. People who we observe going to university after this window are dropped from the analysis. We observe HESA data up until 2015/16, which allows us to remove mature students starting university up until the year they turn 29.

The most common route through university is to attend one institution for an undergraduate degree and to study one subject (although several students study joint degrees with more than one subject). Full-time degrees are usually three years, though some degrees such as languages or sciences are four year degrees. In the HESA records we observe subject, university and course intensity (part-time vs full time), and we are able to link people over time to determine whether they graduate.[10]

Degree subjects are recorded in meticulous detail, with more than 1,500 different subject categories provided. We aggregate these up to around 30 broad subject areas (for example mechanical engineering and civil engineering are aggregated to engineering) based on the official 'Common Aggregation Hierarchy'.[11] To summarise our findings, we sometimes further group these subjects in three groups: LEM (Law, Economics and Management[12]), STEM (Science, Technology, Engineering and Mathematics), and Other, which consists of other social sciences, arts and humanities subjects. A complete list of the subjects in each group is provided in the Appendix.

Individuals attend one of more than 100 UK universities which provide undergraduate degrees. For some analysis we classify universities into five broader groups based on prestige and selectivity. The four most selective and prestigious universities in the UK (the University of Oxford, the University of Cambridge, Imperial College London and the London School of Economics) are put together into the 'Elite Russell Group'. These four universities have notably higher prior attainment than any other universities in the country.[13] The next most selective group of universities are the 'Russell Group' which is a well-known self-defined collective of 24 (including the aforementioned four) high-status institutions. This group is followed by the 'Old universities' which includes the remaining 31 institutions which predate the large expansion of universities that occurred in England in 1992. The remaining universities are non-traditional universities, such as art colleges, or are former technical colleges which converted to university status in 1992. This group of around 80 typi-

[10] For people who did not graduate from their first degree and switched to a second degree, we take their second degree as their undergraduate qualification, so long as it was taken as a full-time, non-mature student.

[11] For a complete list of these, see https://www.hesa.ac.uk/innovation/hecos.

[12] This is common terminology - in practice for our subject classifications this is law, economics and business.

[13] This can be seen in Figure B.4 in the Appendix which show the average GCSE score of universities' student intake.

cally less selective institutions is divided into two equal groups ('more selective' and 'less selective') based on the average GCSE points scores of their students. A complete list of the universities in each group is provided in an Online Appendix.

### 3.3.4 Earnings

Individuals' earnings are obtained from Her Majesty's Revenue and Customs (HMRC) tax records. Earnings from conventional employment are recorded in Pay As You Earn (PAYE) records, which we have for the 2005/06 - 2016/17 tax years.[14] Earnings from self employment and profits from partnerships are recorded separately in Self Assessment (SA) records. We only have these latter records from 2013/14 - 2016/17. To avoid missing a substantial fraction of total earnings,[15] we only make use of the data from 2013/14 onwards. This has the additional advantage of avoiding the immediate labour market fallout from the 2008 recession. The tax data only includes information only on total annual earnings, and we observe no measures of hours worked.

Tax records have been matched to university and school records by the UK Department for Work and Pensions. They employed fuzzy matching using National Insurance Number,[16] first name, surname, date of birth, postcode and gender. The first cohort for whom this link exists are those who took their age 16 exams in 2002, who were born between 1st September 1985 and 31st August 1986.[17] These individuals will be approximately aged 30 in the last tax year for which we have earnings records (2016/17).

Due to concerns about early career earnings not being representative of later life earnings, we only include earnings from individuals aged 25 or older. As

[14] In the UK tax years run from April 6th to April 5th of the following year.

[15] By age 30, around 10% of individuals in our sample have self-employment income.

[16] Equivalent to the US Social Security Number.

[17] In practice, individuals who sat their age 16 exams in 2002 but skipped a year in school or were held back a year might be born before or after this. Skipping a year, or being held back a year is however a very rare occurrence in England.

our complete earnings records run from 2013/14 to 2016/17, the age restrictions mean our analysis will include individuals born between 1st September 1985 and 31st August 1991.

### 3.3.5 Data descriptives

In Table 3.1 we show some background characteristics, demographics and prior attainment of individuals who passed their GCSEs split by whether or not they studied for an undergraduate degree. Undergraduates are more likely to have attended a private secondary school, and are more likely to come from higher socio-economic backgrounds. They are also more likely to be non-white, reflecting the higher participation rates of Asian students in particular. As expected, they also have higher prior attainment, being much more likely to have achieved high grades both in age 11 and age 16 exams.

Table 3.2 summarises our undergraduate sample by the different university and subject groups. More women than men attend university, but slightly more men than women attend the most selective universities. We see around 20,000 men attending one of the four Elite Russell group universities compared to around 17,000 women.[18] Between 100,000 and 170,000 individuals of each gender attend each of the four other university groups, with women outnumbering men in each of these groups. There are also large gender differences in the broad subject areas studied, with more than half of women studying arts, humanities and other social science degrees (labelled 'Other') compared to just 40% of men. Men are more likely to do both LEM (21% vs 16%) and STEM degrees (39% vs 30%).[19] A comparison of the earnings distribution of graduates and non-graduates, and earnings across subjects and institution groups is

[18]One reason for this is that two the four Elite Russell Group universities (Imperial College and LSE) specialise in only a subset of subject areas that are more commonly chosen by men. However, it is still the case that women attend Russell Group or Elite Russell Group universities at a lower rate than men do.

[19]For the full list of subjects in each of these categories, see Table B.3 in the Appendix.

shown in Appendix B.3.

Table 3.1: Background characteristics by attainment group

| | Women | | Men | |
|---|---|---|---|---|
| | No UG | UG | No UG | UG |
| *Background* | | | | |
| State school | 0.95 | 0.85 | 0.94 | 0.83 |
| *of which:* | | | | |
| SES Q1 (richest) | 0.23 | 0.33 | 0.25 | 0.35 |
| SES Q2 | 0.25 | 0.25 | 0.26 | 0.25 |
| SES Q3 | 0.22 | 0.20 | 0.22 | 0.19 |
| SES Q4 | 0.18 | 0.13 | 0.16 | 0.12 |
| SES Q5 (poorest) | 0.13 | 0.09 | 0.11 | 0.08 |
| FSM | 0.07 | 0.05 | 0.06 | 0.04 |
| EAL | 0.04 | 0.10 | 0.04 | 0.10 |
| SEN | 0.03 | 0.02 | 0.05 | 0.03 |
| *Ethnicity* | | | | |
| White | 0.91 | 0.80 | 0.91 | 0.80 |
| Black | 0.01 | 0.04 | 0.01 | 0.03 |
| Asian | 0.03 | 0.09 | 0.02 | 0.09 |
| Other | 0.06 | 0.08 | 0.06 | 0.08 |
| *Attainment* | | | | |
| Age 11 Maths level 5+ | 0.23 | 0.40 | 0.34 | 0.50 |
| Age 11 English level 5+ | 0.33 | 0.52 | 0.24 | 0.41 |
| Age 16 Maths A/A* | 0.07 | 0.34 | 0.09 | 0.40 |
| Age 16 English A/A* | 0.14 | 0.49 | 0.08 | 0.38 |
| N | 329,079 | 602,169 | 320,506 | 500,086 |

Note: UG indicates the individual is treated as an undergraduate in our sample. The No UG group excludes people who did not get five A*-C grades in their GCSE exams. We pool here pooled across the six GCSE cohorts. EAL = English as an additional language, FSM = free school meals, SEN = non-statemented special educational needs. Most of the shares here are based on the state school sample only, except the state-educated share, the age 16 (GCSE) results and some of the age 11 (SAT) exam results (as described in the data section above). The attainment section shows the share of individuals who obtained at least level 5 in their age 11 exams, and the share who obtained an A or A* (the two highest grades) in their maths or English age 16 exams. N is based on the full sample including the independently educated.

Table 3.2: Number of students by university and subject groups

| | Women | | Men | |
|---|---|---|---|---|
| | N | Share | N | Share |
| *University group* | | | | |
| Elite Russell | 16,965 | 0.03 | 20,362 | 0.04 |
| Russell Group | 158,549 | 0.26 | 138,453 | 0.28 |
| Old universities | 106,063 | 0.18 | 100,149 | 0.20 |
| Other (more selective) | 162,466 | 0.27 | 127,865 | 0.26 |
| Other (less selective) | 156,376 | 0.26 | 112,363 | 0.23 |
| Total | 600,419 | 1.00 | 499,192 | 1.00 |
| *Subject group* | | | | |
| LEM | 96,526 | 0.16 | 103,372 | 0.21 |
| STEM | 182,378 | 0.30 | 194,828 | 0.39 |
| Other | 323,265 | 0.54 | 201,886 | 0.40 |
| Total | 602,169 | 1.00 | 500,086 | 1.00 |

Note: includes individuals in the undergraduate group, pooling across six GCSE cohorts. A very small number of graduates are missing information on the university attended, hence the slightly lower sample size in the top panel. LEM indicates 'Law, Economics and Management' and STEM indicates 'Science, Technology, Engineering and Mathematics'.

## 3.4 Earnings model

Our identification strategy relies on selection on observable characteristics. The basic premise follows much of the returns to education literature (for example, Blundell et al., 2000) by estimating a regression model:

$$ln(y_{it}) = \alpha + X_i'\gamma + \sum_j \beta_j D_{ji} + \epsilon_{it} \tag{3.1}$$

where $D_{ji}$ is an indicator for the type of degree ($j$) the individual ($i$) has graduated from and $X_i'$ is a vector of observable characteristics. The outcome measure of interest, $ln(y_{it})$, is the log of annual earnings at time $t$.[20]

The key assumption here is that there are no variables omitted from this equation that are related to both the higher education choice and subsequent

[20] We do not adjust for where in the country people are living when they are working as we consider this to be part of the causal pathway from going to university.

earnings outcomes. Put differently, the assumption is that:

$$cov(D_{ji}, \epsilon_{it}|X_{it}) = 0 \quad \forall \ j \tag{3.2}$$

which says that conditional on the control variables $X$ there is no correlation between the earnings residual and the decision to enter higher education. The challenge in estimating the earnings return to university is therefore to account for all the differences between individuals that might affect both their decision to enrol and their earnings prospects. In what follows we set out our approach to dealing with this challenge.

### 3.4.1 Pooled earnings model

We start by documenting the regression specification that we use, which extends the model given by equation 3.1. The oldest cohort in our sample, the 2002 GCSE cohort, has a median age of 30 in 2016/17, our last year of data. For our headline estimates we use age 30 in order to allow for growth in returns with age as much as possible while keeping our estimates within sample. However, to avoid relying solely on observations from one cohort of students, we include several cohorts of students and also multiple earnings observations per individual in a pooled cross-sectional model. This is important because when we look at the degree (subject interacted with institution) level, sample sizes can be small. The pooled model allows us to estimate returns at age 30 while smoothing across several cohorts, reducing the chances of us over-fitting the model.

Specifically, for individual $i$ from GCSE cohort $c \in \{2002, ..., 2007\}$ at time $t \in \{-5, ..., 0\}$, where $t$ is the number of years since the individual took their GCSEs (normalised to zero for the tax year 14 years after GCSEs, or approximately age

30), we model log real earnings as follows:

$$
\begin{aligned}
ln(y_{ict}) = \quad & X_i'\gamma + I(age_{start} > 18) + \omega_1 t + \omega_2 t^2 + \sum_{c=2003}^{2007} c \\
& + \sum_j \beta_j D_{ji} + \sum_j \beta_{1j}(D_{ji}t) + \sum_j \beta_{2j}(D_{ji}t^2) + \epsilon_{ict}
\end{aligned} \tag{3.3}
$$

That is, we model log earnings as a function of observable characteristics $X_i'$ (see more on this below), a dummy for the individual not starting their degree at age 18 (that is, straight after leaving school), a quadratic in $t$, a set of cohort dummies based on GCSE year (with 2002 the omitted category), the treatment of interest ($D_i$), a treatment-specific quadratic trend in age ($D_i f(t)$) and a random component ($\epsilon_{ict}$).

We exclude individuals still in education or with earnings below £1,000.[21] We further windsorise earnings at the 99th percentile. The latter restriction is to reduce sensitivity to large outliers, while the former is because we are concerned that people with very low earnings in a given tax year are likely to only be working part of the tax year, or a very low number of hours.[22] All earnings data are put into 2018/19 tax year prices to adjust for inflation.

Our main results focus on earnings at age 30, or $t = 0$. We therefore extract our estimates for the different treatments of interest by plugging $t = 0$ into equation 3.3.[23] These estimates are point-in-time gross earnings returns meaning they do not adjust for taxes or student loan payments, nor foregone work experience and other costs incurred during study.

We estimate two main sets of models for equation 3.3. To estimate the overall returns to attending higher education, the treatment of interest $D_i$ is simply a dummy for whether the individual attended higher education. In this case, the

[21]We check robustness of our findings to this restriction and find that our results do not quantitatively change if we instead restrict on earnings above £0, nor if we restrict on earnings above £5,000.

[22]We solely observe annual earnings in the tax records, and do not observe hours worked.

[23]This means for our headline estimates, the coefficient of interest is $\beta_j$. We also investigate returns at other ages by plugging in different values of $t$. However, we do not look later than age 30 because this would involve predicting returns out of sample.

control group is individuals who did not attend university.[24] When estimating the returns to different subjects, institutions and degrees, however, we only include individuals who attended higher education and estimate returns relative to a base case.[25] For the subject and institution estimates, $D_{ij}$ in equation 3.3 is a set of dummies for each of the different subjects and institutions, all included in the same regression additively. For the degree estimates, $D_{ij}$ is a set of dummies for all of the interactions between subjects and institutions. The specification outlined in equation 3.3 means that we allow all of these treatments to have their own independent time effects.

### 3.4.2 Control variables

We believe that the full set of control variables included in the vector $X_i$ in equation 3.3 is plausibly giving us causal estimates for attending different universities because of the uncommonly rich information we have on each individual in our administrative data.

Specifically, the vector $X_i$ includes three sets of characteristics, all of which are obtained from the NPD data. First, for all children who attended a state secondary school (about 93% of each cohort) we have a comprehensive set of background controls which includes individual and area based measures of socio-economic background, ethnicity, an indicator for English as an additional language and special educational needs eligibility (see Section 3.3 for more detail). Second, $X_i$ includes individual secondary school identifiers, which we include as fixed effects. Third, and most importantly, it incorporates extremely detailed information on the prior attainment of each student, specifically the student's grades in specific subjects in national examinations taken at age 11, 16 (GCSE)

[24]As outlined above we always restrict to individuals who passed their age 16 high school exams.

[25]These are history, Sheffield Hallam, and history at Sheffield Hallam for subject, institution and degree returns respectively. These were chosen as they have relatively large numbers of students and have earnings close to the middle of the distribution. In practice our estimates are not sensitive to the base case.

and 18 (A-level) as described earlier, as well as number of subjects taken and subject mix.[26] Finally, we interact A-level attainment and subject choice variables with quadratic time trends to allow, for example, maths A-levels to have an impact on earnings which grows over time.[27] We do not condition on degree outcomes or on people progressing onto postgraduate study. The estimates therefore include the option value of a good degree and of progressing to postgraduate study (which is not necessarily positive by age 30).

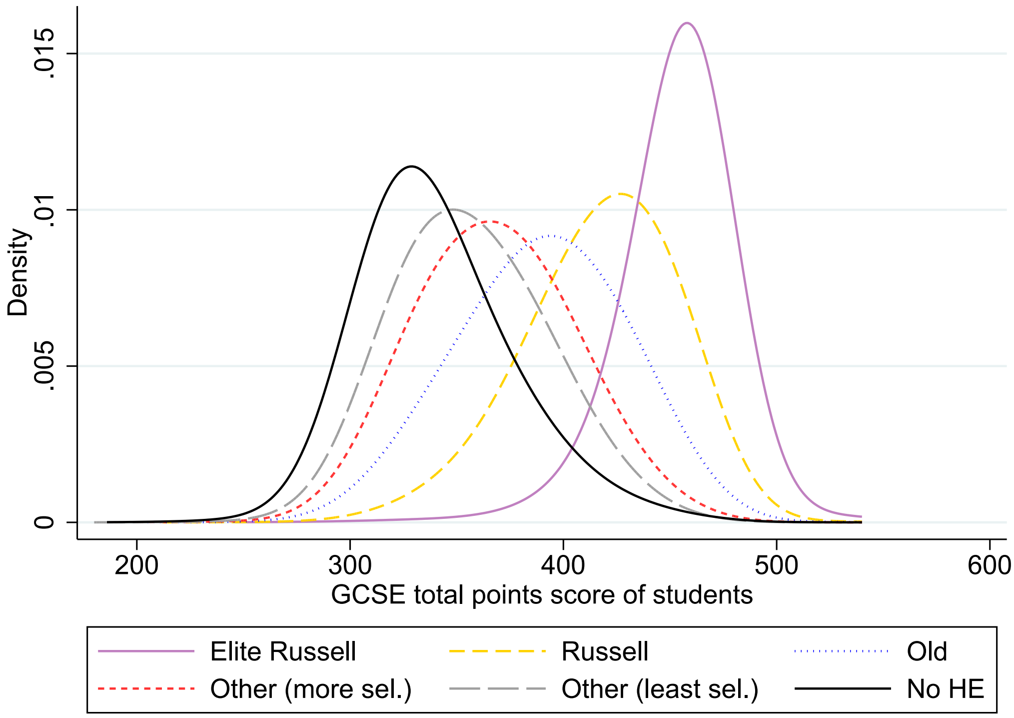
A key issue that we face here is that there is considerable sorting on ability across universities, as we can see in Figure B.4 in the Appendix. This raises the question of how we identify returns for the elite institutions for whom there are not many people with similar characteristics who attended the least selective institutions. Figure 3.1 gives an intuitive idea of how we identify the effects by showing the density of GCSE (age 16) point scores for the different university groups. While there is not a great deal of overlap between the Elite Russell Group and the least selective institutions, there *is* considerable overlap between the Elite Russell Group and the rest of the Russell Group, the rest of the Russell Group and the Old Universities, the Old University and Other (more selective) institutions, and the Other (more selective) institutions and the Other (least selective) institutions. Of course this is at the university group level - in practice there is much more overlap between institutions within these broader groups. This means that we essentially build sequential common support, and depend on functional form assumptions for identification of returns to elite institutions

[26]At age 11 we control separately for scores in all subjects taken (maths, English and science). At age 16 we control for a cubic in total score; scores in maths and English, and scores in science, history, geography, modern languages and vocational courses for those who took these courses; total number of exam entries, as well as total number of full GCSE entries; number of GCSEs with each grade A* to G. At age 18 we control for having any KS5 qualifications; a cubic in total KS5 score; score in vocational courses; number of subjects taken for AS and A-level; number of AS-levels, academic and vocational A-levels with grade A; dummies for taking A levels in maths, sciences, social sciences, arts, humanities, languages and vocational subjects.

[27]We are unable to control for non-cognitive skills. However Buchmueller and Walker (2020) estimate returns to higher education model that is similar to ours using the Longitudinal Survey of Young People in England (LSYPE) and show that the inclusion of rich non-cognitive variables has no effect on the returns estimates conditional on including prior attainment measures.

compared to attending the least selective universities.[28]

Figure 3.1: Distribution of GCSE points score by university group




Notes: Uses the 2004 GCSE cohort only. No HE consists of individuals who did not take an undergraduate degree, but passed their age 16 exams (obtaining at least 5 A*-C GCSEs).

Several papers in this literature - most recently Mountjoy and Hickman (2020) - have argued that it is necessary to control for the set of higher education institutions individuals apply to, following Dale and Krueger (2002). It is argued that such controls capture both the ability and the preferences of the students to help extract causal estimates. While we are unfortunately not able to observe these choice sets,[29] we have much richer controls than previous papers in this literature. Indeed, it is very common to have just a single measure of attainment prior to college, such as the SAT examination in the United States,[30]

[28] A very similar argument to this is made in Hoxby (2018). As a robustness check, we narrow the set of institutions that we include in individual regressions and we get extremely similar estimates of relative returns to when we include the full set.

[29] The data exist as all applications to university are through the centralised University and Colleges Admissions Service (UCAS). Although this dataset could in principle be merged into the LEO dataset, it has unfortunately not been possible to obtain it.

[30] Mountjoy and Hickman (2020) also have scores from an additional 10th grade test taken in

whereas we have scores from high-stakes standardised exams taken in multiple subjects at age 11, 16 and 18. We also have information on subjects taken, which universities factor into their entry requirements. Particularly in the UK, where individuals typically only choose three subjects to study up to age 18, the subject choices will also capture subject-specific skills and preferences. The rich information on the local-level deprivation of the student combined with individual school fixed effects means we are able to effectively compare students from similar backgrounds who attended the same school and chose the same subjects to study in school, and obtained the same results in their exams. We argue that once all of these factors are controlled for, the drivers of differences in choices between different universities are driven by idiosyncratic preferences which are unrelated to subsequent earnings outcomes.

A concern with this argument is that the decision of whether or not to go is distinct from the decision of *where* to go, with the former decision more likely to be driven by factors related to latent earnings potential. To alleviate this concern, our main results are based on 'relative returns', where we show returns relative to a base level degree.

## 3.5 Overall returns and variation by subject and institution

In this section we present our findings on the overall returns to higher education in the UK, before turning to how how these returns vary across different universities and subjects (we look at the interaction of subject and institution in the following section).

Texas. However, this is a low-stakes exam.

### 3.5.1 Overall returns

We start by presenting our OLS estimates of the overall returns to higher education in Table 3.3. This shows the estimated impact of university on gross earnings at age 30, in log points by gender. Column 1 displays the unconditional differences in earnings.[31] Unsurprisingly, graduates earn considerably more on average at age 30 than non-graduates. The coefficient estimate for men is 26 log points, or an earnings premium of around 30%, while the equivalent figure for women is 47 log points (60%).

Table 3.3: Overall returns to university at age 30

| | (1)<br>Unconditional | (2)<br>+ age 16 attain | (3)<br>+ full attain | (4)<br>+ background | (5)<br>+ school FE |
|---|---|---|---|---|---|
| Estimate | 0.261 | 0.089 | 0.057 | 0.068 | 0.065 |
| | (0.004) | (0.004) | (0.005) | (0.005) | (0.005) |
| $R^2$ | 0.0418 | 0.0799 | 0.0965 | 0.1068 | 0.1189 |
| Adj. $R^2$ | 0.0418 | 0.0799 | 0.0964 | 0.1068 | 0.1174 |
| Individuals | 718,339 | 718,339 | 718,339 | 718,339 | 718,339 |
| Observations | 2,206,994 | 2,206,994 | 2,206,994 | 2,206,994 | 2,206,994 |
| *Women* | | | | | |
| | (1)<br>Unconditional | (2)<br>+ age 16 attain | (3)<br>+ full attain | (4)<br>+ background | (5)<br>+ school FE |
| Estimate | 0.473 | 0.286 | 0.222 | 0.221 | 0.216 |
| | (0.004) | (0.004) | (0.005) | (0.005) | (0.005) |
| $R^2$ | 0.0802 | 0.1254 | 0.1386 | 0.1521 | 0.1646 |
| Adj. $R^2$ | 0.0802 | 0.1254 | 0.1386 | 0.1520 | 0.1633 |
| Individuals | 807,131 | 807,131 | 807,131 | 807,131 | 807,131 |
| Observations | 2,501,733 | 2,501,733 | 2,501,733 | 2,501,733 | 2,501,733 |

Note: Table reports derived estimates of the overall impact of HE on annual earnings at age 30 based on the 2002-2007 GCSE cohorts, conditional on at least five A*-C GCSEs. Estimates are in log points (/100), standard errors, clustered at the individual level, are in parentheses. All estimates are statistically significant at the 1% level.

We start by controlling for prior attainment, which has a dramatic impact on our returns estimates. In Column 2 we add controls for maths, English and overall GCSE test scores, which roughly halves the returns for women while

[31] We include in this specification an age-adjustment for those who started university at age 19 or 20, as well as cohort dummies.

cutting them by around two-thirds for men. In Column 3 we include the full set of controls we have for prior attainment.[32] One of the key advantages of our data is that we are able to see very rich information on the test scores of students, taken at three different ages (11, 16 and 18) and in different subjects. This is a significant advantage over much of the literature which often relies on a single measure of prior attainment. We see that the inclusion of these additional attainment controls meaningfully affects our estimates for both genders.

However, the same is not true for the remaining columns. In Column 4 we add background characteristics, including socio-economic status, ethnicity and region, while in Column 5 we add school fixed effects.[33] In each case we see that the estimates do not change very much, despite the overall fit of the model improving. The stability of the estimates to the inclusion of rich, relevant conditioning variables adds weight to our assumption that selection into higher education on unobservable factors is not an important driver of the results.

This final estimate for men suggests a return to university of 6.5 log points, or around 7%, at age 30. This estimate on the face of it seems quite low relative to the previous evidence for the UK, most notably Blundell et al. (2000), who estimate a return of around 12 log points for men using data from the British National Child Development Survey, a panel survey of individuals born in a specific week in 1958. However, these estimates actually align fairly closely. Blundell et al. (2000) estimate returns at slightly later age (33), and focus on *graduates*, while we estimate returns for higher education entrants without conditioning on graduation. In Table 3.4 we show that the estimates are clearly growing quickly with age, and that the returns estimates are considerably higher when we condition on graduates rather than entrants, at 10 log points compared to 7.[34] Given this, estimates for men appear to align quite closely with those from

[32]See Section 3.4.2 for the complete list of controls.

[33]The individuals in our sample attend more than 4,000 different English secondary schools.

[34]This result suggests that outcomes for university dropouts are particularly bad, as only 10-15% of students typically do not complete their degrees in the UK. This finding aligns with Ost et al. (2018), which find large causal effects to dropping out based on a regression discontinuity

Blundell et al. (2000). The estimates from Blundell et al. (2000) are based on a cohort born 30 years earlier, who went through higher education at a time when higher education attendance was much lower. It is therefore notable that returns have kept up for more recent cohorts, despite a considerable expansion in higher education attendance. This pattern aligns with some recent work using different data which suggests that the graduate earnings premium in the UK has held up through the rapid expansion of the 1990s (Blundell et al., 2016).

The final estimates for women suggest a much larger return of 22 log points (24%). These estimates are quite a lot smaller than those in Blundell et al. (2000), who estimate a returns of 34 log points. An important difference between their approach and ours is that we are estimating returns in terms of annual earnings and are therefore unable to adjust for differences in labour supply. We think this issue is likely to be particularly important for women, as women who do not go to higher education typically have children earlier and are therefore much more likely to be working reduced hours. This is especially true when considering their earnings across a whole year.[35] Based on this, we interpret our overall returns estimates for women with extreme caution, and for the rest of the paper focus only on *relative* returns to different higher education options amongst the set of people who go. Differential labour supply is likely to be a much less important issue when making comparisons across different degrees than when comparing people who go to higher education with people who do not.

design from a set of 13 public universities in Ohio.

[35] Figure B.1 in the Appendix shows that women who did not go to higher education are much more likely to have very low earnings (say, below £8,000 a year) than women who did go, while the same is not true for men.

Table 3.4: Overall returns to university by age and with dropouts

| | (1) Age 25 | (2) Age 26 | (3) Age 27 | (4) Age 28 | (5) Age 29 | (6) Age 30 | (7) Excl. Dropouts |
|---|---|---|---|---|---|---|---|
| *Men* | | | | | | | |
| Estimate | 0.008 | 0.020 | 0.031 | 0.042 | 0.054 | 0.065 | 0.103 |
| | (0.014) | (0.011) | (0.007) | (0.005) | (0.003) | (0.005) | (0.005) |
| *Women* | | | | | | | |
| Estimate | 0.126 | 0.144 | 0.162 | 0.180 | 0.198 | 0.216 | 0.248 |
| | (0.014) | (0.010) | (0.007) | (0.004) | (0.003) | (0.005) | (0.005) |

Note: Table reports estimates from the same model as column 5 of Table 3.3 by age (sample size and fit is the same). The final column shows the age 30 estimates but with university dropouts excluded from the analysis sample. Estimates are in log points (/100), standard errors, clustered at the individual level, are in parentheses. All estimates, other than those in columns 1 and 2 for men, are statistically significant at the 1% level.

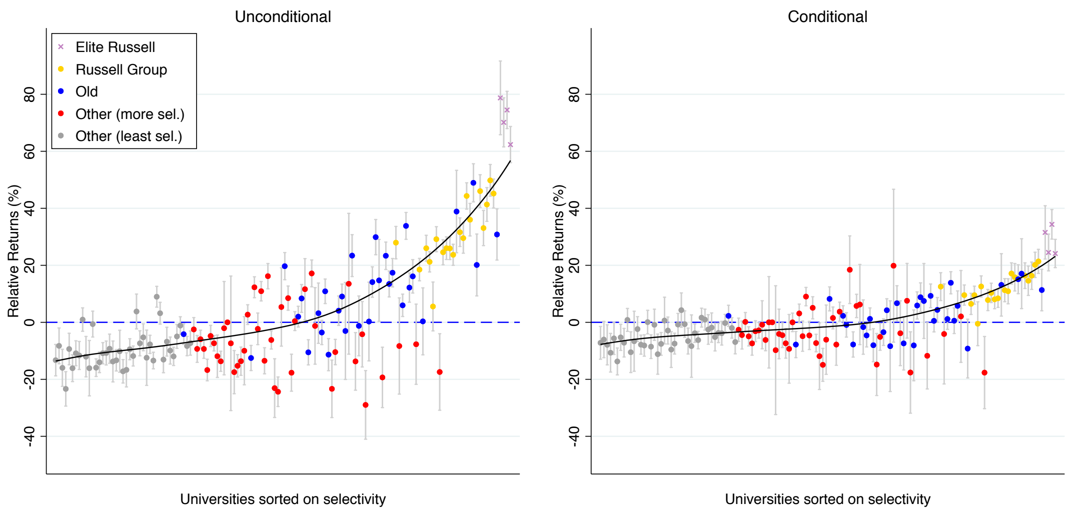
### 3.5.2 Relative returns by university

We now turn to the estimates of the relative returns to different higher education institutions. Figure 3.2 displays the institution fixed effects estimates, which are all shown in *relative* terms, with Sheffield Hallam University - a large, mid-ranking institution in the 'Other (more selective)' group - the omitted category. For these estimates, and for all remaining estimates in the paper, we include men and women in the regressions, controlling for gender (separate results by gender are provided in the Appendix). The estimates have been converted into percentage terms from log points, and institutions are sorted on their selectivity rank, as measured by the average GCSE point score of their students.[36] Results are shown both 'unconditionally' in the left hand panel and 'conditionally' on the right. The unconditional estimates include only the university fixed effects in the model, while the conditional estimates include all of the background controls included in column 5 of Table 3.3, but with controls for subject studied also included. All of the point estimates for this and for subsequent results are

[36]While school grades are not the only factors based on which universities select their applications, for most universities it is the most important one. For specialist institutions which tend to rely more on interviews or portfolios, such as music colleges, our measure will reflect their selectivity less well.

provided in an Online Appendix.

Figure 3.2: Estimated returns at age 30 by institution




Note: Figure reports estimates of the impact of studying at different institutions on annual earnings at age 30 relative to Sheffield Hallam University. Conditional estimates control for year, background, prior attainment and subject. Results have been converted to percentage differences using a log point conversion. Universities are ranked on the average GCSE results of their intake. The black line shows the relationship between returns and selectivity from a locally weighted polynomial regression. 95% confidence intervals are shown by the whiskers and standard errors are clustered at the individual level.

The inclusion of controls substantially flattens the relationship between earnings outcomes and university selectivity. In fact, we see that in the conditional model, the relationship between returns and selectivity is quite weak for universities in the bottom two-thirds of the selectivity distribution. However, the relationship is much steeper amongst the top institutions - returns for the four so-called 'Elite Russell Group' institutions in the conditional model are between 25 and 35% higher than the baseline, while returns for the other Russell Group universities are mostly between 5 and 20% higher than the baseline. This suggests that accessing the very elite institutions can boost outcomes considerably over the next tier of institutions.

At the lower end of the scale, returns amongst the least selective institutions are, on average, around -5% relative to the baseline, which is very similar to the average of the more selective other institutions and a few percentage points below the returns for the 'Old' (more established) institutions. Interestingly, only four of the bottom ten institutions for returns are from the set of least selective institutions - while six of the bottom ten institutions are specialist arts colleges.

Our findings on the relationship between selectivity and returns aligns with the previous UK evidence. Chevalier and Conlon (2003); Hussain et al. (2009); Broecke (2012) and Walker and Zhu (2018) all report similar results. There are some inconsistencies in prior findings in that Broecke (2012) suggests that there is a linear relationship between returns and selectivity, while the other papers align more closely with our findings of a stronger relationship amongst the more elite universities.[37] Broecke (2012) suggest one possible explanation for this is the comparison of a broad range of institutions all within one model. We assess the robustness of our result by re-running our specification using only subsets of universities and find that the estimates are extremely highly correlated across the alternative samples (we show this, plus the fact that the results are robust to some other alternative specifications in Appendix Table B.5).

The only previous paper to estimate returns for individual universities in the UK is Walker and Zhu (2018), which uses the Labour Force Survey (LFS), which relies on self-reported earnings and only allows for the most basic control variables. It is notable that our raw differences in earnings are much greater than their estimates, and unsurprisingly, our control variables make a much larger difference to the university fixed effects. Nevertheless, we end up with a similar range of final estimates.

Finally, we note that in our final specification we condition on the subject studied, unlike most of the previous evidence on institution returns from the US (for example, Mountjoy and Hickman, 2020). This does not dramatically change the final set of results, with the relationship between selectivity and re-

[37]When we plot returns on the average GCSE scores of the intake rather than selectivity rank, this non-linearity is less clear. This is because the most selective universities are effectively shifted over to the right as they are much more selective than the rest. However, we still observe a steeper relationship at the top end of the selectivity distribution than at the bottom.

turns, as well as the standard deviation of estimates changing very little. However, there are some institutions that experience very large changes to their estimates. Many of these are specialist arts institutions at the bottom end of the returns distribution which perform considerably better when subject controls are included, reflecting the low returns for creative arts degrees.

### 3.5.3 Relative returns by subject

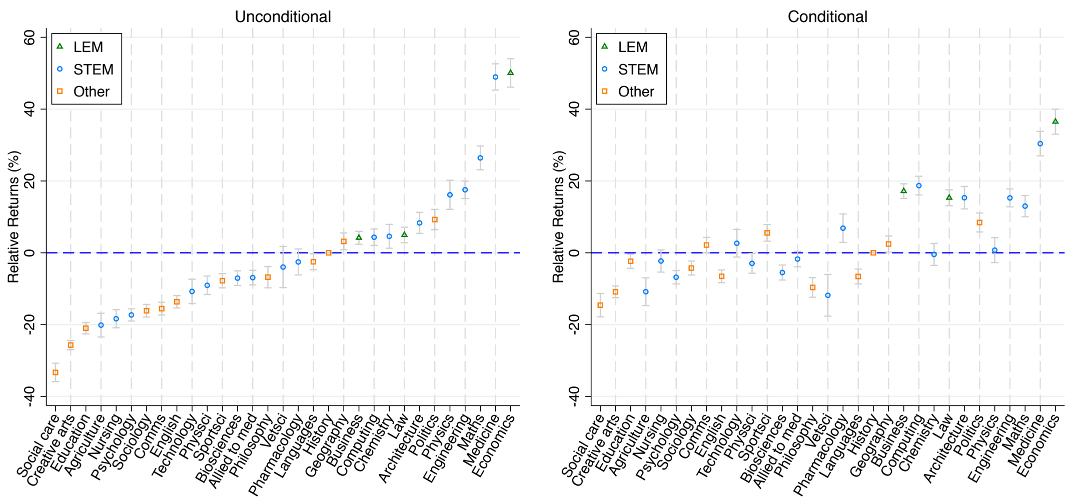
We explore the returns for specific subjects in Figure 3.3. The estimates are again converted into percentage terms and are now reported relative to the returns for history, which is the omitted category. We again show the unconditional estimates and the fully conditional estimates in the Figure to show the full effect of the control variables. We see that they again make a substantial difference to the distribution of subject fixed effects, although not to the same extent as for the institution fixed effects. This is not entirely unexpected as sorting on ability is less strong across subjects than across institutions. Nevertheless, we observe some big changes. For example, at the top end, relative returns for medicine and economics drop from close to 50% to 30% and 36% respectively once background controls are included (the patterns when we look at men and women separately are extremely similar - see Figure B.7 in the Appendix).

We also see fairly large upward shifts in relative returns estimates for some of the lowest earning subjects, most notably social care, creative arts, communications and education. We also see business and computing returns increase considerably, highlighting the fact that these degrees often admit students with relatively low prior attainment.

For the conditional estimates in the right hand panel, we still see significant variation in relative returns across subjects. Outside of economics and medicine, we see very good returns of around 15% for computing, business, architecture and law. At the bottom end, social care, veterinary sciences, creative arts and agriculture all have estimated returns of -10% lower than history or worse. Philosophy, psychology, English, languages and biological sciences all also perform poorly. The correlation of the returns for men and women is very high (0.91).

We also observe an interesting pattern in the returns across our different broad subject groups. In general, it is the case that the three LEM subjects do very well, the 'Other' subjects (which mostly consist of arts, languages and humanities) tend to do quite poorly, while the returns for STEM subjects are very mixed. Medicine, computing, engineering and maths all do well, while veterinary sciences, agriculture, psychology and biological sciences do not. This is particularly worth noting as some of these subjects - especially psychology and biological sciences - are very popular amongst women. This suggests policies encouraging women to study STEM subjects might not actually always result in positive earnings impacts.

Figure 3.3: Estimated returns at age 30 by subject




Note: Figure reports derived estimates of the impact of studying different subjects on annual earnings at age 30 relative to studying History. Conditional estimates control for age, background, prior attainment and institution. Results have been converted to percentage differences using a log point conversion. Subjects are ranked based on raw earnings differences. 95% confidence intervals are shown by the whiskers and standard errors are clustered at the individual level.

Although (as discussed above) we are very cautious about treating these subject estimates as causal, we still consider it useful to report them. The identification strategy is similar to the most comparable previous evidence from the UK on subject returns (Chevalier, 2011), which estimates the annual earnings effects of different subjects based on data from 2006, three and a half years after graduation.[38] Broadly speaking, he finds similar estimates, with medicine doing very well and creative arts doing poorly. One major difference is economics, which we have as the highest-returning subject while he has it much further down the distribution. This could be because he is only observing earnings quite soon after graduation, although it mostly likely to be because he is working with much smaller samples - for example, he only observes 110 economics graduates.

## 3.6 Returns to different degrees

We now turn to the main contribution of the paper and focus on estimates of returns at the degree level. We explore the overall distribution of earnings outcomes and returns before looking at the variation within selectivity bands and the relationship between returns and selectivity. Finally, we consider how well other indicators of university quality are correlated with returns.

### 3.6.1 Returns by degree

Figure 3.4 shows the overall distribution of the more than 1,900 degree fixed effect (these are estimated relative a base case of history at Sheffield Hallam).[39]

[38]The most directly comparable results to ours are reported in column 7 of Table 2 in his paper, noting that his base case is physical sciences, whereas ours is history.

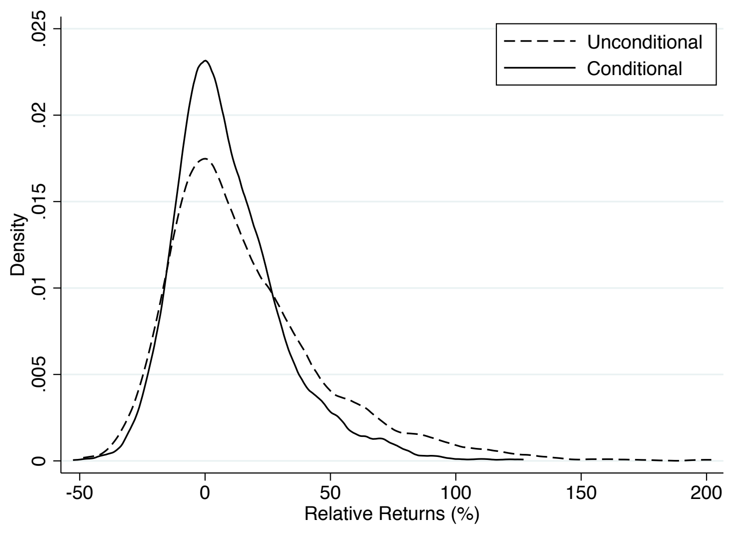
[39]All individual returns estimates can be found in the Online Appendix. For sample size reasons, not all degrees offered are included in this analysis. Specifically, for inclusion we require the degree to have at least 10 individuals with earnings observations at age 30, and 50 unique individuals with earnings observations at any of the ages 25 to 30. This is to ensure data disclosure requirements are met and that we are not predicting earnings returns 'out-of-sample', which would significantly increase the uncertainty and importance of the underlying assumptions in our estimates.

We show the distribution of degree-level fixed effects unconditionally[40] and with the full set of controls. The figure shows substantial variation in the raw estimates, which range from 50% below the base case to 200% above it. The inclusion of the controls considerably reduces this variation. This is summarised in Table 3.5, which shows the standard deviation and range of the returns estimates for the conditional and unconditional degree level estimates. The standard deviation of our degree returns estimates drops from 32 percentage points to a still very large 22 percentage points with the inclusion of controls. Similarly, the 90:10 range drops from 75 percentage points to 52 percentage points.

In columns 3 and 4 of Table 3.5 we compare the degree level returns estimates with equivalent estimates from regression models where *only* subject (column 3) or institution (column 4) fixed effects are included, to provide a comparison with estimates when only subjects or institutions are observed. The degree level returns are much more variable, with a standard deviation and 90:10 range around twice as large as the institution and subject estimates. This shows that the variation in institution or subject level returns dramatically understates the variation in returns to higher education degrees. The table also highlights that more of the variation in earnings is explained in the degree-level regressions, with the (adjusted) $R^2$ increasing from around 0.15 for the subject and institution fixed effects regressions to 0.18 for the degree fixed effects regression with controls.

[40]We do include a dummy for not starting university straight after school, as well as cohort dummies, in this specification.

Figure 3.4: Estimated returns at age 30 by degree




Note: Figure reports derived estimates of the impact of studying different degrees (subject-institution combinations) on annual earnings at age 30 based on the 2002-2007 GCSE cohorts controlling for age, background and prior attainment. Results have been converted to percentage differences using a log-point conversion.

Table 3.5: Summary of degree, subject and institution estimates

| | (1)<br>Degree<br>Unconditional | (2)<br>Degree<br>Conditional | (3)<br>Uni | (4)<br>Subject |
|---|---|---|---|---|
| $\sigma$ | 32.03 | 21.89 | 10.50 | 12.33 |
| 90:10 Range | 75.35 | 51.95 | 26.26 | 27.42 |
| Adj. $R^2$ | 0.15 | 0.18 | 0.15 | 0.16 |
| Controls | No | Yes | Yes | Yes |

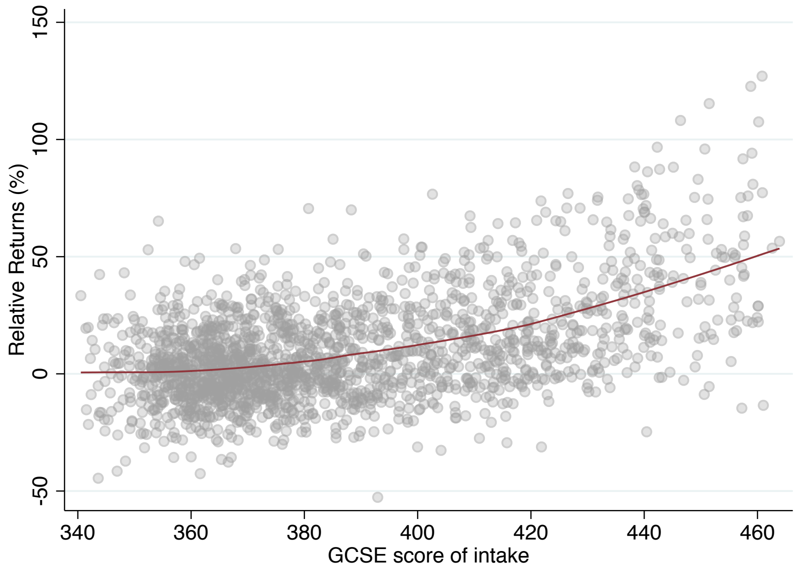
Note: $\sigma$ is the standard deviation of degree returns, the range is the 90th percentile return minus the 10th percentile (all in percentage terms) and the adjusted $R^2$ is from the underlying earnings regression with degree/HEI/subject fixed effects. The conditional university results *exclude* subject controls and similarly the conditional subject results *exclude* university controls.

To get a sense of the types of degrees that give particularly high or low returns, Table B.4 in the Appendix lists the best and worst performing degrees. We find that the top degrees are heavily dominated by law and economics. The top end is also heavily dominated by the high-status Russell Group universities. The worst performing degrees include a wider range of subjects, with social care, philosophy, politics and subjects allied to medicine all appearing in the bottom ten. Most of the lowest performers are from the least selective

'Other' group of universities, although humanities degrees from higher-status institutions do appear. This broad pattern holds throughout the distribution. LEM degrees, and degrees at the most elite institutions perform best, while arts and humanities degrees, and those at low ranked universities perform worst on average.

We show the full set of returns estimates, plotted against their selectivity (as measured by the average age 16 test scores of students), in Figure 3.5. The first point to note is that average returns increase considerably as we move from the least selective to the most selective degrees, with a difference of more than 50 percentage points in average returns. Again, the relationship between returns and selectivity gets stronger as we move up the selectivity distribution. This is documented more explicitly in Table 3.6 which reports the slope coefficient from a regression of returns on selectivity within selectivity bands. We see that this increases from -0.02 (meaning a 100 point increase in GCSE points is associated with a 2% decrease in returns) in the least selective band of degrees, to around 1 (meaning a 100 point increase in GCSE points is associated with a 100% increase in returns).

Figure 3.5: Course returns against selectivity




Notes: Degree level estimates plotted against average GCSE scores of intake. Red line plots the relationship with a locally weighted polynomial.

The second point is that despite this relationship, selectivity by no means explains all the variation in returns. Table 3.6, highlights the considerable variation in returns across different bands of similarly selective degrees. The standard deviation of returns amongst the least selective band is around 15 percentage points, and this doubles to around 30 percentage points for the most selective band of degrees. For reference, this compares to an estimated overall return to higher education of around 7% for men at the same age.

Table 3.6: Summary of degree, subject and institution estimates

| | GCSE score of intake | | | | | |
|---|---|---|---|---|---|---|
| | 340-359 | 360-379 | 380-399 | 400-419 | 420-439 | 440-459 |
| *Main results* | | | | | | |
| Standard deviation | 16.02 | 15.31 | 17.79 | 20.11 | 23.84 | 29.29 |
| Share within subject | 0.47 | 0.52 | 0.46 | 0.64 | 0.71 | 0.77 |
| Selectivity slope | -0.02 | 0.19 | 0.28 | 0.24 | 1.34 | 0.72 |
| *Excl. school FE* | | | | | | |
| Standard deviation | 18.09 | 17.21 | 19.77 | 21.85 | 25.79 | 31.75 |
| Share within subject | 0.48 | 0.51 | 0.48 | 0.61 | 0.70 | 0.77 |
| Selectivity slope | -0.10 | 0.27 | 0.32 | 0.21 | 1.48 | 0.97 |
| *Excl. school FE and background* | | | | | | |
| Standard deviation | 16.95 | 16.23 | 18.84 | 21.09 | 24.68 | 30.55 |
| Share within subject | 0.51 | 0.57 | 0.50 | 0.64 | 0.71 | 0.78 |
| Selectivity slope | -0.06 | 0.18 | 0.31 | 0.24 | 1.38 | 0.83 |
| *Excl. dropouts* | | | | | | |
| Standard deviation | 16.42 | 16.21 | 18.21 | 20.50 | 23.92 | 29.15 |
| Share within subject | 0.50 | 0.57 | 0.48 | 0.62 | 0.72 | 0.77 |
| Selectivity slope | -0.14 | 0.20 | 0.28 | 0.30 | 1.34 | 0.69 |
| *Cross-sectional* | | | | | | |
| Standard deviation | 14.84 | 14.61 | 16.87 | 19.59 | 23.51 | 27.62 |
| Share within subject | 0.40 | 0.42 | 0.34 | 0.60 | 0.67 | 0.76 |
| Selectivity slope | 0.11 | 0.22 | 0.24 | 0.23 | 1.25 | 0.69 |
| *Shrinkage* | | | | | | |
| Standard deviation | 11.40 | 11.57 | 13.27 | 15.80 | 18.07 | 21.94 |
| Share within subject | 0.44 | 0.53 | 0.50 | 0.63 | 0.72 | 0.74 |
| Selectivity slope | -0.06 | 0.15 | 0.20 | 0.22 | 1.01 | 0.58 |
| Number of courses | 303 | 687 | 362 | 280 | 182 | 106 |

Notes: Table shows the standard deviation of returns, and the slope of a regression of returns on selectivity (average GCSE score of student intake) within 20 point selectivity bands. The very few degrees with average GCSE scores of the student intake of 460 or more are not shown. These statistics are shown both for the main degree returns, as well as for a further three specifications. 'Excl. dropouts' estimates returns when individuals who do not finish their degree of study are excluded. 'Cross-sectional' estimates returns on individuals at age 30 only, rather than using the panel model used in the main specification. 'Shrinkage' applies shrinkage to the main returns estimates, where estimates are shrunk towards the average degree returns.

As confirmed by Figure 3.5, the very highest return degrees are dominated by the most selective degrees, yet we also find a number of extremely selective degrees at elite institutions with very low relative returns. Table 3.6 also shows that these two conclusions are robust to removing subsets of the control variables from our regression models - when school fixed effects and then additional background controls are excluded from the models, the qualitative patterns of the estimates are almost identical. We interpret this as promising evidence that relevant variables that we are excluding from the model would not dramatically change our headline findings. The table further shows that these two findings are robust to the exclusion of dropouts, to using a cross sectional rather than a panel estimation for the regression, and to the application of a shrinkage estimator.

### 3.6.2 Within subject returns

As discussed above, we believe that one should be cautious about interpreting the variation in degree level returns across different subject areas, as previous evidence has highlighted the importance of selection on comparative advantage into different fields. However, Table 3.6 also presents the share of the variation in returns within each selectivity band that occurs within subject. This shows that at least half of the variation is *within* subject, across institutions for all selectivity bands. This increases to around three-quarters of the variation for the most selective degrees.

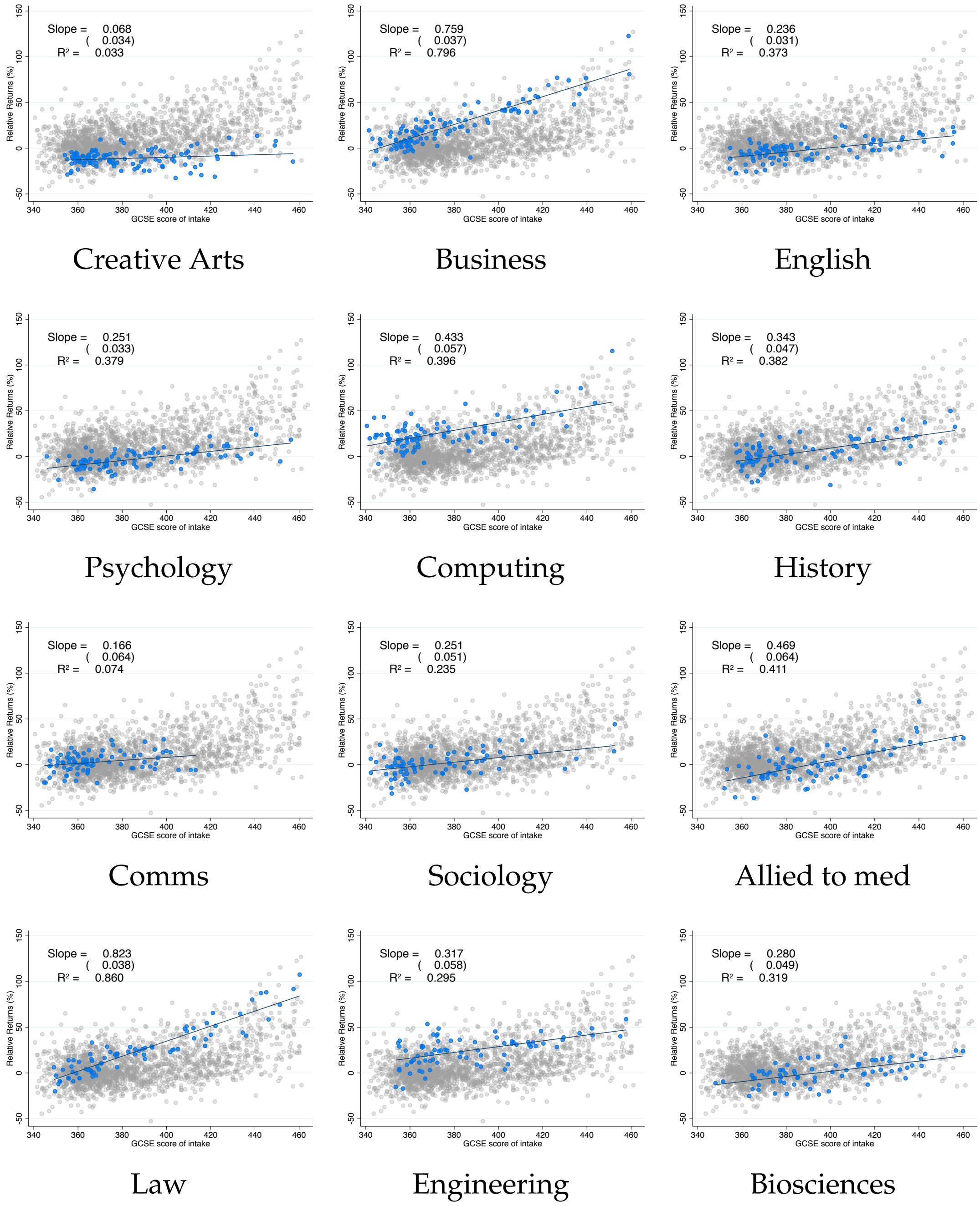
Figure 3.6 then highlights individual institution estimates for each of the 12 largest subjects.[41] The figure supports the point that there is substantial variation in returns even within given subject areas, and across institutions that are similarly selective. Holding subject choice fixed, attending one university over another, similarly selective university can often lead to 40 percentage point difference in returns. This holds true right through the selectivity distribution.

The figure also documents the relationship between returns and selectivity,

[41] Correlations with selectivity for all subjects can be found in Table B.6 in the Appendix. This Table also shows that the within-subject correlations with selectivity are robust to the precise set of control variables we include in the regression models, again adding weight to the argument that unobserved factors are unlikely to change our qualitative findings here (specifically, we see that the within-subject correlations between returns and selectivity are almost identical when we exclude school fixed effects and other student background characteristics).

as well as the share of the variation in returns that can be explained by selectivity within subject (for a summary of the relationship between selectivity and returns for all subjects, see Figures B.8 and B.9 in the Appendix). For business and law, the relationship between returns and selectivity is strong, and selectivity can explain more than 75% of the variation. For the other LEM subject, economics, the relationship is also strong and the share of the variation explained is just over 60%. For most STEM and arts and humanities subjects, less than 50% of the variation can be explained by selectivity, however, and moving into more selective universities to study these subjects has a much lower pay off. Finally, for subjects which lead to professions with regulated earnings, such as education and medicine, less than 10% of the variation in returns is explained by selectivity.

Figure 3.6: Relationship of returns and selectivity at the subject level




Notes: Degree level estimates plotted against average GCSE scores of intake.

### 3.6.3 Correlates with degree returns

So far we have seen evidence suggesting that degree choices can make a substantial difference to earnings outcomes at age 30 and that selectivity can only explain some of this variation. For many subjects, selectivity only explains a very small share of the variation in returns. In this section we therefore consider whether other characteristics of degrees are predictive of returns outcomes. Continuing the theme of the previous sub-section, we do this *within* subject area.

In addition to showing the correlation between selectivity and returns for each subject, Table 3.7 shows the relationship between returns and the following set of degree characteristics:

- **League table ranking:** this is the subject-specific league table ranking of universities. We take these from the Complete University Guide (CUG) from 2010, which was the most relevant year we could collect data for. These rankings combine several characteristics of the degree, including student-staff ratios and research intensity and should therefore capture aspects of degree quality reflected in returns which are unrelated to selectivity. Gibbons et al. (2015) highlights the importance of such rankings in driving institution choices of prospective students.

- **Student satisfaction:** this is taken from the National Survey of Students, focussing on overall satisfaction, again based on data from 2010. This measure has recently been included as an input into the governments' Teaching Excellence Framework, a measure of teaching quality.

- **Age 22 returns:** this is the very early-career earnings of students, usually in the first year after graduating from university. We report this as early-career outcomes are often used as a measure of degree quality. For example, labour market outcomes six months after graduation (taken from a graduate survey) have frequently been included as inputs into league table rankings.

- **Completion rate:** this is the share of students starting a degree who complete it. People who start a degree and switch to another full-time degree before the age of 21 are neither treated as completers nor dropouts from the degree they started first (as we just take the degree they switched to as their main degree in those cases).

- **First class degree rate:** this is the share of students who achieved 'First class honours', the highest degree classification. This is not regulated and so varies across different subjects and universities.

Table 3.7: Correlates with age 30 degree returns

| | (1) Selectivity | (2) League table | (3) Student satisfaction | (4) Age 22 returns | (5) Completion rate | (6) First class degree rate |
|---|---|---|---|---|---|---|
| *LEM* | | | | | | |
| Business | 0.888 | 0.806 | 0.286 | 0.602 | 0.772 | 0.456 |
| Economics | 0.792 | 0.765 | -0.087 | 0.760 | 0.658 | 0.550 |
| Law | 0.928 | 0.836 | 0.073 | -0.342 | 0.789 | 0.601 |
| *STEM* | | | | | | |
| Allied to med | 0.625 | 0.565 | 0.323 | 0.069 | 0.434 | 0.555 |
| Architecture | -0.056 | 0.228 | 0.357 | 0.273 | 0.214 | 0.180 |
| Biosciences | 0.575 | 0.571 | 0.257 | 0.365 | 0.500 | 0.449 |
| Engineering | 0.521 | 0.515 | 0.365 | 0.022 | 0.499 | 0.269 |
| Maths | 0.696 | 0.518 | -0.066 | 0.581 | 0.590 | 0.430 |
| Medicine | 0.088 | 0.254 | 0.182 | . | 0.455 | -0.123 |
| Physsci | 0.541 | 0.306 | 0.038 | 0.235 | 0.369 | 0.323 |
| *Other* | | | | | | |
| Comms | 0.264 | 0.425 | 0.272 | 0.157 | 0.264 | 0.147 |
| Creative arts | 0.184 | 0.197 | 0.090 | 0.266 | 0.145 | 0.081 |
| Education | -0.081 | 0.037 | 0.057 | 0.101 | 0.104 | 0.020 |
| History | 0.610 | 0.380 | 0.018 | 0.194 | 0.557 | 0.517 |
| Languages | 0.541 | 0.490 | 0.233 | -0.049 | 0.453 | 0.334 |
| Sociology | 0.489 | 0.407 | -0.096 | -0.197 | 0.501 | 0.115 |

Note: Descriptions of each of the variables are given in the text. Numbers report the raw correlations. Only subjects for which we could obtain league table rankings and student satisfaction scores are shown.

Column (1) of Table 3.7 repeats the result from above that for many subjects, though not all, returns at age 30 are strongly correlated with selectivity. We then see that league table rankings, completion rates and first class degree rates are positively correlated with returns. This correlation is quite strong for LEM subjects but weaker for STEM subjects and in particular 'Other' subjects. The correlations of returns and student satisfaction ratings with early career (age 22) returns are much noisier and weaker across the board. In fact we even see negative correlations between student satisfaction and returns for economic, maths and sociology, suggesting students studying towards these degrees do not value or appreciate things that are well correlated with their subsequent labour market success. For age 22 returns, we see that this is a very unreliable measure of subsequent success in many cases - for example, there is virtually no correlation at all between returns at 22 and returns at 30 for education, and even a negative correlation for law. This suggests that there are large cross-subject differences in the time it takes for career paths to become established.

Table 3.8: Correlates with age 30 degree returns, controlling for selectivity

| | (1) League table | (2) Student satisfaction | (3) Age 22 returns | (4) Completion rate | (5) First class degree rate |
|---|---|---|---|---|---|
| *LEM* | | | | | |
| Business | 0.010 | 0.145 | 0.158 | 0.044 | 0.109 |
| Economics | 0.195 | 0.001 | 0.235 | -0.096 | 0.055 |
| Law | 0.029 | -0.054 | -0.045 | -0.025 | 0.095 |
| *STEM* | | | | | |
| Allied to med | 0.001 | 0.019 | -0.013 | -0.115 | 0.126 |
| Architecture | 0.272 | 0.377 | 0.257 | 0.238 | 0.204 |
| Biosciences | 0.054 | -0.064 | 0.248 | 0.043 | 0.098 |
| Engineering | 0.076 | 0.113 | -0.016 | 0.076 | -0.057 |
| Maths | -0.234 | -0.409 | 0.242 | 0.038 | 0.073 |
| Medicine | 0.227 | 0.209 | . | 0.438 | -0.158 |
| Physsci | -0.320 | -0.373 | 0.159 | 0.053 | -0.061 |
| *Other* | | | | | |
| Comms | 0.209 | 0.229 | 0.187 | 0.065 | 0.007 |
| Creative arts | 0.043 | 0.023 | 0.293 | 0.001 | -0.040 |
| Education | 0.108 | 0.079 | 0.111 | 0.165 | 0.069 |
| History | -0.194 | 0.007 | 0.213 | 0.034 | -0.047 |
| Languages | 0.026 | 0.124 | 0.007 | -0.010 | -0.013 |
| Sociology | 0.001 | -0.259 | -0.133 | 0.111 | -0.258 |

Note: Descriptions of each of the variables are given in the text. Numbers report the partial correlations, after taking out selectivity. Only subjects for which we could obtain league table rankings and student satisfaction scores are shown.

In Table 3.8 we then look at how much of the correlations in Table 3.7 are driven by the correlation of these variables with selectivity. To do this, we regress returns on selectivity and correlate the residual with the variables of interest. We see that conditioning on selectivity removes almost all of the correlations between the university characteristics and returns. This suggests that there is no additional meaningful information in these measures over and above what you get from a simple measure of the selectivity of the degree. This is a disappointing result from the point of view of policy, as it suggests that the information available to students making their choices about where to study is not very well related to their likely outcomes. This could be particularly damaging as our evidence suggest that these choices matter a lot for earnings. It also has concerning implications for the incentives of universities who are competing for students and for regulators trying to incentivise universities to boost the labour market prospects of their students.[42]

## 3.7 Conclusion

This paper uses a novel administrative data linkage from the UK to investigate the returns to higher education and how they vary across different degrees. Our key finding is that there is substantial variation in returns at the degree level even within relatively tight selectivity bands. We find that a large share of the variation within selectivity bands is *within* subject, mitigating any concerns that the variation in returns across degrees might be overstated by selection into different fields based on comparative advantage (Kirkeboen et al., 2016). Our results therefore suggest that degree choice matters a lot for earnings outcomes at age 30. We provide suggestive evidence that this finding is robust to the empirical specification used, the exact sample of students included, and to unobserved

[42]The findings in this section are robust to the exact specification used to estimate returns. In Appendix Table B.7 we show that returns are extremely highly correlated across specifications, with correlations of more than 0.95 for most subjects when we compare our main returns estimates to estimates excluding dropouts, using age 30 only, or with shrinkage applied. In the final column, we show the results are essentially the same when we estimate degree returns for each subject completely separately.

selection. Since age 30 is still a relatively early point to assess returns to higher education, considerable variation at this age is likely to be indicative of even greater variation later on.

While degree choice appears to matter a lot, we find that once we control for a simple measure of the selectivity of a degree (specifically, the average GCSE scores of the students), many other measures of degree quality, including subject-specific league table rankings of universities, are not at all well correlated with returns. This has important implications, as students are making choices that can have enormous implications for their future outcomes with poor information on which to base those choices. This is likely to drive up the costs of higher education, to damage the productivity of the economy and to increase inequality, as poorer students are likely to be more reliant on publicly available information. It is also likely to create perverse incentives for universities, which may wish to target factors such as student satisfaction or first class degree shares when those things might not be beneficial in the long term.[43]

One potential solution to this could be to make information on the earnings outcomes of students more readily available when prospective students are making their higher education choices. In the UK this is increasingly plausible given the data linkage created for this work, and other countries may wish to develop similar data sources. A more extreme solution would be for the government to use the returns estimates to protect or boost funding where returns are high and restrict it where they are not. However, there are a few reasons why caution should be exercised before using degree level returns estimates to justify funding cuts. First, there is a long lag between changes to university practice and changes to earnings returns. The current estimates are based on people who started university between 10 and 15 years ago, and we have seen that looking at early earnings outcomes can be misleading. Second, a univer-

[43]A notable example of this is the dramatic increase in first class degree shares that have occurred at UK universities in recent years as competition for domestic students has increased following the removal of student number caps.

sity degree may have important positive impacts that might not be reflected in our earnings returns estimates. Third, it is also possible that the returns do not reflect university productivity and are instead a product of peers, labour market signalling or both. Understanding what drives the very large differences in returns is an important topic for future research.

Future research should also look in more detail at what drives the higher returns of certain universities within given subject areas. This could include investigations of the specific practices of the successful universities, such as the style of teaching and the content included. Other channels to explore would be the location of students after graduation, and the occupations they enter. Future work could also investigate the signalling component of these returns by surveying employers about how they value degrees from certain universities amongst job applicants. Such a study would help to highlight any labour market biases that might need to be addressed, and would also promote practices that are associated with good outcomes of students that could potentially boost teaching quality throughout higher education.

# Chapter 4

# Which university degrees are best for intergenerational mobility?

## 4.1 Introduction

Intergenerational mobility looks at the link between parent and child outcomes. A stronger relationship is worse for mobility, as it implies that the rich are more likely to stay rich, while the poor are more likely to stay poor. By this metric, the UK is one of the worst performers in the OECD (Corak, 2013). This has been a long-term concern for successive UK governments, and policymakers have often seen the higher education sector - which has been expanded dramatically over the last 30 years - as a crucial vehicle for addressing the problem.

However, recent work (Belfield et al., 2018b) has shown that not all degrees actually boost the earnings of students by age 30. While selective courses at high-status universities generally see high earnings returns, for a non-negligible share of arts and humanities degrees at less selective universities, earnings returns are low or even negative compared with not going to university at all. This has led to a debate over whether all parts of the sector really provide 'value for money', both for students and for the government. In this paper, we contribute a new angle to this debate - which has largely focused on the impact of degrees on average student outcomes in terms of earnings or employment - by documenting the extent to which universities, subjects and 'courses' (subject-institution combinations) promote social mobility. Specifically, we create 'mobility rates' which show the share of students in each university, subject or course who were both eligible for free school meals (FSM) *and* reach the top 20% of the income distribution at age 30.

This paper contributes to a substantial literature that documents inequalities in UK higher education. At each stage of the system, from application and acceptance into university to subsequent performance and progression, previous work has found large disparities between students from low-income backgrounds and those from better-off families. Crawford et al. (2016a) highlighted that students from low-income households are much less likely to attend university than their wealthier peers and, conditional on attending, they are less likely to attend high-status universities. The authors also showed that virtually all of the attendance gaps can be explained by differences in school attainment, suggesting that improving the school outcomes of children from low-income backgrounds is likely to be crucial for equalising access to university. However, Campbell et al. (2019) found that even when we compare individuals with the same A-level attainment, there are still differences in the quality of university they attend. That is, students from lower-income backgrounds are more likely to 'undermatch' and attend less selective universities than their wealthier peers with equivalent A levels. Belfield et al. (2018b) found large differences in the earnings impact of different degrees, with more selective degrees increasing earnings by more than less selective ones, implying that these differences in access to different types of degrees will matter for later-life outcomes. Other work has found that equalising access rates to university, and selective universities in particular, is an important first step, but unlikely to be sufficient to equalise outcomes between children from the most and least disadvantaged backgrounds. Even conditional on attending university, there are large gaps in performance, with lower-income students being more likely to drop out and less likely to graduate with a 'good' - that is, an upper-second or first-class - degree (Crawford, 2014b). Finally, even after studying the same subject at the same institution, students from more disadvantaged backgrounds *still* earn less than their more affluent peers (Crawford et al., 2016a; Britton et al., 2019).

While this literature provides insight into specific issues around access and subsequent progression, data limitations have meant that existing studies have generally focused on these individual components of mobility in isolation. They have also generally focused on the sector as a whole or on aggregated groups of universities, which could mask important variation.

Instead, the most comprehensive picture on social mobility and higher education comes from recent work from the United States. Chetty et al. (2017) construct statistics on social mobility for each college in the US, focusing on the participation and labour market outcomes of students from the bottom 20% of the parental earnings distribution. A key contribution of this work is the ranking of institutions by mobility, as measured by the proportion of students who come from the bottom income quintile and move into the top earnings quintile. Importantly, they find that the elite Ivy League colleges do little to promote mobility as, despite offering high returns, these colleges admit so few low-income students. In contrast, mid-ranking institutions tend to have the largest share of mobile students as they are both accessible to low-income students and offer reasonable earnings prospects.

In this paper, we exploit detailed, individual-level data from the Longitudinal Education Outcomes (LEO) dataset to estimate mobility rates for all English universities[1] and, unlike Chetty et al. (2017), individual subjects and courses. We document these estimates and then investigate how well correlated they are

[1]We do not include results for individual Welsh, Northern Irish or Scottish universities in this study. This is because low-income students are much less likely to cross borders within the UK and unfortunately our dataset does not allow us to identify students from low-income backgrounds who grew up in Scotland, Wales or Northern Ireland.

with estimates of average returns from Britton et al. (2021c). We also investigate how our mobility rates are affected by adjustments for where people live after leaving university and for the prior attainment and background characteristics of low-income students on different courses. Finally, we document more recent trends in access and look at the implications of those trends for mobility rates. Taken together, this work makes an important contribution to the debate about value in higher education, and may lead to more scrutiny of universities that are doing little for social mobility. Finally, it provides an evidence base upon which future work can draw to explore the key drivers of mobility.

The rest of the paper is set out as follows. In Section 4.2, we describe the LEO dataset. Section 4.3 then investigates access rates and Section 4.4 studies labour market 'success' rates. We then put the access and success rates together to study mobility rates in Section 4.5. Section 4.6 looks at the robustness of our main estimates, and Section 4.7 looks at the characteristics associated with higher mobility rates. Section 4.8 considers the likely implications of more recent trends in access rates for mobility rates, and finally Section 4.9 concludes.

## 4.2 Data

We use individual-level Longitudinal Education Outcomes (LEO) dataset, linked to address records. The LEO dataset links administrative school records from the National Pupil Database (NPD), university records from the Higher Education Statistics Agency (HESA), employment and tax records from Her Majesty's Revenue and Customs (HMRC) and benefits data from the Work and Pensions Longitudinal Study (WPLS), for individuals who attended school in England. This dataset has recently been linked to data on individuals' home addresses from the Department for Work and Pensions' (DWP) Customer Information Spine (CIS).

This dataset provides detailed, individual-level information on individuals' background, as well as their educational and labour market outcomes, making it well suited to our purpose. The detailed university records allow us to estimate mobility rates at the institution, subject and course (institution-subject) level. The dataset includes rich prior attainment measures from school records, including test scores and subjects taken at ages 11, 16 and 18, as well as demographics such as gender and ethnicity. This extensive set of background characteristics allows us to account for selection into different university degrees, enabling us to explore how mobility rates are affected by differences in student intake. The address records - which we have at the Lower Layer Super Output Area (LLSOA) level[2] and for all years from 2012 - allow us to check the robustness of our results to adjusting earnings for differences in living costs across the country.

These fully linked data are available for individuals who took their GCSEs in 2002 or after, or equivalently were born approximately 1 September 1985 onwards. We have earnings data up to the 2018/19 tax year. We face a challenge in balancing the desire to observe individuals' labour market outcomes only once they are well established in the labour market (rather than immediately after leaving higher education) against the need to maintain sufficiently large sample sizes for detailed course-level analysis. Our solution to this challenge is to focus our institution- and subject-level analysis on the three oldest cohorts for which we have complete data. These individuals took their GCSEs between 2002 and 2004, and were mostly born between 1 September 1985 and 31 August 1988. We focus on their earnings at approximately age 30, which is around nine years after graduation for those who go straight to university at 18 and do a three-year degree. Due to the smaller sample sizes involves in the course level analysis, we focus on earnings at age 28 for that part of the analysis, which allows us to boost sample sizes by additionally including the two cohorts who took their GCSEs in 2005 and 2006. In the remainder of this section, we define our key variables

[2] LLSOAs contain approximately 700 households each.

and show some summary statistics of our analysis sample.

### 4.2.1 Family background

We use eligibility for free school meals (FSM) at age 16 as our measure of family background. Children are recorded as being on FSM when their parents are in receipt of means-tested benefits[3] and have annual gross income below a given threshold, currently £16,190. FSM eligibility is therefore a good indicator of family-level deprivation. In our cohorts of analysis, 12.5% of students are recorded as being on free school meals. These students broadly represent the students from the lowest-income families.

For some descriptive analysis, we show results for the whole population. For this we rank state school students who are not eligible for FSM according to the local-area-level deprivation of the Lower Layer Super Output Area (LLSOA) where they live at age 16. We specifically use the Income Deprivation Affecting Children Index (IDACI), which measures the proportion of children aged 0 to 15 living in income-deprived households in each LLSOA. While it is true that some children from lower-income families will live in areas with low deprivation and vice versa, on average children in areas with low deprivation will be from much wealthier families than children who grew up in areas with higher deprivation. Many of our descriptive figures will show results separately for children on FSM, children not on FSM split into quintiles of IDACI, and children who attended private (fee-paying) schools. Private school students represent around 6.5% of students.

[3]These benefits are: income support, income-based jobseeker's allowance, income-related employment and support allowance, support under Part VI of the Immigration and Asylum Act 1999, the guaranteed element of pension credit and child tax credit (provided the parents are not also entitled to working tax credit).

### 4.2.2 University attendance

We obtain information on university attendance from the HESA records. We observe the individuals who attend university in the UK, and can see their degree subject, institution, whether they study full-time or part-time, and whether they obtain their final qualification. We focus on individuals who took an undergraduate degree (UG), and include both those who graduate from such a degree and those who start but do not finish the degree. We restrict our definition of UG students to full-time students who started university before the age of 21, in order to ensure individuals' have been in the labour market for a sufficient number of years for their earnings to stabilise by age 30 (the latest age we observe for all our three main analysis cohorts).

Degree subjects are recorded at the four-digit Joint Academic Coding System (JACS) code level, which means we observe around 1,500 different possible subjects of study. For our analysis, we aggregate these into 35 subjects, based on the CAH2 subject classification.[4] We might, for example, see individuals studying community nursing or palliative care nursing, and will combine these under 'nursing'. When individuals study multiple subjects in their degree, we assign individuals to all the subjects in proportion to the share of their degree which is devoted to each subject. An individual studying French and history with equal weight given to both subjects would therefore be counted as 0.5 of a person in both the French and the history results.

Students attend around 150 different universities across the UK. In the analysis, we will often classify these into broader groups. Oxford, Cambridge, Imperial College London and the London School of Economics are put together into the 'Most selective Russell Group' category, due to the much higher average prior attainment of students at these universities.[5] All other universities

[4] For more information on this subject classification, and how it maps to JACS codes, see https://www.hesa.ac.uk/innovation/hecos.

[5] For all four of these universities, the average GCSE score of their students exceeds 550 points (equivalent to five A*s and five As at GCSE), based on the 2004 GCSE cohort.

that are part of the Russell Group, which is a group of 24 high-status and often research-intensive universities, are grouped together in the 'Russell Group' category. The remaining traditional universities which attained university status prior to the 1992 conversion of many technical colleges to university status are grouped together as 'pre-1992 universities'. The remaining universities, referred to as the 'post-1992 universities', are non-traditional universities such as arts colleges and institutions which gained university status after 1992. Based on the average KS4 score of their students, these universities are split into two equal-sized groups: 'post-1992 (more selective)' and 'post-1992 (least selective)'. The full list of universities and their groups can be found in the Online Appendix.

Where individuals have attended multiple undergraduate courses, we assign them to the first course they graduated from or, if they never graduate, the first course they attend. For example, someone who studies architecture at the University of East Anglia for one year, before dropping out and switching to English at the University of Bedfordshire and graduating from this course will be assigned to the latter course, while someone who has graduated from (dropped out of) both courses will be assigned to the course they graduated from (dropped out of) first.

### 4.2.3 Labour market outcomes

The HMRC tax records contain earnings from conventional employment (from PAYE records), as well as earnings from self-employment and profit from partnerships (from Self Assessment records). We combine these two sources of income to create a measure of total income at age 30, in real terms and 2018/19 prices. We look at annual income and do not distinguish between full-time and part-time work, as hours worked are not recorded in the data. We focus on age 30 as this is the oldest age at which we observe all three of our main analysis cohorts in the tax data, which we have up to the 2018/19 tax year.[6]

In most of the analysis, we measure an individual's income in terms of their rank within their cohort's earnings distribution, defining labour market success as making it to the top 20% of the overall distribution of earnings for people of the same age and in the same cohort. We show robustness to different measures of labour market success.

### 4.2.4 Summary statistics

Table 4.1 shows some basic descriptives of our analysis sample.[7] The first column of the table includes all individuals in our analysis sample, the subsequent columns then split this sample into those who did and did not attend higher education,[8] both among Free School Meal eligible and uneligible students.

Around 34% of individuals in the 2002-2004 GCSE cohorts start a university course on a full-time basis before the age of 21. Of those students, fewer than a third attend the high status Russell Group universities, and more than half attend post-1992 institutions, with the remainder attending pre-1992 institutions. At age 30, 17% of individuals in these cohorts do not record any income from employment or self-employment. Mean earnings (combining employment and self-employment income) are £20,700, with median earnings below that at £18,000.

Focusing on those who were eligible for Free School Meals, we see that fewer than 17% attend university, confirming the findings from previous research that they are much less likely to attend university than their peers from wealthier backgrounds. Those who do attend are also much less likely to attend higher status and more selective universities than students who were not on FSM. FSM

[6]As discussed above, for the course-level analysis we use total income at age 28 instead in order to be able to use five cohorts in the analysis.

[7]Section C.1 in the Appendix shows more detail on the sample selection process, and how we get to the main analysis sample from the raw school records.

[8]Individuals who attended university as part-time or mature students are included in the first column, but are not included in the "UG" nor in the "no UG" sample.

pupils from ethnic minority backgrounds attend university at a far greater rate than white FSM pupils. Only 40% of FSM pupils who attend university are white, compared to close to 80% of FSM pupils who do not attend university. Graduate earnings are on average £7,500 lower for FSM than non-FSM students. The likelihood of FSM graduates ending up in the top 40%, top 20% or top 5% of earnings is barely higher than the population average, but this needs to be put in the context of the extremely poor outcomes of the FSM pupils who do not attend university. Less than one in five individuals in this group reach the top 40%, 6% reaches the top 20%, and only 1% reaches the top 5% of earnings. This makes them around half as likely to reach these points in the earnings distribution as non-FSM individuals who did not attend university.

Table 4.1: Summary statistics

| | All | FSM | | non-FSM | |
|---|---|---|---|---|---|
| | | no UG | UG | no UG | UG |
| *Family background* | | | | | |
| FSM eligible | 0.12 | 1.00 | 1.00 | 0.00 | 0.00 |
| *Ethnicity* | | | | | |
| White | 0.84 | 0.78 | 0.40 | 0.89 | 0.81 |
| Black | 0.03 | 0.05 | 0.15 | 0.02 | 0.03 |
| Asian | 0.06 | 0.10 | 0.34 | 0.03 | 0.10 |
| Missing/Other ethnicity | 0.06 | 0.07 | 0.11 | 0.06 | 0.06 |
| *University attendance* | | | | | |
| Attended UG | 0.34 | 0.00 | 1.00 | 0.00 | 1.00 |
| Elite Russell Group | 0.01 | - | 0.01 | - | 0.03 |
| Other Russell Group | 0.09 | - | 0.11 | - | 0.26 |
| Pre-1992 | 0.06 | - | 0.16 | - | 0.18 |
| Post-1992 (more sel.) | 0.10 | - | 0.25 | - | 0.29 |
| Post-1992 (least sel.) | 0.08 | - | 0.46 | - | 0.23 |
| *Earnings and employment* | | | | | |
| Not in employment | 0.17 | 0.31 | 0.15 | 0.18 | 0.11 |
| Mean earnings (£s) | 20,700 | 11,500 | 21,700 | 17,300 | 29,300 |
| Median earnings (£s) | 18,000 | 8,300 | 20,100 | 15,600 | 26,300 |
| In top 40% of earnings | 0.40 | 0.18 | 0.45 | 0.33 | 0.59 |
| In top 20% of earnings | 0.20 | 0.06 | 0.22 | 0.13 | 0.36 |
| In top 5% of earnings | 0.05 | 0.01 | 0.05 | 0.02 | 0.11 |
| N | 1691294 | 165820 | 33386 | 857040 | 534931 |

Notes: The last four columns do not add up to the total in the first column, as there are also a small set of individuals who are not included in either the UG or the no UG groups. These individuals either only have a postgraduate record, attended university as a mature student, or attended university on a part-time basis. Private school students are excluded from the ethnicity statistics, as this is only recorded for state school students. Earnings are shown in 2018/19 real terms.

## 4.3 University access by socio-economic background

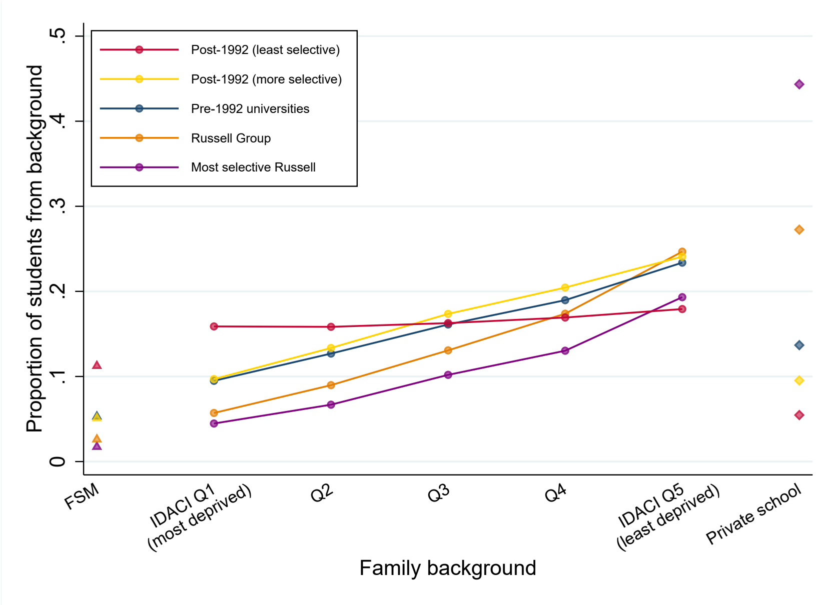
There is now a large body of UK work studying the issue of access to university for students from low-income backgrounds (Crawford, 2014a; Crawford and Greaves, 2015; Crawford et al., 2016a,b). Much of this work has focused on the extent to which socio-economic gaps in access to university can be explained by the background characteristics of students, such as their prior attainment, ethnicity or the school they attended. That work has mostly studied access to university as a whole, or access to groups of institutions, such as the Russell Group. However, there has been relatively little work looking at differences in access to *individual institutions* or different subjects. Since our overall aim is to focus on the extent to which individual institutions and degrees promote mobility, it is natural for us to start by documenting access at a more refined level.

Figure 4.1 starts by presenting the distribution of family background for each of our broad university groupings. At the least selective institutions, individuals from all socio-economic backgrounds are broadly evenly represented,[9] but high-status and selective institutions are dominated by those from the most affluent backgrounds. At the 'Most selective Russell' universities (Oxford, Cambridge, LSE and Imperial), private school students make up more than 44% of the student body, despite representing only 7% of the overall population, while FSM students account for less than 2% of the student body. This means that the privately educated are around 50 times more likely than the poorest students to attend a 'Most selective Russell' Group institution.[10] These numbers are even starker when just looking at the universities of Oxford and Cambridge: privately educated students are nearly 100 times more likely to go to Oxford or Cambridge than pupils who were on FSM.

[9] FSM students represent around 12.5% of individuals in a cohort, private school students around 6.5%, and each quintile of IDACI hence around 16% of individuals. For an equal representation across family background groups, we would expect to see a lower proportion from private schools and slightly lower proportion on FSM than of the IDACI groups.

[10] For comparison, Chetty et al. (2017) find that pupils in the US from the top 1% of the parental income distribution are around 77 times more likely to attend an elite ('Ivy Plus') college than pupils from the bottom 20% of the parental income distribution.

Figure 4.1: University access rates for the 2002-04 GCSE cohorts, by family background and university group

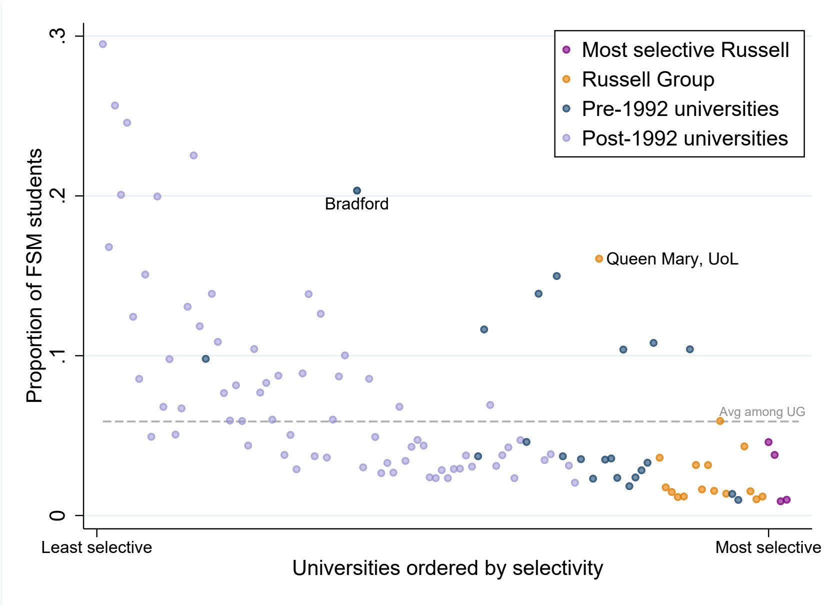



Notes: IDACI quintiles are defined based on state school students who are not eligible for FSM.

There is considerable variation within university groups, as shown in Figure 4.2 which plots the share of FSM students attending each individual university, with universities sorted by their 'selectivity', as measured by the average KS4 score of their students.[11] At the least selective universities, as many as 20-30% of students were on FSM at age 16. At more selective universities, these shares are much lower. For the top 10 most selective universities, this is below 2% on average. Among the universities with the lowest FSM shares in the country - Oxford, Bath and Cambridge - the share of FSM students is less than 1%. With the exception of Queen Mary University of London, all of the Russell Group universities have access rates at or below the national average.

[11]We focus on GCSE grades rather than A-level qualifications as the latter are difficult to accurately compare across subjects. It should be noted that the small number of universities selecting on criteria such such as musical talent or art portfolios may appear to be less selective than they are in practice.

Figure 4.2: University access rates for the 2002-04 GCSE cohorts, by university selectivity




Notes: Selectivity is measured as the average GCSE points score of a university's students. Royal Agricultural College, Harper Adams University College and Leeds City College are not plotted due to sample size restrictions.

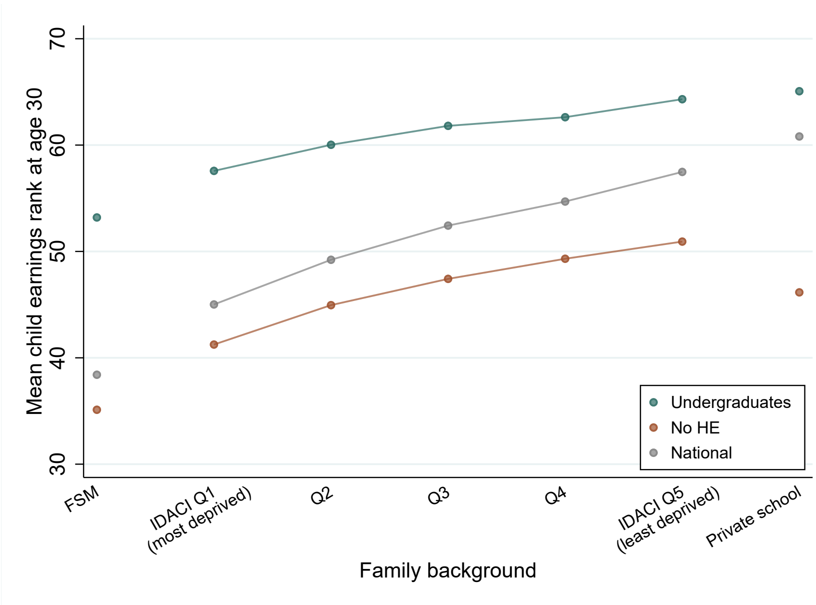
## 4.4 Labour market success

An important motivator of this work is the widely held belief that university can play a key role in increasing social mobility. Figure 4.3 explores this issue by plotting the relationship between the earnings rank of individuals and their parental background among both HE attendees and non-attendees. The figure provides suggestive evidence that this belief is justified.[12] It shows the average earnings rank of people in different family background groups for the whole sample and then split by whether people went to higher education or not. Overall there is a strong positive relationship between family background and earnings rank at age 30, but this relationship is much weaker amongst those who went to university than among those who didn't. Among individuals who

[12]We can only say suggestive because the chart does not deal at all with differential selection into university across groups. For instance, it could be the case that only the very highest-ability poorer students enter university, while a much broader set of wealthier students go. This would result in a flatter curve amongst the university attendees even if university itself did nothing to boost earnings outcomes.

did not go to university the difference in age 30 earnings rank between state school students from the most and least disadvantaged backgrounds is around 16 percentage points, while this gap is only around 12 percentage points for HE attendees. The gap between private school students and FSM students is similar in both cases, but this is driven by the relatively very bad outcomes of non-UG private school children, who are a small and strongly negatively selected part of the private school students due to the extremely high HE participation rates among this group.

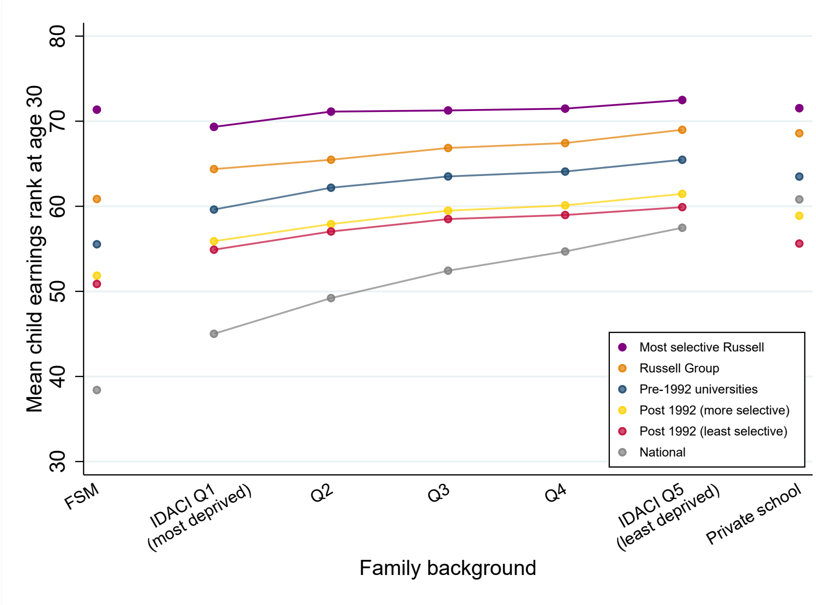
Figure 4.3: Dependence between family background and child earnings rank at age 30 for the 2002-04 GCSE cohorts




Notes: IDACI quintiles are defined based on state school students who are not eligible for FSM. National plot includes children with a linked KS4 and HMRC record. Earnings ranks are calculated in each cohort and are equivalent to percentiles in the earnings distribution.

Figure 4.4 then divides universities up into the five broad university groupings we introduced in the previous section. This shows that the relationship gets weaker still as you look at more selective universities. Amongst the most selective Russell Group universities, there is almost no gap in the average earnings rank for those eligible for FSM and the least-deprived state school students, while the equivalent gap for the 'Other' universities is close to 10 percentiles.

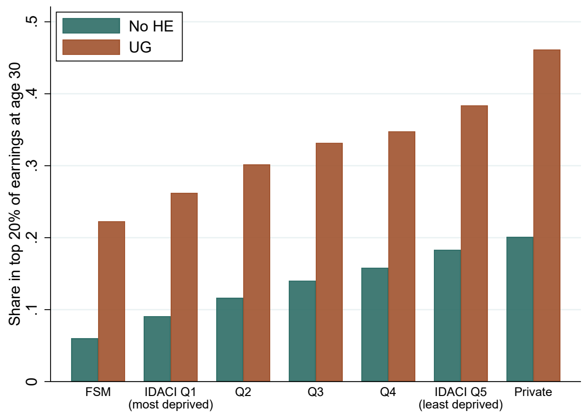
Figure 4.4: Dependence between family background and child earnings rank at age 30 for the 2002-04 GCSE cohorts




Notes: IDACI quintiles are defined based on state school students who are not eligible for FSM.

The socio-economic (SES) gradient in average earnings rank displayed in Figure 4.3 translates into differences accross SES groups in the likelihood of making it into the top 20% of the earnings distribution, our main measure of labour market success. Figure 4.5 looks at the share of individuals who make it to the top earnings quintile at age 30 by family background and higher education status. The likelihood that an individual makes it to the top of the earnings distribution is once again closely tied to family background. For instance, state school students from the least deprived areas are around 10 percentage points more likely to enter into the top earnings quintile at 30 than their counterparts in the most deprived areas with the same higher education status. Across the board, undergraduates have a much greater chance of reaching this high earning group than those who did not attend university. For example, an FSM university student is 16 percentage points more likely to enter the top quintile of earnings at age 30 compared to FSM pupils who did not attend university. University attendees also face a smaller SES gap in the probability of being a high earner: amongst non-HE individuals, the most affluent state school students are around three times more likely to reach the top 20% of the earnings distribution than FSM students, compared to around one-and-a-half times amongst undergraduates. These charts therefore suggest that overall, university education does seem to level the playing field between those coming from different parental backgrounds. This relationship might however mask important variation across institutions or courses.[13]

Figure 4.5: Share in top earnings quintile at age 30, by family background




Notes: Includes the 2002-04 GCSE cohorts. IDACI quintiles are defined based on state school students who are not eligible for FSM.

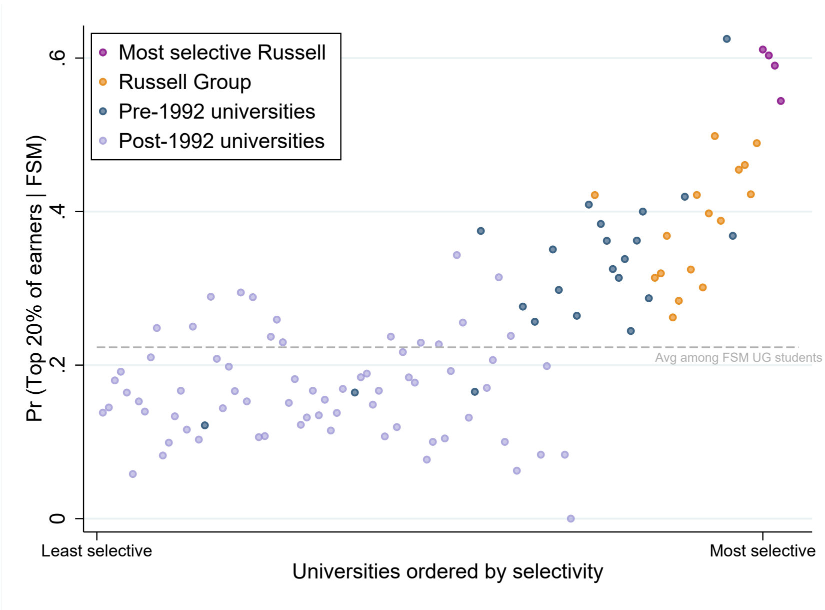
We take this investigation a step further in Figure 4.6 by plotting labour market success rates for individual universities, which we order by selectivity along the horizontal axis. As in the previous section, we follow Chetty et al. (2017) and define labour market success based on whether an individual makes it to the top 20% of their cohort's earnings distribution by age 30. Importantly, these success rates are for the FSM eligible only. Aligning with Figure 4.4, it is the most selective universities which have the highest success rates for their poorer students. We see that all Russell Group universities have success rates above the average,

[13]While the focus of this paper is on differences in mobility *within* the higher education sector, it is important not to forget that many low-income children do not attend university, and that the outcomes of this group are very poor, both relative to their wealthier counterparts, and compared to low-income students who attend university. Policy aimed at improving the outcomes of this group is therefore an important area of further research.

and five institutions - the four 'most selective Russell' Group universities as well as the Royal Veterinary College - have success rates of more than 50%, meaning more than half of their students from low-income families make it to the top 20% of the earnings distribution in their cohort. At the bottom end, we see a set of institutions - mostly arts colleges or among the least selective universities - with success rates below 10%.

While the most selective institutions typically have the best success rates for their poorer students, we saw in the previous section that they also have the lowest access rates. In other words, the poorer students who do get in often do quite well, but not many get in to start with. This makes it difficult to draw conclusions about how well these institutions are performing in terms of contributing to social mobility. In the following section, we explore this issue by putting together access and success rates to create a mobility rate for each university.

Figure 4.6: Labour market success rates for the 2002-04 GCSE cohorts, by university selectivity




Notes: Includes FSM eligible only. Some institutions are excluded for sample size reasons.

## 4.5 Mobility rates

In this section we combine the share of students at an instutition who come from lower income backgrounds with the labour market success of those same students, to create "mobility rates". We first define these "mobility rates" more precisely, before estimating them for each individual institution. We then explore heterogeneity in mobility rates *within* institutions by field of study.

### 4.5.1 Defining mobility rates

We follow Chetty et al. (2017) and define the upward mobility rate of an institution, subject or course as the fraction of its students who (1) were on FSM at age 16 *and* (2) make it to the top 20% of the earnings distribution in their cohort. For institutions and subjects we use earnings at age 30, while for courses we use earnings at age 28 (in order to boost sample sizes).

We calculate this as the product of the proportion of students who were on FSM - which we call the 'access rate' - and the probability for those same FSM students of making it to the top 20% of earnings - the 'success rate'. We can write this as follows:

$$Mobility\ rate = Access\ rate * Success\ rate \tag{4.1}$$

$$P(FSM\ and\ Q5^{child}) = P(FSM) * P(Q5^{child}|FSM) \tag{4.2}$$

where $FSM$ indicates whether a child was on FSM, and $Q5^{child}$ indicates whether a child is in the top 20% of their cohort's earnings distribution at age 30 (or 28 for the course-level analysis). $P(Q5^{child}|FSM)$ indicates the probability of the child making it to the top quintile, conditional on having been eligible for FSM. We will refer to students who were on FSM as a child *and* reach the top 20% of the income distribution at age 30 as 'mobile students'. We also explore different definitions of success, such as the proportion of students on FSM who make it into the top 40% or top 5% of earnings. Higher mobility rates indicate a greater contribution of that university, subject or course to intergenerational mobility.

A potentially useful benchmark for what follows is the mobility rate we would get if access to university was equal for all income groups *and* there was equal labour market success amongst all undergraduates, irrespective of family background. We estimate this benchmark to be 4.4%. That is equal to 12.5% (the share of the population who were FSM eligible in the cohorts we look at) multiplied by 35% (the share of university graduates reaching the top 20% of the earnings distribution).

### 4.5.2 Mobility by institution

We show the variation in mobility rates across institutions in Figure 4.7, by plotting access and success rates against each other. We see that universities with the highest success rates often have low access rates and vice-versa. There are however also a group of universities (mostly from more selective post-1992 institutions) with both low access and low success, leading to an overall correlation between access and success of only -0.24.[14] The figure also plots isoquants which show success-access combinations that yield the 10th, 50th and 90th percentiles of mobility rates. These isoquants reveal that mobility is generally low in UK institutions: the median mobility rate is 0.9%, meaning that for half of all institutions less than 1 in 100 of their students is both from an FSM background and reaches the top 20% of earnings. This is considerably below the benchmark of 4.4% highlighted above. However, there is still considerable variation in mobility rates across institutions. The bottom performing 10% of universities have mobility rates of less than 0.5%, compared with more than 3.3% for the top 10%. The latter means that at least seven times as many graduates from the most mo-

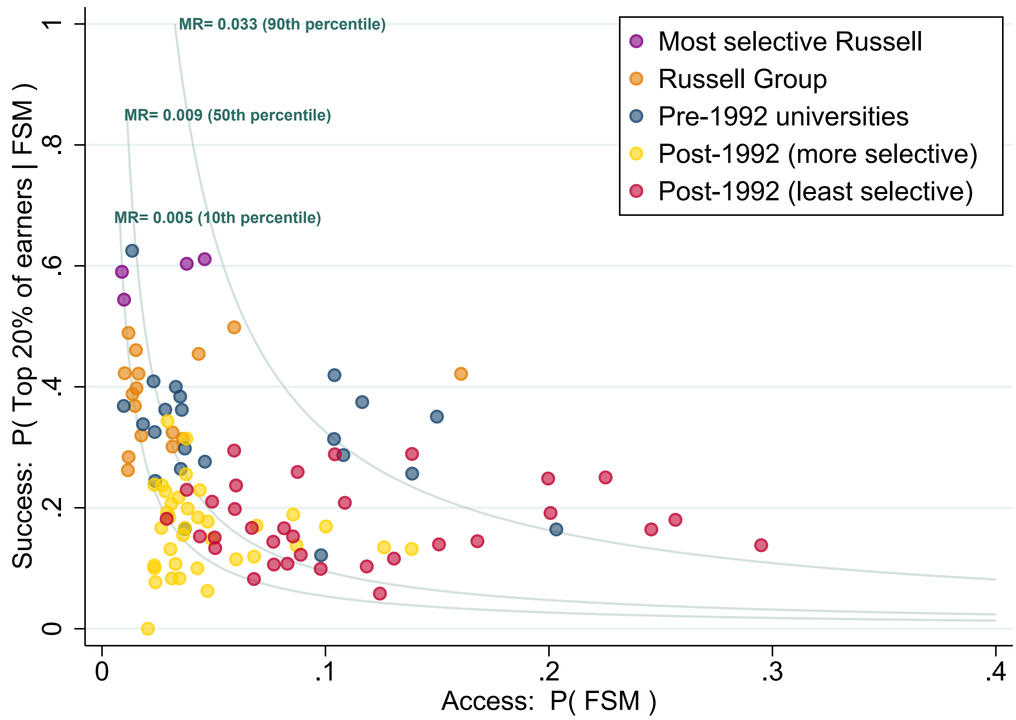
[14]For comparison, Chetty et al. (2020) finds a much stronger correlation of -0.5 in the US.

bile institutions as from the least mobile ones both grew up in households poor enough to be eligible for FSM *and* reached the top 20% of the income distribution.

There are notable differences between our five university groups. Despite the high success rates amongst the Russell Group universities, the low access rates at these institutions mean they do not perform particularly well in terms of overall mobility rates. Only one-third of Russell Group universities are in the top half of mobility scores, while Oxford and Cambridge are close to the 10th percentile. This is not due to poor success rates, but rather to the extremely low access rates at these universities of below 1%. Queen Mary's University of London (QMUL) is an outlier amongst the Russell Group, with an access rate of 16% and a success rate of 42%, meaning it accepts a high share of poor students, and those students do very well, with almost half making it to the top of the earnings distribution. This results in the highest mobility rate across all universities of just under 7%.

The 'pre-1992' universities do a bit better than the Russell Group in terms of mobility, combining slightly higher access rates with decent success rates, resulting in more than half of these institutions being in the top half of the mobility distribution. Among the remaining institutions, the picture is somewhat mixed. A number of institutions in this group, particularly among the 'post-1992 (more selective)' institutions, fail to achieve better access rates than many Russell Group universities, despite lower academic entry requirements. Combining these with relatively low success rates in this group results in a number of institutions that perform poorly in terms of both access and success and consequently have very low mobility rates. Many of the institutions that perform less well are specialist agricultural or arts and music colleges. On the other hand, there are also some institutions in this group where the access rates are so high that despite lower than average success rates, they are among the highest mobility institutions. Notably, the 'post-1992 (least selective)' institutions generally perform very well in terms of mobility, making up over half of the top 10% of institutions for mobility.

Figure 4.7: Institution success, access and mobility rates for the 2002-04 GCSE cohorts




Notes: Universities with at least 250 students are included. Royal Agricultural College, Harper Adams University College and Leeds City College are not plotted due to low sample size.

Table 4.2 lists the top 20 universities for mobility.[15] As already mentioned, QMUL is the university with the highest mobility rate, of 6.8%. The University of Westminster - which belongs to the 'post-1992 (least selective)' group - has the second-highest mobility rate, of 5.6%, followed by City and Greenwich universities with around 5% of mobile students. Overall, seven institutions have mobility rates that match (to one decimal place) or exceed the benchmark mobility rate of 4.4%.

[15] Appendix Table C.2 lists the bottom 20 universities for mobility.

Table 4.2: Top 20 universities for mobility (2002-04 GCSE cohorts)

| Rank | University | Group | Mobility % | Access % | Success % |
|---|---|---|---|---|---|
| 1 | QMU | Russell Group | 6.8 | 16.1 | 42.2 |
| 2 | Westminster | Post-1992 (least selective) | 5.6 | 22.5 | 25.0 |
| 3 | City | Pre-1992 university | 5.3 | 15.0 | 35.1 |
| 4 | Greenwich | Post-1992 (least selective) | 5.0 | 20.0 | 24.8 |
| 5 | London South Bank | Post-1992 (least selective) | 4.6 | 25.7 | 18.0 |
| 6 | Brunel | Pre-1992 university | 4.4 | 11.6 | 37.5 |
| 7 | St George's Hospital | Pre-1992 university | 4.4 | 10.4 | 41.9 |
| 8 | East London | Post-1992 (least selective) | 4.1 | 29.5 | 13.8 |
| 9 | London Met | Post-1992 (least selective) | 4.0 | 24.6 | 16.4 |
| 10 | Kingston | Post-1992 (least selective) | 4.0 | 13.9 | 28.9 |
| 11 | Middlesex | Post-1992 (least selective) | 3.8 | 20.1 | 19.1 |
| 12 | Goldsmiths | Pre-1992 university | 3.6 | 13.9 | 25.6 |
| 13 | Bradford | Pre-1992 university | 3.3 | 20.3 | 16.4 |
| 14 | Aston | Pre-1992 university | 3.3 | 10.4 | 31.4 |
| 15 | SOAS | Pre-1992 university | 3.1 | 10.8 | 28.7 |
| 16 | Hertfordshire | Post-1992 (least selective) | 3.0 | 10.4 | 28.9 |
| 17 | KCL | Russell Group | 2.9 | 5.9 | 49.8 |
| 18 | LSE | Most selective Russell | 2.8 | 4.6 | 61.1 |
| 19 | West London | Post-1992 (least selective) | 2.4 | 16.8 | 14.5 |
| 20 | Imperial | Most selective Russell | 2.3 | 3.8 | 60.3 |

A very noticeable feature of the table is the clear dominance of the London-based institutions. Out of the top 20 universities, only two (Bradford and Aston) are not located in or around London. These London institutions tend to achieve high mobility rates both by taking in a lot of students on FSM and by being relatively likely to send them to the top of the earnings distribution. Many of the graduates of London institutions will end up working in or around London where wages, but also costs of living, are higher than in many other parts of the country. In Section 4.6, we consider the robustness of our findings to adjusting earnings for living costs.

### 4.5.3 Heterogeneity across fields of study

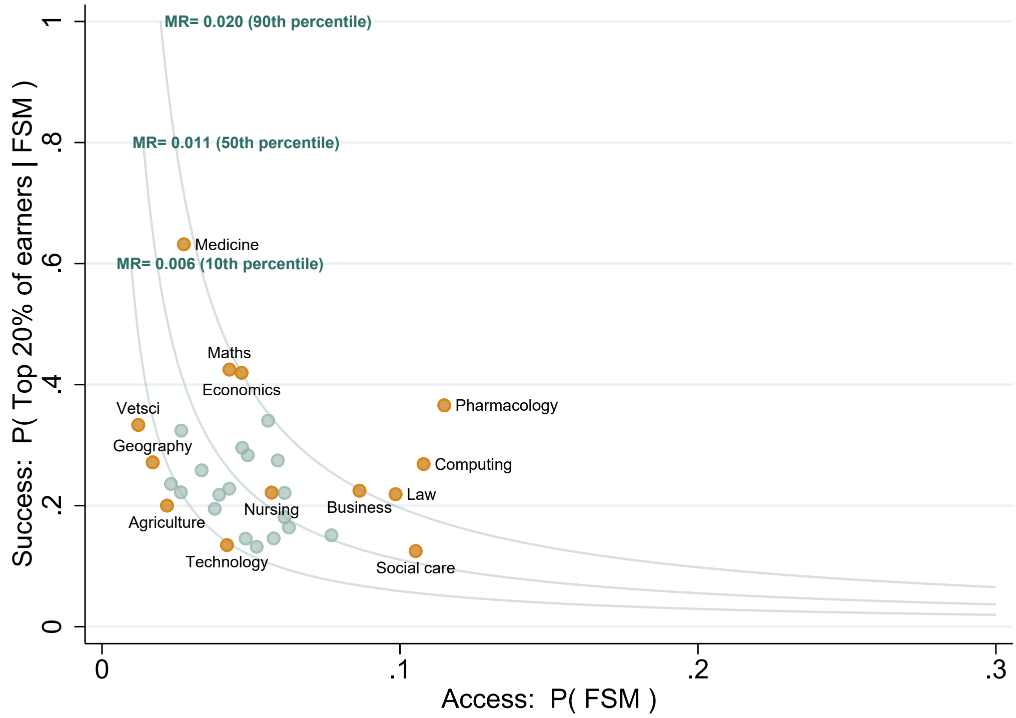
In the above we have explored differences in mobility *across* universities. These institution level mobility rates may hide heterogeneity *within* each institution based on the many different fields of study offered. Previous work (Britton et al., 2021c,d) has shown important heterogeneity in average earnings returns across fields of study within institutions, and large differences in the relative propensity to study subjects across socio-economic background. This suggest potentially large differences in mobility rates across fields of study.

To set the scene, we explore mobility rates at the aggregate subject level, combining all those who study a given subject, regardless of their institution. Figure 4.8 plots access against success rates for each subject, with a subset of subjects labelled.[16] As with institutions, we see significant variation in both access and success rates. Access ranges from 1% for veterinary science to more than 10% for pharmacology, computing and social care. Success ranges from 12.5% for social care to 63% for medicine. The correlation between access and success is nearly identical to that found among institutions at -0.23. As a result, there is a wide range of mobility rates. Pharmacology is the standout performer, having both the highest access rate, and the fourth highest success rate, with around 4 in 100 students coming from the poorest families *and* moving to the top 20% of the earnings distribution. Law, computing and business also combine high access rates with above-average or close to average success rates. In contrast, for veterinary sciences, geography and agriculture, fewer than 0.5% of students were on FSM and reach the top 20% of earnings. These three subjects are also the three subjects with the lowest access rates and have just below or slightly above average success rates. In general, arts and humanities subjects seem to perform poorly, LEM subjects do well and STEM subjects are more mixed.[17] Computing, maths and pharmacology perform well, but agriculture, veterinary science and technology have very low mobility rates.

[16] Appendix Table C.4 lists mobility, access and success rates for all subjects. As with the institution analysis, we include subjects with at least 250 students. We also drop small subjects (Celtic, Humanities non-specific and Combined).

[17] LEM subjects are Law, Economics and Management subjects (the last of which is called 'business' in our subject classification). STEM subjects are Science, Technology, Engineering and Maths subjects.

Figure 4.8: Subject success, access and mobility rates for the 2002-04 GCSE cohorts



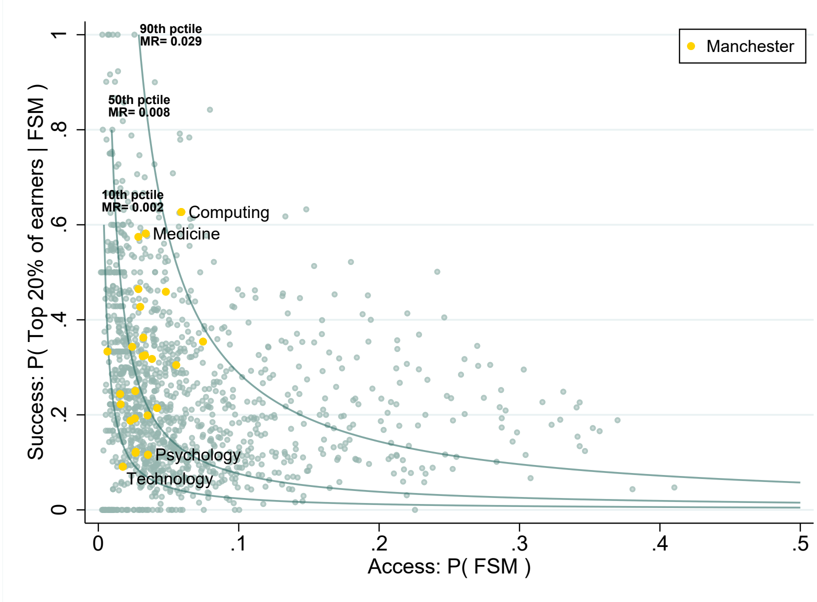

It is worth noting that we observe subjects that have similar mobility rates despite considerable differences in terms of their access and success rates. For instance, mobility rates for business and economics are both at the 90th percentile, but economics has a high success rate and a below-average access rate, while business has one of the highest access rates and an average success rate.

We now explore whether these large differences in mobility rates by field of study also hold *within* institutions. Figure 4.9 plots success versus access at the course (the interaction of institution and subject) level, for over 1,250 courses.[18] There is much greater variation in mobility at the course level than there is at the subject or institution level. For some courses, almost *1 in 10* students were FSM eligible and reached the top 20% of earnings, while for others this number is less than *1 in 100*. Overall, as with subject and institutions, there is only a small (-0.22) negative correlation between success and access rates.

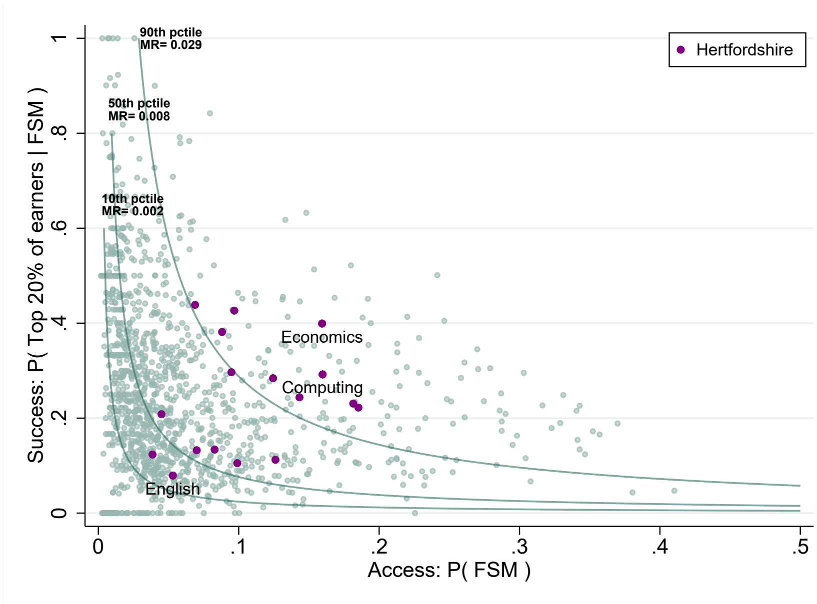
[18]We include all courses with more than 250 students across the five cohorts included in our course-level analysis.

Figure 4.9: Success, access and mobility rates by course for the 2002-04 GCSE cohorts - highlighting institutions

(a) More selective institution




(b) Less selective institution



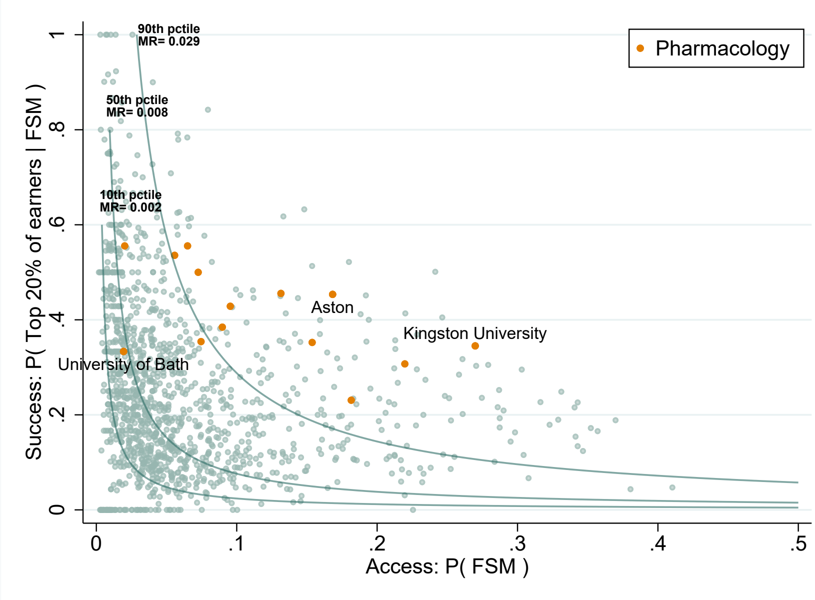

Panels (a) and (b) of Figure 4.9 focus on the different subjects within a more selective and less selective institution, respectively, highlighting the considerable variation *within* institutions. The University of Manchester, highlighted in panel (a), has a relatively narrow range of access rates, which never exceed 8%.

Yet, it has a wide range of success rates, with courses such as technology and psychology towards the lower end, and computing and medicine being among the top. As a result, Manchester has courses in both the top and the bottom 10% of mobility rates. The University of Hertfordshire, shown in panel (b), has a similarly large range, with substantial variation in both its access *and* success rates. Economics at Hertfordshire, for example, is one of the best-performing courses in terms of mobility with 16% of students having been on FSM, and 40% of those students reaching the top 20% of earnings, resulting in an overall mobility rate of 6.4%. At the other end of the scale, English performs much less well with an access rate of only 5% and a success rate of 8% (resulting in a mobility rate of just 0.4%).

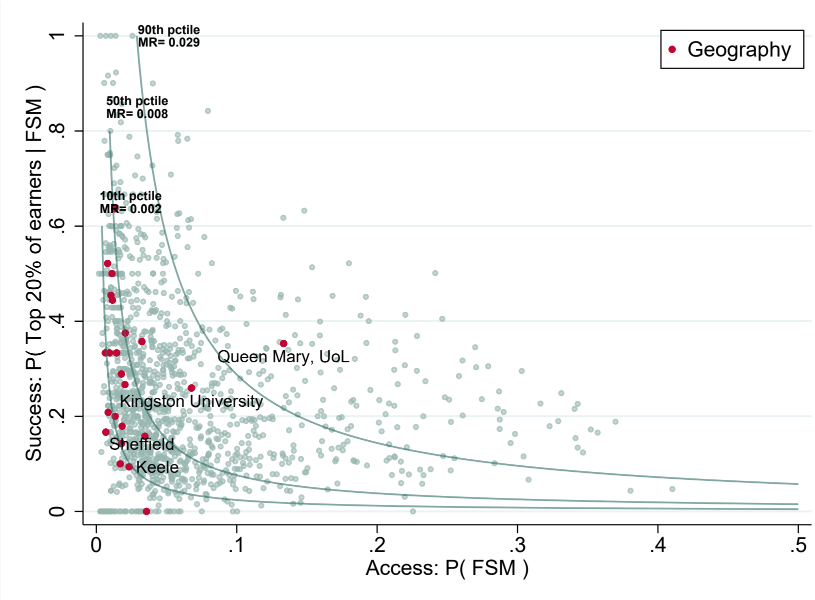
Figure 4.10 repeats the same set of estimates, but instead highlights variation in mobility rates across institutions within two subjects. Panel (a) shows pharmacology, the highest mobility subject, while panel (b) shows geography, a low-mobility subject. Pharmacology performs well across the board, with nearly all of its courses above the median. That said, there is still a lot of variation - the University of Bath is the institution that is below the median for pharmacology, with a mobility rate of 0.8%, while Kingston University is the best performer for pharmacology, with a mobility rate of 9.3% (this comes from an access rate of 27%, multiplied by a success rate of 35%). We also see a large amount of variation in mobility rates for geography in panel (b). Several geography degrees are in the bottom 10% of all courses for mobility, and the vast majority of courses have below-median mobility. However, geography degrees at Kingston University and Queen Mary (two of the most mobile universities) perform very well.

Figure 4.10: Success, access and mobility rates by course for the 2002-04 GCSE cohorts - highlighting subjects

(a) High-mobility subject




(b) Low-mobility subject




Notes: Only pharmacology and geography courses with a sufficient number of FSM students are highlighted.

We list the 20 highest-mobility courses in Table C.3 in the Appendix. The top 20 is largely populated with courses at low-selectivity institutions, in subjects that combine high access with moderate success. QMUL - which dominates the top of the mobility ranking with six of the top 20 courses - is the only Russell

Group institution represented in the top 20. We see the high representation of computing courses, which is a very good subject for mobility in general. Finally, the dominance of London is again clear, with all but one of the highest-mobility courses in or around London.

We use a Shapley-Owen decomposition (Huettner and Sunder, 2012) to determine whether subjects or institutions matters more for mobility rates. This indicates that around 73% of the variance in course-level mobility rates is explained by institution and the remaining 27% is explained by subject. Looking at success and access rates separately we find that the importance of subject and institution is more balanced for success (institutions explain around three-fifths of the variation in success rates), but that institutions are by far the strongest driver of access rates, explaining 87% of the variation compared to just 13% for subjects. Overall, this indicates that while there is indeed substantial heterogeneity across fields of study, institution differences are the main driver of differences in mobility across courses. This further highlights the importance of investigating the policies and characteristics of the highest mobility institutions to learn how mobility may be improved across the higher education sector.

## 4.6 Robustness

In this section we explore the sensitivity of our conclusions to changes in our sample and our definitions of access and success. We also explore the impact of adjusting for differences in cost of living and student composition.

### 4.6.1 Changes in sample and mobility rate definition

*Alternative measures of success.* We start by looking at alternative definitions of labour market 'success'. Columns 2 and 3 of Table 4.3 show how the mobility rates of different institution types vary if we respectively relax our definition of success to include all those in the top 40% of the income distribution, or tighten it to include only those in the top 5%. Figures C.1a and C.1b in the appendix plot these mobility rates against the original 20% mobility rate for each institution. Moving to a top 40% success rate increases the mobility rates of less selective institutions more than those of the more selective ones, such as the Russell Group institutions. This change affects the ranking of institutions relatively little, as many of the best institutions for taking in FSM students and moving them into the top 20% of the income distribution were already the least selective institutions. Overall, rankings are pretty robust to this change, and the correlation between 20% and 40% mobility rates for institutions is very high at 0.94.

If we instead focus on top-tail mobility (moving poor students into the top 5% of earnings), this correlation is somewhat lower at 0.81. Several mid-ranking universities that manage to get a high share of their students into the top 20% are less successfull at getting them to the *very* top of the distribution, while a number of selective universities do relatively better under the 5% definition. For example, Newman University College, a less selective post-1992 university based in Birmingham, ranks 25th in terms of 20% mobility rate but drops to 80th in terms of top 5% mobility rate. Conversely, Oxford University moves up from 95th to 30th place when going from the top 20% to the top 5% mobility rate as it sends a disproportionately large share of students to the very top of the distribution.

*Excluding dropouts.* Throughout the paper we have included all students who started a degree, irrespective of whether they graduated or dropped out. Column 4 in Table 4.3 and Figure C.3 in the Appendix show that our results are virtually unchanged if we instead exclude dropouts. The intuition behind this is that focusing on graduates reduces access rates as students from lower income families are disproportionately likely to drop out,[19] but boosts success rates as drop outs have worse labour market outcomes than graduates. As the mobil-

[19]See Figure C.2 in the Appendix.

Table 4.3: Sensitivity of mobility estimates

| | (1) Main est. | (2) Top 40% | (3) Top 5% | (4) Grads only | (5) Age 32 | (6) CoL | (7) Ability |
|---|---|---|---|---|---|---|---|
| Elite Russell | 1.0 | 1.2 | 0.5 | 1.0 | 0.8 | 0.9 | 0.9 |
| Russell Group | 1.0 | 1.6 | 0.3 | 1.1 | 1.0 | 0.9 | 0.9 |
| Pre-1992 | 1.6 | 2.8 | 0.4 | 1.6 | 1.6 | 1.3 | 1.5 |
| Post-1992 (more sel.) | 0.9 | 2.2 | 0.2 | 1.0 | 1.0 | 1.0 | 1.0 |
| Post-1992 (least sel.) | 2.1 | 4.7 | 0.3 | 2.2 | 2.2 | 1.5 | 2.2 |
| Total | 1.4 | 2.7 | 0.3 | 1.4 | 1.4 | 1.2 | 1.4 |
| Corr with main est. | 1.000 | 0.936 | 0.817 | 0.998 | 0.963 | 0.856 | 0.983 |

Notes: Columns show average mobility rates (in %) for each of the university groups under our main specification and various robustness checks. 'Corr with main est.' shows the correlation of the institution level mobility rates shown in each column with those of the main estimates given in the first column. For consistency we only include institutions with sufficient sample sizes in all specifications. Fifteen smaller institutions are excluded from this exercise as a result.

ity rate combines these two measures, the overall impact from this change on mobility rates is very limited.

*Moving to age 32.* To be able to robustly estimate mobility rates at the institution level, we aggregate over three cohorts in our main estimates. The latest age we observe all these three cohorts is age 30. Column 5 in Table 4.3, and Figure C.4 in the Appendix show how our results change if we instead look at age 32 for our oldest cohort only (which is the only cohort we observe at this age). Again, this change affects our estimates very little. The correlation of the institution level mobility rates is extremely high at 0.96. The largest change is for the Most Selective Russell group institutions. This group consists of only four institutions, all with very low numbers of FSM students. When we only use a single cohort, the estimates of their mobility rates get very noisy.

### 4.6.2 Adjusting for cost of living

One of the most striking results from the previous section is very strong performance of the London-based universities when it comes to mobility rates. One potential challenge to this finding is that our success rates are based on earnings outcomes, but that the same salary may translate to much higher living standards in some parts of the country than in others. Recent work (Britton et al., 2021a), which exploited the new linkage of information on where people live after leaving education to the LEO data, has shown that the geographical distribution of graduates varies dramatically depending on university attended. Graduates from London-based and more selective institutions are much more likely to end up in London after graduation.

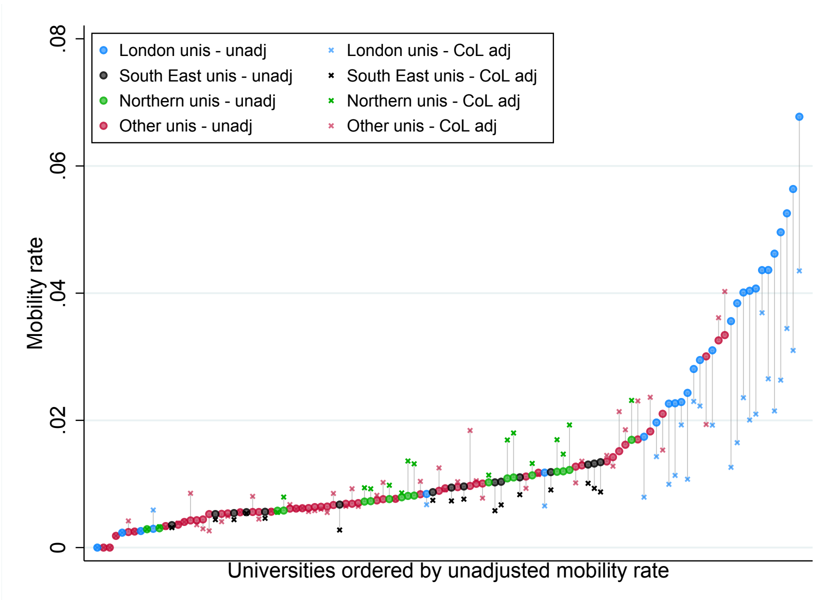
Here we exploit the same data to assess the extent to which our results are affected by adjusting earnings for average differences in the cost of living across the country. To do this, we first combine measures of house prices, rents and the price of goods and services into cost of living indices for each area. We then adjust individuals' earnings in each year according to the cost of living index in their area of residence in the same year.[20] We then define success based on making it to the top 20% of the adjusted earnings distribution, and recalculate mobility rates based on this adjusted measure of success.

We show the impact of this adjustment on the mobility rates of universities in Figure 4.11.This ranks universities by their unadjusted mobility rate and compares this with their adjusted mobility rate, separately colouring universities by their region. Overall mobility rates decrease due to graduates living in more expensive areas than non-graduates on average. Universities at the bottom of the mobility rate distribution see little change, which is usually because access rates are so low at these places that any adjustment to success rates has little effect on mobility rates. In the middle of the distribution, the adjustment increases mobility rates for many of the Northern universities, while it decreases them for universities in the South East. Overall the changes to mobility rates

[20]This method follows Britton et al. (2021a), who provide more detail on the adjustments made. That work highlighted a concern with the regional data which is that students do not appear to always update their home address on their tax forms. As a result, we observe an unrealistic share of students still living at the same address as they did when they were 16. This is likely to overstate the downward adjustment to outcomes for universities situated in more expensive areas, and vice versa. As a result, we believe the adjustments we show below are likely to be an upper bound on the changes we might expect to see if the data more accurately reflected people's home addresses.

in this part of the distribution are still relatively modest, with virtually all adjustments increasing or decreasing mobility rates by less than 0.5 percentage points. However, at the top of the distribution we do see some large changes. Many of these institutions are located in London and a large share of their graduates live in London or the South East. Several of the London institutions at the top of the distribution see downwards adjustments to their mobility rates of over 2 percentage points. Despite this, London institutions perform extremely well even after the adjustment, and still account for seven of the top 10 highest mobility universities. Column 6 in Table 4.4 also shows that the broad rankings of university groups remain the same. Less selective post-1992, and pre-1992 institutions still do much better on average than the Russell Group, and more selective post-1992 institution have the lowest mobility rates.

Figure 4.11: Cost of living (CoL) adjusted mobility rates by university



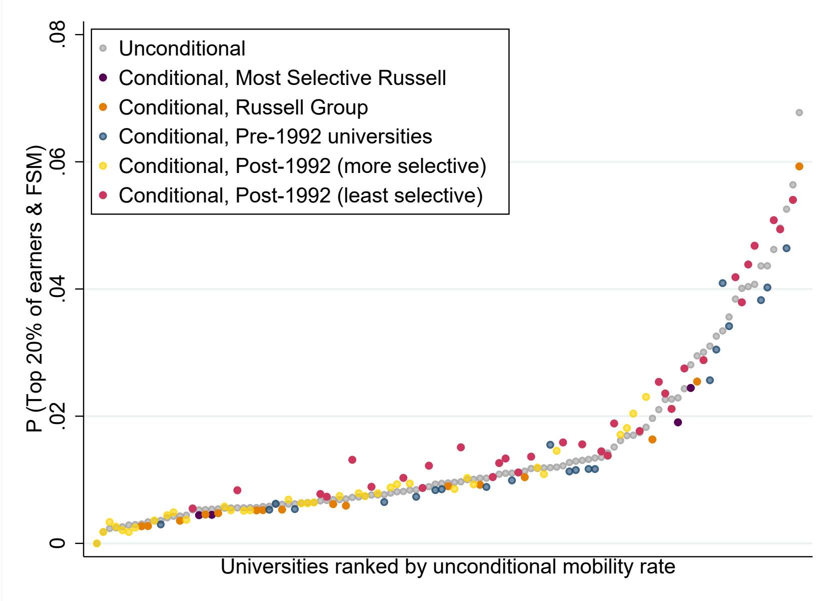

Although the results here do suggest living cost adjustments can make quite large differences to the mobility rate for individual institutions, we would emphasise caution before taking the adjusted results as being a more accurate reflection of mobility rates. As mentioned in footnote 20, there are reporting problems in the regional data we use. More conceptually, it is also important to note that adjusting earnings for costs of living will likely understate differences in living standards. Higher living costs tend to reflect areas having better amenities. A more expensive but greener, safer area, with better schools or better public services, will be preferred by many to cheaper areas with less good amenities. We therefore made the choice to not adjust for living costs in our main results but to show how the results might be affected by these adjustments. And overall, although we see some large changes in individual estimates, we conclude that adjusting earnings for costs of living does not significantly alter our conclusions on which degrees are good for mobility, as it does not dramatically change the ranking of university types.[21]

### 4.6.3 Student composition

Differences in success rates (and hence mobility rates, which are equal to access rates multiplied by success rates) may be driven at least partly by differences in student composition. For example, it could be the case that universities which do very well in terms of mobility are just very good at identifying the highest ability FSM students, who are relatively likely to make it to the top 20% of the earnings distribution. To check whether this drives our results, we test the impact on our findings when we adjust mobility rates to take into account some of these differences in student characteristics. To do so, we calculate *conditional success rates*, which can be interpreted as the probability of a student from a given subject, institution or course making it to the top 20% of the earnings distribution, conditional on their prior attainment and other background characteristics. We then multiply this conditional success rate by the access rate to obtain conditional mobility rates. Section C.4.2 in the Appendix describes our methodology in more detail. Figure 4.12 plots these conditional mobility rates on top of our main institution level mobility estimates.

[21] Nevertheless, we provide both sets of results in our Online Appendix, allowing people to draw upon either in any follow-up to this work.

Figure 4.12: Conditional mobility rates by university




Notes: Universities with at least 250 students and 6 FSM students are included. Negative conditional mobility rates are set to zero.

Although adjusting for differences in student characteristics and attainment can substantially reduce the success rates of the more selective universities (see Appendix Figure C.5) this adjustment only has a fairly minor effect on their mobility rates, due to the low access rates of those institutions. For example, despite the adjustment resulting in a fall of 10 percentage points in the success rate at the University of Cambridge, this barely affects its position in the mobility ranking.

Many of the post-1992 institutions see increases in their success rates. This makes little difference to the mobility rates of the most selective among them. Like the Russell Group institutions, this groups of universities tend to have very low access rates. Some the least selective post-1992 institutions see substantial increases in their mobility rates. Overall, this exercise only reinforces the message that many low-selectivity and low-returns universities do well when it comes to promoting mobility, while many of the more prestigious universities perform poorly.

## 4.7 Correlates of mobility rates

In the previous sections we estimated the mobility rate for all universities and found large differences across institutions. This section will delve into those differences in more detail and to try and understand better what characteristics high mobility universities share.

We start by exploring the relationship between average earnings returns and mobility. Previous work in the UK has used the same LEO data used in this paper to estimate which universities did most to improve the average earnings of their students, and this has been an important input in the debate around the 'value' of different university degrees. In Table 4.4, we show the correlation between the estimated average earnings returns from Britton et al. (2021c) and our mobility rates for universities. We see virtually no relationship between the two measures (correlation = -0.14). This lack of relationship is arrived at via a strong positive correlation between returns and success rates, but a strong *negative* correlation between returns and access rates. Institutions which are good at improving the average earnings of their students also have very good earning outcomes for their FSM students, but take in relatively very few of them. Perhaps unsurprisingly, given Britton et al. (2021c) found selectivity to be strongly related to average earnings returns, we find a similar pattern for the relationship between access, success and mobility with the selectivity of an institution. Several low-selectivity institutions with low earnings returns are at the same time contributing to social mobility by taking in poor students and moving them up the income ladder. For example, the University of West London has amongst the lowest average earnings returns, but is in the top 20 when it comes to mobility rates. Similarly, some of the best institutions in terms of returns do very little to promote mobility among low-income pupils. This highlights that a focus on characteristics often used to rank universities, such as selectivity and average earnings returns, would fail to identify the institutions that make important contributions in terms of helping children from low-income families move up the income ladder.

Looking at the demographics of the student population we find that high mobility institutions are much more likely to have high shares of non-white pupils. While there is only a very weak positive relationship (or even no relationship in the case of the share of Black students) between the share of non-white students and success rates, higher access rates are strongly related to higher shares of non-white students. This is in line with prior findings of ethnic minority pupils being both more likely to be FSM eligible than their white British counterparts, but also considerably more likely to attend university (DfE, 2021). These correlations are however not solely driven by the underlying differences between ethnic groups. Mobility rates for white students only are very highly correlated (corr = 0.91) with our main mobility estimates, and retain a strong positive correlation with the share of non-white students at the university (corr = 0.76). Exploring what drives these very strong correlations would be a fruitful avenue for further research.

Not only characteristics of the university itself may be important in explaining the variation in mobility rates, but also those of the local area the university is located. This is especially true as we know that students from lower socio-economic backgrounds attend university closer to home, and are less likely to move after graduation than those from more affluent backgrounds (Donnelly and Gamsu, 2018; Britton et al., 2021b). The bottom half of Table 4.4 explores this in more detail by correlating characteristics of the Travel to Work Area (TTWA) a university is located, with its mobility, access and success rates.

We start by estimating the correlation between mobility and the share of pupils from the TTWA who are on FSM at age 16. The share of FSM pupils in the local area is positively related to university access rates, but this relationship is relatively weak, and there is no relationship between FSM rates in the local area and success rates. If we focus on the share of students who are both on FSM and high-achieving or of non-white ethnicity (as both these groups are more likely to attend university than the average FSM pupil), there is still no correlation with success rates, but the correlation mobility rates and access rates gets stronger, to around 0.6-0.7. However, being located in an area with a large 'pool' of high-achieving and non-white FSM students by no means explains all the differences in mobility or access rates. For example, while London institutions have the highest access rates on average, the region has universities that are both in the top 10% and bottom 10% in terms of access.

In the final row of the table we estimate the relationship between a university's mobility rates and the mean pay in the area the university is located. We find a relatively strong correlation at 0.7, but interestingly this relationship with mobility rates is mostly driven by universities located in areas with higher mean pay having higher access rates on average, rather than by the weaker relationship between average area pay and success rates. This could both be explained by areas with higher average pay having higher shares of highly attaining FSM students, as is the case in London, or by peer effects, where FSM students interacting with peers from wealthier families might be more likely to attend university.

Table 4.4: Correlates of university level mobility, access and success rates

| | (1) Mobility rate | (2) Access | (3) Success |
|---|---|---|---|
| *University characteristics* | | | |
| Avg earnings returns | -0.14 | -0.48 | 0.81 |
| Selectivity | -0.33 | -0.66 | 0.73 |
| % Asian students | 0.83 | 0.71 | 0.23 |
| % Black students | 0.79 | 0.81 | -0.01 |
| % Other non-white students | 0.84 | 0.64 | 0.34 |
| *Characteristics of TTWA of university* | | | |
| % on FSM | 0.40 | 0.45 | -0.08 |
| % on FSM and 5 A*-C GCSEs | 0.57 | 0.57 | 0.03 |
| % on FSM and non-white ethnicity | 0.68 | 0.63 | 0.12 |
| Mean gross pay | 0.70 | 0.52 | 0.30 |

Notes: Correlates are estimated on the students at each university (for university characteristics) and the population of children from our birth cohorts who lived in a given Travel to Work Areas at age 16 (for TTWA characteristics).

## 4.8 Recent trends

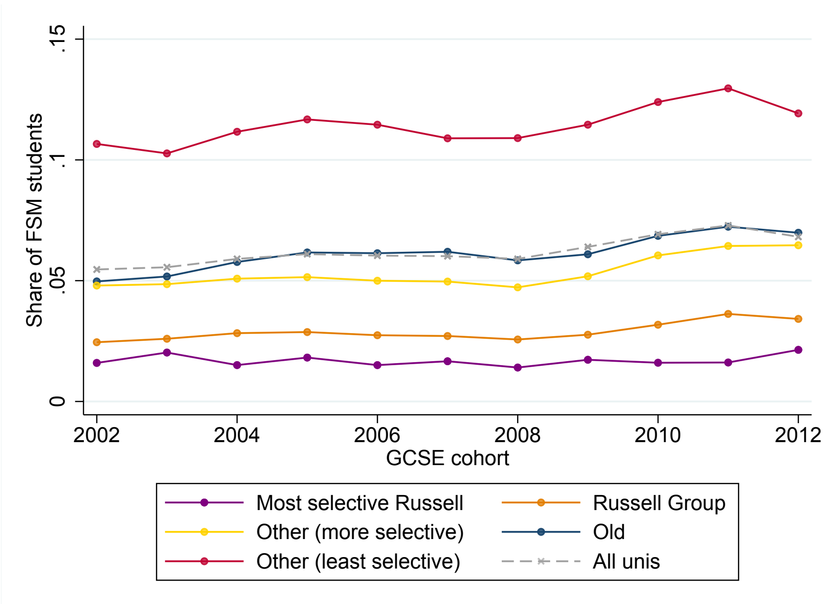
The results so far have focused on students who entered university in the mid to late 2000s. This section attempts to take into account progress universities have made since by looking at more recent trends. We focus first on trends in access and then carefully consider the potential impact of changes in access on success rates, which we do not actually observe for these later cohorts. For comparability over time we focus throughout this section on 18 and 19 year-old entrants.[22] As in the analysis above, we use the share of FSM students at an institution as our measure of access. While this better represents the socio-economic background of students, it is worth noting that universities often target other measures of access such as POLAR rates, which are based on the share of students from areas with low higher education participation. We show the trends in access in terms of POLAR in Figure C.6 in the Appendix.

[22] This means that access rates for the 2002-04 cohorts in this section differ somewhat from the main estimates where we also include 20 year old entrants, though in practice the two are very similar.

### 4.8.1 Access rates

We start by considering changes to access rates. Figure 4.13 plots trends by university groups over the 10-year period up to the 2012 GCSE cohort, the last year we have comparable data for. During that decade, we see a gradual improvement in access in the sector as a whole with the share of FSM students rising from around 5.5% for the 2002 GCSE cohort to around 6.8% for the 2012 GCSE cohort. This is driven by progress across all university groups, though it is particularly pronounced for 'Pre-1992' and 'Post-1992' universities. Progress at Russell Group universities has been slower over the period. Overall, however, there is not strong evidence of either convergence or divergence in access rates between university groups over this period.

Figure 4.13: Access over time by university groups




Notes: Excluding 20-year-olds. Universities with at least 100 students in each year are included.

While most university groups have displayed gradual improvements in access over time, progress is much more varied at the individual university level. We illustrate this by plotting changes in access over time for a selection of universities. Focusing first on our own definition of access rates (the FSM share), panel (a) in Figure 4.14 shows access rates for three of the most selective universities: Cambridge, Oxford and Imperial. Although the single year estimates are noisy, we can see that from similar starting points, Cambridge has made more progress than Oxford in improving access for FSM-eligible students. Access for FSM students at Imperial, meanwhile, has even slightly decreased though is still significantly higher than at Oxford or Cambridge. This highlights the variability in progress towards greater access even amongst similarly selective institutions.

Panels (b) and (c) respectively show changes in FSM-eligible access rates for the three universities with the lowest[23] and highest access rates at the beginning of the period. Newcastle, Bath and Bristol, some of the lowest-access institutions, show flat to modest growth in access, and remain amongst the ten universities with the lowest access rates. At the other end of the scale, access rates at the most accessible universities have largely remained flat or, in the case of London South Bank, even declined. Yet they retain their position as some of the highest-access universities by the end of the period. Finally, panel (d) highlights the universities with the largest growth in access rates from 2002 to 2012. All start the period as universities with above-average access and see significant improvements over this period. This contextualises the very limited growth at institutions such as Cambridge and Bristol.

The figure also plots the trend in the share of students coming from the 20% of areas with the lowest higher education participation (the dashed lines). This is the POLAR statistic that many universities specifically target as part of their widening participation and access schemes. Tracking progress in this measure is more difficult as its definition has changed over time (as indicated by the vertical lines in the figure). While the levels of the POLAR and FSM measures of access are different, it seems to be the case that for the set of universities included in Figure 4.14, the trends are generally quite similar over the 2002 to 2012 period. Looking beyond 2012, universities including Cambridge, Oxford

[23]Excluding universities of Oxford and Cambridge as these are plotted under panel (a), and Harper Adams, Leeds City College and Royal Agricultural College due to small sample sizes.

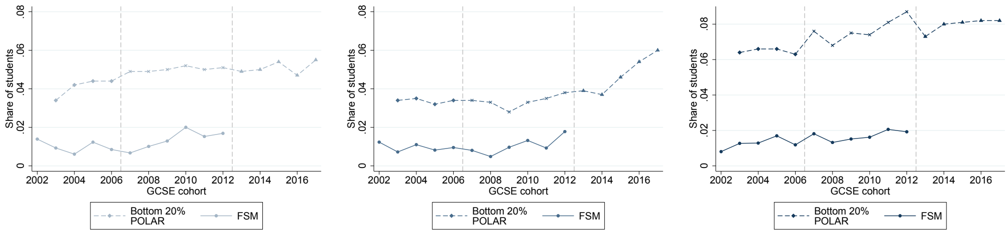
and Bristol have seen shares from the lowest POLAR quintile areas increase. If this reflects an improvement in access for FSM students as well, this would be a welcome development.

Figure 4.14: Access over time for individual institutions

(a) Most selective

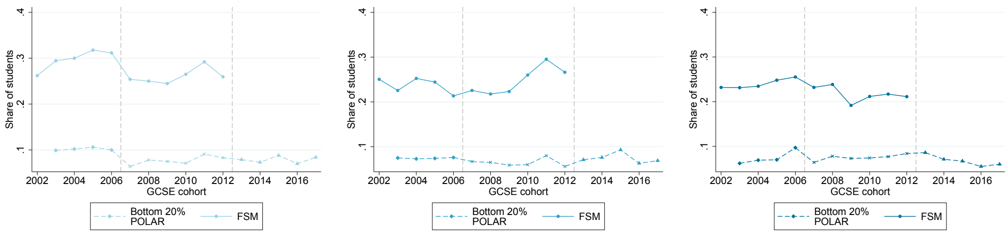

Cambridge Oxford Imperial

(b) Lowest access




Bath Bristol Newcastle

(c) Highest access




East London London Met London South Bank

(d) Highest growth


UC Birmingham Middlesex Newman UC, Birmingham

Notes: Excluding 20-year-olds. Most selective excludes LSE due to small sample sizes. Lowest access excludes Oxford and Cambridge (plotted under most selective) and Harper Adams, Leeds City College and Royal Agricultural College (due to disclosivity). POLAR2 is plotted between 2004 and 2006. This is based on people who were 18 between 2000 and 2004 and who started a course, aged 18 or 19, between 2000-01 and 2005-06. POLAR3 (based on 18- and 19-year-olds starting between 2005-06 and 2009-10 and between 2006-07 and 2010-11, respectively) is plotted up to 2012. 2013 onwards plots POLAR4 (based on 18-year-olds starting between 2009-10 and 2013-14 and 19-year-olds starting between 2010-11 and 2014-15). While POLAR rates are generally higher than FSM rates, many London universities have lower POLAR rates as London has high HE participation.

### 4.8.2 Do higher access rates affect success?

A challenge in updating our mobility rates to account for recent trends is that we do not observe success rates for more recent cohorts as they are too young to have any usable labour market outcomes. We therefore have to predict success rates, and in doing so we want to take into account the possibility that changes to access rates might be associated with changes to success rates. For example, universities that see dramatic improvements in their access rates might also experience corresponding drops in success, offsetting the implied increase to mobility rates. Alternatively, as shares of FSM students rise, universities might improve their ability to cater for such students through support services such as career guidance, thereby boosting success. This exercise will provide us both with valuable information to predict future mobility rates, and answer an important question on whether there is a trade-off in terms of reduced success rates for institutions increasing their access rates.

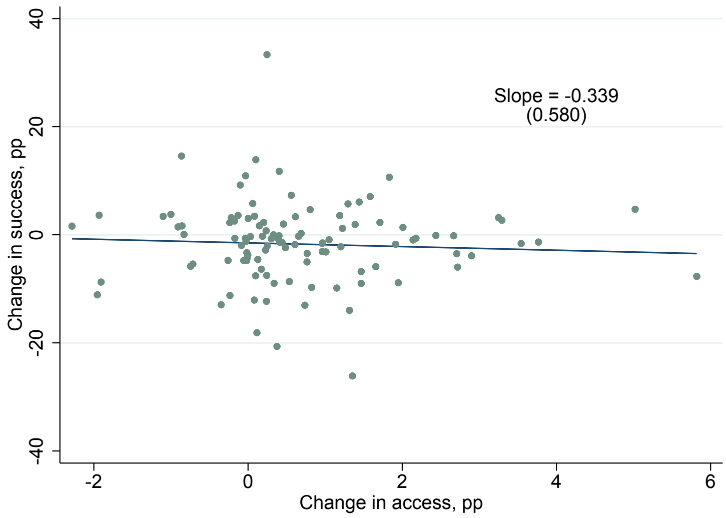
In Figure 4.15 we look at how changes in access relate to changes in success, within institution. Specifically, we plot the change in success rates between the 2002-03 and the 2005-06 GCSE cohorts against the corresponding change in access rates. The figure also shows the regression line, slope and standard error from an (enrollment-weighted) OLS regression. We do not find a significant relationship between changes in access and success, and we fail to reject the null hypothesis of no relationship between the two.[24] In Appendix Figure C.7 we additionally explore whether there is any evidence of changes in access affecting some shorter-term outcomes. We find no significant (at the 5% level) relationship between changes in access and changes in completion, higher class degree, progress to postgraduate study or early-career earnings of FSM students.

Taken together, we don't find any evidence for a 'success-access' trade-off, nor of increases in access actually improving success rates of FSM students.

[24]Chetty et al. (2020), who find that a 1 percentage point improvement in access rates is associated with a 0.17 percentage point reduction in success rates, come to the same conclusion.

Changes in access don't seem strongly related to success rates or shorter terms outcomes. Based on these findings we keep the success rates from the 2002-04 GCSE cohort when we predict mobility rates for more recent cohorts in the next subsection.

Figure 4.15: Changes in success rates and access rates by institution, GCSE cohorts 2002-05




Notes: Success measured at age 29. Excluding 20-year-olds. Institutions with at least 100 students and at least 6 FSM students in 2002 and 2005 are included in the regression.

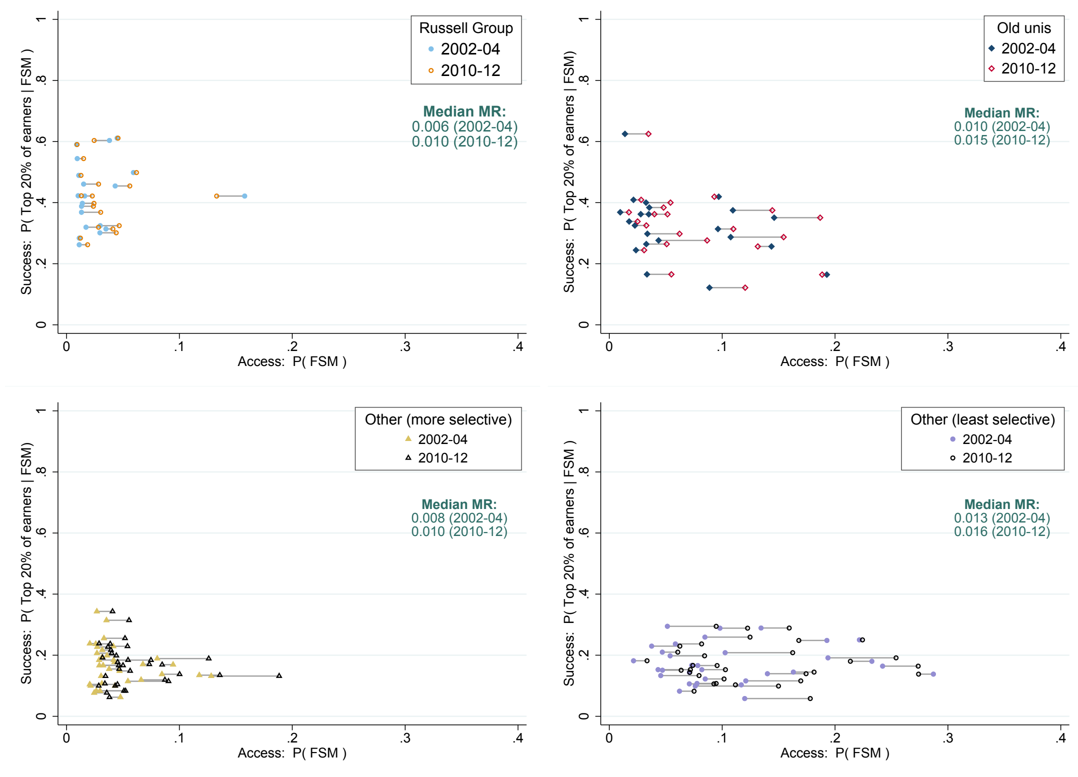
### 4.8.3 Consequences of recent trends for mobility

Bringing together the updated access rates with the success rates, we now project mobility rates for each university in 2012. Figure 4.16 plots the original mobility rates (GCSE cohorts 2002-04) and the projected rates (GCSE cohorts 2010-12), separately for each university group.

Over the 10-year period, the median mobility rates for all groups have increased due to improvements in access. Though each university group contains both institutions that have seen large increases in access and those that have seen very little improvement, on the whole, Russell Group universities have tended to increase access the least while 'Pre-1992' universities and the least selective universities have increased access the most. Despite some differences in growth of access between different types of universities, the ranking of universities in terms of mobility is largely unchanged. Table 4.5 summarises these changes in access rates and subsequent mobility rates for different groups of universities. We show access and mobility rates for 2002-04, our 2010-12 projection based on our most recent data on access, and a projection of access and mobility rates in 2017. Access rates for the 2017 GCSE cohort are projected by drawing on access statistics published by HESA that use an alternative measure of deprivation - namely, the share of students from areas in the bottom 20% of higher education participation (bottom POLAR quintile). While the access levels of POLAR and our FSM measure differ, changes over time are broadly similar. This allows us to use the growth in the POLAR access statistic over recent years to approximate how the FSM access rate is likely to have evolved over these years.[25]

[25]Specifically, for each university we take the average percentage change in the POLAR access measure between cohorts entering university in years 2015 and 2016 and years 2018 and 2019 (equivalent to cohorts taking their GCSEs in summer 2013 and 2014 and summer 2016 and 2017, respectively) and apply this change to the FSM access rate in 2012 to project the FSM access rate in 2017.

Figure 4.16: Updated mobility in 2012, by institution




Notes: Figure adjusts access rates but holds success rates fixed at previous levels. 20-year-olds are excluded resulting in small differences in the estimates for the 2002-04 cohorts relative to Section 4.5.

The table shows improvements in access and hence mobility rates across the higher education sector in the decade following the mid 2000s (notably this period included the large tuition fee reforms in 2012). Under the assumption that success rates will remain the same, we estimate that the average mobility rate across the whole sector has improved from 1.3% to 1.6%, and that there have been modest improvements for all university groups. That said, for the most selective Russell Group universities, almost all of this improvement has been in more recent years, which is a period for which we have less reliable data.

These recent trends compare quite favourably with equivalent estimates from Chetty et al. (2017), who show that high-mobility colleges in the United States have become less accessible over time. We find increases in mobility over time, both among high and low mobility institutions. Nonetheless, access rates at many English institutions - especially the most selective institutions - remain extremely low, and the latest estimate of a mobility rate of 1.6% across the whole sector is still well below our benchmark rate of 4.4%. This suggests there is much progress still to be made.

Table 4.5: Trends in mobility rates from the 2002 to 2017 GCSE cohorts

| | *2002-04* | | | *2010-12 projection* | | *2017 projection* | |
|---|---|---|---|---|---|---|---|
| | Access (%) | Success (%) | Mobility (%) | Access (%) | Mobility (%) | Access (%) | Mobility (%) |
| **University type** | | | | | | | |
| Elite Russell | 1.7 | 59.0 | 1.0 | 1.8 | 1.0 | 2.2 | 1.3 |
| Russell Group | 2.6 | 38.4 | 1.0 | 3.4 | 1.3 | 3.7 | 1.4 |
| Old universities | 5.3 | 27.8 | 1.5 | 7.0 | 2.0 | 7.2 | 2.0 |
| Other (more selective) | 4.9 | 17.7 | 0.9 | 6.3 | 1.1 | 6.6 | 1.2 |
| Other (least selective) | 10.7 | 18.5 | 2.0 | 12.4 | 2.3 | 11.9 | 2.2 |
| **High mobility unis (2002-04)** | | | | | | | |
| Low selectivity | 10.0 | 19.3 | 1.9 | 11.6 | 2.2 | 11.3 | 2.2 |
| High selectivity | 6.6 | 32.7 | 2.1 | 8.0 | 2.6 | 8.1 | 2.6 |
| **Low mobility unis (2002-04)** | | | | | | | |
| Low selectivity | 4.2 | 14.9 | 0.6 | 5.6 | 0.8 | 5.9 | 0.9 |
| High selectivity | 1.7 | 34.2 | 0.6 | 2.6 | 0.9 | 2.8 | 1.0 |
| All | 5.6 | 22.3 | 1.3 | 7.0 | 1.6 | 7.1 | 1.6 |

Notes: High- and low-mobility universities are universities with above- and below-median mobility rates, respectively. High-selectivity universities include 'Pre-1992' and Russell Group universities. Low-selectivity universities include 'Other' universities. 2017 projection uses student numbers from 2010-12 to weight universities; 2002-04 and 2010-12 use contemporaneous student numbers as weights. Excludes those starting university at age 20 for comparability over time. We are not able to project mobility rates in 2017 for subject groups as HESA does not report POLAR statistics by subject.

## 4.9 Conclusion

This paper provides new evidence on the contribution of different institutions, subjects and courses to social mobility. We show large variation in both the proportion of low-income students taken in, and in the proportion of these students who reach the top of the earnings distribution. Despite having extremely high success rates, we see that elite institutions do very poorly in terms of mobility rates, as they let in so few low-income students. Instead, low- to mid-ranking institutions, often based in London, are the best performers in terms of mobility. We also find considerable variation in the mobility rates of subjects, and indeed across subjects within the same institution. Many institutions have courses both in the top and bottom 10% in terms of mobility. This suggests policy might be more appropriately focused at specific combinations of universities and subjects rather than at universities as a whole.

This work can feed into the discussion of 'value' in higher education. The results are important for documenting the value of universities beyond average earnings returns. Indeed, a key finding is that many of the institutions and institution-subject combinations with high mobility rates do not have very high average returns. This implies that policies that restrict funding for low-returning courses could come at a cost in terms of social mobility.

These results can also directly promote mobility themselves. Publicly available rankings of mobility rates might encourage universities to focus on boosting these rates. Finally, this research will hopefully motivate future work that helps us to better understand what drives mobility.

# Chapter 5

# The impact of parental job loss on the long-run outcomes of their children

## 5.1 Introduction

The negative impacts of job loss on those affected have been extensively documented. It has been shown to lead to large and persistent decreases in earnings and consumption (Jacobson et al., 1993; Couch and Placzek, 2010; Eliason and Storrie, 2006; Chetty and Szeidl, 2007; Aguiar and Hurst, 2005), worse health outcomes, higher mortality (Sullivan and von Wachter, 2009; Kuhn et al., 2009), and even an increased probability of divorce (Charles and Stephens, 2004). In light of these large negative impacts on the outcomes of the displaced, we may expect children's outcomes to be affected by the job loss of their parents, yet few papers have looked at the intergenerational impact of displacement. This paper adds to this very limited evidence base and estimates the impact of parental job loss on children's later life outcomes.

To isolate the impact of parental job loss from that of other parental characteristics, this paper makes use of the large scale closures of coal mines in England and Wales in the late twentieth century as an exogenous shock to parental employment. In the second half of the twentieth century mines in the United Kingdom (UK) were nationalised and managed by a single central entity, the National Coal Board. Virtually all UK coal mines closed during the course of the twentieth century[1] and the exact timing of mine closures was driven largely due to geological factors determing their productivity such as the condition and location of coal seams (Glyn and Machin, 1997; Glyn, 1988). In addition, mine workers were a relatively low-skilled group, usually drawn from the surrounding villages. These factors make strong selection of workers into mines closing at different times unlikely, and I show descriptive data which supports this.

I focus on a cohort of children born between 1973 and 1981, all of whom have fathers who worked in mining in 1981. I then compare the outcomes of children whose father lost his job in mining between 1981 and 1991, to the outcomes of children whose father lost his job in mining after 1991. My estimates show that women are 13 percentage points (around 33%) less likely to obtain a degree following paternal job loss in their childhood, and 12 percentage points less likely to own their home. I find no evidence of any impact of paternal job loss in childhood on later life incomes for men. Around half of the impact of job loss on educational attainment, and around 20% of the impact on home ownership can be explained by the loss in fathers' earnings. This suggests an important role for unemployment insurance in mitigating the negative consequences of parental job loss on children.

This paper makes two main contributions to the existing literature. Firstly, it provides new evidence of the long term impact of parental job loss on the outcomes of their children. The small existing literature on the intergenerational impacts of job loss has mostly looked at health (Mork et al., 2014; Schaller and Zerpa, 2019), educational attainment (Pinger, 2016; Rege et al., 2011; Stevens and Schaller, 2011; Kalil and Ziol-Guest, 2008; Coelli, 2011) or earnings (Hilger,

[1] In 2017 only three active coal mines were recorded.

2016; Page et al., 2007; Fradkin et al., 2019) in childhood or early adulthood, which may not be reflective of lifetime impacts. The unique dataset used in this paper, which links population census data over 40 years, allows me go beyond this and look at a wide set of outcomes up to the age of 38, including educational attainment, employment, earnings and housing tenure. Secondly, this paper improves our understanding of the long-run impacts of the decline of the mining industry. A century ago, coal mining accounted for nearly one in ten male workers in the UK, but the industry has now virtually disappeared in this country. Other countries such as Spain, the United States and Germany have seen similar declines. Previous work has shown persistent increases in unemployment, disability benefit receipt and early retirement (Beatty and Fothergill, 1996; Black et al., 2002) in areas affected by the strong decline in the mining industry. To the best of my knowledge, however, this is the first paper exploring the intergenerational impacts of mining closures.

The rest of the paper proceeds as follows. Section 5.2 covers the methodology and Section 5.3 discusses the data used. Results are shown in Section 5.4, before concluding in Section 5.5.

## 5.2 Methodology

To estimate the impacts of parental job loss on children's outcomes, I follow the approach taken by much of the previous literature and use an exogenous shock to employment to isolate the impact of job loss from that of other parental characteristics. The exogenous employment shock I exploit is the coal mine closures that took place in the UK in the latter half of the 20th century.

Historically, mining was one of the UK's main industries, employing close to 1.2 million workers - or 8% of the male labour force - at its peak in 1920. As cheaper alternative energy sources, such as oil, nuclear power or imported coal, became available, the industry entered a steady decline. This decline was particularly rapid in the 1980s. While there were 200 active deep coal mines in 1981, only 50 remained a decade later. Employment losses mirrored these mining closures, decreasing from 170,000 to 40,000 in the same period.[2] By the end of the 20th century, virtually all UK coal mines had closed.[3] After the Second World War, the coal industry in the UK was nationalised and all mines where managed by a single entity, the National Coal Board.[4] The timing of the closure of each mine was centrally determined by the National Coal Board using an assessment of the productivity of each mine, driven mainly by geological factors and market access (Glyn, 1988; Glyn and Machin, 1997). As a result, the timing of mining closures was independent of the characteristics of the individuals working in them.

I exploit these exogenous differences in the timing of mine closures to identify the impact of parental job loss on child outcomes. Coal mining is a predominantly male industry and the overwhelming majority of individuals directly affected by mine closures were men. Father's lay off is also likely to have a larger impact on family income, as men were usually the main breadwinner in the period of study. Following much of the previous literature, I therefore focus on father's lay off. Using linked decennial census data on a cohort of children born between 1973 and 1981 I identify all children with a father who was employed in mining in 1981. I then compare the outcomes of children whose father lost his job in mining between 1981 and 1991 to the outcomes of children with fathers who lost their job in mining after this period. The difference in the outcomes of these two groups informs us about the impacts on children of paternal job loss in childhood. This impact is estimated using a reduced form regression of the outcome of interest on an indicator for whether the father was displaced between 1981 and 1991 and baseline controls from 1981. More formally, I run

[2] Data from the Department for Business, Energy and Industrial Strategy's publication "Historical coal data: coal production, availability and consumption".

[3] See Appendix Figure A1 for the number of active UK coal mines over time. Only three active coal mines were recorded in 2017.

[4] The National Coal board managed the UK mines from 1947 until 1994.

the following regression:

$$Y_{it} = \alpha + \beta D_{it}^{81-91} + \gamma X_{it} + \epsilon_{it} \tag{5.1}$$

where $Y_{it}$ is the outcome of interest, $D_{it}$ is a dummy variable which is equal to one if the father lost his job in mining between 1981 and 1991 and zero otherwise and $X_{it}$ is a vector of baseline controls, which includes region, number of children in the family, housing tenure and parental education in 1981.

The first step of the estimation will be to estimate the impact of paternal job loss on family circumstances. The outcomes in this part of the analysis, or $Y_{it}$ in the above equation, will be economic status and earnings of both parents and housing tenure of the family in 1991. As the job loss of the displaced fathers may have taken place anytime between 1981 and 1991, for some families we will be measuring outcomes soon after displacement, while for others the displacement will have happened a few years prior. The overall effect I will be estimating is the average effect of paternal job loss in the ten years prior to 1991. As mine closures were relatively evenly spread in the decade prior to 1991,[5] job loss will on average have been five years prior to 1991.

The second step of the estimation will look at the impact of paternal displacement on the long-run outcomes of children. This will be estimated using the same reduced form equation shown in equation 5.1. The main outcomes of interest $Y_{it}$ will be degree attainment, earnings and employment, and housing tenure. The main results will focus on the impact of paternal job loss on these four outcomes when individuals are between 30 and 38 years old. Appendix Section D.3 shows the impact on a wider set of outcomes, and the impact at earlier ages.

There are three main potential threats to the identification, which I will address in turn. First, displacement might be mismeasured. The data only ob-

[5]See Figure D.1 in the Appendix for the number of active mines in each year from 1970 to 2018.

serves individuals every ten years, and does not have employment histories for the intervening years. Individuals who have been made redundant after 1981, but found another job in the mining industry by 1991 will be included in the 'survivor' sample. Individuals who managed to find a new job in a severely declining industry would likely be positively selected on observables, and unobservables such as motivation. If these characteristics also lead to better outcomes for their children, my estimates of parental job loss would overestimate the true impact of job loss. This source of measurement error is likely to be small. The mining industry was in continuous and severe decline throughout the 1980s, and it is unlikely many laid off miners would have been able to find a new job in another mine. I also control for differences in observable characteristics between the displacement and survivor samples. Another possible source of mismeasurement is individuals who voluntarily left mining between 1981 and 1991 being erroneously included in the displacement sample. This is unlikely to be an important source of mismeasurement in the context of this study. In the period I look at, the median wage for coal mine workers was more than 30% higher than in other low skilled occupations.[6] These higher wages will at least partly reflect a compensation for relatively dangerous work, but the predominantly rural mining areas also had much higher levels of unemployment than other areas and the local mine was often the main employer. This makes it unlikely that many individuals in the displacement sample had voluntarily left the industry between 1981 and 1991. In Section 5.4, I explore this further and show a large and persistent decrease in employment and earnings among the displacement sample, which aligns with involuntary redundancy. Even if this source of mismeasurement would be an issue, this would be expected to attenuate the estimated impact of job loss and any effect found would provide a lower bound on the true impact.

[6] The New Earnings Survey in 1991 shows median hourly earnings in 1991 prices for coal mine labourers of £7.19, compared to £5.40 for the median manual worker.

The second potential threat to the identification would be miners laid off between 1981 and 1991 being a select sample of individuals working in mining in 1981. The identification relies on fathers in the displacement and survivor samples being identical on other characteristics that influence child outcomes. Previous work has documented selection into closing firms (Brown and Matsa, 2016; Abowd et al., 1999), which would suggest survivor and displacement fathers might be different in other ways that influence child outcomes. This issue is less likely to be an issue in the context of mining closures studied in this paper than it has been in other papers looking at firm or plant closures. In the period of study, UK coal mines were all nationalised and controlled by a single entity, the National Coal Board. The timing of mine closures was largely driven by geological factors (Glyn, 1988; Glyn and Machin, 1997). In addition, mines usually drew their labour force from the surrounding villages, and miners were a fairly homogeneous group. This makes it unlikely that assortative matching between mines and workers and selection into closing mines is a significant issue in this context. I will also control for observable characteristics of the father in 1981 to further alleviate any concerns that selection may drive the estimates.

Finally, as mines were often the main employer in the local area, local labour market conditions will likely have been worse for the children of displaced fathers than for those of non-displaced fathers. Any difference between the outcomes of children in these two groups will therefore be a combination of the direct effect of paternal lay-off and changes in local labour market conditions related to the paternal lay off. Based on human capital theory, we would expect the disappearance of highly paid low-skilled mining jobs - which increases the return to education - to lead to increases in educational achievement. Results from prior work looking at fracking (Cascio and Narayan, 2015) and the 1970s coal boom and 1980s coal bust (Black et al., 2005) are consistent with this prediction. Similar effects have been found by papers looking at the impact on education of local labour market conditions more generally. Betts and McFarland

(1995) find that a 1 percent increase in the unemployment rate is associated with a 4 percent rise in community college enrollment. Clark (2011) finds large positive enrollment effects of higher local youth unemployment in the UK. Based on this existing evidence, we would expect that growing up in areas with large numbers of mine closures - which will be strongly related to having a displaced father - would lead to an increase in educational attainment. Any impact of paternal lay off on educational attainment is therefore likely to be of lower bound of the direct effect of parental lay off.

## 5.3 Data

The analysis uses linked census data from the Office for National Statistics (ONS) Longitudinal Study (LS) combined with occupation based earnings data from the New Earnings Survey (NES) and Annual Survey of Hours and Earnings (ASHE). This section will describe each of these data sets in turn, before showing some descriptives of the analysis sample.

### 5.3.1 Longitudinal Study

The Longitudinal Study (LS) is a random sample clustered by date of birth of the decennial census in England and Wales, where individuals are followed across successive censuses. It includes all individuals born on four dates in the year leading to a sample of approximately 1.1% of the population. The census takes place every ten years, and individuals are followed across successive censuses, starting from the 1971 census. The data currently covers all censuses from 1971 up to 2011,[7] allowing us to track individuals for up to 40 years. Individuals are added the the LS sample at each census through births or immigration of individuals born on one of those four dates in the year. The Longitudinal Study now contains information on more than 1.1 million individuals.

[7]The most recent 2021 census has not been incorporated yet.

Completing the census is a legal requirement in the UK, and those who do not fill in their census forms are contacted by census officers to encourage them to complete it. Those who still do not complete the census can be fined up to £1000. Response rates are extremely high as a result.[8] Forward linkage rates - the share of individuals observed in a census and not registered to have died or emigrated who are found in the subsequent census - are close to 90%.[9]

The census includes rich information on household tenure, educational qualifications, marital status, economic status, hours of work, and extremely detailed measures of occupation and industry of work. Income or wages are not recorded, and are estimated in this paper by matching on gender and detailed occupation using a secondary dataset, the Annual Survey of Hours and Earnings (ASHE)[10], which is discussed in more detail in the next section. The rich information contained in the LS is available both for the sample member, and all individuals living in the household of the sample member at a given census, including any co-habiting partners, children, siblings, and - crucially for this study - parents. While household members are not tracked across censuses themselves, we will observe them in successive censuses if they still live in the same household as the sample member. With the census taking place every ten years, and most children living with their parents for longer than that, this enables us to track changes in employment status, occupation and industry for children's parents over a period of at least ten years during childhood. In addition to the data contained in the census itself, the Longitudinal Study is also linked to administrative data on life events, such as births to sample mothers, deaths and cancer registrations.

[8]The response rate in the most recent census was 94%. Data from the December 2012 ONS report "Response rates in the 2011 census".

[9]ONS report "Longitudinal Study 2001 - 2011: Completeness of census linkage"

[10]Prior to 1997, the ASHE was known as the New Earnings Survey (NES).

### 5.3.2 Earnings

To impute earnings for the LS sample members and their parents, I make use of the New Earnings Survey (NES) and Annual Survey of Hours and Earnings (ASHE). The ASHE, called NES prior to 1997, is based on a representative 1% sample of UK employees, covering around 300,000 individuals. It records earnings and hours worked for employees in all industries and occupations. Questionnaires are sent to and completed by employers each April and employees are followed through time, even when changing employer.

Each year, the Office for National Statistics publishes tables containing estimates of employee earnings broken down by gender and occupation based on the ASHE data.[11] For each gender and year, these tables contain the levels and distribution of earnings for more than 350 unique occupations as determined by 4-digit SOC codes. This occupational coding is identical to that used in the 1991, 2001 and 2011 censuses. This allows me to perfectly match census occupation data to occupation wage data at this very detailed level. I assign individuals in the LS the median earnings of workers of the same gender and detailed occupation in the same year in the ASHE/NES.

### 5.3.3 Sample selection and descriptives

The population census used in this paper observes individuals every ten years. To capture the period of quickest decline in the UK mining industry,[12] and be able to follow up children in adulthood, I focus on fathers who lost their job in mining between 1981 and 1991. The displaced or 'job loss' group consists of fathers who were working in mining in 1981, but were working in a different industry or were out of work in 1991. Fathers who were working in mining in 1981, and were still working in mining in 1991, make up the 'survivor' group

[11]These can be found on the ONS website as "Earnings and hours worked, occupation by four-digit SOC: ASHE Table 14".

[12]See Figure A1 in the Appendix for the rate of mine closures between 1970 and the present day.

used as the control group.

In order to get at the impact of paternal job loss in childhood - before the individual makes further education decisions - I restrict the sample to children born between 1973 (aged 18 in 1991) and 1981. Children born in this period will be aged between zero and eight in 1981, and between 10 and 18 when we observe them again in 1991. Sample sizes do not allow me to separately estimate results for each cohort, hence the main effect estimated will be the impact of father's job loss at any point before the age of 18.

This sample selection leaves an analysis sample of 733 father-child pairs. Of those father-child pairs there are 276 where the father still works in mining in 1991 (the 'survivor' sample), and the remaining 457 are in the 'job loss' sample. Table 5.1 shows some descriptives of the job loss and survivor samples in the first two columns. To allow us to see how the analysis sample compares to the overall population, the final column of the table shows these same descriptives for children who are born in the same years (1973-1981) as our analysis sample, but where the father did not work in mining.

Parents in our analysis sample are slightly younger than those in the the rest of the population with children of similar age. Fathers in the sample are also less likely to have a degree, consistent with mining being a relatively low skilled occupation. Mothers in the analysis sample are less likely to be in work, and the families are less likely to own their house than those of the general population. Comparing fathers who worked in mining and lost their job with those that still worked in mining in 1991, we see much smaller differences in terms of characteristics in 1981. This indicates that mining fathers who lost their job in this period do not seem to be a very select sample of all mining fathers. Table D.1 in the Appendix shows the result of a t-test of the difference in means between the groups, and only the small difference in father's age, and the share of families owning their house is significantly different between the two groups at the 5% level. To alleviate concerns of selection into the job loss group, I will

Table 5.1: Summary statistics in 1981 and 1991

| | Job loss | Survivor | Other |
|---|---|---|---|
| *Father outcomes 1981* | | | |
| Age | 31.5 | 30.5 | 33.4 |
| Degree | <0.04 | <0.04 | 0.06 |
| *Mother outcomes 1981* | | | |
| Age | 29.4 | 28.7 | 30.8 |
| Degree | <0.04 | <0.04 | 0.02 |
| Employed | 0.23 | 0.17 | 0.27 |
| *Family outcomes 1981* | | | |
| Number of children | 2.3 | 2.2 | 2.3 |
| Own house | 0.55 | 0.63 | 0.69 |
| Social renter | 0.32 | 0.29 | 0.25 |
| *Father outcomes 1991* | | | |
| Employed | 0.60 | 1.00 | 0.87 |
| Unemployed | 0.17 | 0.00 | 0.07 |
| Inactive | 0.23 | 0.00 | 0.06 |
| Long-term sick | 0.15 | 0.00 | 0.04 |
| Retired | 0.03 | 0.00 | 0.01 |
| Earnings | £14,150 | £29,470 | £23,910 |
| Earnings (excl £0) | £23,520 | £29,470 | £27,390 |
| *Mother outcomes 1991* | | | |
| Earnings | £8,980 | £10,010 | £11,690 |
| Employed | 0.58 | 0.66 | 0.66 |
| *Family outcomes 1991* | | | |
| Own house | 0.74 | 0.86 | 0.82 |
| Social renter | 0.22 | 0.12 | 0.14 |
| N | 457 | 276 | 36718 |

Notes: Table shows summary statistics for children born between 1973 and 1981, for whom we observe parent outcomes in both 1981 and 1991. This sample is split into three groups: the "Job loss" column shows descriptives for children whose father lost his job in mining between 1981 and 1991; the "Survivor" column shows descriptives for children whose father worked in mining in both 1981 and 1991, the "Other" column gives descriptives of individuals where the father did not work in mining in 1981 and 1991. The individuals in the "Other" group are not used in our analysis, but descriptives are shown for comparison purposes. All figures are in percentage terms except age and earnings. Earnings are mean annual earnings in 2018 £s and includes zeros unless noted otherwise. Earnings are assigned based on the median wage in the individual's occupation for each gender. For statistical disclosure reasons, descriptives are not given when underlying sample sizes are too small.
Data source: ONS LS

control for all these pre-displacement characteristics in the estimation.

Ten years later, after displacement, we see large differences between the groups. Only 60% of fathers who lost their job in mining between 1981 and 1991 are in work in 1991. Just under half of those out of work are unemployed, while the rest has left the labour market all together, often reporting long-term sickness. This is consistent with previous work (Beatty and Fothergill, 1996) that has shown that many miners leave the labour force and end up on (disability) benefits or take early retirement after lay-off. Earnings of displaced fathers, even among those in work, are much lower than the earnings of those still working in mining. These high rates of unemployment and the loss of earnings among the displaced group indicate that we are indeed identifying involuntary lay offs rather than individuals voluntarily quitting the mining industry.

## 5.4 Results

### 5.4.1 Direct effect of job displacement

Before looking at the impact of paternal job loss on child outcomes, I first estimate the impact this job loss had on family resources. In particular, I look at the impact on earnings and employment of those affected and their spouses, as well as on housing tenure. Table 5.2 shows the results of this estimation. The first column shows the results of a regression of the outcomes given in each row on a an indicator for job loss. The second column shows the same coefficient when we include controls for education and age of the parent, region of residence and housing tenure in 1981, to try and account for any pre-exisiting differences between the job loss and survivor samples.

The loss of a job in mining leads to a 35 percentage point drop in employment. The size of this drop in employment is particularly noteworthy if we consider that many of these individuals will have been laid off a number of years before the point at which we measure their outcomes.[13] Table D.2 in the Appendix looks at a wider range of economic activity outcomes and shows that more than half of this reduction in employment is accounted for by fathers exiting the labour market altogether, mostly driven by an increase in the rate of long term sickness. This is in line with findings from Beatty and Fothergill (1996) who found that many miners end up on disability benefits after being laid off. They also found many displaced miners entering early retirement, but this seems to be much less important in our sample, which is still relatively young in 1991.[14] In line with the employment impacts, displacement also leads to a large negative earnings shock of nearly £14,000, equivalent to around half of the earnings of the non-displaced group.

After controlling for parent characteristics, I do not find a significant impact of father's job loss on the earnings and employment of their spouse, contrary to what theory would predict (Ashenfelter, 1980). Closure of the local mine - often the main employer in the area - leads not just to a loss of employment for the men working in the mine, but also a worsening of the local labour market conditions in the area. Under these conditions, it will be very difficult for women to increase their labour supply and earnings to offset the loss of income from their spouse. The large and persistent effects on fathers' earnings together with the lack of increased earnings for women suggest a substantial reduction in household resources following paternal job loss.

### 5.4.2 Intergenerational effects of job displacement

The previous section established that paternal job loss due to mining closures led to a large and persistent negative shock to family resources. This section will first explore the overall impact on paternal job loss on child outcomes, and

[13]The rate of mine closures between 1981 and 1991 was relatively constant, as can be seen in Figure D.1. This implies that fathers who lost their job in mining between 1981 and 1991 will have lost their job on average around 5 years before we observe then again in 1991.

[14]The average age of fathers in our sample in 1991 is 41.

Table 5.2: Effects of father's job loss on family outcomes 1991

| | (1)<br>No controls | (2)<br>Controls |
|---|---|---|
| Father employed | −0.395*** | −0.347*** |
| | (0.030) | (0.030) |
| Father earnings | −15318.277*** | −13746.054*** |
| | (783.386) | (771.108) |
| Mother employed | −0.081** | −0.041 |
| | (0.037) | (0.038) |
| Mother earnings | −1033.574 | −261.871 |
| | (632.375) | (622.259) |
| Own house | −0.120*** | −0.084*** |
| | (0.031) | (0.029) |
| N | 733 | 733 |

Notes: The table shows the coefficients on father's job loss from a regression of the outcome listed in each row on job loss. Estimates under (1) do not include any controls, estimates under (2) include controls for education and age of the parent, region of residence, number of children and housing tenure in 1981. Earnings are annual earnings in 2018 £s. Standard errors are in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% level.
Data source: ONS LS

will then try and disentangle how much of this overall impact can be explained by the loss in family resources.

Table 5.3 shows the estimated impact of paternal job loss on the outcomes of their children in adulthood, measured in 2011 when individuals are between 30 and 38 years old. As before, the table first shows the estimates without any controls, then the estimates once we control for family characteristics pre-displacement. To allow for heterogeneous impacts by gender, I run the estimation entirely separately for men and women.

Looking first at daughters, we find a large and substantial decrease in educational attainment. Women are 13 percentage points (or around 33%) less likely to obtain a degree after experiencing paternal job loss in childhood.[15] Despite the large negative impact of paternal job loss on educational achievement, we do not find an impact on employment and once we control for pre-determined characteristics, only a small and insignificant negative impact on earnings. The probability of owning a home decreases substantially after experiencing paternal job loss. Appendix Table D.3 shows a concordant rise in the probability of living in socially rented accomodation, which is only available to those on low incomes. In light of these findings, and the reduction in educational attainment combined with the high female returns to higher education in the UK (Blundell et al., 1999; Belfield et al., 2018a), the lack of any significant reduction of earnings is surprising. One potential explanation of this may be the small sample sizes and the measurement error introduced in the earnings measure by having to impute earnings from occupation.

For men, unlike for women, we do not find any significant impacts of paternal job loss on any of our outcomes. This aligns with previous work (Rege et al., 2011; Pinger, 2016) which has found stronger impacts of parental job loss on the

[15]Table D.3 in the Appendix shows the results for a wider set of outcomes, including age 16 and 18 qualifications. I use obtaining a degree as my main educational outcome as the sample consists of individuals who experienced paternal job loss between the ages of 8 and 18. Age 16 and 18 qualifications would have been achieved before the paternal job loss for some of this sample.

Table 5.3: Effects of father's job loss on child outcomes in 2011 - by gender

| | Women | | Men | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | No controls | Controls | No controls | Controls |
| Has degree | −0.162*** | −0.132** | 0.006 | 0.032 |
| | (0.052) | (0.052) | (0.046) | (0.046) |
| Employed | −0.042 | 0.007 | −0.024 | −0.020 |
| | (0.051) | (0.052) | (0.040) | (0.042) |
| Earnings | −2176.182 | −507.705 | −388.619 | −221.445 |
| | (1606.830) | (1617.042) | (1540.218) | (1573.353) |
| Own house | −0.147*** | −0.117** | 0.000 | 0.017 |
| | (0.054) | (0.054) | (0.051) | (0.052) |
| N | 319 | 319 | 315 | 315 |

Notes: The table shows the coefficients on father's job loss from a regression of the outcome listed in each row on job loss. Estimates under (1) do not include any controls, estimates under (2) include controls for education and age of the parents, region of residence, number of siblings and housing tenure in 1981. Earnings are annual earnings in 2018 £s. Standard errors are in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% level.
Data source: ONS LS

educational attainment of daughters than of sons. In our context, this difference may also be explained by the presence of two potential counteracting effects for men. For women, the main implications of local mine closures and paternal job loss are the loss in family resources, and any other effects of job loss such as psychological impacts on the family. For men on the other hand, the local mine closures also imply that these low-skilled and highly paid mining jobs are now not available to them anymore, which increases their returns to education.

### 5.4.3 Mechanisms

Job loss can affect children's outcomes through its impact on household resources as well as through non-financial channels, such as increases in parental depression and anxiety (Eales, 1988; Kuhn et al., 2009). While we have no way of directly measuring non-financial impacts of job loss in the current data, Section 5.4.1 did show the large reduction in family resources following paternal

Table 5.4: Effects of father's job loss and earnings on child outcomes

| | Women (1) Main controls | (2) + 1991 father earnings |
|---|---|---|
| Has degree | $-0.132$** | $-0.0604$ |
| | (0.052) | (0.0630) |
| Own house | $-0.117$** | $-0.097$ |
| | (0.054) | (0.066) |
| N | 319 | 319 |

Notes: The table shows the coefficients on father's job loss from a regression of the outcome listed in each row on job loss. Estimates under (1) include our main controls: education and age of the parents, region of residence, number of siblings and housing tenure in 1981. Estimates under (2) add father's earnings in 1991. Standard errors are in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% level.
Data source: ONS LS

job loss. In this section I explore how much of the overall impact of job loss on child outcomes can be accounted for by this large negative financial shock. If the impact of job loss on child outcomes were purely caused by the loss in family income following displacement, we would expect the inclusion of controls for family income to drive the coefficients on job loss to zero. This exercise will necessarily be quite tentative. Family income post displacement may be endogenous to child outcomes, for example if more motivated fathers are more likely to find a job soon after displacement. Furthermore, controlling for earnings in 1991 will be multiple years after displacement for some families. The shorter-term loss in family resources may well have been higher, which I am not able to control for. Nevertheless, this exercise can give us some indication of the importance of the reduction in family resources associated with job loss for explaining the overall impact of paternal job loss on children.

Table 5.4 shows the previously estimated coefficients on job loss, as well as the coefficient when we additionally control for father's earnings in 1991, post displacement. Estimates are shown for the subset of child outcomes for which we found significant effects of paternal job loss - namely the share of daughters with a degree and who own a home. Estimates for other child outcomes are shown in Appendix Table D.4. The inclusion of controls for post-displacement earnings reduces our estimates of the effect of job loss on child outcomes, rendering them statistically insignificant from zero. Looking at the point estimates, the loss in income can explain around half of the drop in degree attainment, and 20% of the drop in home ownership. These findings suggest the loss in family resources play an important role in accounting for the overall impact of job loss. This aligns with findings from related work in Canada (Oreopoulos and Stevens, 2008), though previous work in Norway found the job loss impact on children's school performance to be largely unrelated to the loss in family income (Rege et al., 2011). The difference between the Norwegian findings and those of this paper and Oreopoulos and Stevens (2008) may be explained by the differences between the Norwegian welfare system and that in Canada and the UK. The Norwegian welfare system - one of the most generous in the world - may reduce the importance of parental investments in children's human capital and the loss in earnings following displacement.

## 5.5 Conclusion

In this paper I exploit the closure of UK mines in the late twentieth century to provide new evidence on the impact of parental job loss on children's long term outcomes. I find substantial negative impacts of paternal job loss on daughter's later life outcomes, though not for sons. Women are 13 percentage points (around 33%) less likely to have a degree after paternal job displacement, and are 12 percentage points less likely to own their home. These impacts of paternal job loss during childhood are long-lasting and persist until at least individuals' thirties. I provide suggestive evidence that the loss in family resources explains around half of the overall effects of paternal job loss on educational attainment and 20% of the effect on home ownership. This suggests unemployment insurance may be an important tool for policy makers to mitigate the negative impacts of parental job loss on children.

The results presented in this paper are estimated using mass layoffs of miners in the UK. To the extent that the impact of job loss on children might differ across groups or countries, caution need to be exercised when applying these estimates to other populations.

# Bibliography

**Abowd, John M., Francis Kramarz, and David N. Margolis**, "High Wage Workers and High Wage Firms," *Econometrica*, 1999, *67* (2), 251–333.

**Acciari, Paolo, Alberto Polo, and Giovanni L. Violante**, ""And Yet, It Moves": Intergenerational Mobility in Italy," *IZA Discussion Papers*, 2019, p. No. 12273.

**Aguiar, Mark and Erik Hurst**, "Consumption versus Expenditure," *Journal of Political Economy*, 2005, *113* (5), 919–948.

**Altonji, Joseph G, Erica Blom, and Costas Meghir**, "Heterogeneity in human capital investments: High school curriculum, college major, and careers," *Annu. Rev. Econ.*, 2012, *4* (1), 185–223.

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

**Anelli, Massimo**, "Returns to elite university education: a quasi-experimental analysis," *Journal of the European Economic Association*, 2018.

**Ashenfelter, Orley**, "Unemployment as Disequilibrium in a Model of Aggregate Labor Supply," *Econometrica*, 1980, *48* (3), 547–564.

**Beatty, Christina and Stephen Fothergill**, "Labour Market Adjustment in Areas of Chronic Industrial Decline: The Case of the UK Coalfields," *Regional Studies*, 1996, *30* (7), 627–640.

**Belfield, Chris, Jack Britton, Franz Buscha, Lorraine Dearden, Matt Dickson, Laura van der Erve, Luke Sibieta, Anna Vignoles, Ian Walker, and Yu Zhu**, "The impact of undergraduate degrees on early-career earnings," *Department for Education Report*, 2018.

**_ , _ , _ , _ , _ , _ , _ , _ , _ , and _** , "The returns of different subjects and university choices in England," *Department for Education Report*, 2018.

**Bell, Brian, Jack Blundell, and Stephen Machin**, "The Changing Geography of Intergenerational Mobility," *CEP Discussion Paper*, 2018, p. No. 1591.

**Betts, Julian R. and Laurel L. McFarland**, "Safe Port in a Storm: The Impact of Labor Market Conditions on Community College Enrollments," *Journal of Human Resources*, 1995, *30* (4), 741–765.

**Black, Dan A and Jeffrey A Smith**, "Estimating the returns to college quality with multiple proxies for quality," *Journal of labor Economics*, 2006, *24* (3), 701–728.

**Black, Dan A., Terra G. McKinnish, and Seth G. Sanders**, "Tight Labor Markets and the Demand for Education: Evidence from the Coal Boom and Bust," *Industrial and Labor Relations Review*, 2005, *59* (1), 3–16.

**Black, Dan, Kermit Daniel, and Seth Sanders**, "The Impact of Economic Conditions on Participation in Disability Programs: Evidence from the Coal Boom and Bust," *American Economic Review*, March 2002, *92* (1), 27–50.

**Blanden, J., A. Goodman, P. Gregg, and S. Machin**, "Changes in intergenerational mobility in Britain," *In M. Corak (Ed.), Generational income mobility in North America and Europe, Cambridge University Press*, 2004, p. pp. 122–46.

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

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

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

—, —, **Costas Meghir, and Barbara Sianesi**, "Human Capital Investment: The Returns from Education and Training to the Individual, the Firm and the Economy," *Fiscal Studies*, 1999, *20* (1), 1–23.

**Britton, Jack, Laura van der Erve, Ben Waltmann, and Xiaowei Xu**, "The impact of living costs on the returns to higher education," *Department for Education Report*, 2021.

—, —, —, **and** —, "London calling? Higher education, geographical mobility and early-career earnings," *Department for Education Report*, 2021.

—, —, **Chris Belfield, Franz Buscha, Lorraine Dearden, Matt Dickson, Luke Sibieta, Anna Vignoles, Ian Walker, and Yu Zhu**, "How much does degree choice matter?," *IFS working paper W21/24*, 2021.

—, **Lorraine Dearden, and Ben Waltmann**, "The returns to undergraduate degrees by socio-economic group and ethnicity," *Department for Education Report*, 2021.

—, —, **Neil Shephard, and Anna Vignoles**, "Is improving access to university enough? Socio-economic gaps in the earnings of English graduates," *Oxford Bulletin of Economics and Statistics*, 2019, *81* (2), 328–368.

**Broecke, Stijn**, "University selectivity and earnings: Evidence from UK data on applications and admissions to university," *Economics of Education Review*, 2012, *31* (3), 96–107.

**Brown, Jennifer and David A. Matsa**, "Boarding a Sinking Ship? An Investigation of Job Applications to Distressed Firms," *The Journal of Finance*, 2016, *71* (2), 507–550.

**Buchmueller, Gerda and Ian Walker**, "The Graduate Wage and Earnings Premia and the Role of Non-Cognitive Skills," 2020.

**Campbell, Stuart, Lindsey Macmillan, Richard Murphy, and Gill Wyness**, "Inequalities in student to course match: evidence from linked administrative data," Discussion Paper CEPDP1647, Centre for Economic Performance, 2019.

**Cascio, Elizabeth U and Ayushi Narayan**, "Who Needs a Fracking Education? The Educational Response to Low-Skill Biased Technological Change," Working Paper 21359, National Bureau of Economic Research July 2015.

**Charles, Kerwin and Melvin Stephens**, "Job displacement, disability and divorce," *Journal of Labor Economics*, 2004, *22* (2), 489–522.

**Chetty, Raj and Adam Szeidl**, "Consumption Commitments and Risk Preferences," *The Quarterly Journal of Economics*, 2007, *122* (2), 831–877.

_ **and Nathaniel Hendren**, "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates*," *The Quarterly Journal of Economics*, 2018, *133* (3), 1163–1228.

_ **, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan**, "Mobility report cards: the role of colleges in intergenerational mobility," Working Paper 23618, National Bureau of Economic Research, 2017.

— , — , — , — , **and** — , "Income segregation and intergenerational mobility across colleges in the United States," *Quarterly Journal of Economics*, 2020, *135* (3), 1567–1633.

— **, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez**, " Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States *," *The Quarterly Journal of Economics*, 2014, *129* (4), 1553–1623.

**Chevalier, Arnaud**, "Subject choice and earnings of UK graduates," *Economics of Education Review*, 2011, *30* (6), 1187–1201.

— **and Gavan Conlon**, "Does it pay to attend a prestigious university?," 2003.

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

**Clark, Damon**, "Do Recessions Keep Students in School? The Impact of Youth Unemployment on Enrolment in Post-compulsory Education in England," *Economica*, 2011, *78* (311), 523–545.

**Coelli, Michael B.**, "Parental job loss and the education enrollment of youth," *Labour Economics*, January 2011, *18* (1), 25–35.

**Corak, Miles**, "Income inequality, equality of opportunity, and intergenerational mobility," *Journal of Economic Perspectives*, 2013, *27* (3), 79–102.

— , "The Canadian Geography of Intergenerational Income Mobility," *The Economic Journal*, 2019, *130* (631), 2134–2174.

**Couch, Kenneth A and Dana W Placzek**, "Earnings Losses of Displaced Workers Revisited," *American Economic Review*, March 2010, *100* (1), 572–589.

**Crawford, Claire**, "The link between secondary school characteristics and university participation and outcomes," CAYT Research Report, Department for Education, 2014a.

— , "Socio-economic differences in university outcomes in the UK: drop-out, degree completion and degree class," Working Paper 14/31, Institute for Fiscal Studies, 2014b.

— **and Ellen Greaves**, "Socio-economic, ethnic and gender differences in HE participation," BIS Research Paper, 2015.

— **, Lorraine Dearden, John Micklewright, and Anna Vignoles**, *Family Background and University Success: Differences in Higher Education Access and Outcomes in England*, Oxford University Press, 2016a.

— **, Paul Gregg, Lindsey Macmillan, Anna Vignoles, and Gill Wyness**, "Higher education, career opportunities, and intergenerational inequality," *Oxford Review of Economic Policy*, 2016b, *32* (4), 553–575.

**Cunha, Jesse M and Trey Miller**, "Measuring value-added in higher education: Possibilities and limitations in the use of administrative data," *Economics of Education Review*, 2014, *42*, 64–77.

**Dale, Stacy B and Alan B Krueger**, "Estimating the effects of college characteristics over the career using administrative earnings data," *Journal of human resources*, 2014, *49* (2), 323–358.

**Dale, Stacy Berg and Alan B Krueger**, "Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables," *The Quarterly Journal of Economics*, 2002, *117* (4), 1491–1527.

**DfE**, "Widening participation in Higher Education," *Department for Education annual statistics*, 2021.

**Dillon, Eleanor Wiske and Jeffrey Andrew Smith**, "The consequences of academic match between students and colleges," *Journal of Human Resources*, 2020, *55* (3), 767–808.

**Donnelly, Michael and Sol Gamsu**, "Home and Away: Social, ethnic and spatial inequalities in student mobility," *Sutton Trust Report*, 2018.

**Eales, M. J.**, "Depression and anxiety in unemployed men," *Psychological Medicine*, 1988, *18* (4), 935–945.

**Eliason, Marcus and Donald Storrie**, "Lasting or latent scars? Swedish evidence on the long-term effects of Job Displacement," *Journal of Labor Economics*, 2006, *24* (4), 343–375.

**Fradkin, Andrey, Frédéric Panier, and Ilan Tojerow**, "Blame the Parents? How Parental Unemployment Affects Labor Supply and Job Quality for Young Adults," *Journal of Labor Economics*, January 2019, *37* (1), 35–100.

**Gibbons, Stephen, Eric Neumayer, and Richard Perkins**, "Student satisfaction, league tables and university applications: evidence from Britain," *Economics of Education Review*, 2015, *48*, 148–164.

**Glyn, Andrew**, *The Economic Case Against Pit Closure*, Cambridge University Press,

_ **and Stephen Machin**, "Colliery Closures and the Decline of the UK Coal Industry," *British Journal of Industrial Relations*, 1997, *35* (2), 197–214.

**Gregg, P., L. Macmillan, and C. Vittori**, "Moving towards estimating sons' lifetime intergenerational economic mobility in the UK," *Oxford Bulletin of Economics and Statistics*, 2017, *79* (1), 79–100.

**Hastings, Justine, Christopher A Neilson, and Seth D Zimmerman**, "The effects of earnings disclosure on college enrollment decisions," Technical Report, National Bureau of Economic Research 2018.

**Hastings, Justine S, Christopher A Neilson, and Seth D Zimmerman**, "Are some degrees worth more than others? Evidence from college admission cutoffs in Chile," Technical Report, National Bureau of Economic Research 2013.

**Heidrich, S.**, "Intergenerational mobility in Sweden: a regional perspective.," *Journal of Population Economics*, 2017, *20*, 1241–1280.

**Hilger, Nathaniel G.**, "Parental Job Loss and Children's Long-Term Outcomes: Evidence from 7 Million Fathers' Layoffs," *American Economic Journal: Applied Economics*, July 2016, *8* (3), 247–283.

**Hoekstra, Mark**, "The effect of attending the flagship state university on earnings: A discontinuity-based approach," *The Review of Economics and Statistics*, 2009, *91* (4), 717–724.

**Hoxby, Caroline M**, "The productivity of us postsecondary institutions," in "Productivity in Higher Education," University of Chicago Press, 2018.

**Huettner, Frank and Marco Sunder**, "Axiomatic arguments for decomposing goodness of fit according to Shapley and Owen values," *Electronic Journal of Statistics*, 2012, *6*, 1239–1250.

**Hussain, Iftikhar, Sandra McNally, and Shqiponja Telhaj**, "University quality and graduate wages in the UK," 2009.

**Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan**, "Earnings Losses of Displaced Workers," *The American Economic Review*, 1993, *83* (4), 685–709.

**Kalil, Ariel and Kathleen M. Ziol-Guest**, "Parental employment circumstances and children's academic progress," *Social Science Research*, June 2008, *37* (2), 500–515.

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

**Kuhn, Andreas, Rafael Lalive, and Josef Zweimüller**, "The public health costs of job loss," *Journal of Health Economics*, 2009, *28* (6), 1099–1115.

**Mork, Eva, Anna Sjogren, and Helena Svaleryd**, "Parental Unemployment and Child Health," *CESifo Economic Studies*, 04 2014, *60* (2), 366–401.

**Mountjoy, Jack and Brent Hickman**, "The Returns to College (s): Estimating Value-Added and Match Effects in Higher Education," *University of Chicago, Becker Friedman Institute for Economics Working Paper*, 2020, (2020-08).

**OECD**, "Education at a Glance 2014: OECD Indicators," 2014.

**Ost, Ben, Weixiang Pan, and Douglas Webber**, "The returns to college persistence for marginal students: Regression discontinuity evidence from university dismissal policies," *Journal of Labor Economics*, 2018, *36* (3), 779–805.

**Page, Marianne, Ann Huff Stevens, and Jason Lindo**, *Parental Income Shocks and Outcomes of Disadvantaged Youth in the United States*, University of Chicago Press, April

**Philip, Marianne Page Oreopoulos and Ann Huff Stevens**, "The intergenerational effects of worker displacement," *Journal of Labor Economics*, 2008, *26* (3), 455–483.

**Pinger, Pia**, "Understanding the mechanisms behind intergenerational effects of economic distress," *University of Bonn Working Paper*, 2016.

**Rege, Mari, Kjetil Telle, and Mark Votruba**, "Parental Job Loss and Children's School Performance," *The Review of Economic Studies*, 2011, *78* (4), 1462–1489.

**Rothstein, Jesse**, "Inequality of Educational Opportunity? Schools as Mediators of the Intergenerational Transmission of Income," *Journal of Labor Economics*, 2019, *37* (S1), S85–S123.

**Saavedra, Juan E**, "The returns to college quality: A regression discontinuity analysis," 2008.

**Schaller, Jessamyn and Mariana Zerpa**, "Short-Run Effects of Parental Job Loss on Child Health," *American Journal of Health Economics*, 2019, *5* (1), 8–41.

**Stevens, Ann Huff and Jessamyn Schaller**, "Short-run effects of parental job loss on children's academic achievement," *Economics of Education Review*, April 2011, *30* (2), 289–299.

**Sullivan, Daniel and Till von Wachter**, "Job Displacement and Mortality: An Analysis Using Administrative Data*," *The Quarterly Journal of Economics*, 08 2009, *124* (3), 1265–1306.

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

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

**Webber, Douglas A**, "Are college costs worth it? How ability, major, and debt affect the returns to schooling," *Economics of Education Review*, 2016, *53*, 296–310.

**Wiswall, Matthew and Basit Zafar**, "Determinants of college major choice: Identification using an information experiment," *The Review of Economic Studies*, 2014, *82* (2), 791–824.

**Zimmerman, Seth D**, "The returns to college admission for academically marginal students," *Journal of Labor Economics*, 2014, *32* (4), 711–754.

—, "Elite colleges and upward mobility to top jobs and top incomes," *American Economic Review*, 2019, *109* (1), 1–47.

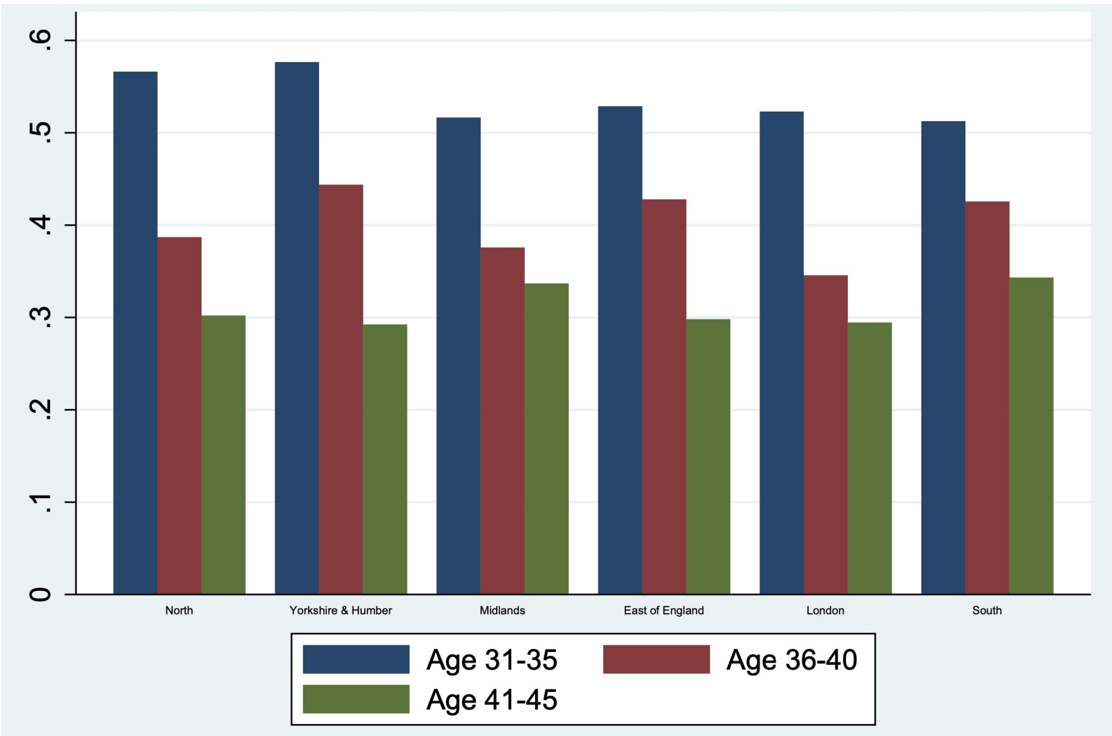
# Appendices

# Appendix A

# Appendix to Chapter 2

## A.1 Correlation of age 28 with later earnings by region

Figure A.1: Correlation of age 28 with later earnings by region for women




Notes: Data from the British Household Panel Survey.

## A.2 Top and bottom areas for mobility

Table A.1: Top 10 areas for mobility - women

| | Region | $R_c^{FSM}$ | $P(Q5\|FSM)$ |
|---|---|---|---|
| Redbridge | London | 41.9 | 0.14 |
| Kensington and Chelsea | London | 41.2 | 0.14 |
| Tower Hamlets | London | 40.7 | 0.13 |
| Ealing | London | 40.5 | 0.13 |
| Harrow | London | 40.5 | 0.15 |
| West Berkshire | South East | 40.4 | 0.11 |
| Hackney | London | 40.2 | 0.12 |
| Wandsworth | London | 40.2 | 0.12 |
| Newham | London | 40.1 | 0.12 |
| Bexley | London | 39.9 | 0.15 |

Notes: Table shows, for the 10 areas with the highest average income rank, average income rank and the share who reached the top 20% of income, for individuals on FSM at age 16. Individuals are assigned to the area they lived at age 16. Only areas with at least 100 children on FSM in our analysis sample are shown. We assign children with zero income the average income rank of that group.

Table A.2: Top 10 areas for mobility - men

| | Region | $R_c^{FSM}$ | $P(Q5\|FSM)$ |
|---|---|---|---|
| Havering | London | 53.0 | 0.25 |
| Wokingham | South East | 50.9 | 0.17 |
| Bracknell Forest | South East | 50.3 | 0.12 |
| Kingston upon Thames | London | 50.0 | 0.22 |
| Hillingdon | London | 49.9 | 0.20 |
| Tower Hamlets | London | 49.8 | 0.19 |
| Surrey | South East | 49.3 | 0.17 |
| Sutton | London | 49.3 | 0.20 |
| Barking and Dagenham | London | 49.3 | 0.19 |
| Hertfordshire | East | 49.2 | 0.19 |

Notes: See footnotes to Table A.1.

Table A.3: Bottom 10 areas for mobility - women

| | Region | $R_c^{FSM}$ | $P(Q5\|FSM)$ |
|---|---|---|---|
| Lincolnshire | East Mid | 27.8 | 0.03 |
| County Durham | North East | 27.8 | 0.02 |
| Stoke-on-Trent | West Mid | 27.6 | 0.02 |
| Wakefield | Yorkshire | 27.5 | 0.02 |
| Middlesbrough | North East | 27.1 | 0.03 |
| Nottingham | East Mid | 27.1 | 0.03 |
| Stockton-on-Tees | North East | 26.5 | 0.02 |
| North East Lincolnshire | Yorkshire | 26.0 | 0.03 |
| Barnsley | Yorkshire | 25.3 | 0.01 |
| Kingston upon Hull, City of | Yorkshire | 25.1 | 0.02 |

Notes: Table shows, for the 10 areas with the lowest average income rank, average income rank and the share who reached the top 20% of income, for individuals on FSM at age 16. Individuals are assigned to the area they lived at age 16. Only areas with at least 100 children on FSM in our analysis sample are shown. We assign children with zero income the average income rank of that group.

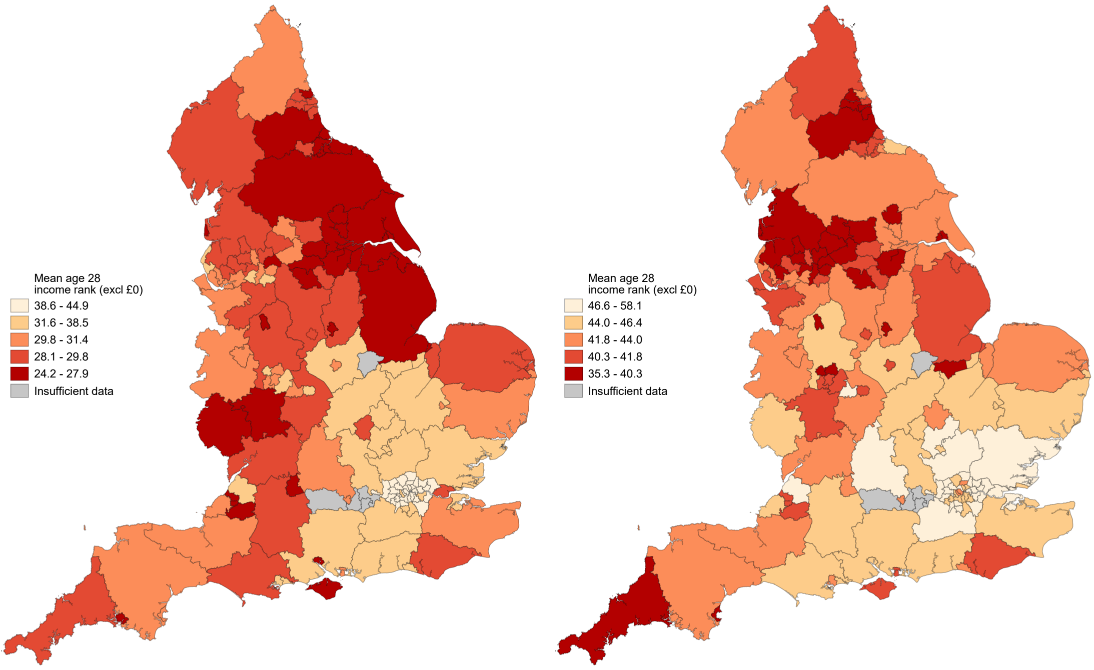
Table A.4: Bottom 10 areas for mobility - men

| | Region | $R_c^{FSM}$ | $P(Q5\|FSM)$ |
|---|---|---|---|
| Calderdale | Yorkshire | 38.3 | 0.05 |
| Newcastle upon Tyne | North East | 38.2 | 0.07 |
| County Durham | North East | 38.2 | 0.07 |
| Bolton | North West | 38.0 | 0.07 |
| Manchester | North West | 37.5 | 0.06 |
| Bradford | Yorkshire | 37.1 | 0.06 |
| Gateshead | North East | 36.8 | 0.05 |
| Blackpool | North West | 36.8 | 0.05 |
| Sheffield | Yorkshire | 35.9 | 0.05 |
| Nottingham | East Mid | 35.8 | 0.06 |

Notes: See footnotes to Table A.3.

## A.3 Mobility rates across local authorities when excluding zero earnings

Figure A.2: Average earnings rank age 28 excluding £0 - by gender

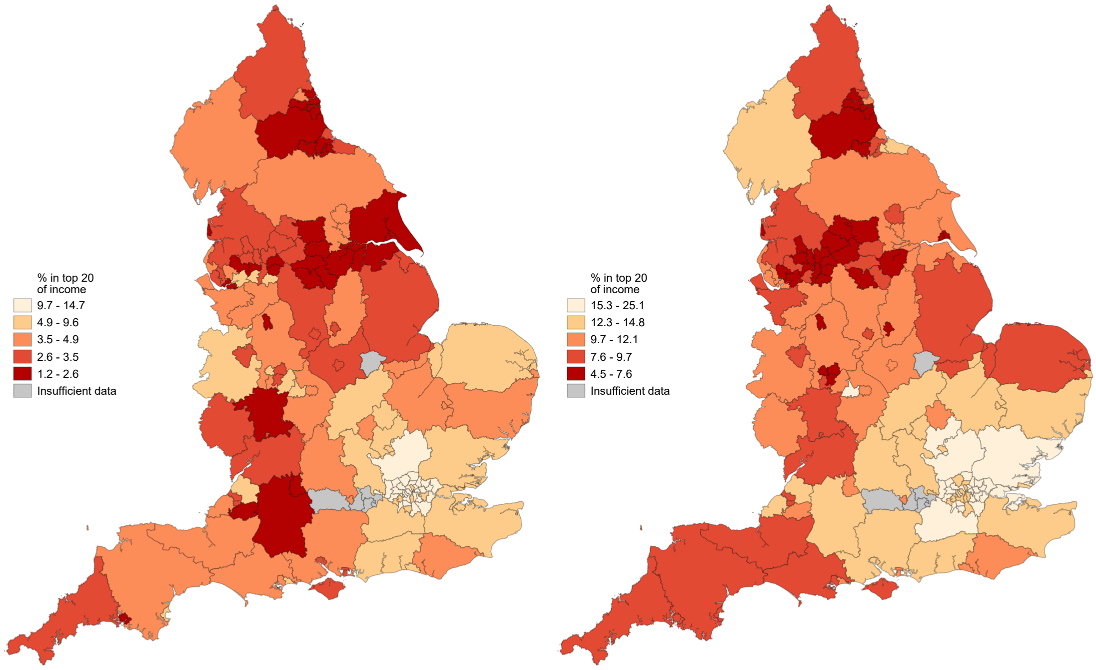



Women - excluding £0 | Men - excluding £0

Notes: Figure shows average earnings rank at age 28 of individuals who were on FSM at age 16. Individuals are assigned to the area they lived at age 16. Only areas with at least 100 children on FSM in our analysis sample are shown. Local authorities are split into quintiles based on the mean earnings rank of FSM children. The darkest colour shows the areas with the lowest average earnings rank. Individuals with zero earnings at age 28 are excluded.

## A.4 Probability for FSM children of reaching the top 20% across local authorities

Figure A.3: Probability of reaching top 20% of income for children on FSM

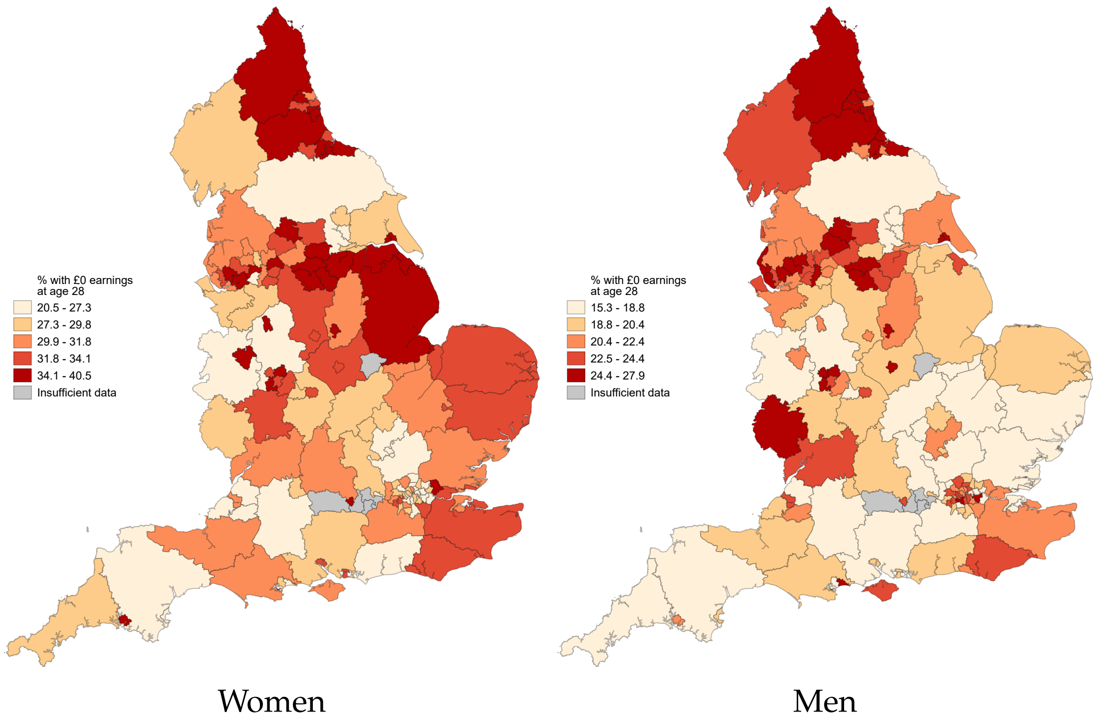



Women Men

Notes: Figure shows share of FSM individuals who reached the top 20% of the income distribution in their cohort at age 28. Only areas with at least 100 children on FSM in our analysis sample are shown. Local authorities are split into quintiles based on the share of individuals with zero earnings. The darkest colour shows the areas with the lowest share of individuals in the top 20%.

# A.5 Zero earnings rates across local authorities

Figure A.4: Zero earnings rates by area and gender




Notes: Figure shows share of FSM individuals with zero earned income recorded at age 28. Only areas with at least 100 children on FSM in our analysis sample are shown. Local authorities are split into quintiles based on the share of individuals with zero earnings. The darkest colour shows the areas with the highest share of individuals with zero earnings.

# A.6 Robustness

## A.6.1 Robustness to allowing returns to vary across regions

In our main specification for decomposing the variance of $R_{i,a}^{FSM}$, we assume the return to education, $\beta$ is constant across areas. In this section we relax that assumption and consider the robustness of our findings to allowing the return to education to vary across the area where an individual grew up. Instead of Equation 2.3 we estimate the equation below:

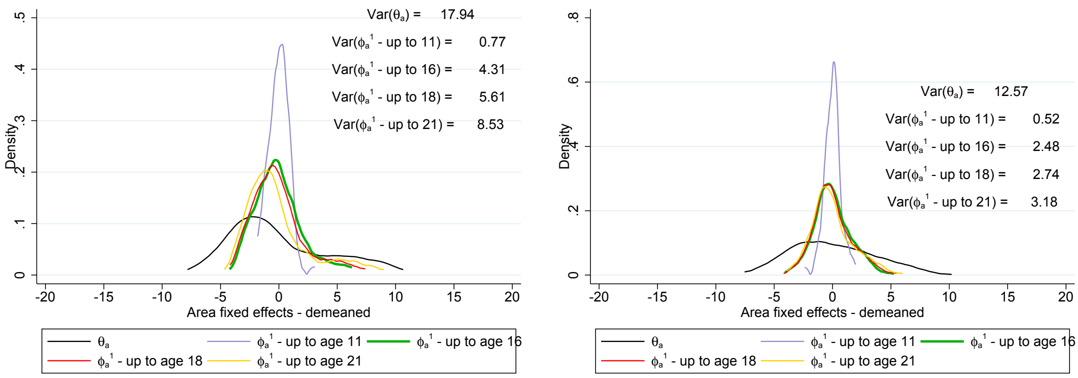
$$R_{i,a}^{FSM} = \beta_a H_{i,a}^{FSM} + \eta_a + w_{i,a} \tag{A.1}$$

This means that our decomposition becomes:

$$\overline{R}_a^{FSM} = (\beta_a \overline{H}_a^{FSM} + \eta_a) \tag{A.2}$$

To estimate how much the variance of $\overline{H}_a^{FSM}$ can explain of the overall variance, we hold $\beta_a$ and $\eta_a$ constant at the mean and allow only $\overline{H}_a^{FSM}$ to vary across areas. Figure A.5 shows the shape of the resulting distributions, both when our measure of human capital includes measures of educational attainment up to age 21, and when we only include attainment up to 11, 16 or 18.

Figure A.5: Decomposing the variance of $\overline{R}_a^{FSM}$




Women Men

Notes: Figure shows the distribution of area fixed effects $\theta_a$ and $\phi_a$. $\theta_a$ is equal to $\overline{R}_a^{FSM}$ demeaned. $\phi_a$ is equal to $(\beta \overline{H}_a^{FSM} + \eta)$ demeaned, where $\eta$ and $\beta$ are the averages of $\beta_a$ and $\eta_a$ across areas. '$\phi_a$ - up to 21' uses our main, most comprehensive, measure of human capital. The figure also shows the same decomposition when instead we only use measures of human capital up to ages 11, 16 or 18.

We estimate the variance of the resulting distributions of area effects to determine how much of the variance in mobility rates $R_{i,a}^{FSM}$ across areas can be explained by the variance of human capital $\overline{H}_a^{FSM}$ across areas. Table A.5 summarises the results from this exercise. The share of variance in mobility across areas which can be explained by differences in human capital across areas are virtually identical to those in our main specification, indicating the robubstness to our main findings to relaxing our assumption of constant returns to education.

Table A.5: Decomposition of $Var(\overline{R}_a^{FSM})$ - $\beta$ varying across area

| | **Share of** $Var(\overline{R}_a^{FSM})$ **Men** | **Women** |
|---|---|---|
| *HC index* | 0.25 | 0.48 |
| *HC index - up to age 11 only* | 0.04 | 0.04 |
| *HC index - up to age 16 only* | 0.20 | 0.24 |
| *HC index - up to age 18 only* | 0.22 | 0.31 |
| Total $Var(\overline{R}_a^{FSM})$ | 12.6 | 17.9 |

Notes: Table shows the variance of distribution where we only allow $\overline{H}_a^{FSM}$ to vary, as a proportion of total variance of $\overline{R}_a^{FSM}$. Local Authorities with fewer than 250 individuals included in the analysis are dropped from the analysis of that gender. Results are shown for multiple measures of human capital, using educational attainment up to age 11, and up to age 16, up to age 18 and up to age 21. Results are shown for men and women separately and human capital measures and area fixed effects are constructed completely separately by gender.

### A.6.2 Robustness to alternative definitions of human capital

Table A.6 shows the results of our decomposition when, instead of measuring human capital at each age as the predicted earnings rank based on educational achievement up to that point, we create human capital measures using a principal component analysis. We run a principal component analysis on educational achievement at each age, and take the first component of this analysis. The human capital index at each age is then constructed by taking the first component of a further PCA analysis, combining all de indices up to the given age. We can explain slightly less of the variation in mobility across areas by variation in human capital across areas, than using our main measure of human capital. as this measure is much less flexible than our main measure, this is unsurprising. The overall message remains the same however: the results point towards a meaningful role for improving educational achievement of low-income students in low mobility areas in order to equalise opportunities across the country, yet other channels, for example differences in labour market practices across areas, are clearly at least as important in explaining variation in mobility across the country. Educational achievement differences seem also more important in explaining mobility differences for women than for men. Again, the results highlight the importance of including a rich set of measures of educational achievement at different ages. Only including measures at younger ages considerably understates the proportion of variation in absolute mobility which can be explained by differences in educational achievement.

Table A.6: Decomposition of $Var(\overline{R}_a^{FSM})$ - different measures of HC

| | | **Share of** $Var(\overline{R}_a^{FSM})$ | |
|---|---|---|---|
| | | **Men** | **Women** |
| *PCA HC index* | $\beta^2 Var(\overline{H}_a^{FSM})$ | 0.24 | 0.36 |
| | $Var(\eta_a)$ | 0.83 | 0.32 |
| | $2\beta Cov(\overline{H}_a^{FSM}, \eta_a)$ | -0.07 | 0.32 |
| *PCA HC index - up to age 11 only* | $\beta^2 Var(\overline{H}_a^{FSM})$ | 0.03 | 0.04 |
| | $Var(\eta_a)$ | 0.92 | 0.92 |
| | $2\beta Cov(\overline{H}_a^{FSM}, \eta_a)$ | 0.04 | 0.04 |
| *PCA HC index - up to age 16 only* | $\beta^2 Var(\overline{H}_a^{FSM})$ | 0.09 | 0.11 |
| | $Var(\eta_a)$ | 0.79 | 0.63 |
| | $2\beta Cov(\overline{H}_a^{FSM}, \eta_a)$ | 0.12 | 0.26 |
| *PCA HC index - up to age 18 only* | $\beta^2 Var(\overline{H}_a^{FSM})$ | 0.15 | 0.21 |
| | $Var(\eta_a)$ | 0.79 | 0.45 |
| | $2\beta Cov(\overline{H}_a^{FSM}, \eta_a)$ | 0.06 | 0.34 |
| | Total $Var(\overline{R}_a^{FSM})$ | 12.6 | 17.9 |

Notes: Local Authorities with fewer than 250 individuals included in the analysis are dropped from the analysis of that gender. Results are shown for multiple measures of human capital, using educational attainment up to age 11, and up to age 16, up to age 18 and up to age 21. Results are shown for men and women separately and human capital measures and area fixed effects are constructed completely separately by gender.

## A.7 Multivariate correlations area characteristics and mobility

Table A.7: Conditional multivariate correlations of area characteristics and mobility (women)

| | Raw area effects | | | Controlling for education | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Strong labour market | -0.006 | 0.075 | 0.070 | 0.238 | 0.345** | 0.389** |
| | (0.088) | (0.088) | (0.102) | (0.155) | (0.159) | (0.192) |
| Good jobs | 0.262*** | 0.197** | 0.101 | 0.326* | 0.241 | 0.105 |
| | (0.097) | (0.095) | (0.109) | (0.172) | (0.172) | (0.204) |
| Immigration | 0.607*** | 0.518*** | 0.477*** | 0.174 | 0.058 | -0.058 |
| | (0.068) | (0.071) | (0.076) | (0.121) | (0.129) | (0.142) |
| Stable families | 0.269*** | 0.240*** | 0.157 | 0.146 | 0.108 | -0.126 |
| | (0.090) | (0.086) | (0.115) | (0.158) | (0.155) | (0.216) |
| Median earnings | 0.339*** | 0.142 | 0.182 | 0.370* | 0.112 | 0.190 |
| | (0.109) | (0.121) | (0.127) | (0.192) | (0.220) | (0.238) |
| Inequality | -0.253*** | -0.185*** | -0.129* | -0.472*** | -0.382*** | -0.294** |
| | (0.062) | (0.063) | (0.072) | (0.109) | (0.114) | (0.135) |
| Segregation by KS4 | 0.051 | 0.068 | 0.052 | 0.058 | 0.080 | 0.031 |
| score | (0.054) | (0.052) | (0.051) | (0.095) | (0.093) | (0.095) |
| Urban | -0.061 | -0.044 | -0.037 | -0.009 | 0.012 | 0.036 |
| | (0.059) | (0.057) | (0.059) | (0.105) | (0.103) | (0.110) |
| Index of dissim - | -0.103* | -0.072 | -0.015 | -0.189** | -0.148 | -0.040 |
| ethnicity | (0.053) | (0.051) | (0.056) | (0.093) | (0.093) | (0.105) |
| School quality | 0.045 | 0.043 | 0.032 | -0.048 | -0.051 | -0.031 |
| | (0.056) | (0.053) | (0.052) | (0.098) | (0.096) | (0.098) |
| London dummy | No | Yes | No | No | Yes | No |
| Region FE | No | No | Yes | No | No | Yes |
| R-squared | 0.841 | 0.858 | 0.884 | 0.546 | 0.572 | 0.629 |
| Adj R-squared | 0.822 | 0.839 | 0.854 | 0.490 | 0.514 | 0.532 |
| N | 93 | 93 | 93 | 93 | 93 | 93 |

Notes: Each column shows the coefficients from a multivariate regression of Local Authority level mobility measures on the area characteristics listed in the rows. Both mobility and area characteristics are standardized and coefficients can therefore be interpreted as correlations. Area characteristics are described in more detail in Appendix Section A.8.

Table A.8: Conditional multivariate correlations of area characteristics and mobility (men)

| | Raw area effects | | | Controlling for education | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Strong labour market | -0.022 | 0.119 | 0.016 | 0.017 | 0.142 | 0.065 |
| | (0.152) | (0.153) | (0.159) | (0.168) | (0.173) | (0.191) |
| Good jobs | 0.052 | -0.049 | -0.182 | -0.085 | -0.175 | -0.209 |
| | (0.166) | (0.162) | (0.164) | (0.183) | (0.183) | (0.197) |
| Immigration | 0.246** | 0.105 | -0.147 | -0.063 | -0.188 | -0.477*** |
| | (0.116) | (0.121) | (0.114) | (0.129) | (0.137) | (0.137) |
| Stable families | 0.547*** | 0.500*** | 0.305* | 0.547*** | 0.505*** | 0.261 |
| | (0.155) | (0.149) | (0.177) | (0.171) | (0.168) | (0.212) |
| Median earnings | 0.346* | 0.049 | 0.176 | 0.249 | -0.015 | 0.085 |
| | (0.175) | (0.196) | (0.179) | (0.193) | (0.222) | (0.215) |
| Inequality | -0.203* | -0.092 | -0.021 | -0.156 | -0.057 | 0.018 |
| | (0.107) | (0.110) | (0.111) | (0.119) | (0.124) | (0.133) |
| Segregation by KS4 | 0.026 | 0.053 | -0.046 | 0.054 | 0.078 | -0.030 |
| score | (0.095) | (0.092) | (0.079) | (0.105) | (0.103) | (0.095) |
| Urban | -0.076 | -0.047 | 0.065 | -0.065 | -0.040 | 0.069 |
| | (0.101) | (0.098) | (0.089) | (0.112) | (0.110) | (0.107) |
| Index of dissim - | -0.328*** | -0.280*** | -0.068 | -0.401*** | -0.358*** | -0.146 |
| ethnicity | (0.090) | (0.088) | (0.085) | (0.100) | (0.099) | (0.102) |
| School quality | 0.072 | 0.066 | 0.066 | -0.056 | -0.061 | -0.059 |
| | (0.094) | (0.090) | (0.078) | (0.105) | (0.102) | (0.093) |
| London dummy | No | Yes | No | No | Yes | No |
| Region FE | No | No | Yes | No | No | Yes |
| R-squared | 0.543 | 0.586 | 0.738 | 0.441 | 0.475 | 0.624 |
| Adj R-squared | 0.488 | 0.530 | 0.671 | 0.374 | 0.404 | 0.527 |
| N | 94 | 94 | 94 | 94 | 94 | 94 |

Notes: Each column shows the coefficients from a multivariate regression of Local Authority level mobility measures on the area characteristics listed in the rows. Both mobility and area characteristics are standardized and coefficients can therefore be interpreted as correlations. Area characteristics are described in more detail in Appendix Section A.8.

## A.8 Description of correlates

We use the following area characteristics from the 2001 population census in England:

- **Economically active**: the share of the population between the ages of 16 and 74 who are economically active.
- **Unemployment**: the share of the economically active population between the ages of 16 and 74 who were looking for work in the week preceding the census.

- **Professional jobs**: the share of the usual resident population who provided a valid occupation and are working in professional jobs according to the ONS' Social Class based on Occupation classification.
- **Manufacturing share**: the share of individuals aged 16 to 74 in employment with a valid industry of work, who work in the manufacturing industry.
- **Share foreign born:** the share of individuals who list a country outside of the United Kingdom as their country of birth.
- **Share Black, Asian and white**: the share of individuals who self-report as each ethnic group. "Asian" includes Indian, Pakistani, Bangladeshi, Chinese and other Asian.
- **% single parent families:** the share of families with dependent children where the head of the household is a single parent.
- **% of married families:** share of families with dependent children headed by a married couple.
- **Urban**: whether the area is classified as "urban" as defined by the ONS classification of areas.
- **Share rural**: the share of the population who live in rural areas (including large market towns), as defined by the settlement type and population density.

Aggregate data on annual gross pay in 2002 by Local Authority published by ONS based on the Annual Survey of Hours and Earnings (ASHE) allows us to construct measures of earnings and earnings inequality:

- **Median income**: 50th percentile of annual gross pay.

- **90-10 and 50-10 ratio**: ratio of the 90th (50th) percentile of gross annual pay in the area to the 10th percentile of gross annual pay in the same area. Due to disclosivity this data is not available for certain small areas. Those areas are excluded when estimating the correlation of mobility and the 90-10 or 50-10 ratios.

From the analysis dataset we obtain from the of students attending state school in England who took their GCSEs in 2002:

- **Index of dissimilarity for ethnicity and FSM**: measures the evenness with which white and non-white students (for ethnicity) and FSM and non-FSM students (for FSM) across state secondary schools within each local authority.

- **Segregation by KS4 score**: KS4 scores in the local authority among state school students are regressed on indicators for all local state secondary schools. As measure of segregation we take the degree of variance in KS4 scores explained by schools (R squared).

- **Average school value-added**: constructed for each school in the local authority by regressing overall KS4 scores on quadratics in KS2 maths, English and sciences scores, gender, FSM status, English as an Additional Language (EAL) status and ethnic group. Value added is then averaged over all state secondary schools in the local authority.

**The percentage of schools rated outstanding** is calculated based on the share of state schools in the area which were rated outstanding by Ofsted in 2002.

The multivariate analysis groups together similar measures into indices using principal component analysis:

- **Strong labour market**: combines 'Economically active', 'Unemployment'. Loads positively on the former and negatively on the latter.

- **Good jobs**: combines 'Professional jobs' and 'Manufacturing share'. Loads positively on the former and negatively on the latter.
- **Immigration**: combines 'Share foreign born', 'Share Black', 'Share Asian', 'Share white'. Loads negatively on 'Share white' and positively on the remaining three measures
- **Stable families**: combines '% single parent families' and '% of married families'. Loads negatively on the former and positively on the latter.
- **Inequality**: combines '90-10 ratio' and '50-10 ratio'. Loads positively on both measures.
- **School quality**: combines '% schools rated outstanding' and 'Avg school value added'. Loads positively on both measures.

# Appendix B

# Appendix to Chapter 3

## B.1 Sample selection

Table B.1 provides details of the LEO dataset, by GCSE cohort (based on the year these exams were taken, as discussed above). The first column shows all individuals with an age 16 GCSE record in the NPD who attended school in England.

In column 2 we drop some people who appear in the baseline sample whom we cannot use for our analysis. This is around 10% of the overall population and primarily consists of people with statemented special educational needs who were unable to take the examinations, people who are in the records but were not in Year 11 at school (for example, people who took some GCSE examinations early or did some retakes) and people with lots of missing background data or exam records. This leaves us with a 'usable sample' of between 520,000 and 600,000 individuals per cohort.

Table B.1: LEO sample by GCSE year

| | Population (1) | Non-missing NPD (2) | Linked (3) | Passed age 16 exams (4) |
|---|---|---|---|---|
| 2002 | 589,663 | 521,153 | 486,717 | 279,409 |
| 2003 | 621,929 | 566,279 | 531,139 | 296,365 |
| 2004 | 644,873 | 601,000 | 569,854 | 312,579 |
| 2005 | 644,345 | 601,300 | 572,970 | 320,643 |
| 2006 | 653,971 | 589,383 | 568,392 | 325,581 |
| 2007 | 662,225 | 598,641 | 577,184 | 332,322 |
| Total | 3,817,006 | 3,477,756 | 3,306,256 | 1,866,899 |

Note: Column 1 is the full sample of English domiciled pupils in the NPD. Column 2 excludes people with incomplete school records. Column 3 shows the number of those individuals who can be matched to the HMRC tax records. Column 4 shows the number of individuals who passed their age 16 exams (obtained at least five A*-C GCSE grades).

In column 3 we document the match rate to the HMRC tax data. Across the six cohorts around 95% of individuals are linked to the tax data, with match rates going up slightly across cohorts. Individuals never matching to the tax data means that there is never a record of them in the 11 years of tax or benefits data, or - more likely - because matching to the tax records was not possible due to incorrect or missing information.[1] The proportion of individuals who do not match to the tax data is approximately twice as large for women as it is for men, suggesting that women are more likely to never be in contact with the tax authorities. Aside from this gender difference, we essentially treat these people as missing at random in our analysis.[2]

Finally, column 4 shows the number of people who passed their age 16 exams, as defined by obtaining at least five A*-C grades in GCSE exams. This level of attainment is a near-universal prerequisite for entry to university[3] and we will therefore focus on this group in our analysis, as we only want to include individuals who conceivably had the option of going to higher education

[1]This step was done separately by the Department for Work and Pensions before we had access to the data.

[2]In practice these people are more likely to be deprived or from an independent school. However it is a very small share of the overall population and therefore unlikely to affect our conclusions.

[3]Less than 10% of those without five good GCSEs start an undergraduate degree by age 21.

in our control sample.[4] We can see that this group represents around 56% of all students with linked HMRC records.

Table B.2 shows how the final sample given in column 4 of Table B.1 breaks down. Column 2 shows that around a third of those who passed their age 16 exams do not start an undergraduate degree. In column 3, we show the individuals who enter university as mature or part-time students.[5] We define mature students as anyone entering their first undergraduate degree more than three years after leaving school at age 18, while part-time status is a variable we observe in the HESA dataset. Combined, this group is about 6% of the individuals who passed their age 16 exams, and we exclude it from our analysis entirely. The primary reason for this is that we only observe earnings data up to age 30, which limits the number of years mature and part-time students with linked NPD records can possibly have been in the labour market after graduation (for example, someone who started a three-year degree at age 25 would only have had one or two years of labour market experience as a graduate by age 30). The focus of our paper is therefore on the impact of graduating from a full-time university degree started soon after leaving school, which is by far the most common route for obtaining an undergraduate degree. Finally, column 4 shows the individuals with high GCSEs whom we observe doing standard undergraduate degrees in UK universities. This is close to 60% of those passing their age 16 exams, and roughly one-third of the overall cohort.[6]

[4]This is less restrictive than Blundell et al. (2005) and Walker and Zhu (2018), who use individuals with at least one A-level as a control group. We take this decision because during our sample period, more than 10% of individuals who attend HE did not take any A levels or other KS5 qualifications.

[5]We also include a very small number of individuals who start their degrees before age 17 in this column, or for whom we only observe a postgraduate qualification. We think it is most likely that the latter individuals have taken an undergraduate qualification abroad and should therefore be excluded from the analysis.

[6]Although it is commonly cited that around half of people go to university, only around one-third of these cohorts start a 'standard' undergraduate degree within three years of leaving school.

Table B.2: LEO sample by GCSE year

| | Baseline (1) | No UG (2) | PT/Mature/PG (3) | UG sample (4) |
|---|---|---|---|---|
| 2002 | 279,409 | 98,524 | 20,091 | 160,794 |
| 2003 | 296,365 | 102,790 | 20,483 | 173,092 |
| 2004 | 312,579 | 110,091 | 21,255 | 181,233 |
| 2005 | 320,643 | 114,130 | 19,691 | 186,822 |
| 2006 | 325,581 | 110,938 | 18,093 | 196,550 |
| 2007 | 332,322 | 113,112 | 15,446 | 203,764 |
| Total | 1,866,899 | 649,585 | 115,059 | 1,102,255 |

Note: Column 1 is taken from Column 4 of Table B.1. Columns 2-4 sum to Column 1. PT indicates part-time, PG indicates postgraduate.

## B.2 Subject groups definition

Table B.3: Subjects included in each subject group

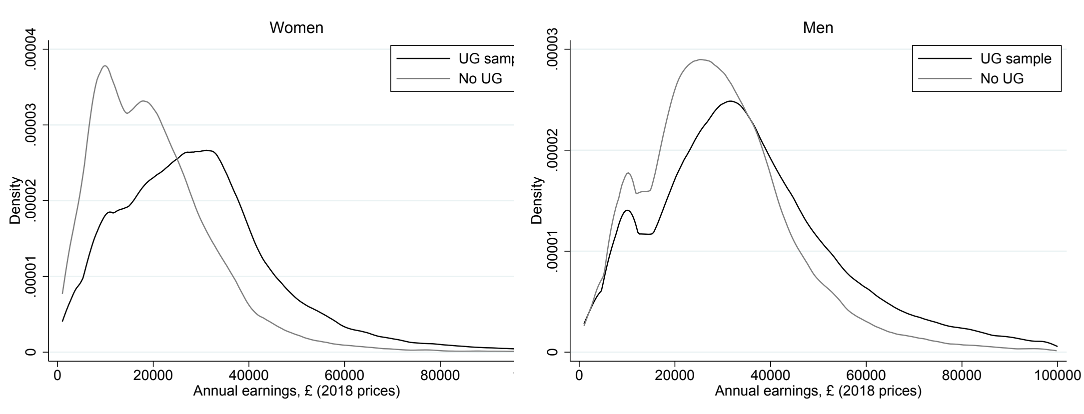
| Subject | Subject group | CAH2 code and description |
|---|---|---|
| Agriculture | STEM | (CAH06-01) agriculture, food and related studies |
| Allied to med | STEM | (CAH02-03) subjects allied to medicine not otherwise specified |
| Architecture | STEM | (CAH13-01) architecture, building and planning |
| Biosciences | STEM | (CAH03-01) biosciences |
| Business | LEM | (CAH17-01) business and management |
| Chemistry | STEM | (CAH07-02) chemistry |
| Comms | Other | (CAH18-01) communications and media |
| Computing | STEM | (CAH11-01) computing |
| Creative arts | Other | (CAH21-01) creative arts and design |
| Economics | LEM | (CAH15-02) economics |
| Education | Other | (CAH22-01) education and teaching |
| Engineering | STEM | (CAH10-01) engineering |
| English | Other | (CAH19-01) English studies |
| Geography | STEM | (CAH12-01) geographical and environmental studies |
| History | Other | (CAH20-01) history and archaeology |
| Languages | Other | (CAH19-03) languages, linguistics and classics |
| Law | LEM | (CAH16-01) law |
| Maths | STEM | (CAH09-01) mathematical sciences |
| Medicine | STEM | (CAH01-01) medicine and dentistry |
| Nursing | STEM | (CAH02-01) nursing |
| Pharmacology | STEM | (CAH02-02) pharmacology, toxicology and pharmacy |
| Philosophy | Other | (CAH20-02) philosophy and religious studies |
| Physics | STEM | (CAH07-01) physics and astronomy |
| Physsci | STEM | (CAH07-03) physical, material and forensic sciences |
| Politics | Other | (CAH15-03) politics |
| Psychology | STEM | (CAH04-01) psychology |
| Social care | Other | (CAH15-04) health and social care |
| Sociology | Other | (CAH15-01) sociology, social policy and anthropology |
| Sportsci | STEM | (CAH03-02) sport and exercise sciences |
| Technology | STEM | (CAH10-02) technology |
| Vetsci | STEM | (CAH05-01) veterinary sciences |

Note: For sample size reasons we do not include individuals studying: (CAH08-01) general and others in sciences; (CAH14-01); humanities and liberal arts (non-specific); (CAH19-02) Celtic studies; (CAH23-01) combined and general studies. See https://www.hesa.ac.uk/innovation/hecos for more information about the CAH2 subject mapping.

## B.3 Earnings descriptives

Figure B.1 shows the earnings of men and women at age 30 for those who did and did not go to higher education, for those with earnings between £1,000 and £100,000 in the given tax year. We see that men earn more than women and that those who attended HE earn more than those who did not, particularly for women.[7] Average annual taxable earnings of those earning more than £1,000 (including those earning above £100,000) for male graduates are £41,000 versus £31,000 for non-graduates, while the equivalent figures for women are £31,000 and £20,000. The medians are around £2,000 below the mean in all cases except for graduate men where the difference is closer to £6,000. This is due to the very long right hand tail of earnings for graduate men (not shown in the figure).

Figure B.1: Real earnings distributions by education level and gender at age 30




Note: Includes the 2002 GCSE cohort in 2016/17, roughly age 30, in the range £1,000 - £100,000. No HE consists of individuals who did not take an undergraduate degree, but passed their age 16 exams (obtaining at least 5 A*-C GCSEs). The tax data includes PAYE and SA earnings.

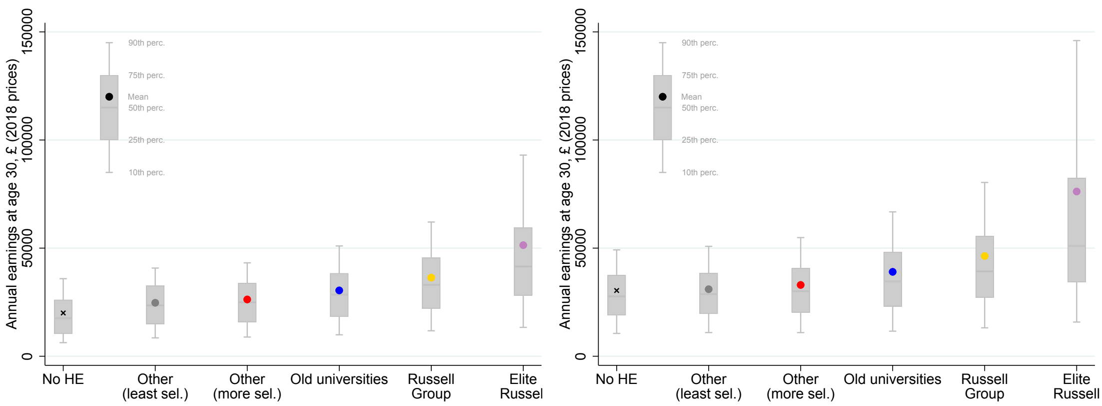
We turn to consider how earnings vary among those who attended HE. Figure B.2 shows the earnings distribution of individuals in the different university groupings for men and women separately. We see that average earnings increase with the selectivity of the institutions, with a significant jump for the Elite Russell Group. For men average earnings of those from the most selective

[7] There is also a clear spike in the distribution at around £10,000. This is due to bunching at the income tax and national insurance contribution thresholds.

universities are more than £75,000, while at the lower end, the earnings distribution for the least selective universities is very similar to that for those who did not attend higher education. For women the differences between the least selective universities and individuals who did not attend HE are larger.

The figure also highlights various points in the earnings distribution, showing that the variance of earnings increases dramatically with institution selectivity. There is also a significant right hand tail for men from the more selective institutions, with the mean more than £25,000 higher than the median for the Elite Russell Group. Excluding those earning below £1,000, just under 10% of men who attended the Elite Russell Group earn more than £150,000 per year.

Figure B.2: Real earnings by HEI type at age 30 - Women (left) and men (right)




Note: 2002 GCSE cohort in 2016/17, conditioning on earnings being above £1,000. No HE consists of individuals who did not take an undergraduate degree, but passed their age 16 exams (obtaining at least 5 A*-C GCSEs).

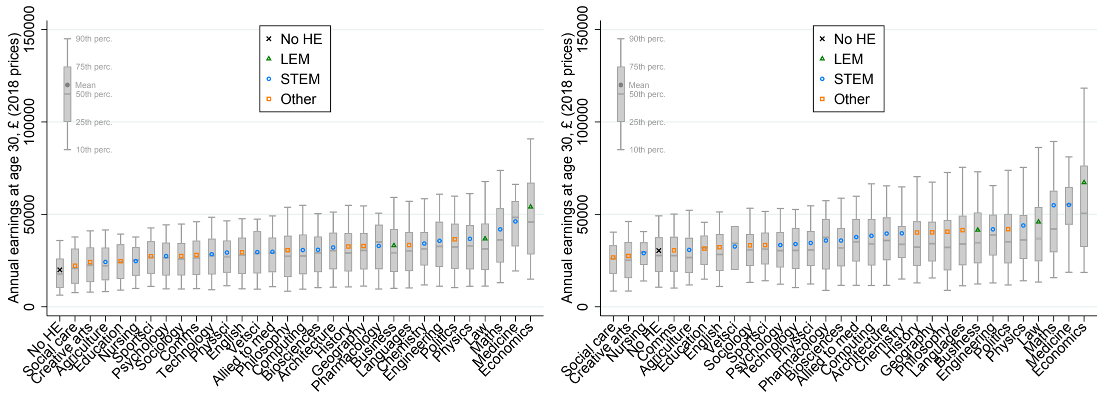
Figure B.3 then shows the equivalent distributions by individual subject,[8] with the broader subject groups also highlighted. Economics, maths, medicine and law are the subjects with the highest earnings at age 30 for both gender, while social care and creative arts and nursing have the lowest earnings.

All LEM subjects have high earnings, while earnings of STEM graduates are a bit more mixed. Maths, medicine, physics and engineering graduates all have high earnings, but nursing, agriculture, veterinary sciences and psychol-

[8]For Veterinary Sciences we drop the tail due to insufficient sample sizes to stay within data disclosure rules.

ogy graduates do not. We see a similar pattern among 'Other' subjects, with politics, languages, history and geography doing reasonably well, but lower earnings for the remaining subjects. We also note that that the spread of earnings is typically lower in the subjects that feed heavily into public sector careers (and centrally regulated pay scales), such as nursing, education and medicine.

Figure B.3: Real earnings by degree subject at age 30 - Women (left) and men (right)




Note: 2002 GCSE cohort in 2016/17, conditioning on earnings being above £1,000. No HE consists of individuals who did not take an undergraduate degree, but passed their age 16 exams (obtaining at least 5 A*-C GCSEs).

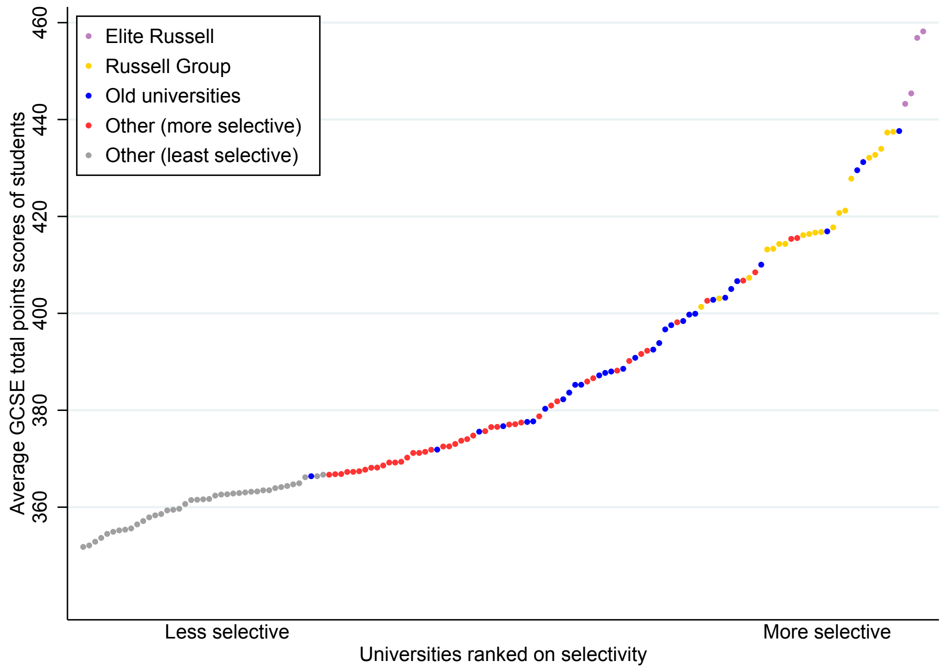
## B.4 Selectivity

As we described in Section 3.3.3, the higher education system in the UK is highly selective, meaning the highest status universities take the students with the highest prior attainment. We display this feature visually in Figure B.4, which plots the average GCSE points scores of the students of the different universities.[9] On the y axis 6 points is one grade higher in one exam - 100 points is therefore a substantial difference of around 17 grades across all GCSEs taken (students typically take around ten). We see that the 'Elite Russell' group is by far the most selective group, followed by the rest of the Russell Group, although

[9]We show total GCSE points, based on the following: A* = 58 points, A = 52 points, B = 46 points, C = 40 points, D = 34 points, E = 28 points, F = 22 points and G = 16 points (there is also an 'Unclassified', U grade which is worth 0 points). The total GCSE points score is then obtained by adding up the points for the different subjects.

there is some overlap with some of the old universities and more selective other institutions. The least selective other institutions are all at the bottom by construction, as we defined the Other group of universities based on GCSE scores.[10] There are also large differences in average prior attainment for people doing different subjects - see Figure B.5.

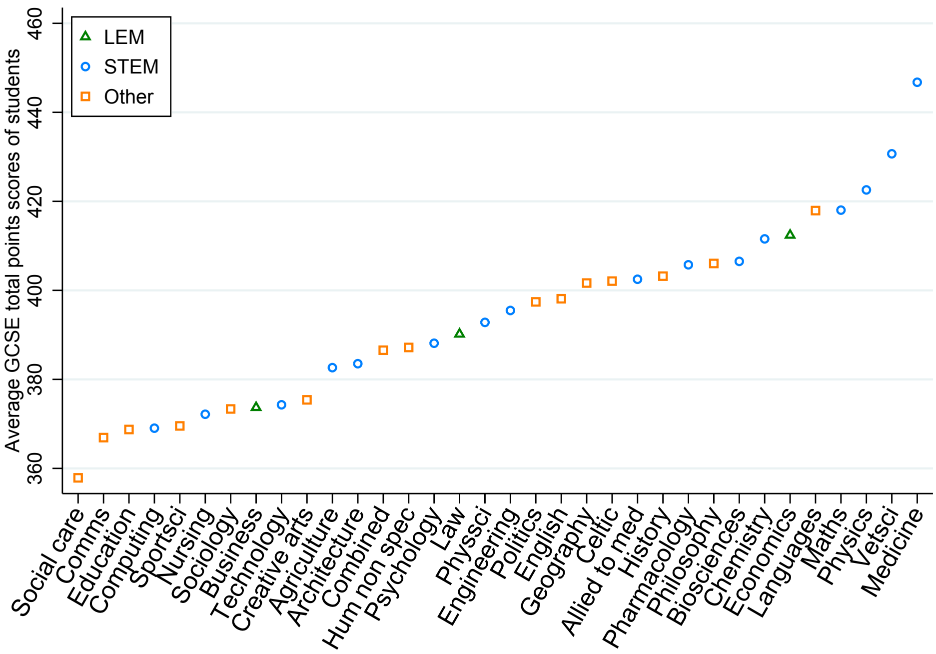
Figure B.4: Selectivity by university at age 30




Note: Selectivity is based on the average total GCSE points scores of each institutions' full-time, non-mature students from the 2004-2007 GCSE cohorts.

[10]It is important to note that we refer to the universities' selectivity taking into account age 16 test scores only. In fact, several of the universities here will be selective on other metrics such as music ability or arts portfolios. We do not account for that here.

Figure B.5: Selectivity by subject at age 30

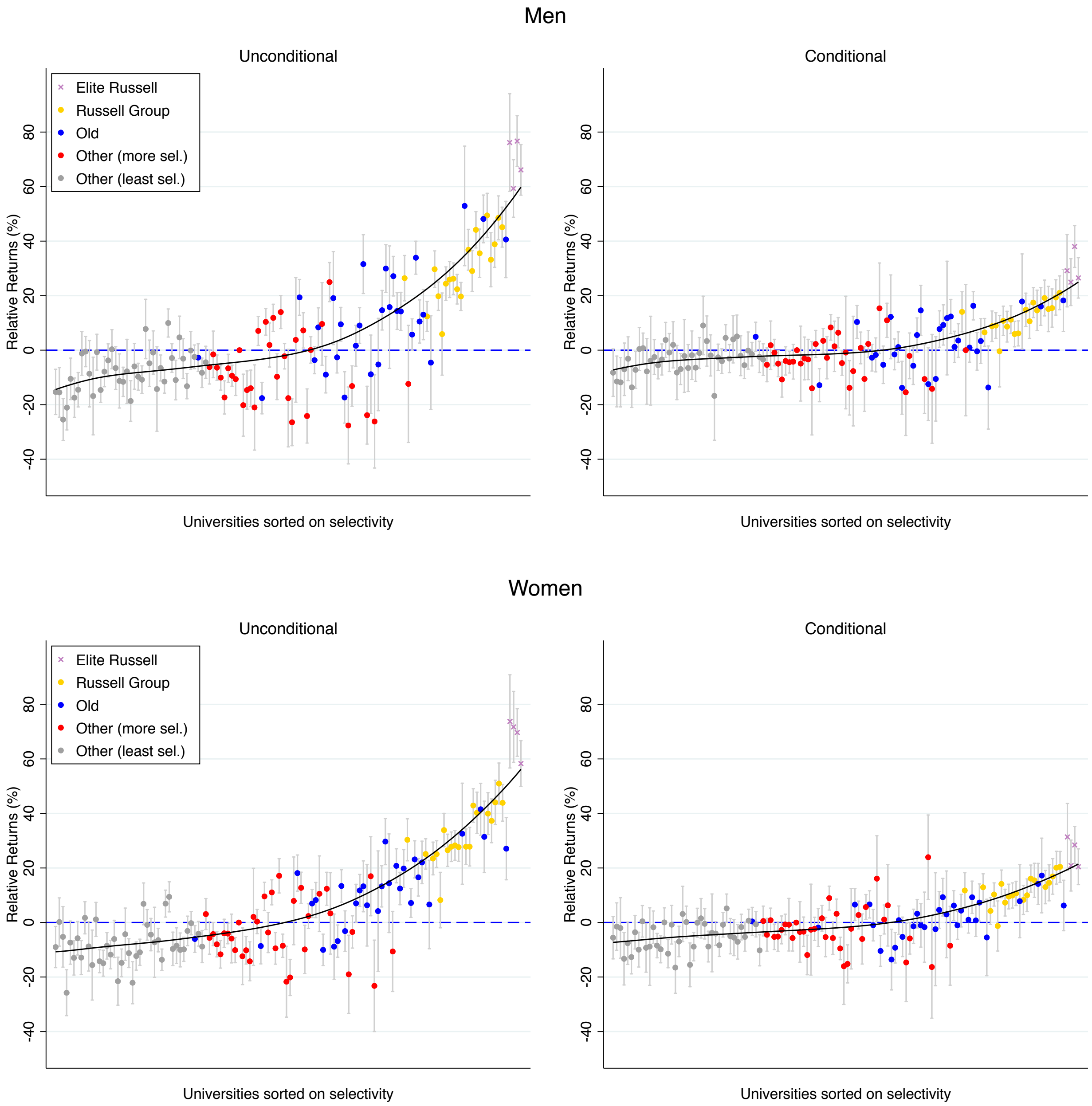



Note: Selectivity is based on the average total GCSE points scores of each subjects' full-time, non-mature students from the 2004-2007 GCSE cohorts.

## B.5 Additional results

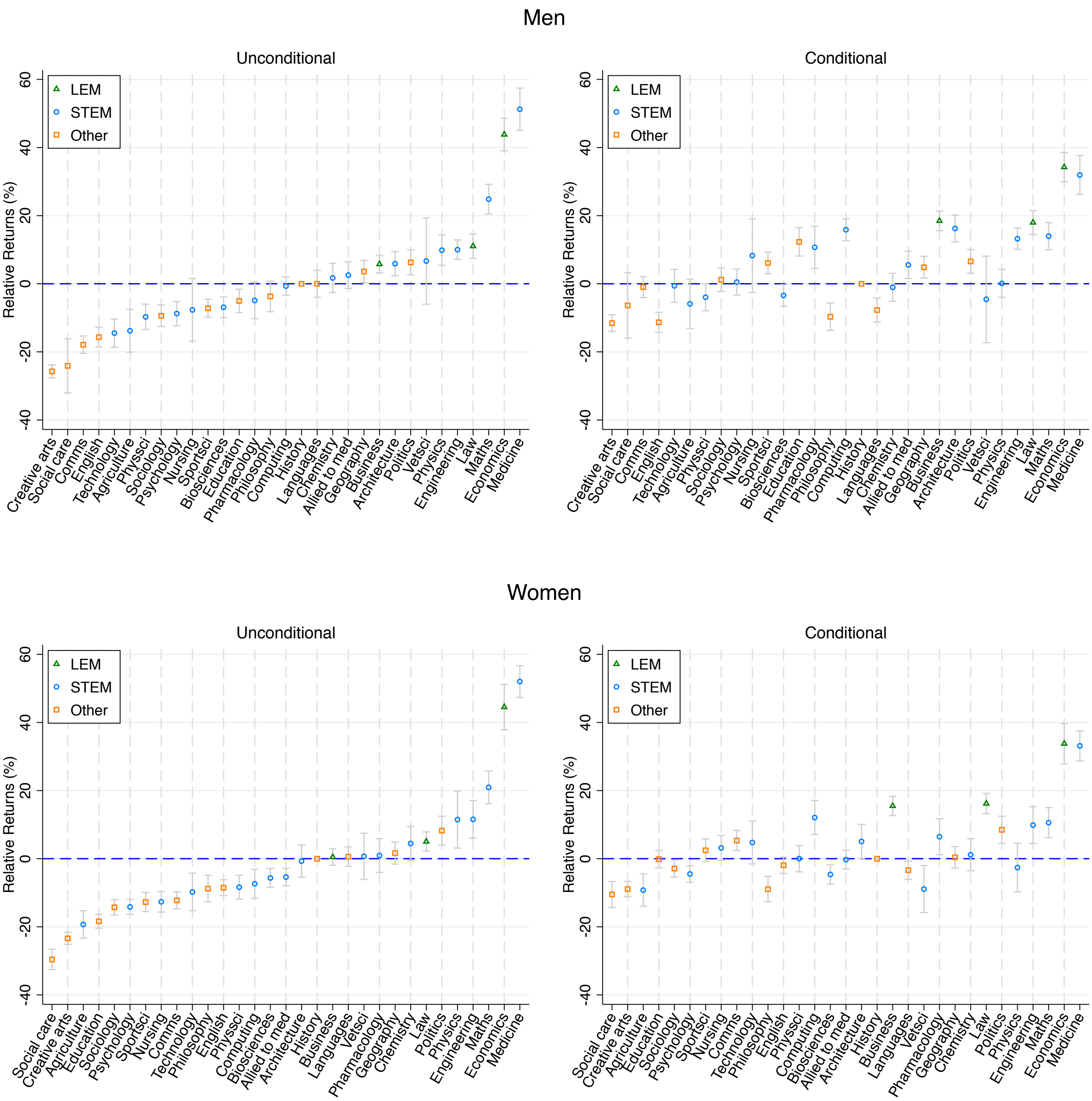
### B.5.1 Relative returns estimates by gender

Figure B.6: Estimated returns at age 30 by institution




Note: Equivalent to Figure 3.2, split by gender.

Figure B.7: Estimated returns at age 30 by subject




Note: Equivalent to Figure 3.3, split by gender.

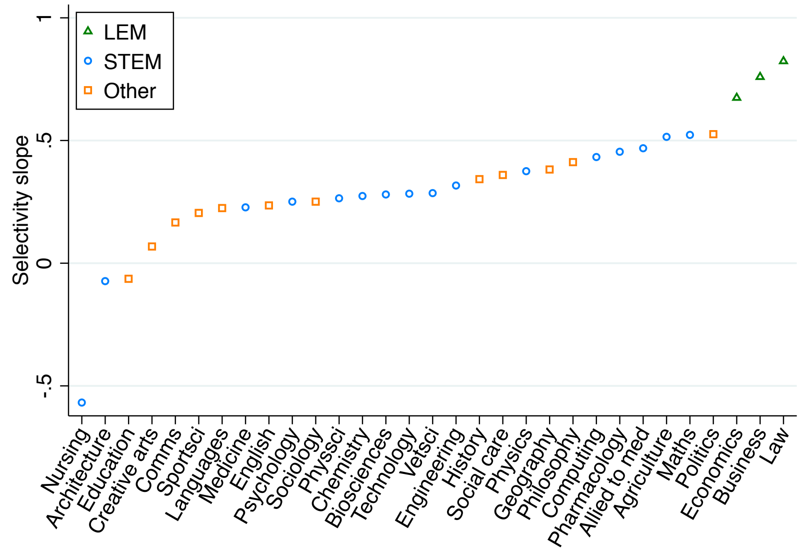
### B.5.2 Additional degree returns results

Table B.4: Best and worst performing degrees

| University | Subject | Relative returns (%) |
|---|---|---|
| University of Cambridge | Economics | 127.0 |
| Oxford University | Business | 122.7 |
| University of Cambridge | Computing | 115.3 |
| University College London | Economics | 108.1 |
| University of Cambridge | Law | 107.5 |
| University of St Andrews | Economics | 96.7 |
| University of Warwick | Economics | 95.9 |
| Oxford University | Economics | 94.1 |
| Oxford University | Law | 91.7 |
| University of Aberdeen | Medicine | 88.3 |
| School of Oriental and African Studies | Philosophy | -52.7 |
| Roehampton University | Social care | -44.5 |
| University of Gloucestershire | Social care | -42.6 |
| University of St Mark & St John | Social care | -41.5 |
| University of Central Lancashire | Philosophy | -37.7 |
| University of Wolverhampton | Politics | -37.2 |
| University of Worcester | Allied to med | -36.5 |
| Roehampton University | Allied to med | -35.7 |
| University of Glamorgan | Psychology | -35.6 |
| London Metropolitan University | Politics | -35.4 |

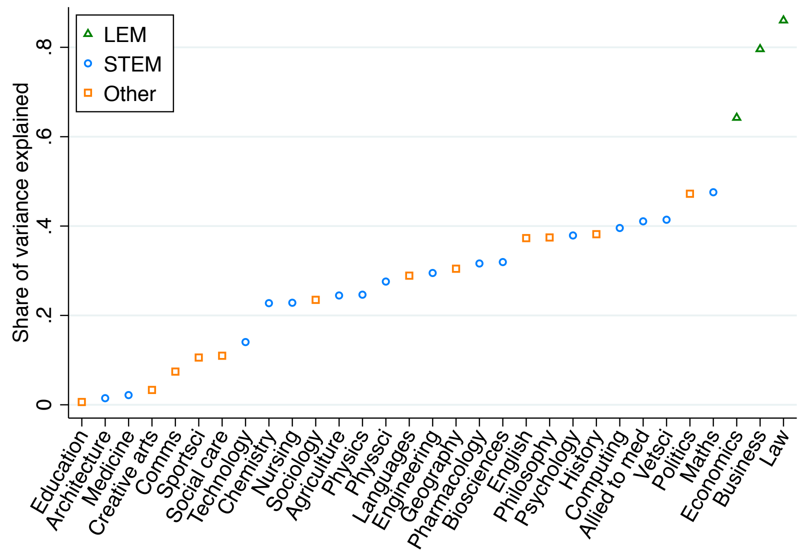
Note: Selected estimates of relative returns (in percentage points) from Figure 3.4. Returns are relative to History at Sheffield Hallam University.

Figure B.8: Returns-selectivity relationship by subject




Notes: Figure shows, for each subject, the slope of a degree-level regression of earnings returns on average GCSE score of the degree intake.

Figure B.9: Goodness-of-fit by subject




Notes: Figure shows, for each subject, the $R^2$ of a course-level regression of earnings returns on average GCSE score of the course intake. Subjects are coloured according to their broad subject group (LEM, STEM, Other). Slope coefficients of these regressions are shown in Appendix Figure B.8.

Table B.5: Correlations with main institution returns of alternative specifications

| | Excl. dropouts | Cross-sectional | Shrinkage | Within uni group |
|---|---|---|---|---|
| Elite Russell | 0.998 | 0.997 | 0.997 | 0.983 |
| Old universities | 0.979 | 0.975 | 0.984 | 0.947 |
| Other (more selective) | 0.980 | 0.956 | 0.938 | 0.981 |
| Other (least selective) | 0.956 | 0.941 | 0.982 | 0.930 |

Note: Column (1) shows the correlation of our main institution returns with returns estimated on a sample which excludes individuals who did not graduate from their degree. Column (2) shows the correlation with institution returns estimated at age 30, on a cross-sectional sample only. Column (3) shows the correlation with the institution returns after shrinkage has been applied, where we shrink degrees returns to the average degree return. Column (4) shows the correlation with institution returns estimated within subsamples for each university group.

Table B.6: Correlations of selectivity with returns for alternative specifications

| | Main results | Excl. school FEs | Excl. school FEs and background |
|---|---|---|---|
| *LEM* | | | |
| Business | 0.892 | 0.869 | 0.890 |
| Economics | 0.801 | 0.838 | 0.822 |
| Law | 0.927 | 0.921 | 0.925 |
| *STEM* | | | |
| Agriculture | 0.495 | 0.496 | 0.491 |
| Allied to med | 0.641 | 0.676 | 0.656 |
| Architecture | -0.122 | -0.102 | -0.112 |
| Biosciences | 0.565 | 0.639 | 0.588 |
| Chemistry | 0.477 | 0.590 | 0.519 |
| Computing | 0.629 | 0.626 | 0.620 |
| Engineering | 0.543 | 0.559 | 0.558 |
| Maths | 0.690 | 0.713 | 0.707 |
| Medicine | 0.147 | 0.205 | 0.176 |
| Nursing | -0.478 | -0.440 | -0.487 |
| Pharmacology | 0.562 | 0.577 | 0.553 |
| Physics | 0.496 | 0.516 | 0.507 |
| Physsci | 0.525 | 0.619 | 0.570 |
| Psychology | 0.616 | 0.649 | 0.639 |
| Technology | 0.375 | 0.370 | 0.360 |
| Vetsci | 0.644 | 0.728 | 0.696 |
| *Other* | | | |
| Comms | 0.273 | 0.285 | 0.288 |
| Creative arts | 0.182 | 0.264 | 0.230 |
| Education | -0.079 | -0.035 | -0.074 |
| English | 0.611 | 0.651 | 0.633 |
| Geography | 0.552 | 0.556 | 0.550 |
| History | 0.618 | 0.637 | 0.628 |
| Languages | 0.538 | 0.555 | 0.554 |
| Philosophy | 0.612 | 0.650 | 0.642 |
| Politics | 0.687 | 0.716 | 0.702 |
| Social care | 0.331 | 0.445 | 0.359 |
| Sociology | 0.485 | 0.553 | 0.516 |
| Sportsci | 0.325 | 0.307 | 0.322 |

Note: Column (1) shows the correlation of selectivity with our main degree returns. Column (2) shows the correlation of selectivity with degree returns estimated when excluding school FEs from the controls. Column (3) shows the correlation of selectivity with degree returns estimated when both school FEs and individual level background characteristics are excluded from the controls.

Table B.7: Correlations with main degree returns of alternative specifications

| | Excl. dropouts | Cross-sectional | Shrinkage | Within subject |
|---|---|---|---|---|
| *LEM* | | | | |
| Business | 0.985 | 0.976 | 0.983 | 0.990 |
| Economics | 0.983 | 0.962 | 0.974 | 0.924 |
| Law | 0.990 | 0.984 | 0.997 | 0.982 |
| *STEM* | | | | |
| Agriculture | 0.987 | 0.973 | 0.996 | 0.947 |
| Allied to med | 0.985 | 0.955 | 0.985 | 0.977 |
| Architecture | 0.934 | 0.916 | 0.985 | 0.949 |
| Biosciences | 0.960 | 0.925 | 0.989 | 0.943 |
| Chemistry | 0.963 | 0.901 | 0.992 | 0.855 |
| Computing | 0.958 | 0.938 | 0.988 | 0.964 |
| Engineering | 0.934 | 0.944 | 0.982 | 0.966 |
| Maths | 0.949 | 0.914 | 0.986 | 0.871 |
| Medicine | 0.973 | 0.914 | 0.925 | 0.860 |
| Nursing | 0.936 | 0.899 | 0.991 | 0.885 |
| Pharmacology | 0.957 | 0.792 | 0.993 | 0.805 |
| Physics | 0.953 | 0.944 | 0.993 | 0.886 |
| Physsci | 0.915 | 0.881 | 0.976 | 0.942 |
| Psychology | 0.932 | 0.932 | 0.989 | 0.957 |
| Technology | 0.951 | 0.933 | 0.990 | 0.874 |
| Vetsci | 0.980 | 0.993 | 0.996 | 0.888 |
| *Other* | | | | |
| Comms | 0.935 | 0.905 | 0.986 | 0.962 |
| Creative arts | 0.954 | 0.903 | 0.955 | 0.907 |
| Education | 0.968 | 0.946 | 0.976 | 0.965 |
| English | 0.960 | 0.892 | 0.986 | 0.960 |
| Geography | 0.983 | 0.963 | 0.987 | 0.978 |
| History | 0.979 | 0.943 | 0.993 | 0.936 |
| Languages | 0.919 | 0.894 | 0.980 | 0.911 |
| Philosophy | 0.970 | 0.942 | 0.986 | 0.952 |
| Politics | 0.972 | 0.960 | 0.985 | 0.982 |
| Social care | 0.906 | 0.932 | 0.968 | 0.886 |
| Sociology | 0.932 | 0.961 | 0.985 | 0.953 |
| Sportsci | 0.944 | 0.882 | 0.995 | 0.952 |

Note: Column (1) shows the correlation of our main degree returns with returns estimated on a sample which excludes individuals who did not graduate from their degree. Column (2) shows the correlation with degree returns estimated at age 30, on a cross-sectional sample only. Column (3) shows the correlation with the degrees returns after shrinkage has been applied, where we shrink degrees returns to the average degree return. Column (4) shows the correlation with course returns estimated within subsamples for each subject.

# Appendix C

# Appendix to Chapter 4

## C.1 Sample selection

Table C.1 shows, for each of the cohorts we use in our main analysis, how we go from the total population of individuals taking GCSEs in England in each year to our analysis sample.

Column (1) shows that around 600,000 students take their GCSEs in England each year. For a small number of these, we are missing crucial data from the school census that we need to construct a measure of their parental background. Column (2) shows the numbers of students having the required background data to be included in our analysis, which is around 98% of the full population from Column (1). Column (3) then shows the people who are successfully merged to the HMRC tax records. Around 7% of the sample cannot be matched to tax records. This group includes individuals with mistakes in the name or address recorded in their education or tax records,[1] as well as some individuals who have never been in touch with the tax and benefit system - for example, because they have moved abroad.

Columns (4)-(6) then takes the people for whom we have all the required in-

[1]The linkage between the different datasets was done using National Insurance numbers or, where unavailable, using fuzzy matching based on name, postcode and gender. This matching was performed by the Department for Work and Pensions before we had access to the dataset.

Table C.1: LEO sample by GCSE cohort

| GCSE cohort | Population (1) | Non-missing KS4 vars (2) | Matched to HMRC (3) | No UG (4) | PT, mature, & PG-only (5) | UG (6) |
|---|---|---|---|---|---|---|
| 2001/2002 | 589,543 | 580,524 | 534,707 | 321,270 | 33,103 | 180,334 |
| 2002/2003 | 621,697 | 614,475 | 567,767 | 344,779 | 33,563 | 189,425 |
| 2003/2004 | 638,242 | 629,207 | 588,820 | 356,811 | 33,451 | 198,558 |
| 2004/2005 | 646,429 | 631,935 | 592,725 | 356,631 | 30,871 | 205,223 |
| 2005/2006 | 648,331 | 636,265 | 602,953 | 356,699 | 28,207 | 218,047 |

formation and splits them in three groups based on their university attainment. Column (4) shows the around 60% of individuals who did not attend university within our sample period.[2] We focus on the implications for social mobility of undertaking an undergraduate (UG) degree. This definition includes both students who graduate from a degree as well as those who drop out of university. We restrict to full-time students who started university by the age of 21, in order to ensure that individuals have had a reasonable number of years in the labour market after graduating by the end of our sample period. Column (5) shows the number of individuals who attended university but are excluded from our analysis as they do not meet one of these criteria. This includes people studying part-time (PT) and individuals who started university as a mature student.[3] Column (6) then shows our main analysis sample, which consists of around 200,000 students in each cohort. For our university- and subject-level analysis, which uses the three oldest cohorts, our analysis is based on around 570,000 individuals. Our course-level analysis is based on almost 1,000,000 individuals.

## C.2 Additional results tables

[2] We observe HESA records, and hence whether someone has attended university, up to 2015-16, which is age 28, 29 or 30 depending on the cohort.

[3] It also includes a very small number of individuals for whom we observe a postgraduate but not an undergraduate degree, likely due to them studying abroad for their undergraduate qualification. As our focus here is on the social mobility impact of undergraduate degrees, we exclude these individuals from our analysis.

Table C.2: Bottom 20 universities for mobility

| Rank | University | Group | Mobility (%) | Access (%) | Success (%) |
|---|---|---|---|---|---|
| 1 | Arts Inst Bournemouth | Post-1992 (more selective) | 0.2 | 2.4 | 7.7 |
| 2 | Rose Bruford | Post-1992 (more selective) | 0.2 | 2.3 | 10.0 |
| 3 | York St John UC | Post-1992 (more selective) | 0.2 | 2.3 | 10.4 |
| 4 | Leeds City | Post-1992 (more selective) | 0.3 | 1.0 | 25.0 |
| 5 | Central Sch Speech/Drama | Post-1992 (more selective) | 0.3 | 3.1 | 8.3 |
| 6 | L'pool Inst Perf Arts | Post-1992 (more selective) | 0.3 | 3.5 | 8.3 |
| 7 | Cons Dance/Drama | Post-1992 (more selective) | 0.3 | 4.7 | 6.2 |
| 8 | Newcastle | Russell Group | 0.3 | 1.2 | 26.2 |
| 9 | Exeter | Russell Group | 0.3 | 1.2 | 28.4 |
| 10 | Winchester | Post-1992 (more selective) | 0.4 | 3.3 | 10.7 |
| 11 | Bath | Pre-1992 university | 0.4 | 1.0 | 36.8 |
| 12 | Bath Spa | Post-1992 (more selective) | 0.4 | 3.1 | 13.2 |
| 13 | Bishop Grosseteste | Post-1992 (more selective) | 0.4 | 4.3 | 10.0 |
| 14 | Bristol | Russell Group | 0.4 | 1.0 | 42.3 |
| 15 | Norwich UC Arts | Post-1992 (more selective) | 0.4 | 2.7 | 16.7 |
| 16 | Writtle C | Post-1992 (least selective) | 0.5 | 2.9 | 18.2 |
| 17 | Oxford | Most selective Russell | 0.5 | 0.9 | 59.0 |
| 18 | York | Russell Group | 0.5 | 1.4 | 38.8 |
| 19 | Cambridge | Most selective Russell | 0.5 | 1.0 | 54.4 |
| 20 | Southampton | Russell Group | 0.5 | 1.5 | 36.8 |

Notes: Three universities with low access and/or success rates are omitted from this list due to low sample size.

Table C.3: Top 20 courses for mobility (2002-04 GCSE cohorts)

| Rank | University | Group | Subject | Mobility % | Access % | Success % |
|---|---|---|---|---|---|---|
| 1 | QMU | Russell Group | Computing | 12.1 | 24.1 | 50.1 |
| 2 | QMU | Russell Group | Maths | 10.0 | 24.6 | 40.5 |
| 3 | City | Pre-1992 university | Nursing | 9.4 | 18.0 | 52.2 |
| 4 | QMU | Russell Group | Economics | 9.4 | 14.8 | 63.2 |
| 5 | Kingston | Post-1992 (least selective) | Pharmacology | 9.3 | 27.0 | 34.5 |
| 6 | City | Pre-1992 university | Computing | 9.1 | 22.0 | 41.4 |
| 7 | Goldsmiths | Pre-1992 university | Computing | 8.9 | 30.3 | 29.5 |
| 8 | City | Pre-1992 university | Economics | 8.7 | 19.2 | 45.1 |
| 9 | Middlesex | Post-1992 (least selective) | Computing | 8.5 | 27.9 | 30.5 |
| 10 | QMU | Russell Group | Engineering | 8.4 | 22.8 | 36.9 |
| 11 | Greenwich | Post-1992 (least selective) | Computing | 8.3 | 26.1 | 31.7 |
| 12 | Westminster | Post-1992 (least selective) | Computing | 8.2 | 33.1 | 24.9 |
| 13 | KCL | Russell Group | Computing | 8.2 | 13.3 | 61.8 |
| 14 | Westminster | Post-1992 (least selective) | Law | 8.2 | 27.1 | 30.3 |
| 15 | QMU | Russell Group | Law | 7.9 | 15.4 | 51.3 |
| 16 | City | Pre-1992 university | Law | 7.8 | 19.4 | 40.0 |
| 17 | Westminster | Post-1992 (least selective) | Biosciences | 7.8 | 34.3 | 22.6 |
| 18 | Brunel | Pre-1992 university | Computing | 7.8 | 18.8 | 41.2 |
| 19 | Aston | Pre-1992 university | Pharmacology | 7.6 | 16.8 | 45.3 |
| 20 | QMU | Russell Group | Business | 7.6 | 21.3 | 35.7 |

## C.3 Subject heterogeneity

Table C.4: Mobility, access and success rates for all subjects

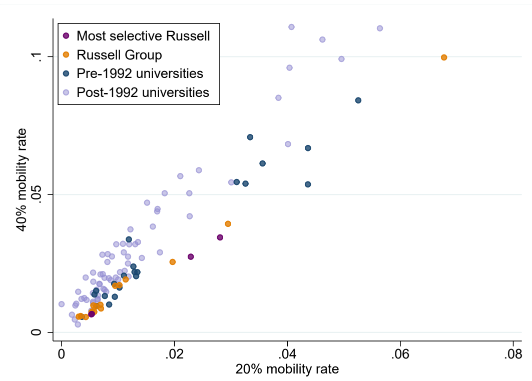
| Subject | Mobility (%) | Access (%) | Success (%) |
|---|---|---|---|
| Pharmacology | 4.2 | 11.5 | 36.6 |
| Computing | 2.9 | 10.8 | 26.9 |
| Law | 2.2 | 9.9 | 21.9 |
| Economics | 2.0 | 4.7 | 41.9 |
| Business | 1.9 | 8.6 | 22.5 |
| Engineering | 1.9 | 5.6 | 34.0 |
| Maths | 1.8 | 4.3 | 42.5 |
| Medicine | 1.7 | 2.7 | 63.2 |
| Allied to med | 1.6 | 5.9 | 27.5 |
| Architecture | 1.4 | 4.7 | 29.6 |
| Chemistry | 1.4 | 4.9 | 28.3 |
| Biosciences | 1.4 | 6.1 | 22.1 |
| Social care | 1.3 | 10.5 | 12.5 |
| Nursing | 1.3 | 5.7 | 22.2 |
| Sociology | 1.2 | 7.7 | 15.1 |
| Comms | 1.1 | 6.1 | 18.1 |
| Psychology | 1.0 | 6.3 | 16.4 |
| Politics | 1.0 | 4.3 | 22.8 |
| Philosophy | 0.9 | 3.3 | 25.8 |
| Physics | 0.9 | 2.7 | 32.4 |
| Sportsci | 0.9 | 3.9 | 21.8 |
| Education | 0.8 | 5.8 | 14.6 |
| English | 0.7 | 3.8 | 19.5 |
| Physsci | 0.7 | 4.8 | 14.5 |
| Creative arts | 0.7 | 5.2 | 13.2 |
| History | 0.6 | 2.6 | 22.2 |
| Technology | 0.6 | 4.2 | 13.5 |
| Languages | 0.5 | 2.3 | 23.6 |
| Geography | 0.5 | 1.7 | 27.1 |
| Agriculture | 0.4 | 2.2 | 20.0 |
| Vetsci | 0.4 | 1.2 | 33.3 |

## C.4 Robustness

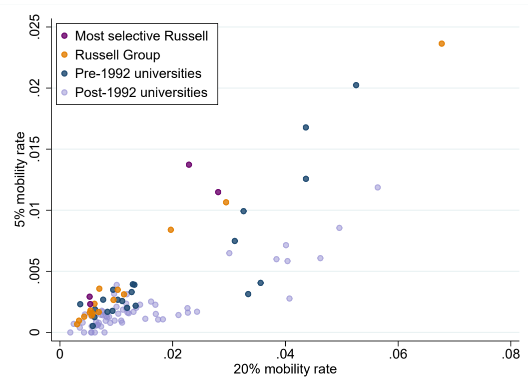
### C.4.1 Changes in sample and mobility rate definition

Figure C.1: Correlation between main estimates and top 40% and top 5% mobility rates

(a) Top 40% mobility rates

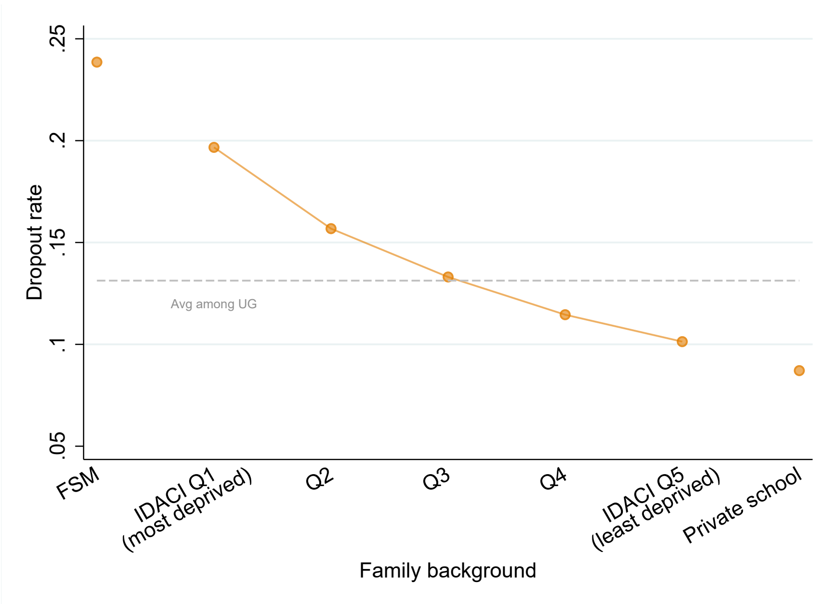



(b) Top 5% mobility rates



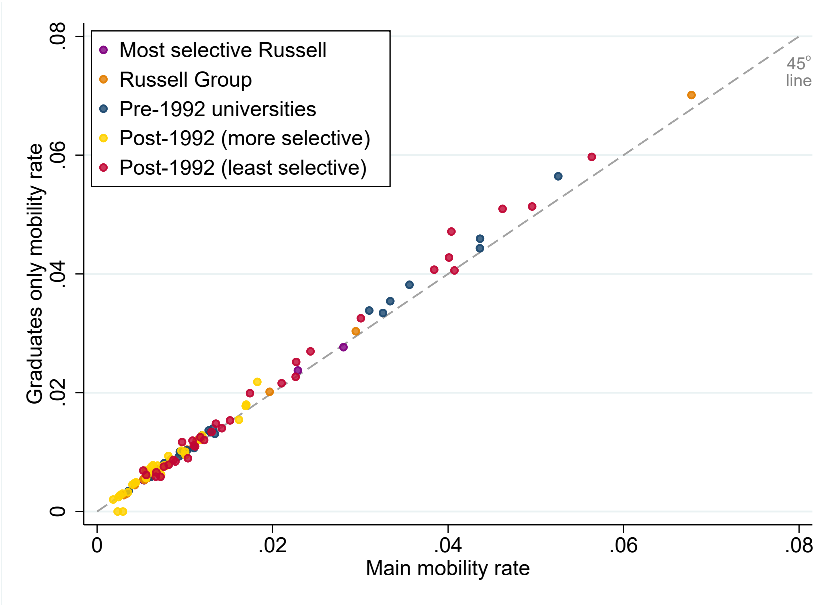

Notes: Includes universities with at least 500 students.

Figure C.2: Student dropout rates by income group for the 2002-04 GCSE cohorts



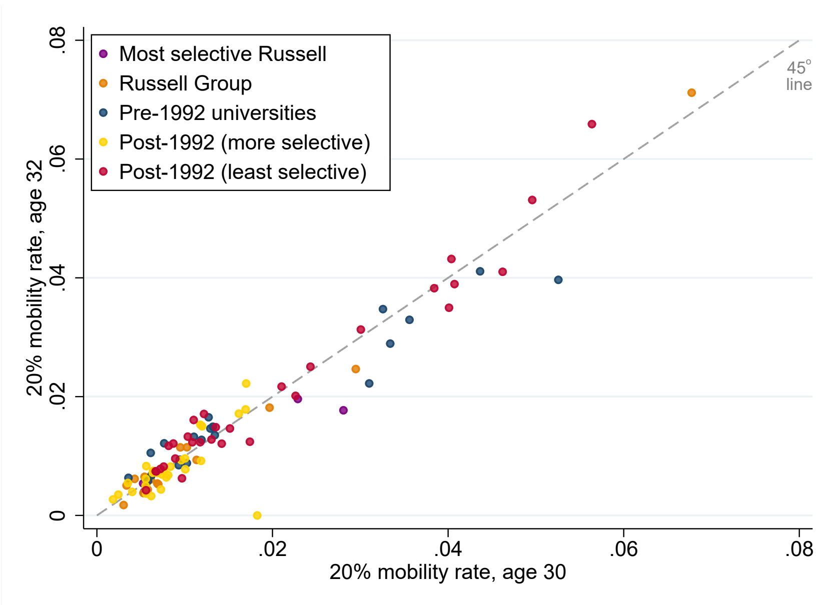

Notes: Figure shows proportion of university entrants who do not graduate, for state school students in each quintile of IDACI score, as well as for FSM and private school students.

Figure C.3: Mobility rates by institution, main estimates vs graduates only




Notes: Harper Adams, Leeds City College, Royal Agricultural College and Trinity LABAN dropped due to small sample.

Figure C.4: Mobility rates by institution, age 30 vs age 32




Notes: Age 32 results use 2002 GCSE cohort only. Universities with at least 250 students and 6 FSM students are included.

### C.4.2 Adjusting for student composition

To compute conditional success rates, we start by regressing an indicator for individuals being in the top 20% of the earnings distribution on an indicator for their institution, subject or course, as well as a set of prior attainment and background variables. We do this *only for FSM students* and use a simple linear probability model (LPM). This can be written as follows:

$$Q5_i = \alpha + HEI_i'\beta + X_i'\gamma + \epsilon_i \tag{C.1}$$

$$Q5_i = \alpha + Subject_i'\beta + X_i'\gamma + \epsilon_i \tag{C.2}$$

$$Q5_i = \alpha + Course_i'\beta + X_i'\gamma + \epsilon_i \tag{C.3}$$

where $Q5_i$ is an indicator for whether the individual is in the top quintile of earnings, at age 30 for institutions and subjects and at age 28 for courses, and $HEI_i'$, $Subject_i'$ and $Course_i'$ are institution, subject and course dummies respectively (we omit the conditioning on individuals being from low-income families from the notation for the sake of clarity). The inclusion of controls $X_i'$ helps to account for differences in success that are due to differences in observable characteristics between students across subjects, courses or institutions. We control for:

- KS4 point score (non-parametrically);[4]
- home region;
- ethnicity;
- gender.

In the presence of these controls, the coefficient $\beta_j$ gives us the difference in probability of having age 30 earnings in the top 20% between students from

[4]We split individuals into deciles of KS4 score within cohort and include dummies for each decile.

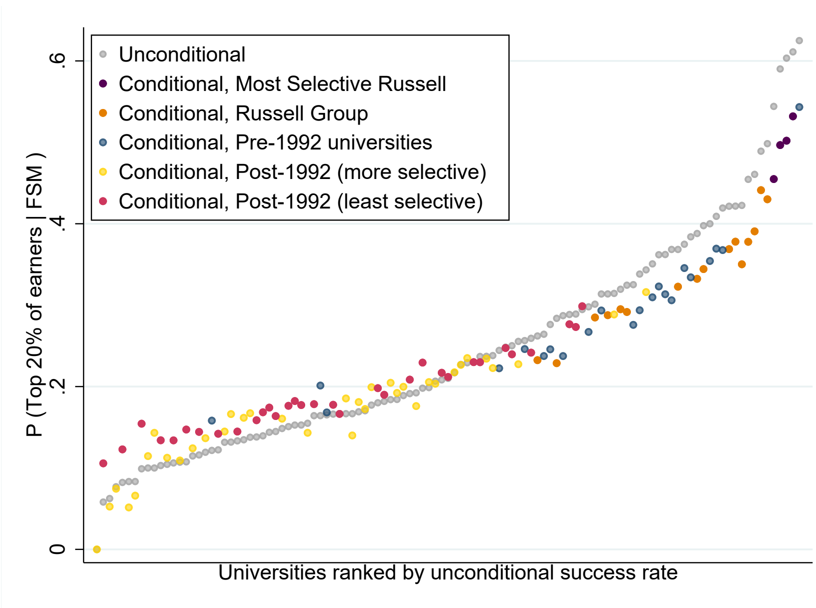
low-income backgrounds who attended university, subject or course $j$ and comparable - in terms of the characteristics contained in $X_i'$ - individuals from low-income backgrounds who did not attend HE (the omitted category). This can be interpreted as the differences in success which cannot be explained by differences in GCSE attainment, ethnicity, gender or region of origin.

To construct conditional success rates, we use the coefficients from equations (3), (4) and (5) to predict the conditional success rate at each institution, subject and course respectively, for the average student from a low-income background. Writing the characteristics of the average poor student as $\overline{X}$, we can thus write the conditional success rate of institution, subject or course $j$ as follows:

$$Success_j^{cond} = \alpha + \beta_j + \overline{X}'\gamma \tag{C.4}$$

We then multiply these conditional success rates by access rates for each university, subject and course to construct conditional mobility rates.

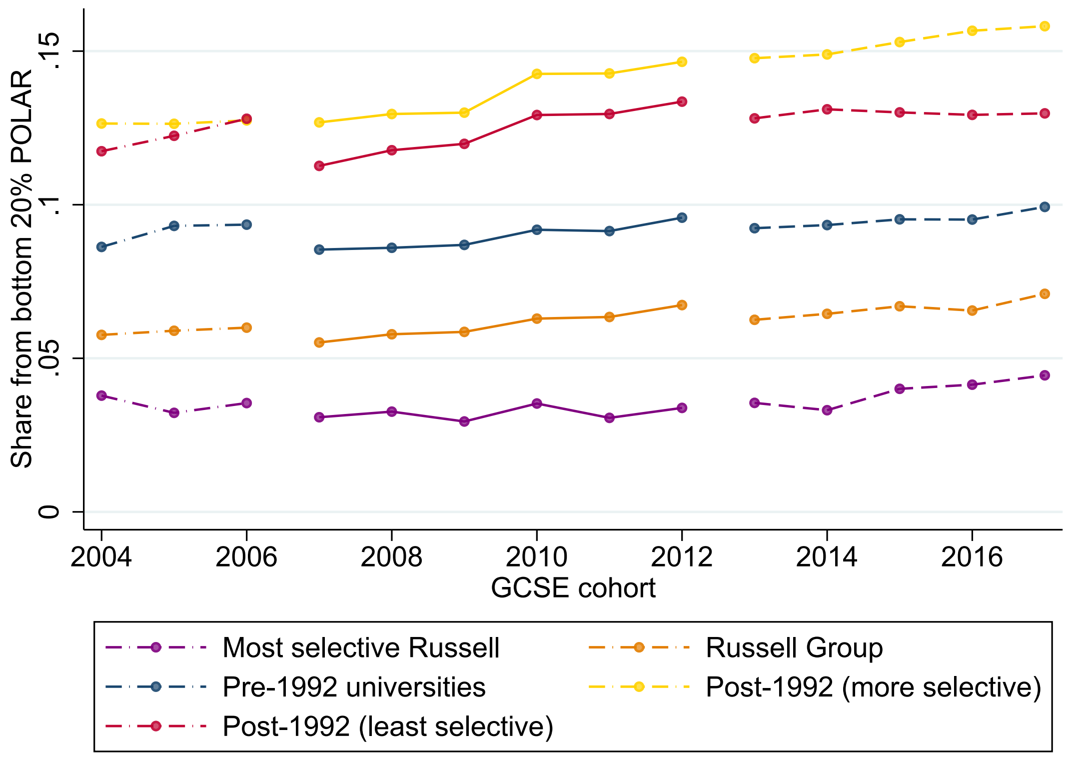
Figure C.5: Conditional success rates by university




Notes: Universities with at least 250 students and 6 FSM students are included. Negative conditional success rates are set to zero.

## C.5 Recent trends in access and mobility

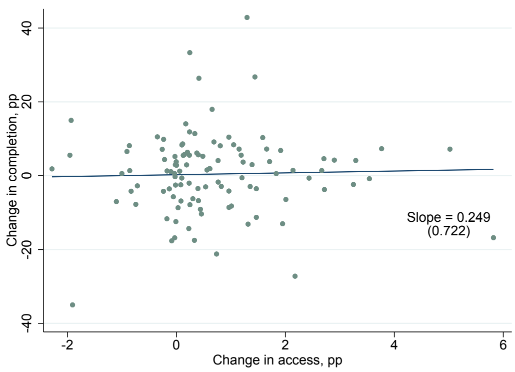
Figure C.6: Access over time using POLAR



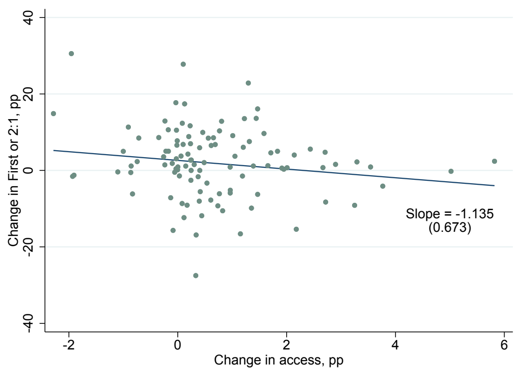

Notes: POLAR2 is plotted between 2004 and 2006. This is based on people who were 18 between 2000 and 2004 and who started a course, aged 18 or 19, between 2000-01 and 2005-06. POLAR3 (based on 18- and 19-year-olds starting between 2005-06 and 2009-10 and between 2006-07 and 2010-11, respectively) is plotted up to 2012. 2013 onwards plots POLAR4 (based on 18-year-olds starting between 2009-10 and 2013-14 and 19-year-olds starting between 2010-11 and 2014-15).

Figure C.7: Relationship between access and intermediate outcomes of FSM students

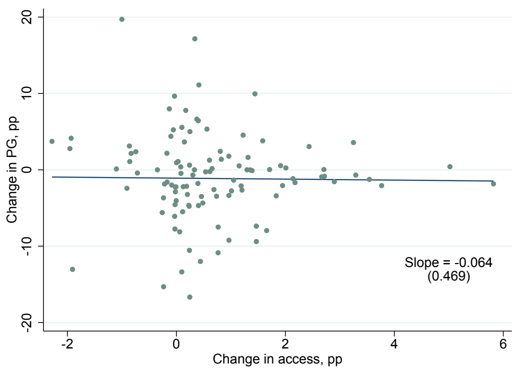
(a) Degree completion



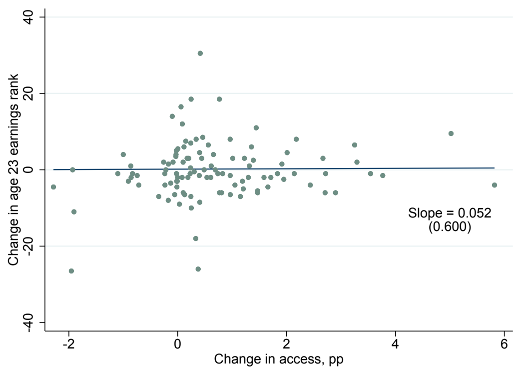

(b) Degree classification




(c) Postgraduate study




(d) Early-career earnings



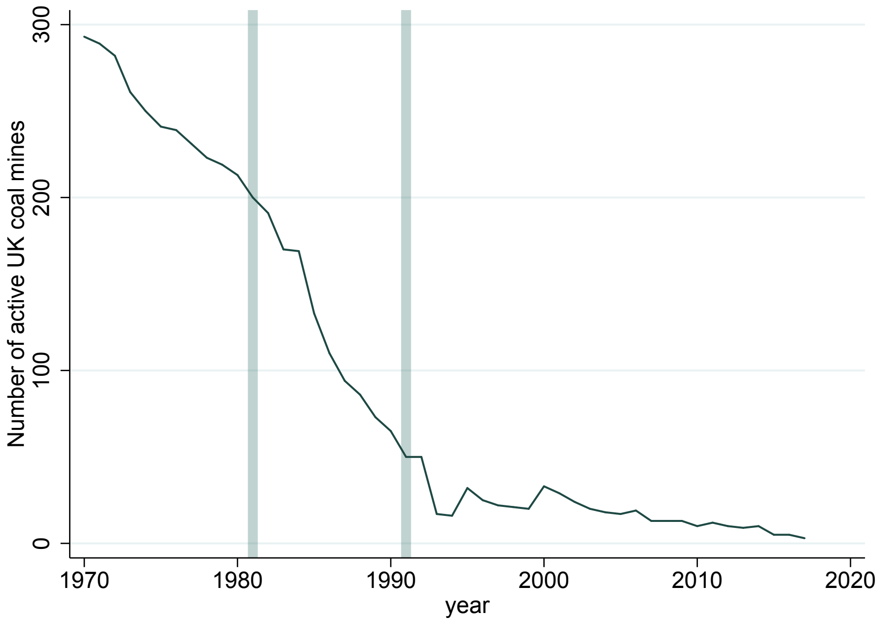

Notes: Figure includes institutions with at least 100 students and at least 6 FSM students in 2002-03 and 2005-06. Degree completion looks at the share of FSM students who graduate from their degree. Degree classification looks at the share of students from FSM backgrounds who obtain a First Class or Upper Second Class degree. Postgraduate study is defined as starting a postgraduate course within four years of starting an undergraduate course for those taking a three-year degree (excluding those taking integrated masters). Our measure of early career earnings is earnings rank at age 23.

# Appendix D

# Appendix to Chapter 5

## D.1 Number of coal mines in the UK over time

Figure D.1: Number of active deep coal mines in the UK 1970-2018




Notes: Figure shows the number of active deep coal mines in the UK. Lines mark the 1981 and 1991 census. Data comes from the Department for Business, Energy and Industrial Strategy's publication "Historical coal data: coal production, availability and consumption" which can be found at https://www.gov.uk/government/statistical-data-sets/historical-coal-data-coal-production-availability-and-consumption.

## D.2 Summary statistics

Table D.1: Comparison of means of family characteristics in 1981 for job loss and survivor samples

| | Job loss | Survivor | Difference in means |
|---|---|---|---|
| *Father outcomes 1981* | | | |
| Age | 31.5 | 30.5 | 1.1** |
| Degree | $< 0.04$ | $< 0.04$ | . |
| *Mother outcomes 1981* | | | |
| Age | 29.4 | 28.7 | 0.7 |
| Degree | $< 0.04$ | $< 0.04$ | . |
| Employed | 0.23 | 0.17 | 0.06* |
| *Family outcomes 1981* | | | |
| Number of kids | 2.3 | 2.2 | 0.1* |
| Own house | 0.55 | 0.63 | -0.08** |
| Social renter | 0.32 | 0.29 | 0.03 |
| N | 457 | 276 | 733 |

Notes: *, ** and *** denote significance at the 10%, 5% and 1% level, based on a t-test of the difference in means. For statistical disclosure reasons, descriptives are not given when underlying sample sizes are too small.
Data source: ONS LS

## D.3 Impact of job loss on parent and child outcomes

Table D.2: Effects of father's job loss on parent outcomes

| | (1) No controls | (2) Controls |
|---|---|---|
| *Father outcomes 1991* | | |
| Employed | −0.395*** | −0.344*** |
| | (0.030) | (0.030) |
| Unemployed | 0.171*** | 0.155*** |
| | (0.023) | (0.023) |
| Inactivity | 0.224*** | 0.189*** |
| | (0.025) | (0.025) |
| Long term sick | 0.153*** | 0.136*** |
| | (0.022) | (0.022) |
| Retired | 0.031*** | 0.013 |
| | (0.010) | (0.009) |
| Earnings | −15318.277*** | −13746.054*** |
| | (783.386) | (771.108) |
| *Mother outcomes 1991* | | |
| Employed | −0.081** | −0.041 |
| | (0.037) | (0.038) |
| Unemployed | 0.015 | 0.013 |
| | (0.011) | (0.011) |
| Inactive | 0.066* | 0.028 |
| | (0.037) | (0.037) |
| Earnings | −1033.574 | −261.871 |
| | (632.375) | (622.259) |
| *Family outcomes 1991* | | |
| Migrated | −0.004 | 0.021 |
| | (0.032) | (0.030) |
| Own house | −0.120*** | −0.084*** |
| | (0.031) | (0.029) |
| Social renting | 0.100*** | 0.073*** |
| | (0.029) | (0.027) |
| N | 733 | 733 |

Notes: The table shows the coefficients on father's job loss from a regression of the outcome listed in each row on job loss. Estimates under (1) do not include any controls, estimates under (2) include controls for education and age of the parents, region of residence, number of siblings and housing tenure in 1981. Earnings are annual earnings in 2018 £s. Standard errors are in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% level.
Data source: ONS LS

Table D.3: Effects of father's job loss on child outcomes - by gender

| | Women | | Men | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | No controls | Controls | No controls | Controls |
| *Education* | | | | |
| Has GCSEs | −0.075 | −0.059 | 0.032 | 0.063 |
| | (0.050) | (0.052) | (0.052) | (0.054) |
| Has A-levels | −0.146** | −0.114** | 0.024 | 0.067 |
| | (0.057) | (0.058) | (0.054) | (0.055) |
| Has degree | −0.162*** | −0.132** | 0.006 | 0.032 |
| | (0.052) | (0.052) | (0.046) | (0.046) |
| *Economic outcomes 2001* | | | | |
| Employed | −0.042 | −0.029 | −0.030 | −0.028 |
| | (0.058) | (0.060) | (0.045) | (0.047) |
| Earnings | −1138.434 | −991.763 | −633.484 | −357.104 |
| | (1515.492) | (1544.321) | (1528.466) | (1568.712) |
| Own house | −0.124** | −0.107* | −0.002 | 0.003 |
| | (0.057) | (0.059) | (0.053) | (0.055) |
| Social renter | 0.098** | 0.087* | −0.031 | −0.043 |
| | (0.0464) | (0.048) | (0.042) | (0.041) |
| *Family formation 2001* | | | | |
| Married | 0.065 | 0.053 | −0.029 | −0.032 |
| | (0.044) | (0.043) | (0.036) | (0.035) |
| Number of kids | 0.112 | 0.029 | 0.039 | 0.014 |
| | (0.093) | (0.085) | (0.065) | (0.056) |
| *Economic outcomes 2011* | | | | |
| Employed | −0.042 | 0.007 | −0.024 | −0.020 |
| | (0.051) | (0.052) | (0.040) | (0.042) |
| Earnings | −2176.182 | −507.705 | −388.619 | −221.445 |
| | (1606.830) | (1617.042) | (1540.218) | (1573.353) |
| Own house | −0.147*** | −0.117** | 0.000 | 0.017 |
| | (0.054) | (0.054) | (0.051) | (0.052) |
| Social renter | 0.129*** | 0.114** | −0.036 | −0.051 |
| | (0.045) | (0.046) | (0.037) | (0.039) |
| *Family formation 2011* | | | | |
| Married | 0.024 | 0.022 | −0.059 | −0.072 |
| | (0.059) | (0.063) | (0.056) | (0.057) |
| Number of kids | −0.051 | −0.115 | 0.014 | −0.002 |
| | (0.131) | (0.128) | (0.130) | (0.137) |
| N | 319 | 319 | 315 | 315 |

Notes: The table shows the coefficients on father's job loss from a regression of the outcome listed in each row on job loss. Estimates under (1) do not include any controls, estimates under (2) include controls for education and age of the parents, region of residence, number of siblings and housing tenure in 1981. Earnings are annual earnings in 2018 £s. Standard errors are in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% level.
Data source: ONS LS

## D.4 Mediating impact of father's loss of earnings

Table D.4: Effects of father's job loss and earnings on child outcomes - all outcomes

| | Women | | Men | |
|---|---|---|---|---|
| | (1) Main controls | (2) + 1991 father earnings | (3) Main controls | (4) + 1991 father earnings |
| *Education* | | | | |
| Has GCSEs | −0.059 | −0.074 | 0.063 | 0.073 |
| | (0.052) | (0.064) | (0.054) | (0.071) |
| Has A-levels | −0.114** | −0.057 | 0.067 | 0.076 |
| | (0.058) | (0.071) | (0.055) | (0.072) |
| Has degree | −0.132** | −0.060 | 0.032 | 0.080 |
| | (0.052) | (0.063) | (0.046) | (0.060) |
| *Economic outcomes 2001* | | | | |
| Employed | −0.029 | 0.084 | −0.028 | −0.042 |
| | (0.060) | (0.072) | (0.047) | (0.063) |
| Earnings | −991.763 | 2812.022 | −357.104 | −97.873 |
| | (1544.321) | (1850.562) | (1568.712) | (2073.445) |
| Own house | −0.107* | −0.048 | 0.003 | −0.023 |
| | (0.059) | (0.071) | (0.055) | (0.072) |
| Social renter | 0.087* | 0.057 | −0.043 | −0.037 |
| | (0.048) | (0.058) | (0.041) | (0.054) |
| *Family formation 2001* | | | | |
| Married | 0.053 | 0.047 | −0.032 | 0.004 |
| | (0.043) | (0.052) | (0.035) | (0.046) |
| Number of kids | 0.029 | −0.104 | 0.014 | 0.032 |
| | (0.085) | (0.103) | (0.056) | (0.073) |
| *Economic outcomes 2011* | | | | |
| Employed | 0.007 | 0.023 | −0.020 | −0.002 |
| | (0.052) | (0.064) | (0.042) | (0.055) |
| Earnings | −507.705 | 2358.650 | −221.445 | 1608.469 |
| | (1617.042) | (1971.564) | (1573.353) | (2077.237) |
| Own house | −0.117** | −0.097 | 0.017 | 0.002 |
| | (0.054) | (0.066) | (0.052) | (0.068) |
| Social renter | 0.114** | 0.064 | −0.051 | −0.030 |
| | (0.046) | (0.056) | (0.039) | (0.051) |
| *Family formation 2011* | | | | |
| Married | 0.022 | −0.001 | −0.072 | −0.062 |
| | (0.063) | (0.077) | (0.057) | (0.076) |
| Number of kids | −0.115 | −0.278* | −0.002 | 0.049 |
| | (0.128) | (0.157) | (0.137) | (0.155) |
| N | 315 | 315 | 319 | 319 |

Notes: The table shows the coefficients on father's job loss from a regression of the outcome listed in each row on job loss. Estimates under (1) include our main controls: education and age of the parents, region of residence, number of siblings and housing tenure in 1981. Estimates under (2) add father's earnings in 1991. Standard errors are in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% level.
Data source: ONS LS